UKERC

UK ENERGY RESEARCH CENTRE

UKERC Review of Evidence for Global Oil Depletion

Technical Report 5: Methods of estimating ultimately recoverable resources

July 2009: REF UKERC/WP/TPA/2009/020

Steve Sorrell[1]
Jamie Speirs[2]

1. Sussex Energy Group, SPRU, University of Sussex
2. Imperial College Centre for Environmental Policy and Technology

THE UK ENERGY RESEARCH CENTRE

The UK Energy Research Centre is the focal point for UK research on sustainable energy. It takes a whole systems approach to energy research, drawing on engineering, economics and the physical, environmental and social sciences.

The Centre's role is to promote cohesion within the overall UK energy research effort. It acts as a bridge between the UK energy research community and the wider world, including business, policymakers and the international energy research community and is the centrepiece of the Research Councils Energy Programme.

www.ukerc.ac.uk

Acknowledgements

The authors are very grateful to Fabiana Gordon of the Imperial College Statistical Advisory Service for her help with Section 5 of this report. Thanks are also due to Jean Laherrère, both for his helpful comments on an earlier draft and for his pioneering work in this area. The usual disclaimers apply

Preface

This report has been produced by the UK Energy Research Centre's Technology and Policy Assessment (TPA) function.

The TPA was set up to address key controversies in the energy field through comprehensive assessments of the current state of knowledge. It aims to provide authoritative reports that set high standards for rigour and transparency, while explaining results in a way that is useful to policymakers.

This report forms part of the TPA's assessment of evidence for **near-term physical constraints on global oil supply**. The subject of this assessment was chosen after consultation with energy sector stakeholders and upon the recommendation of the TPA Advisory Group, which is comprised of independent experts from government, academia and the private sector. The assessment addresses the following question:

What evidence is there to support the proposition that the global supply of 'conventional oil' will be constrained by physical depletion before 2030?

The results of the project are summarised in a *Main Report*, supported by the following *Technical Reports*:

1. Data sources and issues
2. Definition and interpretation of reserve estimates
3. Nature and importance of reserve growth
4. Decline rates and depletion rates
5. Methods for estimating ultimately recoverable resources
6. Methods for forecasting future oil supply
7. Comparison of global supply forecasts

The assessment was led by the Sussex Energy Group (SEG) at the University of Sussex, with contributions from the Centre for Energy Policy and Technology at Imperial College, the Energy and Resources Group at the University of California (Berkeley) and a number of independent consultants. The assessment was overseen by a panel of experts and is very wide ranging, reviewing more than 900 studies and reports from around the world.

Each technical report examines one set of issues relevant to the assessment of global oil depletion. *Technical Report 5: Methods of estimating ultimately recoverable resources* examines the methods for estimating the size of oil resources in a region, focusing in particular on the extrapolation of historical trends. It also summarises and evaluates the estimates that have been produced for size of global resources and assesses their implications for future oil supply.

Executive Summary

The date of ultimate exhaustion of the oil resource is largely irrelevant to the 'peak oil' debate. Instead, the primary focus of this debate is the *rate* of production (typically measured in barrels per day) and the reasons why that rate must eventually decline. But while the absolute size of an oil resource is less important than the potential rate of extraction of that resource, disputes over the former nevertheless play a prominent role in the peak oil debate. This is especially the case for conventional oil which continues to dominate global oil supply. Other things being equal, larger estimates of the resource size for conventional oil lead to more optimistic forecasts for future global oil supply – and vice versa. Hence, the 'pessimists' and 'optimists' about future global supply often have very different views on the volume of conventional oil resources that are likely to be economically recoverable.

A central concept in this debate is the *ultimately recoverable resources,* or URR, for a field or region, or the amount of oil estimated to be economically extractable over all time. A variety of methods may be used to estimate *URR* and these may be applied at levels of aggregation ranging from a single well to the entire world. One group of methods relies more upon geological information and is more appropriate for less explored regions, while a second group relies more upon the extrapolation of historical trends and is more appropriate to well-explored regions. In both cases, the methods can either be extremely simple, relying solely upon aggregate data from a region, or highly complex, requiring either detailed geological information or data from individual fields. As with the URR estimates themselves, the relative merits of these different methods is the subject of intense and frequently polarised debate.

The primary objective of this report is to describe and evaluate these different methods. Primary attention is paid to the methods based upon the extrapolation of historical trends, since these are widely used by the analysts concerned about global oil depletion. A second objective is to summarise and evaluate the estimates that have been produced for the global URR of conventional oil and to assess the implications for future oil production. Of particular interest is the relative plausibility of the optimistic and pessimistic estimates and the implications of both for medium-term oil supply.

The main findings of this report are as follows

Methods and principles

- There are a variety of methods for estimating URR and many variations on the basic techniques. 'Geological' techniques are more appropriate for relatively explored regions while 'extrapolation' techniques are more appropriate where exploration is advanced. The confidence bounds on these estimates are commonly very large and the few studies that compare different techniques show they can lead to quite different results. Accuracy can be improved through analysing disaggregate regions, but this is resource intensive and generally requires access to proprietary data. All estimation techniques have identifiable limitations and it is important that estimates are accompanied by confidence intervals and full details about the methodology and assumptions made.
- The extrapolation techniques differ in degree rather than kind and share many of the same strengths and weaknesses. But a key practical difference is that *field-size distribution* and *discovery process* techniques require data on individual fields, while simple *curve-fitting*

only requires aggregate data. All assume a skewed field size distribution and diminishing returns to exploration, with the large fields being found relatively early. But these assumptions will only hold if depletion outweighs the effect of technical change and if the region is geologically homogeneous and has had a relatively unrestricted exploration history. This is frequently not the case.

- Assumptions about the field size distribution and discovery process underlie most of the extrapolation techniques. It is generally acknowledged that the majority of oil resources are contained in a small number of large fields, with around 100 oil fields accounting for up to half of global oil production and up to 500 fields accounting for two thirds of cumulative discoveries. Most of these fields are relatively old, many are well past their peak of production and most of the rest will begin to decline within the next decade or so. The remaining reserves at these fields, their future production profile and the potential for reserve growth is therefore of critical importance for future global supply.
- The proportion of total resources contained within small, undiscovered fields continues to be disputed. While the observed lognormal size distribution of discovered fields is likely to be the result of sampling bias, there is insufficient evidence to conclude whether a 'linear' or 'parabolic fractal' better describes the population size distribution. While technical improvements and higher prices should make more small fields viable, many will remain uneconomic to develop and the exploitation of the rest will be subject to rapidly diminishing returns. As a result, the competing estimates of the resources contained in small fields should be of less significance to future supply than the potential for increased recovery from the giant fields.

Curve fitting techniques

- The popularity of curve-fitting techniques to estimate URR derives from their simplicity and the relative availability of the required data. But many applications of curve-fitting take insufficient account of the weaknesses of these techniques, including: the inadequate theoretical basis; the sensitivity of the estimates to the choice of functional form; the risk of overfitting multi cycle models; the inability to anticipate future cycles of production or discovery; and the neglect of economic political and other variables. In general, these weaknesses appear more likely to lead to underestimates of the URR and have probably contributed to excessively pessimistic forecasts of oil supply.
- Curve fitting to discovery data introduces additional complications such as the uncertainty in reserve estimates and the need to adjust estimates to allow for future reserve growth. The common failure to make such adjustments is likely to have further contributed to underestimates of resource size.
- Tests of curve fitting techniques using illustrative data from a number of regions has shown how different techniques, functional forms, length of time series and numbers of curves can lead to inconsistent results. But although the results raise concerns about the reliability of curve-fitting estimates, the degree of uncertainty may be expected to decline in the future as exploration matures. Also, accuracy may be improved by using the lowest possible level of spatial aggregation, distinguishing between onshore and offshore regions and adjusting for future reserve growth using functions derive from the technical literature.
- The literature on curve-fitting techniques has generally paid insufficient attention to the statistical issues involved, such as goodness of fit, missing variables and serial correlation of the error terms. Where data is available, some of the limitations of curve fitting may be

overcome with the use of hybrid models that incorporate relevant economic and political variables. But despite their better fit to historical data, such models may not lead to substantially different estimates of the URR.

- These limitations do not mean that curve fitting should be abandoned, but do imply that its applicability is more limited than commonly assumed and that the confidence bounds on the results are wider than is commonly assumed. Where possible, resource assessments should employ multiple techniques and sources of data and acknowledge the uncertainty in the results obtained.

Global Estimates

- Estimates of the global URR for conventional oil vary widely in their methods, assumptions and results. Comparison is complicated by the differing definitions of 'conventional oil' and the more pessimistic estimates of the global URR result in part from an excessively narrow definition. Further difficulties arise from the use of competing reserve definitions and differing time-frames for the definition of URR, together with uncertainty over OPEC reserves and the inconsistent treatment of reserve growth. The information currently available does not allow strong constraints to be placed on the last two variables.

- Estimates of the global URR of conventional oil have been trending upwards for the last 50 years and this trend shows little sign of diminishing. Contemporary estimates fall within the range 2000-4300 Gb, while the corresponding estimates of the quantity of remaining resources fall within the range 870 to 3170 Gb. This wide range leads to a corresponding uncertainty in the projections of future global oil supply and the date of peak production.

- The USGS estimated a global URR of 3345 Gb in 2000 and in 2008 the IEA revised this upwards to 3577 Gb. Despite being much larger than previous estimates, the repeated assertions that the USGS estimates are 'discredited' or 'over-optimistic' appear at best premature. Global reserve growth appears to be matching the USGS assumptions, the size of recent discoveries may have been underestimated, there are continuing restrictions on exploration in the most promising areas and a more recent study by Aguilera *et al's* comes to comparably optimistic conclusions. However, the IEA estimate relies upon a large contribution from EOR that they anticipate will take decades to be realised while some of Aguilera *et al's* assumptions appear questionable.

- In a simple logistic model, increasing the global URR by one billion barrels would delay the date of peak production by only 4.7 days. This result is not substantially changed if a more sophisticated model is used, that allows for varying degrees of asymmetry in the production cycle (Kaufmann and Shiers, 2008). For a range of assumptions about the size of the global URR and the rate of change of production before and after the peak, the date of peak production is found to lie between 2009 and 2031. Delaying the peak beyond 2030 requires optimistic assumptions about the global URR combined with a relatively steep post-peak decline rate and/or slower rates of demand growth than are conventionally assumed. Forecasts that predict no peak before 2030 should be evaluated on this basis.

- Even if the larger URR estimates are correct, it does not necessarily follow that the resource can or will be accessed at the rate required to maintain global production at a particular level. If these resources can only be accessed relatively slowly at high cost, supply constraints could inhibit demand growth. Furthermore, if producers lack the incentive to maximize production, demand growth could be constrained further – especially in the importing countries. Hence, the primary issue for the period to 2030 is the *rate* at which the resource can be accessed and produced.

Contents

1 INTRODUCTION ... 1

1.1 WHY DO WE NEED TO ESTIMATE RESOURCE SIZE? ... 1
1.2 STRUCTURE OF THE REPORT ... 3

2 CONCEPTS, DEFINITIONS AND METHODS ... 5

2.1 INTRODUCTION ... 5
2.2 WHAT ARE ULTIMATELY RECOVERABLE RESOURCES? ... 5
2.3 LEVELS OF AGGREGATION FOR ESTIMATES OF ULTIMATELY RECOVERABLE RESOURCES ... 9
2.4 CUMULATIVE DISCOVERIES AND RESERVE GROWTH ... 12
2.5 FIELD SIZE DISTRIBUTIONS ... 15
2.5.1 Why big fields matter ... 21
2.6 METHODS OF ESTIMATING ULTIMATELY RECOVERABLE RESOURCES ... 24
2.6.1 Geological assessment ... 25
2.6.2 Expert assessment ... 26
2.6.3 Field-size distributions ... 26
2.6.4 Historical extrapolation ... 28
2.6.5 Comparison of methods ... 30
2.7 SUMMARY ... 30

3 EXTRAPOLATION METHODS – CLASSIFICATION, DESCRIPTION AND EVALUATION ... 33

3.1 INTRODUCTION ... 33
3.2 EXPLAINED AND EXPLANATORY VARIABLES FOR CURVE-FITTING TECHNIQUES ... 34
3.2.1 The production cycle ... 35
3.2.2 The discovery cycle ... 36
3.2.3 Backdated discovery estimates ... 39
3.2.4 Growth functions ... 40
3.2.5 The backdated discovery cycle ... 41
3.2.6 Discovery as a function of effort ... 42
3.2.7 Summary of explained and explanatory variables ... 45
3.3 PRODUCTION OVER TIME TECHNIQUES ... 46
3.3.1 Production projection ... 47
3.3.2 Production decline curves ... 55
3.3.3 Summary ... 61
3.4 DISCOVERY OVER TIME TECHNIQUES ... 62
3.4.1 Discovery projection using current data ... 62
3.4.2 Discovery projection using backdated data ... 66
3.4.3 Summary ... 70
3.5 DISCOVERY OVER EFFORT TECHNIQUES ... 71
3.5.1 Creaming curves ... 72
3.5.2 Yield per effort curves ... 78
3.5.3 Summary ... 80
3.6 DISCOVERY PROCESS MODELS ... 81
3.6.1 Arps-Roberts model ... 82
3.6.2 Barouch-Kaufman model ... 84
3.6.3 Summary ... 86
3.7 SUMMARY ... 87

4 CONSISTENCY OF CURVE FITTING TECHNIQUES ... 89

4.1 INTRODUCTION ... 89
4.2 HUBBERT LINEARISATION ... 93
4.2.1 Background and approach ... 93
4.2.2 Results - consistency over time ... 94
4.3 DISCOVERY PROJECTION ... 104
4.3.1 Background and approach ... 104
4.3.2 Results – consistency over functional form ... 106

4.3.3 Results - consistency over time 111
4.4 CREAMING CURVES 119
4.4.1 Background and approach 119
4.4.2 Results - consistency over functional form 119
4.4.3 Results - consistency over the number of curves 125
4.5 COMPARISON OF TECHNIQUES 128
4.6 SUMMARY AND IMPLICATIONS 129

5 STATISTICAL ROBUSTNESS OF CURVE-FITTING TECHNIQUES 132

5.1 INTRODUCTION 132
5.2 OVERVIEW OF STATISTICAL ISSUES 132
5.2.1 Specification of time-series models 132
5.2.2 Missing variables in model specification 135
5.2.3 Serial correlation in the error terms 136
5.2.4 Forecasting 137
5.2.5 Summary 138
5.3 ILLUSTRATION - GLOBAL PRODUCTION PROJECTION 139
5.3.1 Model 1: global cumulative production projection 139
5.3.2 Model 2: global production projection 141
5.3.3 Model 3: global production projection with lagged dependent variable 142
5.3.4 Model 4: ARIMA model of global production: 144
5.4 RECONCILING ECONOMETRICS AND CURVE-FITTING 145
5.4.1 A two-stage production projection 146
5.4.2 Production projection using cointegration techniques 148
5.4.3 Production projection with variable URR 149
5.4.4 Hybrid modelling of yield per effort 149
5.4.5 Modelling technical change 150
5.4.6 The challenge of hybrid modelling 153
5.5 SUMMARY AND IMPLICATIONS 153

6 GLOBAL ESTIMATES OF ULTIMATELY RECOVERABLE RESOURCES AND THEIR IMPORTANCE FOR FUTURE OIL SUPPLY 156

6.1 INTRODUCTION 156
6.2 A BRIEF HISTORY OF GLOBAL ESTIMATES OF ULTIMATELY RECOVERABLE RESOURCES 157
6.2.1 Campbell and Laherrère 162
6.2.2 Miller (1992) 163
6.2.3 Odell 164
6.3 THE USGS WORLD PETROLEUM ASSESSMENT 2000 165
6.3.1 Methods 165
6.3.2 Results 167
6.3.3 Evaluation 172
6.4 RECENT MODIFICATIONS TO THE USGS ESTIMATES 173
6.4.1 The IEA World Energy Outlook 2008 173
6.4.2 Colorado School of Mines 176
6.5 THE IMPLICATIONS OF GLOBAL URR ESTIMATES FOR FUTURE GLOBAL SUPPLY 178
6.6 SUMMARY 182

7 SUMMARY AND CONCLUSIONS 185

7.1 METHODS AND PRINCIPLES 185
7.2 CURVE FITTING TECHNIQUES 185
7.3 GLOBAL ESTIMATES 186

REFERENCES 189

Figures

FIGURE 2.1 RESOURCES CLASSIFICATION IN THE PETROLEUM RESOURCE MANAGEMENT SYSTEM 8
FIGURE 2.2 ILLUSTRATION OF CUMULATIVE RESERVE GROWTH 14
FIGURE 2.3 OIL AND GAS FIELD SIZE DISTRIBUTION FOR THE DENVER BASIN IN 1958 16
FIGURE 2.4 OBSERVED FIELD SIZE DISTRIBUTION FOR THE FRIO-STRANDPLAIN PLAY IN TEXAS AT THREE DIFFERENT POINTS IN TIME 18
FIGURE 2.5 CUMULATIVE FREQUENCY PLOT OF FIELD SIZES FOR THE FRIO STRANDPLAIN PLAY IN TEXAS, EXCLUDING SMALLER FIELD SIZES 20
FIGURE 2.6 CUMULATIVE FREQUENCY PLOT OF FIELD SIZES FOR THE NIGER DELTA 21
FIGURE 2.7 THE ESTIMATED CONTRIBUTION OF GIANT OILFIELDS TO GLOBAL CRUDE OIL PRODUCTION................ 23
FIGURE 2.8 CLASSIFICATION OF METHODS OF ESTIMATING URR 24
FIGURE 2.9 ESTIMATING URR FROM A CUMULATIVE FIELD SIZE DISTRIBUTION THAT IS ASSUMED TO FOLLOW A PARETO LAW.. 27
FIGURE 2.10 ESTIMATING URR BY PLOTTING CUMULATIVE DISCOVERIES AS A FUNCTION OF FIELD RANK........... 28
FIGURE 3.1 HUBBERT'S 1956 PROJECTION OF THE FORTHCOMING PEAK IN US OIL PRODUCTION 48
FIGURE 3.2 LOGISTIC MODEL OF CUMULATIVE PRODUCTION CYCLE 50
FIGURE 3.3 LOGISTIC MODEL OF PRODUCTION CYCLE 51
FIGURE 3.4 A FIT OF THE LOGISTIC MODEL TO US CUMULATIVE PRODUCTION DATA (CRUDE OIL +NGLS) 52
FIGURE 3.5 A FIT OF THE LOGISTIC MODEL TO US PRODUCTION DATA (CRUDE OIL +NGLS) 52
FIGURE 3.6 CUMULATIVE NORMAL MODEL OF CUMULATIVE PRODUCTION CYCLE 53
FIGURE 3.7 GOMPERTZ MODEL OF A CUMULATIVE PRODUCTION CYCLE... 54
FIGURE 3.8 PRODUCTION VERSUS CUMULATIVE PRODUCTION AS AN IDEALISED PARABOLA 57
FIGURE 3.9 'HUBBERT LINEARISATION' OF PARABOLIC RELATIONSHIP BETWEEN PRODUCTION AND CUMULATIVE PRODUCTION 57
FIGURE 3.10 HUBBERT LINEARISATION OF US OIL PRODUCTION 57
FIGURE 3.11 LINEARISATION OF EXPONENTIAL PRODUCTION DECLINED FOR THE UK FORTIES FIELD................... 60
FIGURE 3.12 PRODUCTION CYCLE OF THEUK FORTIES FIELD 60
FIGURE 3.13 HUBBERT'S IDEALISED RELATIONSHIP BETWEEN CUMULATIVE DISCOVERIES, CUMULATIVE PRODUCTION AND PROVED RESERVES AS A FUNCTION OF TIME 63
FIGURE 3.14 HUBBERT'S IDEALISED RELATIONSHIP BETWEEN RATE OF DISCOVERY, RATE OF PRODUCTION AND RESERVE ADDITIONS AS A FUNCTION OF TIME 63
FIGURE 3.15 US CUMULATIVE PROVED DISCOVERIES, CUMULATIVE PRODUCTION AND PROVED RESERVES FROM 1900 TO 1962.. 64
FIGURE 3.16 US RATE OF DISCOVERY, RATE OF PRODUCTION AND RATE OF CHANGE OF PROVED RESERVES FROM 1900 TO 1962.. 65
FIGURE 3.17 DISCOVERY PROJECTION FOR THE PERMIAN BASIN USING BACKDATED DISCOVERY ESTIMATES THROUGH TO 1964 68
FIGURE 3.18 DISCOVERY PROJECTION FOR THE PERMIAN BASIN USING BACKDATED DISCOVERY ESTIMATES THROUGH TO 1964 69
FIGURE 3.19 EXAMPLE OF A CREAMING CURVE.. 72
FIGURE 3.20 EXPLORATION HISTORY OF THE MICHIGAN BASIN 75
FIGURE 3.21 LAHERRÈRE'S CREAMING CURVE ANALYSIS OF THE UNITED STATES.. 76
FIGURE 4.1 HUBBERT LINEARISATION OF LOGISTIC GROWTH IN CUMULATIVE PRODUCTION 93
FIGURE 4.2 HUBBERT LINEARISATION OF PRODUCTION DATA FOR REGION A 94
FIGURE 4.3: HUBBERT LINEARISATION OF PRODUCTION DATA FOR REGION B 95
FIGURE 4.4 HUBBERT LINEARISATION OF OFFSHORE PRODUCTION IN REGION B. 96
FIGURE 4.5 HUBBERT LINEARISATION OF ONSHORE PRODUCTION IN REGION B.. 96
FIGURE 4.6 HUBBERT LINEARISATION OF GOMPERTZ GROWTH IN CUMULATIVE PRODUCTION 97
FIGURE 4.7 HUBBERT LINEARISATION OF OFFSHORE PRODUCTION DATA FOR REGION C. 97
FIGURE 4.8 SUMMARY OF CONSISTENCY OVER TIME TESTS FOR HUBBERT LINEARISATION TECHNIQUE 99
FIGURE 4.9 COMPARISON OF BACKDATED CUMULATIVE DISCOVERY TRENDS IN 'PIONEER' AND 'YOUNG' REGIONS 107
FIGURE 4.10: LOGISTIC DISCOVERY PROJECTION FOR REGION D... 108
FIGURE 4.11: GOMPERTZ DISCOVERY PROJECTION FOR REGION D 108
FIGURE 4.12 SUMMARY OF CONSISTENCY OVER FUNCTIONAL FORM TESTS FOR DISCOVERY PROJECTION 110
FIGURE 4.13: REGION E – SENSITIVITY OF URR ESTIMATES FROM LOGISTIC DISCOVERY PROJECTION TO THE TIME THROUGH TO DISCOVERY (T_D) 112

Figure 4.14: Region E – sensitivity of URR estimates from Gompertz discovery projection to the time through to discovery (T_D)........ 112
Figure 4.15: Region B – sensitivity of URR estimates from logistic discovery projection to the time through to discovery (T_D) 113
Figure 4.16: Region B – sensitivity of URR estimates from Gompertz discovery projection to the time through to discovery (T_D)........ 113
Figure 4.17:Change in R^2 estimates for discovery projections in Region E using different lengths of time series 114
Figure 4.18 Change in R^2 estimates for discovery projections in Region F using different lengths of time series 114
Figure 4.19 Summary of consistency over time for discovery projection 116
Figure 4.20: Hyperbolic and exponential creaming curves for Region A 120
Figure 4.21 Backdated discoveries as a function of exploratory effort in Region B........ 121
Figure 4.22 Rate of discovery over time for Region B with smoothed 5 year average 121
Figure 4.23: Hyperbolic and linear creaming curves for Region L 122
Figure 4.24: Rate of discovery over time for Region L with smoothed 5 year average........ 122
Figure 4.25 Summary of consistency over functional form tests for creaming curves........ 124
Figure 4.26: Creaming curve data for Region E fitted with a single hyperbola 126
Figure 4.27: Creaming curve data for Region E fitted with two sequential hyperbola 126
Figure 4.28: Example of the difference between simple linear model and overfitted polynomial. 127
Figure 4.29: Creaming curve data for Region H fitted with two sequential hyperbola........ 127
Figure 5.1 Gobal (a) and local (b) trends in time-series data 133
Figure 5.2 Structural breaks in time series........ 136
Figure 5.3 Model 1 curve fit to cumulative production........ 140
Figure 5.4 Model 1 curve fit to cumulative production post 2002 140
Figure 5.5 Model 2 curve fit to rate of production 141
Figure 5.6 Autocorrelation function (ACF) and partial autocorrelation function (PACF) for Model 2........ 142
Figure 5.7 Model 3 curve fit to rate of production 144
Figure 5.8 Autocorrelation function for Model 3 144
Figure 5.9 Model 4 – time series model of rate of production........ 145
Figure 5.10 Model 4: (a) Autocorrelation function (ACF) (b) Partial autocorrelation function (PACF)........ 145
Figure 5.11 Kaufmann's econometric model (solid line) of US lower 48 oil production (dots) as compared to logistic model (dashed line) 147
Figure 5.12 The long run average cost of oil production in the lower 48 US states 148
Figure 5.13 Yield per effort for oil exploration in the Gulf of Mexico 1947-98 151
Figure 5.14 Individual effect of technical change and depletion on yield per effort for oil exploration in the Gulf of Mexico 1947-98........ 152
Figure 5.15 Net effect of technical change and depletion on yield per effort for oil exploration in the Gulf of Mexico 1947-98........ 153
Figure 6.1: Comparison of global URR estimates over the last 70 years 162
Figure 6.2: Presentation of extrapolation methods for Libya oil data as presented in Campbell and Heapes (2008)........ 163
Figure 6.3 The global oil system model as presented by Miller (1992) 164
Figure 6.4: Odell's estimates of world ultimate reserves of crude oil from conventional sources (with extrapolation to the year 2000) 165
Figure 6.5 USGS 2000: components of the estimated global URR for conventional oil 169
Figure 6.6 Comparing historical trends in backdated 2P discoveries with those implied by the USGS 2000 for the period 1995-2025 169
Figure 6.7 IEA 2008: components of the estimated global URR for conventional oil........ 175
Figure 6.8 The peaking of global conventional oil production under different assumptions about the global URR - simple logistic model 179
Figure 6.9 Sensitivity of the date of global peak production of conventional oil to different assumptions about the global URR – simple logistic model........ 180
Figure 6.10 Illustrative scenarios for future global oil production with a URR of 3 Gb........ 181

Tables

TABLE 2.1: IVANHOE AND LECKIE'S ESTIMATES OF THE SIZE DISTRIBUTION OF THE WORLD'S OILFIELDS............. 22
TABLE 3.1 CLASSIFICATION OF CURVE-FITTING METHODS BY EXPLAINED AND EXPLANATORY VARIABLES.......... 35
TABLE 3.2 MATHEMATICAL NOTATION FOR CURVE-FITTING TECHNIQUES 35
TABLE 3.3 CLASSIFICATION OF CURVE-FITTING METHODS BY EXPLAINED AND EXPLANATORY VARIABLES – NOTATIONAL SUMMARY.. 46
TABLE 3.4 COMPARISON BETWEEN HUBBERT LINEARISATION AND EXPONENTIAL DECLINE CURVE 61
TABLE 3.5 CLASSIFICATION OF DISCOVERY PROCESS MODELS BY EXPLAINED AND EXPLANATORY VARIABLES.... 82
TABLE 3.6 MATHEMATICAL NOTATION FOR DISCOVERY PROCESS MODELS.. 82
TABLE 3.7 COMPARISON OF THE ARPS-ROBERTS AND BAROUCH-KAUFMAN MODELS.. 86
TABLE 4.1 CONSISTENCY TESTS ON CURVE FITTING TECHNIQUES 90
TABLE 4.2 NOTATION FOR EXPLAINED AND EXPLANATORY VARIABLES... 91
TABLE 4.3 SUMMARY OF CONSISTENCY TESTS OF HUBBERT LINEARISATION TECHNIQUE..................................... 98
TABLE 4.4 SUMMARY OF CONSISTENCY OVER FUNCTIONAL FORM TESTS FOR DISCOVERY PROJECTION 109
TABLE 4.5 SUMMARY OF CONSISTENCY OVER TIME TESTS FOR DISCOVERY PROJECTION 115
TABLE 4.6 SUMMARY OF CONSISTENCY OVER FUNCTIONAL FORM TESTS FOR CREAMING CURVES...................... 124
TABLE 4.7 SUMMARY OF CONSISTENCY OVER THE NUMBER OF CURVES TESTS FOR CREAMING CURVES 128
TABLE 4.8 SUMMARY OF CONSISTENCY BETWEEN TECHNIQUES TESTS 129
TABLE 4.9 RESULTS OF CONSISTENCY TESTS - SUMMARY 131
TABLE 5.1 PARAMETER ESTIMATES AND GOODNESS OF FIT FOR MODEL 1 140
TABLE 5.2 PARAMETER ESTIMATES AND GOODNESS OF FIT FOR MODEL 2 141
TABLE 5.3 PARAMETER ESTIMATES AND GOODNESS OF FIT FOR MODEL 3 143
TABLE 6.4 HISTORICAL ESTIMATES OF THE GLOBAL ULTIMATELY RECOVERABLE RESOURCE OF CONVENTIONAL OIL 158
TABLE 6.5 USGS WPA 2000: MEAN ESTIMATES OF GLOBAL URR FOR PETROLEUM LIQUIDS (GB) 168
TABLE 6.6 USGS WORLD PETROLEUM ASSESSMENT 2000: SUMMARY OF GLOBAL URR ESTIMATES FOR PETROLEUM LIQUIDS 170
TABLE 6.7 USGS WPA 2000: MEAN ESTIMATES OF UNDISCOVERED RESOURCES BY REGION............................. 171
TABLE 6.8 IEA 2008 WEO: MEAN ESTIMATES OF GLOBAL URR FOR PETROLEUM LIQUIDS (GB)........................ 174

1 Introduction

1.1 Why do we need to estimate resource size?

Concerns about global oil depletion are often misleadingly characterised as concerns about 'running out of oil'. The image is one of a tank being slowly drained and eventually running dry, which implies that the main concern is precisely when this will occur. But while oil is clearly a finite resource, the date of ultimate exhaustion of this resource is largely irrelevant to the 'peak oil' debate. Instead, the primary focus of this debate is the *rate* of production (typically measured in barrels per day) and the reasons why that rate must eventually decline.

There are well-established physical and geological reasons why the rate of production from both individual fields and oil-producing regions typically rises to a peak and subsequently declines (Bentley, 2009). However, these physical determinants are mediated by a multitude of technical, economic and political factors that make forecasting future supply a hazardous undertaking. While the estimated size of the resource is an important variable in such forecasts, it is not necessarily the most important one. For example, the global resource of 'non-conventional oil' is acknowledged to be several times larger than that of 'conventional oil' (IEA, 2008),[1] but these resources are costly and difficult to exploit, require significant amounts of energy to extract, transport and refine and are associated with serious environmental impacts. Most importantly, if these resources can only be accessed relatively *slowly*, they may not compensate for the decline in production from more conventional sources and hence may not have much influence on the date of global peak production.

But while the absolute size of an oil resource is less important than the potential rate of extraction of that resource[2], disputes over the former nevertheless play a prominent role in the peak oil debate. This is especially the case for conventional oil which continues to dominate global oil supply. Other things being equal, larger estimates of the resource size for conventional oil lead to more optimistic forecasts for future global oil supply – and vice versa (Bartlett, 2000; Bentley, *et al.*, 2009). Hence, the 'pessimists' and 'optimists' about future global supply often have very different views on the volume of conventional oil resources that are likely to be economically recoverable. This disagreement is compounded by confusion and disagreement over the meaning of key terms and concepts (e.g. 'conventional') and even over whether the physical size of the resource is relevant at all (Adelman, 1993).

A central concept in this debate is the *ultimately recoverable resources*, or *URR*, for a field or region. This is defined as *the amount of oil estimated to be economically extractable from a field or region over all time*. The *URR* can be broken down into a number of different components, as summarised in Box 1.1. Current estimates of the global *URR* for conventional oil fall within the range 2000 to 4300 Gb which compares to cumulative production through to 2007 of 1128 Gb.[3] This represents a quite remarkable range of uncertainty for such a

[1] There is no single definition of these terms and ambiguity over their meaning is a major source of confusion in the 'peak oil' debate. Conventional oil is taken here to include crude oil, condensate, and natural gas liquids (NGLs) and to *exclude* oil sands, shale oil and extra heavy oil, as well as substitute liquids derived from natural gas, coal and biomass. For more background on the definitions of these terms, see the companion report by Speirs and Sorrell (2009).
[2] Often stated as: "…it's the size of the tap, not the size of the tank".
[3] These figures include natural gas liquids (NGLs).

fundamental quantity and in turn contributes to a corresponding uncertainty in the projections of future global oil supply.

Box 1.1 Components of ultimately recoverable resources

At any point in time, the *URR* for a region may be broken down into the sum of the following:

- *Cumulative production*: the total amount of oil that has been produced from the region since production began.
- *Reserves*: the volume of oil estimated to be extractable from known deposits in the region under defined technical and market conditions.
- *Yet to find*: the volume of oil estimated to be economically extractable from unknown deposits in the region (i.e. those that have yet to be discovered).

While cumulative production should be known relatively accurately, estimates of reserves and yet to find resources are inherently uncertain. For example, the level of confidence in reserve estimates is typically indicated by the terms *proved* reserves (1P), *proved and probable* reserves (2P) and *proved, probable and possible* reserves (3P). Similar distinctions can be made for estimates of yet to find resources, although this is less common. All such estimates rely upon assumptions about the geological features of the region, the technology of resource extraction and the economics of oil production.

The sum of cumulative production and reserves in a region is commonly referred to as *cumulative discoveries*. Estimates of cumulative discoveries tend to grow over time, as a result of improved technology and other factors. This is commonly referred to as *reserve growth* although it is more accurately described as *cumulative discovery growth*, as it is the estimates of cumulative discoveries that are growing, rather than declared reserves. While poorly understood, reserve growth is of critical importance for future oil supply.

For individual fields, the URR represents the sum of cumulative discoveries and estimates of future reserve growth. For a geographical region, the URR represents the sum of cumulative discoveries, future reserve growth and yet to find resources. The *remaining resources* for a region are all the resources that have yet to be produced, calculated by subtracting cumulative production from the estimate of URR.

A variety of methods may be used to estimate *URR* and these may be applied at levels of aggregation ranging from a single well to the entire world. One group of methods relies more upon geological information and is more appropriate for less explored regions, while a second group relies more upon the extrapolation of historical trends and is more appropriate to well-explored regions. In both cases, the methods can either be extremely simple, relying solely upon aggregate data from a region, or highly complex, requiring either detailed geological information or data from individual fields. As with the URR estimates themselves, the relative merits of these different methods is the subject of intense and frequently polarised debate.

The primary objective of this report is to describe and evaluate these different methods. Primary attention is paid to the methods based upon the extrapolation of historical trends, since these are widely used by the analysts concerned about global oil depletion. We seek to identify the relative strengths and weaknesses of these methods, the degree of uncertainty in the associated URR estimates and the conditions under which they are more or less likely to produce reliable results. A second objective is to summarise and evaluate the estimates that have been produced for the global URR of conventional oil and to assess the implications for future oil production. Of particular interest here is the relative plausibility of the optimistic and pessimistic estimates and the implications of both for medium-term oil supply.

As with other elements of the UKERC study, the report is based upon a *systematic review* of the academic and technical literature, in this case drawing upon more than 900 studies from around the world. To supplement the literature review, we have also analysed data from a number of oil-producing regions in order to assess the reliability of extrapolation methods under different conditions and to highlight a number of the relevant statistical issues. As well as drawing conclusions relevant to the UKERC study we hope that this report can provide a reference source for future work in this area.

1.2 Structure of the report

The report is structured as follows. Section 2 introduces some key concepts and definitions and summarises the methods available to estimate ultimately recoverable resources (URR). Particular attention is paid to the phenomena of reserve growth and to the distribution of petroleum resources between different sizes of field. It shows how the field size distribution underpins many of the methods for estimating URR and how global oil resources tend to be concentrated in a small number of large fields. The methods of estimating URR are grouped into four categories, namely geological assessments, expert assessments, field size distribution approaches and historical extrapolation techniques. The latter are widely used by those concerned about peak oil and form the primary focus of the remainder of the report.

Section 3 is the core of the report. It describes and evaluates the *extrapolation methods* of estimating ultimately recoverable resources, which involve analysing historical data on production or discoveries in a region and extrapolating this to derive an estimate of the *URR*. While these techniques vary greatly in their data requirements and level of sophistication, they share the common assumptions that: a) the field size distribution is highly skewed, with the majority of oil being located in a small number of large fields; and b) these large fields tend to be discovered early in the exploration process, with subsequent discoveries being progressively smaller and the product of increasingly greater effort. The extrapolation techniques are shown to fall into two broad groups, namely *curve-fitting* techniques which use aggregate data for a region and *discovery process models* which require data on individual fields. Curve-fitting techniques, in turn, are classified into three groups, namely *production over time*, *discovery over time* and *discovery over effort*, which each encompass three individual techniques. Section 3 describes each technique, identifies its historical origins and contemporary application, evaluates its strengths and weaknesses, clarifies its relationship to other techniques and identifies the conditions under which it is more or less likely to be reliable. It also introduces a standard mathematical notation that is used throughout the remainder of the report and which can assist the interpretation of the empirical literature.

Section 4 uses data from ten regions to investigate the *consistency* of *URR* estimates from curve-fitting techniques; that is, the extent to which one estimate differs from another. For each region, it compares the estimates obtained from different extrapolation techniques, and also from the same technique using different length of time series, different choices of functional form and different choices for the number of curves. The results raise serious concerns about the reliability of these techniques, at least when (as is often the case) they are applied at the country or regional level. Some reasons for these inconsistencies are discussed and the conditions under which more reliable estimates may potentially be obtained are highlighted. In particular, it is recommended that the techniques are best applied in well-explored regions at the lowest possible level of spatial aggregation, distinguishing between onshore and offshore regions and (if possible) between different types of exploratory activity. It is also important that the discovery estimates are adjusted to allow for future reserve growth.

Section 5 explores some of the statistical issues raised by curve-fitting techniques and argues that much of the current literature fails to address these issues adequately. It introduces problems of model specification and comparison, missing variables and serial correlation of the error terms and uses a case study to both illustrate these issues and show how they may potentially be addressed. Using examples from the literature, it shows how the inclusion of economic and political determinants of discovery and/or production can improve the model fit and allow the dependence of *URR* on energy prices and other factors to be directly explored. However, there are relatively few examples of this type and it is not obvious that such 'hybrid' models will lead to substantially different estimates of the regional *URR*.

Section 6 provides an overview and evaluation of global *URR* estimates and assesses their implications for future global oil supply. It first summarises and compares some global *URR* estimates that have been made in the past, illustrates how these have grown over time and looks in more detail at three of the more prominent estimates. It then summarises the methods and results of the US Geological Survey (USGS) World Petroleum Assessment 2000, evaluates whether the subsequent experience is consistent with these estimates and examines how they have recently been updated by the IEA and Colorado School of Mines. It then examines the implications of the uncertainty in global URR estimates for the date of peak global production and argues that delaying the peak beyond 2030 requires very optimistic assumptions about the size of the global *URR* and also implies a relatively steep post-peak decline rate.

Finally, Section 7 provides a brief summary of the main findings.

2 Concepts, definitions and methods

2.1 Introduction

This section introduces some key concepts and definitions relevant to ultimately recoverable resources (*URR*) and introduces the main methodological approaches that are available to estimate the size of those resources. It argues that *URR* estimates are necessarily uncertain and dynamic and subject to a wide range of institutional, economic and technological influences. Estimates of *URR* may be derived for levels of aggregation ranging from a single reservoir to the entire world and for both unexplored and heavily explored areas. They may also be obtained by using either very simple or highly complex techniques. In all cases, however, such estimates of best expressed as a probability distribution rather than a 'most likely' value.

The structure of this section is as follows. Section 2.2 clarifies the definition of URR and relates this to a standard method for classifying petroleum resources, namely the Petroleum Resources Management System (PRMS). Section 2.3 identifies the different levels of aggregation for which estimates of URR may be developed and provides some relevant background on oil formation. Section 2.4 introduces the concept of cumulative discoveries and examines the tendency of these estimates to grow over time - so called 'reserve growth'. Section 2.5 investigates how petroleum resources are distributed between different sizes of field within a region and shows how this fact underpins many of the methods of estimating URR. Finally, Section 0 examines these methods and classifies them under four categories, namely: a) geological assessments; b) expert assessments; c) field size distribution approaches; and d) historical extrapolation. While each approach is summarised, it is the extrapolation methods that form the primary focus of the remainder of the report.

2.2 What are ultimately recoverable resources?

As with oil and gas reserves (Thompson, 2008), the concept of ultimately recoverable resources[4] (*URR*) is defined and interpreted in different ways by different individuals and organisations. Since those holding optimistic views on the future global oil supply frequently interpret the term differently from those holding more pessimistic views, quantitative estimates of URR are an enduring focus of dispute. The BP Statistical Review defines URR as follows:

> "*URR* is an estimate of the total amount of oil that will ever be recovered and produced. It is a subjective estimate in the face of only partial information. While some consider *URR* to be fixed by geology and the laws of physics, in practice estimates of *URR* continue to be increased as knowledge grows, technology advances and economies change. Economists often deny the validity of the concept of ultimately recoverable resources as they consider that the recoverability of resources depends upon changing and unpredictable economies and evolving technologies."(BP, 2008)

[4] Some authors use the term 'ultimately recoverable reserves'. However, this is misleading since it fails to acknowledge the basic distinction between reserves and resources that is reflected in the majority of classification schemes. An alternative and more accurate term is Estimated Ultimate Recovery (EUR).

Reflecting the economists' viewpoint, Adelman (1991) rejects the notion that estimates of *URR* can play a useful role in forecasting future oil supply.

> "Mineral resources are essentially inexhaustible……how much remains in the ground is an amount unknown, probably unknowable and ultimately unimportant. 'Finite limited resources' is therefore an empty slogan. Only cost and price matter." (Adelman, 1991)

In contrast, estimates of *URR* play a central role in Hubbert's forecasts of future US and global oil supply and in the work of subsequent authors such as Campbell (1997) and Laherrère (2003; 1999b). These authors forecast future production from a region by fitting a curve to historical data on oil production and projecting this forward into the future (see Section 3). Estimates of the *URR* for the region are used to constrain these forecasts by setting limits to the area under the curve. Without this constraint, such projections would be more difficult to perform, especially in regions that have yet to reach their peak of production (Caithamer, 2008). However, such 'curve-fitting' techniques can also be used to *estimate* the URR for the region. Two key assumptions of this approach are that the *URR* for a region can be estimated reasonably accurately from the historical pattern of discovery or production in that region and that these estimates will be relatively unaffected by future changes in costs, prices and technology. Critics strongly dispute both of these assumptions (Lynch, 2004; Nehring, 2006a; b; d), leading to a highly polarised debate:

> "….In general, [*URR*] estimates produced by analysts who stress the physical aspects of oil discovery and production are well below those produced by analysts who stress the economic aspects. Each group generates estimates that are heralded by adherents and ridiculed by opponents, regardless of the merits of the estimation process itself. If the estimate confirms one's *a priori* expectations about the scarcity or abundance of remaining oil resources then adherents argue that the estimate is accurate and unbiased, the methodology is rigorous and scholarly, and the estimators' integrity and qualifications are beyond reproach. If the estimate contradicts one's *a priori* expectations then opponents argue that the data used to make the estimate are inappropriate, the methodology is fraught with bias, and the analysts obviously have an 'axe to grind'." (Cleveland, 1991)

To assist in the interpretation of *URR* estimates, it is helpful to review a typical classification scheme for petroleum resources and reserves. As described by Thompson (2008), a variety of such schemes have been used over the years, but international standardisation has yet to be achieved. The chosen scheme is the Petroleum Resources Management System (PRMS), which was introduced in 2007 by the Society for Petroleum Engineers (SPE), the American Association of Petroleum Geologists (AAPG), the World Petroleum Council (WPC) and the Society of Petroleum Evaluation Engineers (SPEE). This system embodies many of the features of earlier classification schemes and is expected to be influential.[5]

The PRMS reflects two variables relevant to resource evaluation, namely: a) varying knowledge about the existence, quality and magnitude of hydrocarbon deposits; and b) the varying extent to which these are likely to be technically and economically recoverable under current and anticipated future conditions. A two-dimensional classification scheme based upon these dimensions was first introduced by Mckelvey (1972). In this (and most other) classification system, 'reserves' are defined as recoverable and commercial volumes of identified hydrocarbons associated with known fields, while the more inclusive term of 'resources' also includes hydrocarbons that have yet to be discovered (sometimes termed 'yet

[5] However, the PRMS is complex with numerous subdivisions, which could be a drawback (Weeks, 1975).

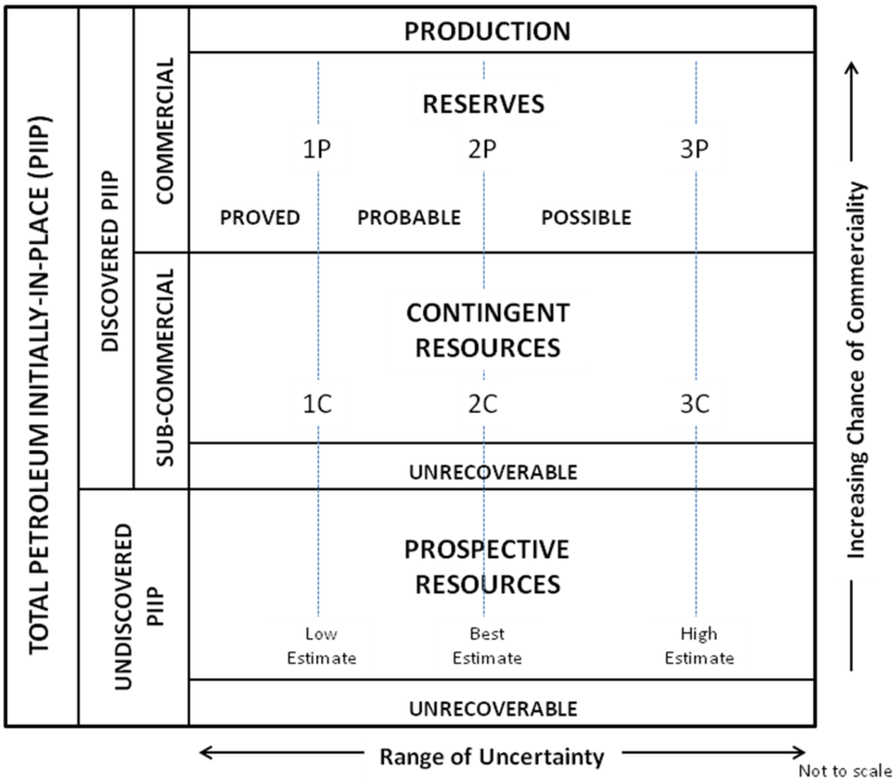
to find' or YTF) as well as those that have been discovered but have yet to become either technically possible or economically viable to recover. The PRMS classification scheme is illustrated in Figure 2.1 and some of the relevant terms are defined in Box 2.1. The main innovation of the PRMS compared to earlier systems is that estimates of recoverable resources are linked to investment in specific projects. The full classification system includes considerably more guidance on issues such as the definition and economics of projects and the methodologies of resource estimation (WPC, 2007).

Box 2.1 Key definitions in the Petroleum Resource Management System

- *Total petroleum initially in place*: includes the quantity of petroleum that is estimated, as of a given date, to be contained in known accumulations prior to production plus estimated quantities in accumulations that have yet to be discovered.
- *Discovered petroleum initially in place:* the quantity of petroleum that is estimated, as of a given date, to be contained in known accumulations prior to production.
- *Undiscovered petroleum initially in place*: the quantity of petroleum estimated, as a given date, to be contained within accumulations yet to be discovered.
- *Production:* the cumulative quantity of petroleum that has been recovered at a given date.
- *Reserves:* the quantities of petroleum anticipated to be commercially recoverable by the application of projects to known accumulations under defined conditions. Reserves must be discovered, recoverable, commercial and remaining and may be further categorised in accordance with the level of certainty associated with the estimates (Thompson, 2008). Proved reserves (1P) are estimated to have a 90% probability of profitable extraction, based upon assumptions about cost, geology, technology and future oil prices. Proved and probable (2P) reserves include additional volumes that are thought to exist in discovered accumulations but are estimated to have only a 50% probability of profitable extraction. Proved, probable and possible (3P) reserves include additional resources that are estimated to have only a 10% probability of being profitable.
- *Contingent resources:* those quantities of petroleum estimated, as of a given date, to be potentially recoverable from known accumulations, but where the applied projects are not yet considered mature enough for commercial development due to one or more contingencies. Contingent resources may include, for example, projects for which there are currently no viable markets or where commercial recovery is dependent upon technology under development. As with reserves, these are further categorised in accordance with the level of certainty associated with the estimates.[6]
- *Prospective resources*: those quantities of petroleum estimated, as a given date, to be potentially recoverable from undiscovered accumulations by application of future development projects. Prospective resources have both an associated chance of discovery and a chance of development. As with reserves, these are further categorised in accordance with the level of certainty associated with the estimates.
- *Unrecoverable*: that portion of discovered or undiscovered petroleum initially in place which is estimated, as a given date, to not be recoverable by future development projects. A portion of these quantities may become recoverable in the future as commercial circumstances change or technological developments occur.

[6] The scheme recognises that some ambiguity may exist between the definitions of contingent resources and unproved (2P and 3P) reserves. Contingent resources are not expected to be developed and placed into production within a 'reasonable' timeframe.

Figure 2.1 Resources classification in the Petroleum Resource Management System



The PRMS uses the term 'Estimated Ultimate Recovery' (EUR) instead of *URR* and clarifies that this is not a resource category in itself, but:

> "…a term that may be applied to any accumulation or group of accumulations (discovered or undiscovered) to define those quantities of petroleum estimated, as of a given date, to be potentially recoverable under defined technical and commercial conditions plus those quantities already produced." (WPC, 2007)

Several points are apparent from this definition. First, resource estimates require specification of the hydrocarbons covered, the classification scheme used, the timeframe for which the estimate is made and/or the associated technical and economic assumptions. It is frequently difficult to compare resource estimates owing to the lack of clarity over such issues (Andrews and Udall, 2003). Even where a single classification scheme is used, the associated estimates may be made using different technological and economic assumptions, which may not be stated explicitly.

Second, all resource estimates, including estimates of *URR* are inherently *uncertain* - although the degree of uncertainty should decline as exploration and production proceeds. Unfortunately, many estimates of *URR* are 'deterministic', in that they present a single point, or 'best guess' estimate of likely outcomes (Rogner, 1997). Such estimates are potentially misleading, since they fail to capture or express the possible range of outcomes (which is likely to be greater for contingent and prospective resources than for reserves). Also, the underlying assumptions may not be reported and it may not be clear whether the 'best guess' represents the mean, median or mode value of a range (NPC, 2007). As discussed in Thompson (2008), a better approach is to present the full range of possible recoverable volumes, together with their estimated likelihood (i.e. a probability distribution).

Third, resource estimates are inherently *dynamic* since they depend upon the economic and technical conditions prevailing at the time the estimate is made, together with assumptions about how those conditions may change over a specified period of time into the future. Increasing prices will make marginal resources (including smaller field sizes)[7] profitable, as well as inducing technical improvements that reduce production costs and boost recovery factors. Increasing prices will also encourage exploration and the development of associated technologies that will help to identify and access prospective resources and allow more accurate assessments of their magnitude. Over time, resources will shift from one category to another and the degree of geological and economic uncertainty should fall. The visual representation in Figure 2.1 could therefore be misleading, since the relative size of each category will vary widely, both over time and from one region to another. For example, in mature regions such as the United States cumulative production and identified reserves should be much larger than contingent and prospective resources, while the opposite should be the case for relatively unexplored regions.

2.3 Levels of aggregation for estimates of ultimately recoverable resources

All estimates of ultimately recoverable resources require specification of the geographical and geological *level of aggregation* to which they apply. The relevant level may be defined through geological, political or economic considerations or a combination of the three. It may range from individual reservoirs to the entire world. Box 2.2 defines some of terms used by geologists and the oil industry for classifying the appropriate level of aggregation of petroleum resource estimates, while Box 2.3 provides some relevant background on the geological formation of petroleum. Different countries and institutions have slightly different definitions of the terms in Box 2.2 and both the definitions themselves and the relative use of these different levels of aggregation has changed over time. For example, Sleipner in Norway is classified as one oil field, but a comparable geological structure in the UK continental shelf (UKCS) would probably be classified as four separate fields (Rosing and Odell, 1984). Similarly, smaller fields that were previously classified as separate in US records have subsequently been merged into larger fields as exploration progressed (Drew, 1997). Inconsistencies such as these can greatly complicate the analysis and interpretation of the relevant data.

[7] As technology improves, extraction costs fall and oil prices increase, it will become economic to recover oil from smaller fields. However, this process will be limited by the 'energy return on investment' (EROI) (Cleveland, 1992a). Since oil exploration and production is necessarily associated with energy consumption, at some point more energy will be required to extract the resource than is obtained from it. Whether or not this coincides with the economic limit on minimum field size will depend upon the relative price of the different energy carriers involved, together with associated economic factors such as the availability of investment subsidies.

Box 2.2 Geological levels of aggregation in petroleum resource assessment

- *Petroleum Well*: A well may be are drilled to find, delineate and produce petroleum, with some wells being drilled to inject fluids to enhance the productivity of other wells. The *URR* of a producing well is typically calculated by extrapolation of its past production performance, using standard formulae for 'decline curves' (Chaudhry, 2003)
- *Petroleum Reservoir/Pool*: A reservoir is a subsurface accumulation of oil and/or gas whether discovered or not, which is physically separated from other reservoirs and which has a single natural pressure system. *Pool* is an older term for reservoir and *accumulation* is an alternative term.
- *Petroleum Field*: A field is an area consisting of a single reservoir or multiple reservoirs of oil and gas, all related to a single geological structure and/or stratigraphic feature. Individual reservoirs in a single field may be separated vertically by impervious strata or laterally by local geological barriers. When projected to the surface, the reservoirs within the field can form an approximately contiguous area that may be circumscribed. However, other sources define a field simply as a contiguous geographic area within which wells produce oil or gas. In either case, the boundary of a field may shift over time and two or more individual fields may merge into one larger field (Drew, 1997). Oil fields are classified on the basis of their oil to gas ratio and may either be discovered (located by exploratory drilling), under development, producing or abandoned. The number of wells in a producing field may range from one to thousands.
- *Petroleum Prospect*: A prospect is a geological anomaly that has some positive probability of containing reservoirs of recoverable hydrocarbon and is considered to be a suitable target for exploration. This generally requires a sufficiently high probability that all four elements of petroleum formation, namely source rock, migration pathway, reservoir rock and viable trap, are likely to exist (Box 2.3). The boundaries of a prospect may also be influenced by legal and economic considerations, such as the availability of leases for exploration.
- *Petroleum Play*: A play is an area for petroleum exploration, containing a collection of oil prospects which share certain common geological attributes and lie within some well-defined geographic boundary. The specific geological attributes may vary from one play to another and may refer to geologic time intervals, rock types, structures or some combination thereof. Plays have varying levels of exploration maturity with the term 'conceptual play' referring to a region where no discoveries have been made.
- *Petroleum Basin*: A basin is a single area of subsidence which filled up with either sedimentary or volcanic rocks and which is known or expected to contain hydrocarbons. Since subsidence is slow and filling is continuous, there may be little surface depression, even when the 'basin' contains many kilometres of accumulated fill. Sedimentary basins are the primary source of petroleum, as a result of organic carbon getting progressively buried, heated and compressed.
- *Petroleum System*: A petroleum system is "….the essential elements and processes as well as all genetically related hydrocarbons that occur in petroleum accumulations whose provenance is a single pod of active source rock" (Magoon and Sanchez, 1995). A petroleum system therefore includes the source rock, migration pathway, reservoir rock and trap (Box 2.3). The components and timing relationships are typically displayed in a chart with geologic time along the horizontal axis and the system elements along the vertical axis. The concept was first introduced by Dow (1972) and now forms the basis of the resource assessments conducted by the USGS.
- *Petroleum Assessment Unit* An assessment unit (AU) is a volume of rock within a petroleum system that is sufficiently homogeneous, both in terms of geology, exploration considerations, accessibility and risk to be examined using a particular resource assessment methodology. For example, fields within a AU should form a sufficiently homogeneous population for historical

extrapolation methods to be reliable. An AU may coincide with a single petroleum system, or the latter may be broken down into several AUs.

- *Petroleum Province*: A province is an area with common geological properties relevant to petroleum formation. Adjacent provinces might have the same original rocks, but be considered separate because they have quite different histories. A province may contain a single petroleum basin or petroleum system or several similar basins/systems. A province typically has an area of several hundred square kilometres and is largest entity defined solely on the basis of geological considerations that is relevant for resource assessment. Globally, the USGS (2000a) identifies 937 provinces, 406 of which are known to contain petroleum. In 1995, 76 provinces were estimated to account for 95% of discovered resources.

Sources: Energy Information Administration (1990); Klett (2004); Magoon and Sanchez (1995)

Box 2.3 The geological formation of petroleum resources

Most petroleum is formed from the remains of marine plankton and algae which settled along with sediments to a sea or lake bottom to form *source rock.* After burial, the combination of heat, pressure and the absence of oxygen leads to chemical reactions which convert the hydrocarbons first into *kerogen* which is found in various *oil shales* around the world and then into oil and natural gas. The term *oil window* refers to a temperature range, below which the hydrocarbons remain in the form of kerogen and above which the oil is converted into gas. This temperature range is found at different depths throughout the world, but typically lies in the range of 4 to 6km.

The chemical reactions responsible for all formation involve expansion, which leads to the fracturing of rocks and migration of the oil to areas of lower pressure. The oil either escapes to the surface or accumulates in porous and permeable *reservoir rock* such as sandstone and limestone that are capable of storing the oil in its pore spaces. High-quality reservoir rocks have high permeability and porosity as a result of the pore space taking up a large percentage of the overall volume, while low quality reservoir rocks have the opposite. High permeability facilitates the movement of oil through the rocks to the producing well, thereby lowering costs and improving productivity. The degree of porosity may vary throughout the reservoir, leading to isolated pockets of oil.

For oil and gas to accumulate and remain, the reservoir rocks need to be sealed by a less porous and largely impermeable rock known as a *trap*. To persist over millions of years, the trap needs relatively unaffected by geophysical changes that could introduce fractures. Typical traps include anticlines, faults and salt domes.

Timing is crucial in oil formation. First, the reservoir must be deposited prior to oil migrating from the source rock; second, the trap must be in place prior to oil migrating; and third, the source rock must be exposed to the appropriate temperature and pressure for a sufficiently long period of time. This combination of conditions is relatively rare, with the result that oil and gas is only found in a few sedimentary basins around the world. Much oil has escaped over geological time, although in some areas (e.g. Alberta), heavy residues remain near the surface and can be mined.

Estimates of *URR* may be derived for any of the levels of aggregation indicated in Box 2.2, but different techniques (or combinations of techniques) may be more or less suitable for each. More aggregate estimates may be derived by summing estimates developed at a lower level, but such estimates need to be summed probabilistically (e.g. via a Monte Carlo simulation) rather than arithmetically and the frequent failure to do this can lead to misleading results (Pike, 2006; Thompson, 2008). Aggregate estimates may also be derived by the extrapolation of discovery and production trends for the aggregate region. While this approach is generally simpler and has fewer data requirements, it may also be less accurate.

While the highest 'geological' level of aggregation is the petroleum province, resource assessments are frequently conducted at the country or regional level. Such geographical boundaries may encompass several distinct petroleum provinces, basins and/or systems, and portions of these may extend into neighbouring countries or regions. The lack of geological homogeneity within such boundaries can lead to difficulties in country/regional level resource assessment, especially when only aggregate data is used (Charpentier, 2003).

At the same time, geological homogeneity is not the only consideration for developing valid resource estimates. If historical data on field discovery is to be used, the stability or homogeneity of the exploration and discovery process must also be considered. For example, a petroleum basin that is shared between two neighbouring countries is unlikely to have a consistent exploration history. But even where a region is located within a single country or jurisdiction, its exploration history can be greatly complicated by economic, political and institutional factors, such as the legal procedures associated with leasing areas for exploration. As Harbaugh *et al* (1995) note: "…the orderly development of plays and prospects is an ideal that is seldom achieved in practice. Exploratory wells may not be part of established plays, and even may be drilled with little or no geological information at locations where there is no perceivable prospect." Hence, the most appropriate level of aggregation for resource assessments is likely to vary from one region to another. The USGS concept of an appropriate *assessment unit* is designed to reflect these different considerations.

2.4 Cumulative discoveries and reserve growth

Cumulative production (Q_t) represents the total amount of oil that has been produced from a region since production began, while reported reserves (R_t) represent the estimated volume of remaining resources at known fields. As indicated above, reserve estimates are normally categorised in accordance with the level of certainty associated with the estimates. So *proved* (1P) reserves are estimated to have a 90% probability of profitable extraction, while *proved and probable* (2P) reserves are estimated to have only a 50% probability of profitable extraction. The sum of cumulative production (Q_t) and reported reserves (R_t) for a region at a particular point in time (t) may be referred to as *cumulative discoveries* (D_t).

$$D_t = Q_t + R_t \tag{2.1}$$

The cumulative discoveries represent all the oil that is known to a given level of confidence to have been discovered in that region. The appropriate interpretation of these estimates will depend upon the particular definition of reserves that is being used (e.g. R^{1P} or R^{2P}). For example, *cumulative 2P discoveries* (D_t^{2P}) May be expected to be larger than *cumulative 1P discoveries* (D_t^{2P}). While cumulative discoveries at a particular point in time could be taken as an estimate of the ultimately recoverable resources (*URR*) for that region, there are two reasons why this is likely to be an underestimate:

- *New discoveries*: New fields will be discovered and subsequently brought into production, thereby adding to cumulative discoveries in a region. In terms of the PRMS (Figure 2.1), this may be interpreted as the conversion of prospective resources into reserves/production.
- *Reserve Growth*: Estimates of cumulative discoveries from <u>known</u> fields will also tend to increase as a result of improved recovery factors, the physical expansion of fields, the discovery of new reservoirs within fields, the re-evaluation of cumulative discovery

estimates in the light of production experience, and other factors (Drew and Schuenemeyer, 1992; Gautier, *et al.*, 2005; Klett and Gautier, 2005; Klett and Schmoker, 2003; Morehouse, 1997). In terms of the PRMS classification, this may be interpreted as the exploitation of more uncertain reserves (2P and 3P) together with the conversion of contingent resources into reserves/production.

The second process is generally referred to as *reserve growth* or 'field growth' or 'recovery growth'). However, more accurate terms would be 'cumulative discovery growth' or 'ultimate recovery growth', since it is the estimates of URR of known fields that are growing rather than the reported reserves for those fields. As a result of reserve growth, the cumulative discovery estimates for individual fields will typically increase over time. For example, a study by Barker, *et al.* (2004) of 99 Canadian oilfields found that estimates of field size (i.e. *URR*) had grown by 97% since the fields were discovered. In turn, this means that the cumulative discovery estimates for a region will also increase over time, even if no new fields are found (Attansi and Root, 1994; Drew, 1997; Muller and Sturm, 2000; Odell, 1973a; Root and Mast, 1993; Verma, 2005).

Reserve growth has been most closely studied in the United States, where it accounted for 89% of the additions to US proved reserves over the period 1978 to 1990 (Attanasi and Root, 1994). While reserve growth also occurs in the other regions of the world, the evidence base is much thinner.[8] However, despite being systematically investigated more than 40 years ago (Arrington, 1960), the phenomenon of reserve growth was relatively neglected before the 1980s (Drew, 1997). An important stimulus to further investigation was the retrospective examination of petroleum discovery forecasts for the US, which were found to have systematically underestimated future discoveries (Drew and Schuenemeyer, 1992). The chosen forecasting methodology relied upon estimates of the size of known fields but failed to adjust these estimates to allow for future reserve growth. Since these fields subsequently doubled in size within less than 10 years, the volume of new discoveries was also underestimated. The USGS World Petroleum Assessment 2000 was the first to systematically incorporate reserve growth into a global assessment of petroleum resources – a move which has generated some controversy (see Section 6). Subsequent evaluation of this forecast suggested that between 1995 and 2003, reserve growth exceeded new-field discoveries as a source of global additions to (2P) reserve estimates by the ratio of three to one (Klett, *et al.*, 2005b).

Future reserve growth may be estimated by analysing the historical growth in the estimated size of individual fields (Root and Mast, 1993). A *growth function* ($G(\tau)$) may be estimated, representing the ratio of the estimated size of a field τ years after it was discovered[9] to the estimated size of the field at the time of discovery. Typically, growth functions are assumed to be independent of the time of discovery of the field. Both annual and cumulative growth functions can be calculated and used to convert current estimates of cumulative discoveries into future estimates for a specified year, with the amount of growth depending upon the age (and sometimes size) of the field. In some cases, reserve growth functions are measured with respect to the date of first *production* from the field, rather than the date of discovery (Thompson, *et al.*, 2009b). This is because a discovered field may 'lie fallow' for many years

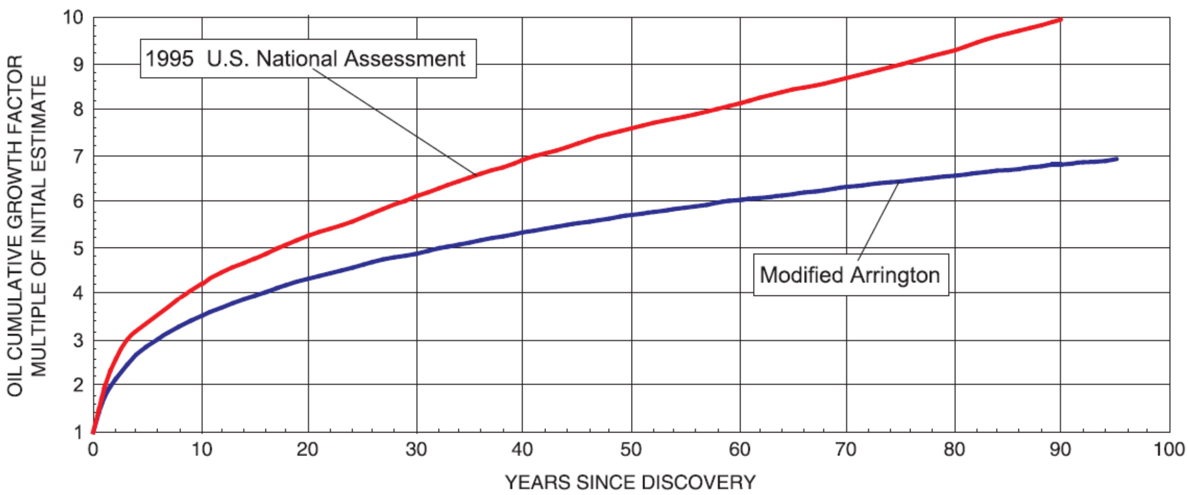
[8] Relevant references include Klett (2005), Klett and Gautier (2005), Gautier and Klett (2005), Gautier, *et al* (2005), Klett and Schmoker (2003), Klett and Verma (2004), Verma (2000; 2003; 2005), Verma and Ulmishek (2003), Verma, *et al.* (2004; 2001), Watkins (2002), Sem and Ellerman (1999) and Odell (1973b)

[9]

prior to first production, with most of the development work that contributes to reserve growth occurring after production has begun.

Figure 2.2 shows two examples of cumulative growth functions estimated for onshore US oil fields: namely the 'Modified Arrington' function developed by Verma (2005) and an earlier function developed by Attanasi and Root (1994) (which was subsequently found to have overestimated reserve growth). Note that the rate of growth is largest in the first ten years after discovery and that the current cumulative discovery estimates for 80-year old fields (based upon 1P reserves) are nearly seven times larger than those made at the date of discovery - and are still growing. While we would expect the annual additions to cumulative discovery estimates to decline over time as more of the uncertain/contingent resources are converted into proved reserves or cumulative production (i.e. $G'(\tau) \rightarrow 0$ as $\tau \rightarrow \infty$), this is not evident from the estimated function. The onshore United States contained fields of widely different ages and Verma (2005) estimates the average growth rate of these fields to be 2.7%/year.

Figure 2.2 Illustration of cumulative reserve growth



Source: Verma (2005)

Reserve growth may partly be a consequence of conservative reserve reporting. For example, Laherrère (1999a) has argued that the bulk of US reserve growth can be attributed to the reporting of only proved (1P) reserves under SEC rules. This is acknowledged to be a highly conservative estimate of future production (Thompson, 2008). As Drew (1997) observes:

> "……Ask a manager in an oil company what the reserves are and he or she will tell you that it depends on who is asking. The manager will also tell you that three sets of books are kept - one that has the optimistic estimates that are used to sell deals to upper management and the stockholders and to give the geologists are good measure for what they have found; another has the conservative estimates that the accountants used to borrow money from the banks; and a third set has the middling numbers calculated by the engineers for internal use in the company."(Drew, 1997)

These numbers may be very roughly interpreted as 3P 1P and 2P reserves estimates respectively. However, Drew (1997) argues that the phenomenon of reserve growth is not confined to cumulative discovery estimates based on 1P reserves:

"…The irony of this summation is that all three sets of numbers are pessimistic - they all grow with the passage of time." (Drew, 1997)

We would expect, nevertheless, that reserve growth would be smaller for cumulative discovery estimates based upon 2P reserves than for those based upon 1P reserves since the former are less conservative. Indeed, using the probabilistic interpretation of 2P reserves (Thompson, 2008), we would expect cumulative discovery estimates based upon 2P estimates to be downgraded as frequently as they are upgraded. However, analysis by Klett and Schmoker (2003) suggests that this is not the case. Klett and Schmoker analysed the reserve growth of 186 giant oil fields located outside of the US and Canada using data from successive editions of the IHS database (Klett, 2005). The cumulative discovery estimates reported in this database represent the sum of cumulative production and 2P reserve estimates. In contrast, the cumulative discovery estimates reported for US fields and used by Verma (2005) and others to develop cumulative growth functions represent the sum of cumulative production and 1P reserve estimates. But despite this important difference, the percentage reserve growth observed in the giant fields between 1981 and 1995 was very close to that predicted from growth functions estimated from US fields. The applicability of these growth functions was further reinforced by Klett *et al*'s (2005a) evaluation of the USGS World Petroleum Assessment which showed that reserve growth over the period 1995 to 2003 was consistent with expectations.

Reserve growth is therefore of major importance for any method of estimating URR that relies upon cumulative discovery data (i.e. *D* or *B*), as well as for future projections of global oil supply. Put simply, if reserve growth is underestimated or overlooked, estimates of *URR* will be too conservative and forecasts of future oil supply will be too pessimistic (Nehring, 2006c; e). But the estimation of future reserve growth is fraught with difficulties. At present, relatively little is known about how patterns of reserve growth vary between different types and sizes of field (e.g. onshore versus offshore) or between different regions. Similarly, little is known about how those patterns are influenced by various institutional, economic and technical factors. Since advances in seismic technology allow resources to be estimated more accurately, it is possible the URR estimates for newly discovered fields may not grow as much as those for old fields. However, this has yet to be established with any confidence empirically. Since the analysis of reserve growth is hampered by lack of data and inconsistencies within the available datasets, the contribution of reserve growth to future oil supply remains a topic of controversy (Drew, 1997). Reserve growth is discussed further in a companion report (Thompson, *et al.*, 2009b) and is a recurring theme in much of what follows.

2.5 Field size distributions

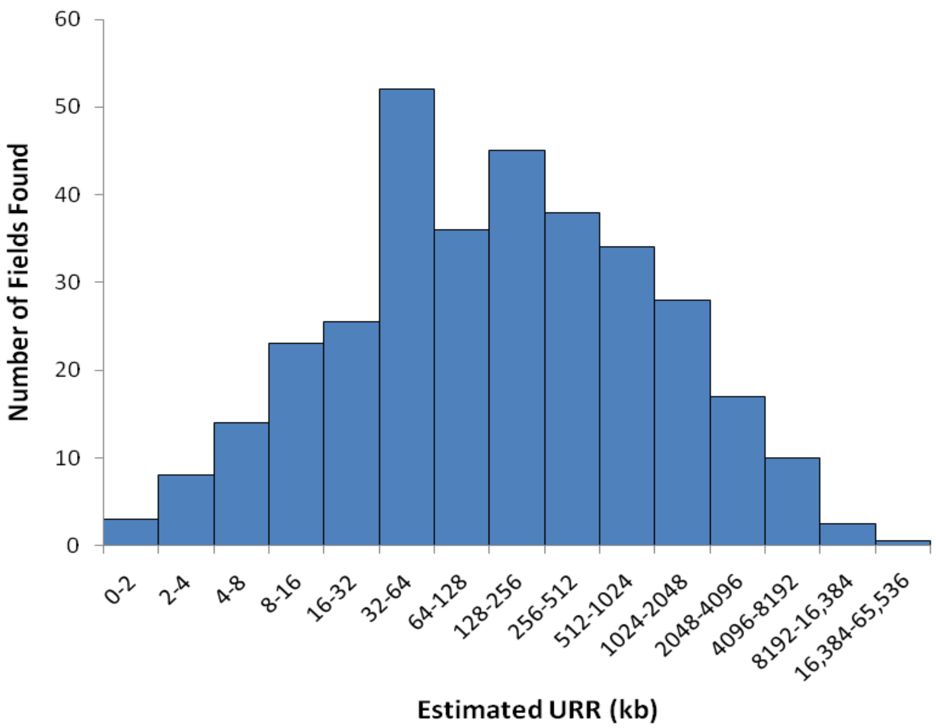
Many methods of resource assessment rely upon assumptions about the size distribution of oil fields within a region, where 'size' refers to current estimates of the ultimately recoverable resource from each field. Two fundamental observations are that:

- Within any given region (e.g. play, basin, system) most of the oil tends to be found in a small number of large fields.
- Large fields tend to be discovered early in the exploration process, with subsequent discoveries tending to be smaller and requiring increasingly greater effort to locate.

These observations are borne out by empirical observations of exploration histories and the size distribution of discovered fields at levels of aggregation ranging from single plays to the entire world. However, the precise form of the field (or reservoir) size distribution varies from one region to another (Laherrère, 2000a) and is a long-standing focus of controversy (Drew, 1997; Kaufman, 2005). For example, Klemme (1984) has highlighted empirical relationships between field size distributions and the morphology and size of petroleum basins. Depending upon the type of basin, the estimated proportion of *URR* contained within the five largest fields could vary from less than 10% to more than 75%.

The size distribution of the underlying population of fields cannot be directly observed, but can only be inferred from the size distribution of discovered fields (however estimated). Arps and Roberts (1958) were one of the first to observe that the latter typically took a *lognormal* form - in other words, the frequency distribution of the natural log of discovered field sizes resembled a normal distribution (Figure 2.3). The *mode* field size therefore occurred in the middle of this size range. This observation was subsequently supported by several empirical studies, including McCrossan's (1969) analysis of reservoir sizes in Western Canada[10] and studies of US data sets by Kaufman (1963) and Drew and Griffiths (1965).

Figure 2.3 Oil and gas field size distribution for the Denver basin in 1958



Source: Adapted from Arps and Roberts (1958) and Drew (1997).

[10] "…it is probably safe to assume on the basis of the present and other published work that a geologically homogeneous group of oil deposits should form a unimodal, lognormal size-frequency distribution" (Mcrossan, 1969). Note that Western Canada is relatively unique in that reserves have historically been reported on a reservoir rather than a field basis. This choice has some important implications. When reservoirs are used as the basis for assessment, the discovery of a new reservoir is counted as a new discovery, but where fields are used the discovery of new reservoirs in a field forms a component of reserve growth. McCrossan's results suggest that skewed size distributions apply as much to reservoirs within a field as to fields within a larger region.

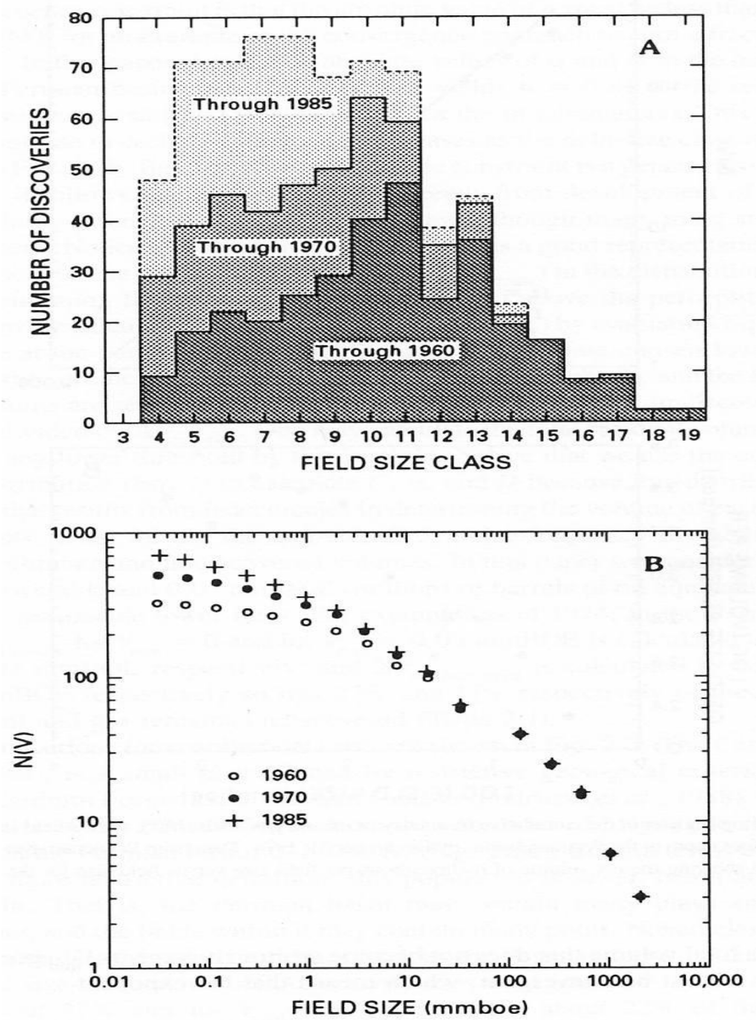
That oil and gas fields typically had a lognormal size distribution became established as conventional wisdom during the 1960s and 1970s and subsequently formed the basis of some highly sophisticated 'discovery process' models which were used to estimate the regional URR (Kaufman, 1975b). This was despite the difficulties in making statistical inferences about the size distribution of the population of fields from a relatively small sample of fields that may not be representative of the population as a whole (Bloomfield, *et al.*, 1979; Drew, 1997; Kaufman, 1993). These difficulties had been anticipated by Arps and Roberts (1958):

> "….the "frequency density distribution" strongly resembles the typical bell shaped normal Gaussian probability curve. Not too much significance should be attributed to its apparent symmetry, because obviously the physical reasons for right-hand and left-hand slopes of this curve are quite different. The diminishing number of fields with growing ultimates on the right-hand side of the mode are, as would be expected, the result of having fewer fields of larger size….The tapering off on the left-hand side of the mode, however, must be largely caused by economic factors. For instance, under 30,000 barrels of ultimate per well, it may be questionable whether an operator should run pipe at all, and many discoveries in this category which were found will probably never completed and therefore escape the statistics." (Arps and Roberts, 1958):

This sampling bias was rediscovered in the 1980s and given the name 'economic truncation' (Attanasi and Drew, 1984; Drew, 1997; Drew, *et al.*, 1988). Schuenemeyer and Drew (1983) observed that smaller fields tended to be underrepresented in the sample of discovered fields because they were not economic to develop. However, as exploration proceeded, technology improved, oil prices rose and/or costs fell, it became increasingly economic to find and exploit the smaller fields. This process is demonstrated graphically in Figure 2.4 (A), which shows the field size distribution for the Frio Standplain play in Texas as estimated in 1960, 1970 and 1985. The mode field size shifts progressively to the left as more small fields are discovered and developed. Importantly, the population of large fields remains largely unchanged.

The same process is illustrated in the cumulative frequency distribution of Figure 2.4 (B). Using log scales, this shows the number of fields (N) in the play that exceed a particular size (V) – sometimes referred to as the field *rank*. As more small fields are discovered and developed, the curvature of the $\ln N(V)$-$\ln V$ plot is reduced. Similar inferences may be drawn from cross-sectional studies that compare field size distributions in different regions, since the costs of developing petroleum resources - and hence the minimum viable field size - can vary widely from one region to another (Drew, *et al.*, 1982b). These differences are especially important when comparing onshore and offshore regions (Drew and Schuenemeyer, 1993).

Figure 2.4 Observed field size distribution for the Frio-Strandplain play in Texas at three different points in time



Source: Cramer Barton and La Pointe (1995)

Power (1992) has shown how the apparently lognormal distribution of observed field sizes may also arise from the second form of sampling bias - namely, that the largest fields tend to be discovered first. Power simulated the discovery process for theoretical populations of fields whose size distribution took a Weibull form. Importantly, the sampling process did not impose any economic truncation. Power found that, as the number of exploratory wells increased, the size distribution of discovered fields evolved towards the parent frequency size distribution. However, the size distribution of discovered fields could not be rejected as being non-lognormal for a wide range of measures of exploratory effort (i.e. number of wells drilled). He concluded that the discovery sequence could compound the sampling bias introduced by economic truncation, thereby potentially reinforcing the misleading conclusion that the population size distribution took a lognormal form.

On the basis of these and similar observations, Drew and colleagues proposed that the population field size distribution was more likely to take a power law form, as follows:

$$N(V) = AV^{-\alpha} \quad (2.2)$$

Where A is a scaling factor and the parameter α defines the shape of the distribution. This type of distribution is sometimes termed a 'Pareto distribution', after Pareto (1987) who represented income distribution in a similar way. It is also part of a family of distributions known as 'probabilistic fractals', which have their roots in the work of Mandlebrot (1977). Indeed, Mandlebrot (1962) was the first to propose that petroleum and mineral resources could be modelled with the Pareto distribution, demonstrating this by an analysis of both the surface area[11] and the ultimate recovery of US oil fields.

The Pareto distribution is also related to 'Zipf's law' which describes a relationship between the size and 'rank' (*N*) of discrete phenomena (Deffeyes, 2005; Merriam, *et al.*, 2004; Zipf, 1949). When oil fields are ranked in descending order of size so that the largest is rank 1, Zipf's law states that the product of the rank and size is approximately constant. Hence, for a field of rank *N*: $V \sim N^{-\beta}$ and $\beta \approx 1.0$. The applicability of Zipf's law is usually investigated by plotting field size as a function of rank, while the applicability of a Pareto distribution is usually investigated by plotting cumulative frequency (*N*(*V*)) as a function of field size. The two approaches are equivalent, since the phrase "the *N*th largest field has a URR of *V*" is equivalent to saying that "*N* fields have a URR equal to or greater than *V*" (Adamic and Huberman, 2002). But the Pareto distribution is more general in that it does not constrain the value of the exponent (α).

While Davies and Chang (1989) have criticised the Pareto model, it appears to be gaining increasing acceptance. A version of the Pareto distribution was first adopted by the USGS in 1989 as part of their regular assessment of US petroleum resources (Houghton, *et al.*, 1993; Mast, 1989). If the size distribution of the underlying population follows a Pareto law, a plot of the natural log of *N* against the natural log of field size (as in Figure 2.4 B) should approximate a straight line with slope equal to $-\alpha$:

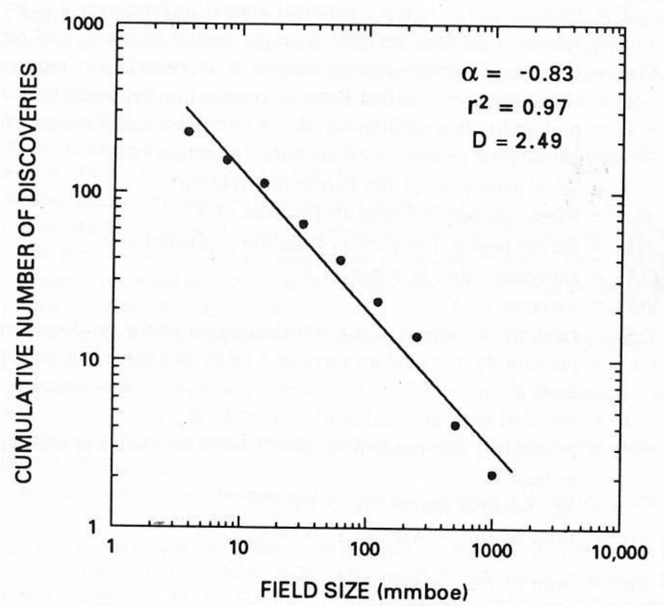
$$\ln N(V) = \ln A - \alpha \ln V \quad (2.3)$$

The observed curvature in the cumulative frequency distribution of discovered fields (Figure 2.4 B) may then be due to biased sampling from a population of fields with a Pareto size distribution. To illustrate this, Figure 2.5 uses the same 1985 data as Figure 2.3, but excludes the smaller fields whose economic viability is marginal. A straight line provides a relatively good fit, with an r^2 of 0.97. If the Pareto rule applies, the estimated value of α can provide a useful basis for comparing field size distributions in different regions.[12] As $\alpha \rightarrow 1.0$, the volume of oil in small fields forms an increasingly large share of *URR*.

[11] Surface area may be easier to estimate than *URR* and is expected to be correlated with it. For example, Arps and Roberts (1958) estimated that the *URR* of a field was approximately proportional to the 1.275 power of its surface area. However, the relationship will vary from one region to another and will depend in part upon how the surface area of a field is defined.

[12] For example, Coustau (1979) has used this to distinguish between concentrated, normal and dispersed petroleum basins.

Figure 2.5 Cumulative frequency plot of field sizes for the Frio Strandplain play in Texas, excluding smaller field sizes



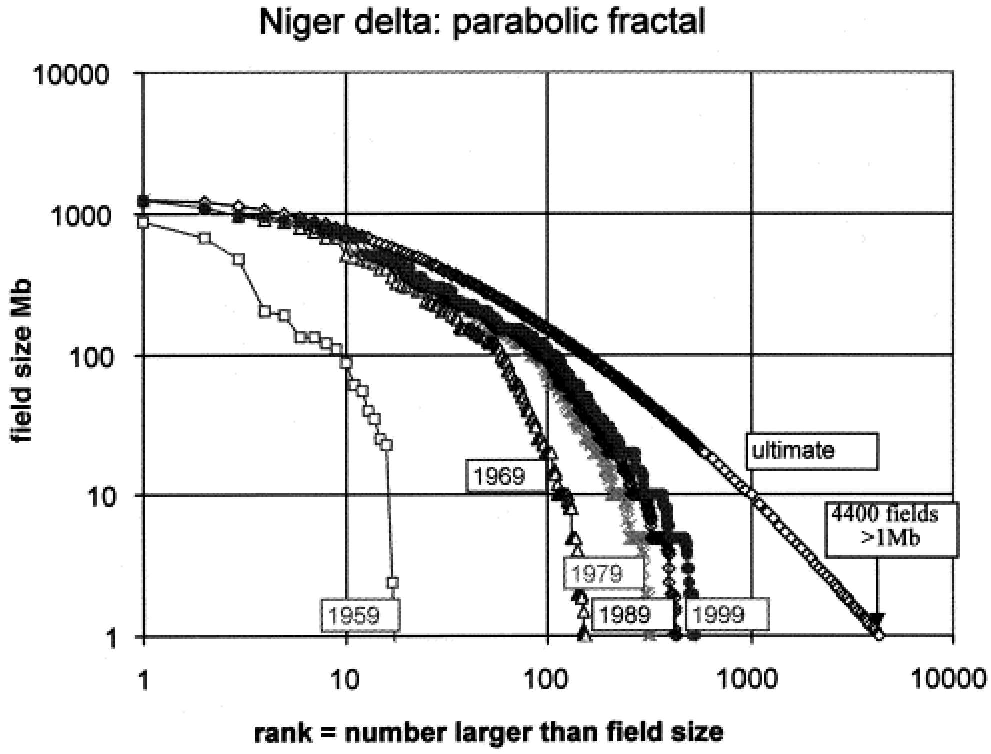
Source: Cramer Barton and La Pointe (1995)

However, Laherrère (1996; 2000a) has questioned whether the Pareto distribution, or 'linear fractal' is the most appropriate one to use:

> "….. A linear fractal implies perfect self-similarity, whereby one complete segment of the distribution describes the whole, but in nature, self similarity is imperfect and limited. Natural data gives rise to curved, not linear plots." (Laherrère, 2000a)

As a result, Laherrère advocates a 'parabolic fractal', which assumes a quadratic relationship between the log of field size and the log of the field rank. Figure 2.6 shows such a plot for the Niger delta, using field size data from different periods. While the larger fields exhibit reserve growth between 1959 and 1969, their estimated size has remained relatively stable since then. The lower part of the curve reveals the discovery of increasingly smaller fields, in a similar manner to Figure 2.4. Laherrère extrapolates the curve to estimate that there are around 4400 fields exceeding 1MB in size - although he acknowledges that there is a 'degree of latitude' in how the curve is drawn (note that this approach reverses the axes). Laherrère argues that the parabolic fractal works well in most cases, but problems can arise with the so-called 'King effect' where the largest field is very much larger than the rest. The 'King' field is an outlier to the distribution, but may contain the bulk of a region's oil.

Figure 2.6 Cumulative frequency plot of field sizes for the Niger delta



Source: Laherrère (2000a)

In summary, while the parabolic fractal (Figure 2.6) may represent the underlying population distribution of field sizes, it is equally possible that the underlying distribution takes a Pareto form, with the curvature resulting primarily from the sampling bias introduced by economic truncation and the discovery process. At present, there appears to be no consensus on this issue, which contributes to the uncertainty in URR estimates. Also, the relative suitability of different functional forms to represent the field size distribution may be expected to vary from one region to another

The proportion of the URR contained in smaller, undiscovered fields may be estimated by fitting one of these functions to the size distribution of discovered fields and extrapolating to smaller field sizes. The proportion will be greater with a power law distribution, smaller with a parabolic fractal and smaller still with a lognormal. For example, Barton and Scholz (1995) fitted a power law to six regions and estimated that undiscovered small fields contained between 9% and 31% of the regional URR (excluding fields smaller than 30 kb). Corresponding estimates are not available at the global level, but are likely to be sensitive to the minimum size threshold assumed.

While technical improvements and higher prices should make more small fields viable, there will always be a lower limit imposed by the energy return on investment. As a result, many small fields will never contribute to global supply, especially in offshore regions.

2.5.1 Why big fields matter

While the precise form of field size distributions remains a source of dispute, there is no question that the majority of oil is found in a small number of large fields. Since this is of profound importance for the future of global oil supply, it deserves a closer look.

One of the first global surveys of crude oil fields was by Ivanhoe and Leckie (1993) who grouped fields into ten size categories on the basis of their estimated URR (Table 2.1). The 370 fields with a URR exceeding 0.5 Gb (i.e. >7 days current global supply of crude oil) represented less than 1% of the total number of fields but accounted for three quarters of cumulative discoveries. Of particular importance were the 42 ‘super-giant’ fields with a URR exceeding 5 Gb (i.e. >73 days current global supply) with the largest (Ghawar in Saudi Arabia) having a URR of ~140 Gb. The 1300 fields with a URR exceeding 0.1 Gb (i.e. >1.5 days current global supply) represented only 3% of the total but accounted for 94% of cumulative discoveries. The remaining 39000 fields accounted for less than 6% of the total and individually contributed only a tiny fraction of global supply.

Table 2.1: Ivanhoe and Leckie’s estimates of the size distribution of the world’s oilfields

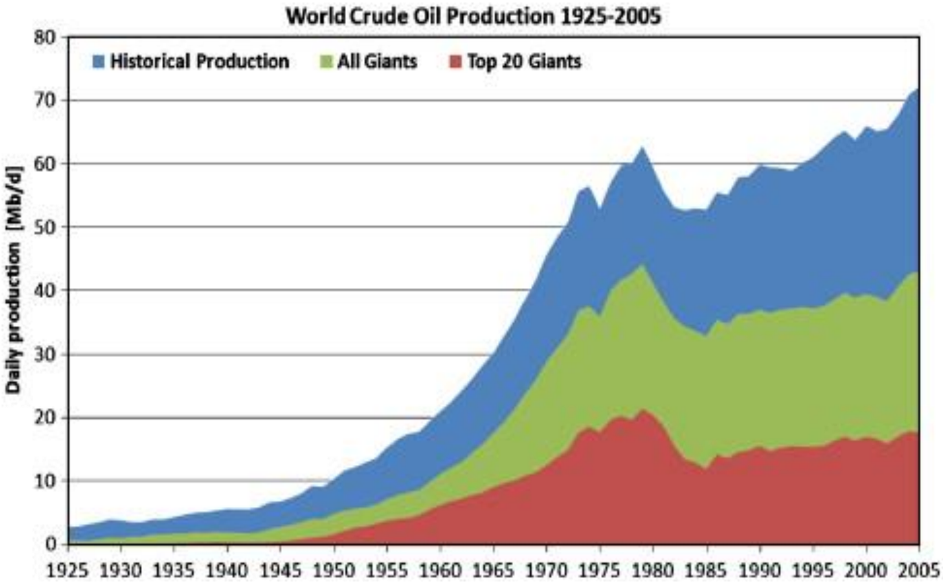
Category	**Estimated URR (mb)**	**No. in world**
Megagiant	>50,000	2
Supergiant	5000-50,000	40
Giant	500-5000	328
Major	100-500	961
Large	50-100	895
Medium	25-50	1109
Small	10-25	2128
Very small	1-10	7112
Tiny	0.1-1	10849
Insignificant	< 0.1	17740
Total		**41164**

Source: Ivanhoe and Leckie (1993)

Similar results were obtained by Robelius (2007), who provided an updated analysis of the world's ‘giant’ oilfields using data from a variety of sources. Robelius estimates that there are ~47500 oil fields in the world, 73% of which are in the United States.[13] Only 507 (<1%) of these are ‘giants’, with an estimated URR of more than 0.5 Gb, of which 430 are in production and 17 under development. Robelius estimates that the 100 largest fields account for 45% of the global production of crude oil (see Figure 2.7) while the giants as a whole account for approximately two thirds of global cumulative discoveries. Half of these giants were discovered more than 50 years ago.

[13] These figures demonstrate that much more exploration has taken place in the US compared to other regions of the world. This has implications for the relative suitability of different resource assessment methodologies (Section 4) and suggests there is considerable unexploited potential outside the US.

Figure 2.7 The estimated contribution of giant oilfields to global crude oil production



Source: Robelius (2007)

Simmons (2002) defines giant fields as those producing more than 100 kb/d (i.e. 0.14% current global supply of crude oil).[14] He estimates that there are 116 giants under this definition which in 2002 accounted for approximately half the global production of crude oil. The smallest 62 of these fields accounted for only 12% of production while the largest 14 accounted for over 20%. In 2007, the average age of the 14 largest fields was 51 years and of the 26 giants discovered since 1980, only four produce more than 200 kb/day.

The most up-to-date estimates are provided by the IEA (2008) who use Ivanhoe's classification system and rely largely upon the IHS database. They estimate that 70,000 oil fields were in production in 2007, but around 60% of crude oil production derived from 374 fields (54 supergiant and 320 giant). An additional 84 giant fields were either under development or 'fallow'. Approximately half of global production derived from only 110 fields, 25% from only 20 fields and as much as 20% from only 10 fields, with Ghawar accounting for a full 7% (**Error! Reference source not found.**). Most of the 20 largest fields have been in production for several decades and 16 of them are past their peak of production. The world's second-largest oil field, Canterell, peaked in 2003 and its' production has since declined by ~70%.

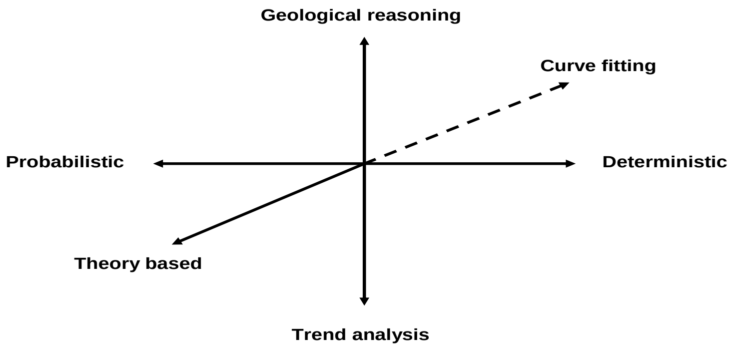
Hence, while the precise numbers may be uncertain, it is clear that around 100 oil fields account for up to half of the global production of crude oil, while up to 500 fields account for two thirds of cumulative discoveries. Most of these fields are relatively old, many are well past their peak of production and most of the rest will begin to decline within the next decade or so. The remaining reserves at these fields, their future production profile and the potential for reserve growth is therefore of critical importance for future global supply.

[14] Höök, *et al.* (2009) estimate there are 20 giant fields under Simmons' definition which are not giant fields under Ivanhoe's definition.

2.6 Methods of estimating ultimately recoverable resources

There are a wide range of methods for estimating ultimately recoverable resources, together with many variations on the basic techniques. The appropriate choice depends upon the nature and level of aggregation of the region under study, the degree of exploration maturity and the data and human resources that are available. Since different authors classify the methodologies in different ways, it may be more appropriate to think of a spectrum of possibilities, involving differing reliance upon: a) geological data and reasoning versus extrapolation of historical trends; b) probabilistic estimates (e.g. Monte Carlo simulation) versus single-value estimates; and c) theoretical reasoning versus simple curve-fitting. This is illustrated in Figure 2.7.

Figure 2.8 Classification of methods of estimating URR



Most of the methods associated with Hubbert may be characterised as producing single value estimates from the extrapolation of curves fitted to historic data on cumulative discovery or cumulative production for aggregate regions such as an oil-producing country. There is relatively little use of geological judgement or information and the methods are simple to apply with data that is often available in the public domain. In contrast, the methods used by the USGS and others produce probabilistic estimates from geological assessments of disaggregate regions, with extensive use of geological judgement and information (USGS, 2000a). These methods are complex and resource intensive and rely upon extensive data sources that are often inaccessible to third parties. This characterisation is oversimplified, however, as there are considerable overlaps between the two approaches, especially for regions that are relatively mature stage of exploration and production (Drew and Schuenemeyer, 1993).

All of the different methods have their strengths and weaknesses and while some methods may be more or less suitable for particular levels of aggregation and stages of exploratory effort, there is unlikely to be a single 'best' method (Divi, 2004). Indeed, the most reliable assessments are likely to be derived from a combination of methods (Ahlbrandt and Klett, 2005). In what follows, the various methods are classified into four groups, namely: a) geological assessments; b) expert assessments; c) field size distribution approaches; and d) historical extrapolation (Charpentier, *et al.*, 1995b). The first two of these rely more upon geological information and judgement and are more appropriate for less explored regions, while the second two rely more upon the extrapolation of historical trends and are more appropriate to well explored regions. The methods associated with Hubbert fall into the last category and are the primary focus of this report.

2.6.1 Geological assessment

These methods primarily rely upon geological analysis of seismic and other data to estimate the resource size. A traditional approach, commonly applied at the basin level, is to estimate hydrocarbon volumes by multiplying the estimated sedimentary volume by an estimated yield in barrels per cubic kilometre (Gautier, 2004; Weeks, 1952; White and Gehman, 1979). For unexplored areas, the values for such calculations are typically based upon measurements or estimates from geologically similar regions (analogs) where more information is available. The accuracy of such calculations depends in part on the suitability of the choice of analog(s),[15] but given the complexity of the geological determinants, there are considerable uncertainties in estimating yield factors even when the geology is relatively well-known. With typical yield factors ranging from 0.01 to 2.0 MBO per cubic mile (Divi, 2004), there is considerable scope for error:

> "….Unfortunately, none of the wide variety of technological approaches to prediction of petroleum resources is trustworthy because each basin is unique.....Not even a geological similarity of 99% between basins is enough to guarantee any similarity in petroleum resources. One crucial difference in geological parameters can completely negate the effect of all the similarities." (Jones, 1975)

More modern approaches are commonly applied at lower levels of aggregation, such as the petroleum play. These typically use Monte Carlo methods to multiply either point estimates or probabilistic distributions of factors such as the volume of the reservoir rocks, the ratio between this and the total volume of the sediments, the pore volume of the rocks, their average porosity and oil saturation and the average recovery efficiency (Capen, 1976). Many of these factors will be highly uncertain and since they are unlikely to be independent of each other, the multiplication of probability distributions may be inappropriate. The resulting estimates are typically adjusted downwards by 'risking' procedures, designed to weigh the likelihood that the relevant geological conditions (including source, migration, reservoir, trap and timing) were sufficiently favourable to generate at least one reservoir larger than the minimum field size. As White and Gehman (1979) note: "….the geological and economic uncertainties inherent in these questions can be awesome". Even when exploration is relatively mature, there may still be only a limited number of measurements of a subset of the relevant parameters. Furthermore, there may be no universally accepted causal relationship between these variables and the size of the petroleum resource (La Pointe, 1995). Such uncertainties led Hubbert (1981) to question the accuracy of such approaches:

> "…. it is easy to show that no geological information exists, other than that provided by drilling, that will permit an estimate to be made of the recoverable oil obtainable from a primary area that has a range of uncertainty of less than several orders of magnitude." (Hubbert, 1982)

As exploration proceeds in the majority of the world's petroleum producing regions, such methods become less suitable and their limitations less relevant. However, in areas that

---

[15] This may be relatively sophisticated. For example, Gess and Bois (1977) use cluster analysis to find that play most like the one being assessed. They describe the plays by 153 parameters and 106 quality judgments transformed into numbers!

remain to be drilled, such as most of the Arctic region, resource estimates must necessarily have large error bounds.

2.6.2 Expert assessment

These methods use formal procedures to combine the expert judgment of several geologists regarding the probability distribution of potential resources in an area (Baxter, *et al.*, 1978; Gautier, 2004). Typically, each geologist first reviews the relevant information and then estimates either a single value or a probability distribution for each of the relevant factors. These estimates can then be statistically combined into a probability distribution that reflects the full range of opinions of the group (Gautier, 2004). In a variant of this approach, the group as a whole reviews all the individual results and makes revisions where appropriate. The group may aim for a consensus estimate, or the final outcome may represent the average of the individual opinions (White and Gehman, 1979).

The advantage of this method is that it is straightforward and probabilistic and can accommodate special situations such as exploration constraints that may be poorly handled by more 'mechanical' methods (Charpentier, *et al.*, 1995b). It is also appropriate for all levels of geological aggregation and data availability. The disadvantage is that it lacks transparency to third parties and relies rather heavily on the knowledge and objectivity of the individual assessors: "…. one must know how expert are the experts in order to assess the assessment." (White and Gehman, 1979) Also, the social processes that govern such assessments could potentially lead to bias.

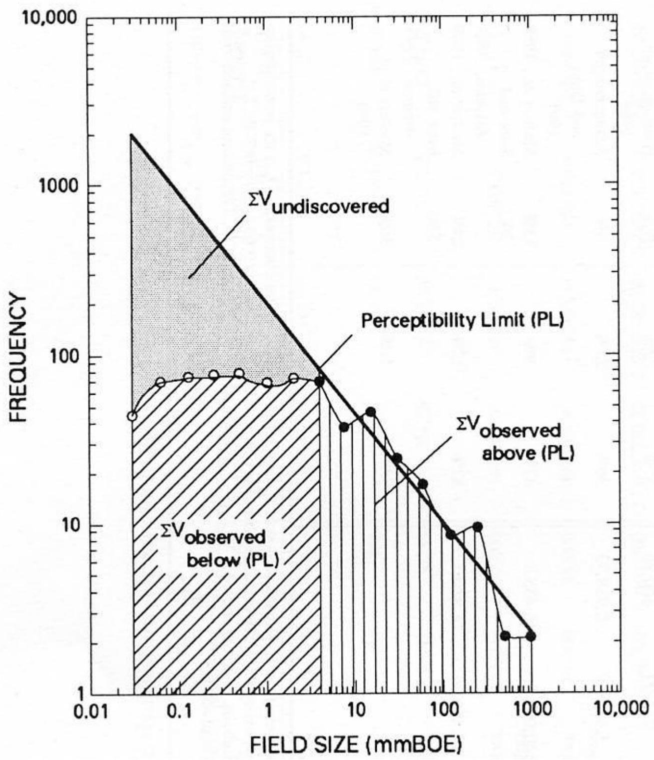
2.6.3 Field-size distributions

For explored regions, estimates of URR may be derived by combining data on the size of discovered fields with assumptions about the size distribution of the underlying population of fields. For example, if a Pareto distribution is assumed, undiscovered resources may be estimated by plotting a cumulative frequency distribution, fitting a linear regression and extrapolating this to smaller field sizes. The URR may then be approximated by the area under the curve (see Figure 2.8).[16] This estimate is sensitive to the point at which the observed size distribution is curtailed when fitting the curve, as well as to assumptions about the minimum viable field size and the future reserve growth in discovered fields. It is also sensitive to the size distribution assumed. For example, the resources contained in small fields may be estimated to be larger if a Pareto distribution is assumed than if a parabolic fractal or lognormal distribution is assumed. Such assumptions will also influence the expected size of resources contained in undiscovered large fields, which could have a major influence on the results (Attanasi and Charpentier, 2002).[17]

[16] Barton and Scholz (1995) use analyses of this type to estimate that around 9% of conventional recoverable oil remain undiscovered in the lower 48 US states and 31% remained undiscovered globally. Both of these estimates neglect reserve growth.

[17] Attanasi and Charpentier (2002) compared oil and gas resource assessments made using Parto and lognormal assumptions for the field size distribution. The use of the lognormal distribution reduced the oil estimates by 16% and the gas estimates by 15%. Nearly all of the difference resulted from the lognormal distribution having fewer larger fields relative to the Pareto distribution.

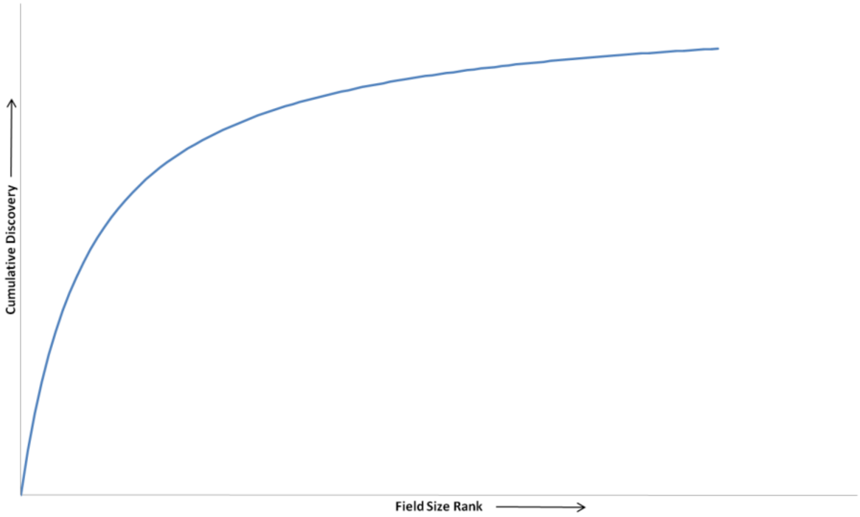
Figure 2.9 Estimating URR from a cumulative field size distribution that is assumed to follow a Pareto law



Source: Barton and Scholz (1995)

An alternative approach plots cumulative discoveries (i.e. the sum of the estimated URR of discovered fields) as a function of the *rank* of the field (where the largest field is rank 1). The resulting curve should flatten out as rank increases (Figure 2.9) as a result of the declining size of fields. If exploration is sufficiently advanced, the curve should trend towards an asymptope which can be taken as an estimate of the regional URR. The asymptope may be identified visually or estimated through non-linear regression. This approach has much in common with the discovery projection technique (described in Section 3) which plots cumulative discoveries as a function of time and estimates URR in a similar manner. Indeed, if the discovery rate was constant and if fields were discovered precisely in declining order of size, then the field-rank and discovery projection curves would be identical. However, unlike the field-rank technique, discovery projection does not require information on the size of each individual field.

Figure 2.10 Estimating URR by plotting cumulative discoveries as a function of field rank



Estimates based upon field size distributions are only suitable in areas that: first, have information available on the estimated size of each field; and second, are 'sufficiently explored' to have a statistically significant sample of field size estimates. The meaning of 'sufficiently explored' is ambiguous however, as Power (1992) has shown that incorrect size distributions can be estimated even after hundreds of fields have been discovered. While larger sample sizes may be obtained from larger regions, this would be inappropriate if the region lacks geological homogeneity. For example, when inhomogeneous regions are combined, the right-hand tail of a cumulative frequency distribution also tends to curve (Barton and Scholz, 1995). Charpentier, *et al.* (1995a) state that this method can give good results in mature regions with a large number of discoveries, since the field size distribution is fairly well-behaved. However, constraints on exploration (e.g. areas that are off-limits for some reason) can have a significant impact on the results and need to be accounted for separately. The difficulty in estimating reserve growth may also introduce uncertainties. Generally, the results may be improved through the use of discovery process models – introduced below.

2.6.4 Historical extrapolation

These methods are based upon the analysis of historical data on production or discoveries and the extrapolation of the identified trends into the future (Harbaugh, *et al.*, 1995). The techniques vary in their level of sophistication, but generally make little reference to geological concepts, information or techniques. The assumptions underlying all the techniques are that: a) the field size distribution is highly skewed, with the majority of oil being located in a small number of large fields; and b) these large fields tend to be discovered early in the exploration process, with subsequent discoveries being progressively smaller and the product of increasingly greater effort. The techniques fall into two broad groups:

- *Curve-fitting*: These use non-linear regression to fit curves to the historic trends in discovery or production and extrapolate these to estimate URR. The explained variable may be cumulative production, the rate of production, cumulative discoveries (measured using either 1P or 2P reserve estimates) or the rate of discoveries ('yield'). The explanatory variable may either be time or some measure of exploratory effort, such as the total area explored, the cumulative number of exploratory wells drilled or the cumulative depth of exploratory wells (in metres). These techniques were pioneered by Hubbert (1956; 1959; 1962; 1982) and have subsequently been adopted and developed by numerous analysts (e.g. Cleveland (1992b), Mohr and Evans (2008) Imam, *et al* (2004)) including in particular those concerned about 'peak oil' (Campbell, 2002; Laherrère, 2003).

- *Discovery process modeling*: These have many similarities with the 'curve-fitting' approaches, but are typically more sophisticated. They are based upon statistical analyses of the number and size of discovered fields as a function of either time, the discovery sequence or exploratory effort. This is sometimes combined with assumptions about the field size distribution and/or information about field location. Several of these models simulate a probabilistic law governing the process of new field discovery and can be used to provide forecasts of the number, size and sequence of future discoveries together with the anticipated success rate of exploratory drilling. They may also be combined with economic models to estimate the anticipated returns to exploratory drilling and improved through the incorporation of economic variables that influence the success of exploratory drilling. Major references include Arps and Roberts (1958), Kaufman (1975b) Schuenemeyer and Drew (1994) and Meisner and Demirmen (1981).

A crucial difference between curve-fitting and discovery process modelling is that the former use *aggregate* data for a region, while the latter require data on *individual fields*. While in many ways a superior technique, the extensive data requirements of discovery process models entirely preclude their use in areas where accurate data on individual fields is not available - which is the case for most regions of the world. In addition, their methodological sophistication makes them less practical for many researchers and they are only suitable for relatively low levels of aggregation. In contrast, since the data requirements of curve-fitting techniques are relatively modest, they can be readily applied in a variety of circumstances.

Both groups of techniques are only applicable to regions where exploration is relatively advanced, thereby providing a sufficiently large data set for statistical analysis. In principle, discoveries should provide a better explained variable than production, because the discovery cycle will be more advanced than the production cycle (see Section 3) and discoveries may be affected by fewer intervening variables than production. Similarly, exploratory effort should provide a better explanatory variable than time, since it reduces the effect of economic and political factors such as difficulties in accessing a region and changes in tax regimes that are unlikely to be stable over time. However, the use of exploratory effort as an explanatory variable does not remove the effect of factors such as advances in exploration technology that may increase the success rate of exploration at the same time as physical depletion is reducing it. Also, exploratory effort is far from immune to economic and political influences and the required data is less readily available (Cleveland, 1991).

Historical extrapolation techniques are best applied to data from geologically homogeneous areas that have had a relatively unrestricted exploration history (e.g. without areas being closed to exploration for legal or political reasons). If this is not done, the opening up of a new area for exploration (e.g. new plays within a basin) can lead to inconsistencies in the time-series and undermine the basis for historical extrapolation (Wendebourg and Lamiraux, 2002).[18] But even if the region is geologically homogeneous, the development of new

[18] The same applies to exploratory depth. Shallow fields are usually more extensively explored and exploited before deep fields, so shallow fields are generally overrepresented and deeper fields underrepresented in the distributions of discovered fields (Harris, 1977).

exploration and production technologies can lead to structural breaks in the time-series – for example, the advent of horizontal drilling of fractured reservoirs in Texas (Charpentier, 2003; Harris, 1977). The required time-series for discovery process models can also be complicated by factors such as the merger of separate fields over time (Drew, 1997). However, if curves are being fit to aggregate data, this is not a problem.

2.6.5 Comparison of methods

Traditionally, the methods based upon geological assessment have been considered to produce relatively optimistic resource assessments, while those based upon historical extrapolation have been considered to produce more conservative estimates. However, there appear to be relatively few published studies that systematically compare the results obtained from these different methods for particular regions. Comparison may be difficult, owing to lack of documentation, variations in the boundaries and depths included, varying assumptions about factors such as minimum viable field size, deterministic versus probabilistic presentation of results, reporting in terms of URR or undiscovered resources and so on. One of the best documented studies is a comparison of seven methods applied to seven regions by Ahlbrandt and Klett (2005). The comparison includes the synthetic method used by the USGS in their World Energy Assessment 2000, which combines geological assessment with discovery process modelling and can therefore be applied to both well-explored and unexplored regions.

The different methods were found to produce widely different results for the same regions. For example, in the case of the Sirte Basin region in Libya, the mean estimates for undiscovered resources from the different methods ranged from 10 Gboe to 48 Gboe. The latter estimate derived from a Pareto field size distribution method and exceeded the discovered resources in this (well explored) region. The corresponding variation in estimates of URR will depend in part upon the level of exploration in the region and hence the estimated proportion of total resources that remain to be discovered.

Ahlbrandt and Klett (2005) conclude that: a) resource assessments should be conducted for regions that are homogeneous in terms of both geology and exploration history; b) assumption about reserve growth and field size distribution can have a major influence on the results (with an assumed Pareto distribution yielding relatively large estimates relative to the other methods); and c) some methods can yield very conservative estimates, especially for mature regions. In the case of the latter, they cite the results for the Nequen Basin in Argentina where the discoveries in the preceding five years had exceeded the total estimate of undiscovered resources provided by the parabolic fractal method.

2.7 Summary

This section has introduced some key concepts and definitions relevant to ultimately recoverable resources (URR) and summarised the main methodological approaches that are available to estimate the size of those resources. The key conclusions are as follows:

- Estimates of URR are inherently uncertain and should be expressed in a probabilistic form. They are also dynamic since they depend upon the economic and technical conditions prevailing at the time, together with assumptions about how these may change in the future. While the level of uncertainty should fall as exploration proceeds, estimates of URR may vary widely even for maturely explored regions.

- Estimates of URR may be produced by a variety of methods for levels of aggregation ranging from a single field to the entire world. Most methods work best when they are applied to regions that are homogeneous in terms of both geology and exploration history. If this condition does not hold, there is a risk of underestimating URR by neglecting exploration constraints and other relevant factors.
- Estimates of the size of fields (i.e. cumulative production plus reserves) tend to 'grow' substantially over time as a result of factors such as the discovery of new reservoirs and conservative initial reporting. This phenomenon is well studied in the United States, where it is commonly attributed to the use of 1P estimates in reserve reporting. However, since comparable growth has been observed in field size estimates based upon 2P reserves, it appears to be a generic and universal phenomenon (although poorly studied for many regions of the world). Assumptions about reserve growth can have a major influence on both URR estimates and projections of future oil supply.
- Within any given region, most of the oil tends to be found in a small number of large fields which tend to be discovered relatively early in the exploration process. Subsequent discoveries tend to be smaller and require increasingly greater effort to locate. These observations underlie many of the methods of resource assessment, but the precise form of the field size distribution remains a focus of dispute. For discovered fields, the size distribution results in part from economic factors and may not be representative of the size distribution of the underlying population of fields.
- Around 100 oil fields account for up to half of the global production of crude oil, while up to 500 fields account for two thirds of cumulative discoveries. Most of these fields are relatively old, many are well past their peak of production and most of the rest will begin to decline within the next decade or so. The remaining reserves at these fields, their future production profile and the potential for reserve growth is therefore of critical importance for future global supply.
- The proportion of total resources contained within small, undiscovered fields is disputed. While the observed lognormal size distribution of discovered fields is likely to be the result of sampling bias, there is insufficient evidence to conclude whether a 'linear' or 'parabolic fractal' better describes the population size distribution. While technical improvements and higher prices should make more small fields viable, many will remain uneconomic to develop and the exploitation of the rest will be subject to rapidly diminishing returns. As a result, the competing estimates of the resources contained in small fields should be of less significance to future supply than the potential for increased recovery from the giant fields.
- There are a wide range of methods for estimating ultimately recoverable resources together with many variations on the basic techniques. Although the appropriate choice depends upon the nature and level of aggregation of the region, there is rarely a single 'best' method. One group of methods relies more upon geological information and is more appropriate for less explored regions, while a second group relies more upon the extrapolation of historical trends and is more appropriate to well-explored regions. The methods associated with Hubbert fall within the second group and typically apply relatively simple techniques to aggregate data to produce single value estimates for URR. However, these have strong parallels with more sophisticated and probabilistic techniques applied to disaggregate data (e.g. discovery process models) that are routinely used by organisations such as the USGS.

- The different methods are rarely compared systematically, but evidence suggests that they can produce widely different results. Assumptions about reserve growth and the field size distribution can be critical, while the application of extrapolation techniques to a non-homogeneous region can be misleading. Hence, estimates of URR at all levels of aggregation should be treated with caution and the uncertainties fully acknowledged.

3 Extrapolation methods – classification, description and evaluation

3.1 Introduction

This section describes and evaluates the *extrapolation methods* of estimating ultimately recoverable resources. All these methods involve analysing historical data on production or discoveries in a region and extrapolating this to derive an estimate of URR. The techniques vary greatly in their data requirements and level of sophistication, but share the common assumptions that: a) the field size distribution is highly skewed, with the majority of oil being located in a small number of large fields; and b) these large fields tend to be discovered early in the exploration process, with subsequent discoveries being progressively smaller and the product of increasingly greater effort. The physical constraints these provide on both the discovery and production cycle for a region can form a basis for estimating the URR for that region. However, the neglect of political and economic factors that may influence discovery and production trends could potentially be a significant source of error.

The extrapolation techniques fall into two broad groups, namely *curve-fitting* and *discovery process models*. While the former use aggregate data for a region, the latter require data on individual fields. While in many ways a superior technique, the extensive data requirements of discovery process models can make them impractical in many circumstances. In contrast, the data requirements of curve-fitting techniques are relatively modest, allowing them to be more readily applied. Both groups of techniques are only applicable to regions where exploration is relatively advanced, thereby providing a sufficiently large data set for statistical analysis. Furthermore, such techniques are best applied to geologically homogeneous regions with a relatively consistent history of exploration and/or production. Many of the difficulties with these techniques arise from the fact that these conditions do not hold for many oil producing regions.

Curve-fitting techniques are widely used by the individuals and groups associated with the peak oil debate and are therefore given the greatest attention here. Section 3.2 classifies these techniques in terms of their choice of explained and explanatory variables and clarifies the mathematical relationships between these variables. This provides a foundation for the rest of the report and helps to interpret the somewhat confusing empirical literature. Particular attention is paid to the appropriate interpretation of discovery data and the complications introduced by reserve growth.

Section 3.2 classifies the curve-fitting techniques into three groups, namely *production over time*, *discovery over time* and *discovery over effort*. Each group encompasses three separate techniques, although the terms used to label these techniques are not standardised. Sections 3.3 to 3.5 describe each group of techniques in turn. In each case, the aim is to describe the techniques, identify their historical origins, clarify the relationships between them and highlight some important strengths and weaknesses. While many issues are common to all the techniques, there are others which are specific to individual techniques.

Following this, Section 3.6 introduces discovery process modelling and summarises two of the most commonly used approaches – namely the Arps-Roberts and Barouch-Kaufmann models. Contrary to the claims of some authors, it is argued that the differences between discovery process models and simple curve-fitting is one of degree rather than kind, with the result that these models have many of the same limitations as simple curve-fitting. Section 3.7 concludes by summarising the main lessons and implications.

3.2 Explained and explanatory variables for curve-fitting techniques

All the curve-fitting techniques relate an *explained* variable to a primary *explanatory* variable. They all proceed by analysing historical trends in these variables and extrapolating these trends to produce an estimate of URR. But the methods vary in both the choice and the particular definition of the relevant explained and explanatory variables. Hence, before describing each method, it is helpful to clarify these definitions and to identify the choices that are available.

For the explained variable, there is a choice between measures of *production* or measures of *discovery*. These may either be measured in *cumulative* terms or in terms of *rates of change*. While cumulative production is always defined in relation to *time*, cumulative discovery may either be defined in relation to *time* or alternatively in relation to some measure of *effort*, such as the number of exploratory wells drilled.

The primary explanatory variable for cumulative measures of production is *time*, while the primary explanatory variable for cumulative measures of discovery may be either *time* or *effort*. The same applies to the rate of change of cumulative production (*rate of production* or more simply *production*) and the rate of change of cumulative discovery (*rate of discovery* or more simply *discovery*). However, the rate of production may also be expressed as a function of *cumulative production* and the rate of discovery may be expressed as a function of *cumulative discovery*.

Table 3.1 classifies the curve-fitting methods into three groups, namely: *production over time*; *discovery over time*; and *discovery over effort*. The relevant methods in each group are identified by their choice of explained and explanatory variables. Note that the terms used to label these methods are not standardised.

Table 3.2 summarises the mathematical notation used in the following sections. It important to note that cumulative discovery may either be measured in *current* terms or *backdated* to the date of discovery of the field. This choice, which is explained further below, can have a significant influence on the results.

The remainder of this section summarises the mathematical relationships between these variables. The discussion is adapted from a number of sources, but draws in particular from Hubbert (1982). Particular attention is paid to the implications of *reserve growth* and the extent to which this may be accommodated by using backdated measures of cumulative discovery.

Table 3.1 Classification of curve-fitting methods by explained and explanatory variables

Group	Technique	Explained variable	Explanatory variable
Production over time	Cumulative production projection	Cumulative production	Time
	Production projection	Rate of production	Time
	Production decline curve	Rate of production	Cumulative production
Discovery over time	Cumulative discovery projection	Cumulative discovery	Time
	Discovery projection	Rate of discovery	Time
	Discovery decline curve (time)	Rate of discovery	Cumulative discovery
Discovery over exploratory effort	Creaming curve	Cumulative discovery	Exploratory effort
	Yield per effort curve	Rate of discovery *wrt* exploratory effort	Exploratory effort
	Discovery decline curve (effort)	Rate of discovery *wrt* exploratory effort	Cumulative discovery *wrt* exploratory effort

Notes:

- The terms used to label these techniques are not standardised.
- Rate of production is the first derivative of cumulative production with respect to time. Alternative terms are the rate of change of cumulative production, or more simply production. Similar comments apply to the rate of discovery, although here the derivative may be with respect to either time or exploratory effort.

Table 3.2 Mathematical notation for curve-fitting techniques

Notation	Definition
t	Time
ε	Effort
t_d	Cumulative time for discovery
ε_d	Cumulative effort for discovery
$Q(t)$	Cumulative production
$Q'(t)$	Rate of change of cumulative production (rate of production, or production)
$R(t)$	Reported reserves
$D(t)$ or $D(\varepsilon)$	Cumulative discovery
$D'(t)$ or $D'(\varepsilon)$	Rate of change of cumulative discovery (rate of discovery, or discovery)
$B(t_d,t)$ or $B(\varepsilon_d,t)$	Backdated cumulative discovery
$B'(t_d,t)$ or $B'(\varepsilon_d,t)$	Rate of change of backdated cumulative discovery
Q_∞ or D_∞ or $B_{\infty,\infty}$	Ultimately recoverable resource

3.2.1 The production cycle

First, let $Q(t)$ represent the *cumulative production* from a region as a function of time. The *rate of change of cumulative production over time* ($Q'(t)$) is given by:

$$Q'(t) = \frac{dQ(t)}{dt} \qquad (3.1)$$

This is frequently termed the *rate of production* or more simply *production* and may alternatively be represented as $P(t)$. In empirical work, *Q'(t)* is frequently measured by annual production. A plot of the rate of production from when production begins to when it finally ends represents a full *production cycle*. As Hubbert (1982) notes:

> "…..The rate of oil production....begins at near zero rate and thereafter commonly increases exponentially for a few decades. Eventually, as the rate of discovery slows down, the rate of production follows. It reaches one or more principal maxima, and finally goes into a slow negative-exponential decline. Then at some definite time, production ceases altogether." (Hubbert, 1982)

Let t_{p_0} represent the start date of production in a region. The cumulative production up to time t (Q_t) is given by the area under the *Q'(t)* curve between t_{p_0} and t:

$$Q_t = \int_{t_{po}}^{t} Q'(t)dt \qquad (3.2)$$

Similarly, let t_{p_f} represent the end date of production in a region. The total cumulative production over the full production cycle ($Q_{t_{p_f}}$) is then given by:

$$Q_{t_{p_f}} = \int_{t_{p0}}^{t_{pf}} Q'(t)dt \qquad (3.3)$$

Or, since *Q'(t)=0* for $t < t_{p_0}$ and $t > t_{p_f}$:

$$Q_\infty = \int_{0}^{\infty} Q'(t)dt \qquad (3.4)$$

Q_∞ represents all the oil that will ultimately be produced from the region. This is equal to the *ultimately recoverable resource* (URR) for the region.

Cumulative oil production may be measured in terms of mass, energy content or volume, although the latter is most commonly employed (Speirs and Sorrell, 2009). Similarly, the rate of oil production is most commonly measured in volume (barrels) per day or per year. But since the quality and composition of crude oil varies between different regions, within the same region and over time, the quantity of production will depend upon the particular measure that is used. Also, the *net* energy production from a region will be less than the *gross* production since energy is required to find and produce oil and to manufacture and distribute oil products (Cleveland, 1992a; 1992b). However, only gross oil production will be considered here.

3.2.2 The discovery cycle

Let *D(t)* represent the *cumulative discovery* (i.e. the total amount that has been discovered) in a region as a function of time. Cumulative discovery (*D(t)*) may be measured in the same units as cumulative production (*Q(t)*). The *rate of change of cumulative discovery* over time is then given by:

$$D'(t) = \frac{dD(t)}{dt} \qquad (3.5)$$

This is frequently termed the *rate of discovery*, or more simply *discovery,* although this term can be misleading as described below. A plot of the rate of discovery against time ($D'(t)$) from when discovery begins to when it finally ends represents a full *discovery cycle*. This need not take the same shape as the production cycle, but may also be expected to begin near zero, increase to one or more maxima and decline again to zero when the resources in the region are exhausted. Since oil has to be discovered before it can be produced, the discovery cycle will precede the production cycle in time. However, the interval between discovery and production may vary from one region to another and also from one period to another.

Let t_{d_0} represent the date at which resources are first discovered in a region. The cumulative discovery up to time t (D_t) is given by the area under the $D'(t)$ curve between t_{d_0} and t:

$$D_t = \int_{t_{d_o}}^{t} D'(t)dt \qquad (3.6)$$

Similarly, let t_{d_f} represent the date at which the rate of discovery ($D'(t)$) falls to zero in the region. It is important to note that *this generally will <u>not</u> coincide with the date at which the last new field is discovered in the region*. Indeed, cumulative discoveries typically continue to grow for many years <u>after</u> the last field is discovered. The reason is the phenomenon of *reserve growth*, which was introduced in Section 2.4 and is discussed further below. The total cumulative discovery over the full discovery cycle ($D_{t_{d_f}}$) is then given by:

$$D_{t_{d_f}} = \int_{t_{d_0}}^{t_{d_f}} D'(t)dt \qquad (3.7)$$

Or, since $D'(t)=0$ for $t < t_{d_0}$ and $t > t_{d_f}$:

$$D_\infty = \int_{0}^{\infty} D'(t)dt \qquad (3.8)$$

D_∞ represents all the oil that will ultimately be discovered in the region. This is equivalent to the ultimately recoverable resource (URR) for the region.

While cumulative discovery (D_t) for a region may not be reported explicitly, it can be estimated from published data on cumulative production (Q_t) and reported reserves at time t (R_t). To form a consistent time series, the data on production and reserves should be reported in the same units using consistent definitions. Unfortunately, this is not always the case, even for industry-standard data sources such as the BP Statistical Review (Laherrère, 2004).[19] The sum of cumulative production at a particular point in time (Q_t) and the reported reserves at that time (R_t) represents all the oil that is estimated to a given level of confidence to have been discovered by that time. Hence cumulative discoveries at time t may be written as:

[19] For example, the BP Statistical Review includes the Canadian oil sands in their production figures but not in their reserve figures.

$$D_t = Q_t + R_t \tag{3.9}$$

At any point in time the reported reserves will be equal to the difference between cumulative discoveries and cumulative production ($R_t = D_t - Q_t$). For most of the discovery cycle, cumulative discoveries will exceed cumulative production ($D_t > Q_t$). But as $t \to \infty$, reserves will be exhausted ($R_\infty = 0$) and the cumulative production will equal the cumulative discoveries ($Q_\infty = D_\infty$). Note further that reserves are only equal to cumulative discoveries prior to the beginning of production (i.e. for $t \le t_{p_0}$ when $Q_t = 0$). However, the term reserves is sometimes used interchangeably with cumulative discoveries (e.g. Laherrère (1997)) which can be a source of confusion.

The rate of change of discovery over time may be written as:

$$\frac{dD(t)}{dt} = \frac{dQ(t)}{dt} + \frac{dR(t)}{dt} \tag{3.10}$$

Or:

$$D'(t) = Q'(t) + R'(t) \tag{3.11}$$

Where $Q'(t) \ge 0$ and $D'(t) \ge 0$ Note that $R'(t) = D'(t) - Q'(t)$. Hence, if the rate of production is less than the rate of discovery ($Q'(t) < D'(t)$), reserves will increase ($R'(t) > 0$). Conversely, if the rate of production exceeds the rate of discovery, reserves will fall. The rate of change of reserves with respect to time is often called *reserve additions* although during the latter part of the production cycle (and perhaps during earlier periods) reserve additions are more likely to be negative ('reserve subtractions'). In the absence of either new discoveries or reserve growth at existing fields ($D'(t) = 0$), reserves will be depleted at the rate of production ($R'(t) = -Q'(t)$).

As described in Thompson (2009a), reserves are typically estimated to three different levels of confidence, namely: proven (1P); proven and probable (2P); and proven, probable and possible (3P) (Bentley, *et al.*, 2007).[20]. Hence, the appropriate interpretation of cumulative discovery estimates depends upon the particular definition of reserves that is being used. Depending upon the data available, it may be possible to estimate either cumulative proved discoveries (D_t^{1P}) or cumulative proved and probable discoveries (D_t^{2P}). In practice, 3P reserve estimates are very rarely available.

As with production, both cumulative discoveries and the rate of discovery may be measured in terms of mass, energy equivalent or volume.[21] [22]

---

[20] If the estimates derive from same source we would expect that $R^{1P} \le R^{2P} \le R^{3P}$, but this may not necessarily be the case if the estimates derive from different sources (Speirs and Sorrell, 2009).

[21] Complications can arise in classifying new discoveries as oil fields, since many oil fields also produced gas - and vice versa. While energy-equivalent units could be used to measure total hydrocarbon discoveries and production, this would not help in estimating the *URR* of liquid fuels in a region.

[22] Cumulative discoveries have also been measured in a variety of other ways, including: a) the total number of fields or reservoirs discovered; b) the number of giant fields discovered (Woods, 1985); c) the surface area of giant fields discovered (Menard and Sharman, 1975) and; d) the number of fields discovered by size category (with size being defined on either a volumetric or surface areas basis) (Arps and Roberts, 1958). These measures may be more or less useful depending upon the data available and the purpose at hand. However, the relationships identified above would no longer apply.

3.2.3 Backdated discovery estimates

It is tempting to interpret the rate of discovery (*D'(t)*) as the rate of discovery of new fields, but this would be incorrect. As discussed in Section 2.4, the increase in cumulative discoveries over time derives from two sources:

- the discovery of *new* fields through exploration;
- reserve growth at *existing* fields through processes such as improved recovery, physical expansion and the discovery of new reservoirs.

The relative contribution of reserve growth to the increase in cumulative discoveries in a region will depend upon a number of factors, including: reporting conventions; the particular definition of reserves being used; the physical characteristics of the region (e.g. onshore versus offshore); the degree of exploration maturity; technological change; and various political, economic and institutional influences (Cleveland, 1992b). Generally, the contribution of reserve growth to reserve additions may be expected to increase over time and play a more important role in the later stages of a region's development.

The existence of reserve growth greatly complicates the analysis of the discovery cycle. However, it needs to be taken into account since additions to cumulative discoveries from reserve growth now exceed those from new discoveries in most of the world's oil producing regions. Also reserve growth is not confined to cumulative discovery estimates based upon 1P reserves (Klett and Schmoker, 2003).

To facilitate the analysis, it is helpful to introduce a measure of *backdated cumulative discoveries* (*B*). This is a function of both the time of discovery (t_d) and the time at which the estimate was made (t): $B(t_d,t)$ with $t \geq t_d$. Hence $B_{t_d,t}$ *represents the cumulative discoveries contained in fields that were discovered before time* t_d *as estimated at a later time t.* These estimates are made with the benefit of hindsight and are typically larger than the estimates made at the time of field discovery.

Backdating the subsequent increase in the estimated field size to the time of discovery of that field (t_d) allows a more accurate estimate of what was 'actually' found at a particular time. In contrast, cumulative discovery estimates (D_t) are not backdated and provide a poor guide to the quantity of resources found at a particular point in time. The relationship between cumulative discoveries and backdated cumulative discoveries is as follows:

$$D_t = B_{t,t} \tag{3.12}$$

$$\frac{dD(t)}{dt} = \frac{\partial B(t,t)}{\partial t} \tag{3.13}$$

Backdated cumulative discoveries for any (t_d,t) can also be written as:

$$B_{t_d,t} = Q_{t_d,t} + R_{t_d,t} \tag{3.14}$$

Where $Q_{t_d,t}$ represents the cumulative production up to time t from the fields discovered before time t_d and $R_{t_d,t}$ represents the reported reserves for those fields as estimated at time t. As with cumulative discoveries, the appropriate interpretation of backdated cumulative discovery estimates will depend upon the particular definition of reserves that is being used (e.g. 1P, 2P or 3P). While backdated cumulative discovery estimates have some advantages over cumulative discovery estimates, they require information on the date of discovery of fields, together with subsequent estimates of reserves and production for those fields. This information is not readily available in the public domain, although it is contained in industry databases such as that provided by IHS Energy. The use of backdated cumulative discovery estimates was pioneered by Hubbert (1967) and plays an important role in the estimation of URR through historical extrapolation techniques.

3.2.4 Growth functions

As indicated, the estimates of backdated cumulative discoveries ($B_{t_d,t}$) will typically increase over time (t) as a result of reserve growth (i.e. $B_{t_d,t+\tau} \geq B_{t_d,t}$). The *rate of change of backdated cumulative discovery estimates with respect to the time of the estimate* is given by:

$$B'_t(t_d,t) = \frac{\partial B(t_d,t)}{\partial t} \tag{3.15}$$

A plot of $B'_t(t_d,t)$ versus t for a particular value of t_d represents a *growth function*. This may be interpreted as the change in the estimated size of fields discovered at time t_d that occurs in each subsequent time interval dt up until time t. While we would generally expect cumulative discovery estimates to increase over time ($B'_t(t_d,t) \geq 0$), in some cases and/or time intervals cumulative discovery estimates may decrease ($B'_t(t_d,t) \leq 0$). The growth function may be normalised and represented in terms of the interval of time (τ) between discovery and the estimate ($\tau = t - t_d$) as follows:

$$G(t_d,\tau) = \frac{B'_t(t_d,t_d+\tau)}{B'_t(t_d,t_d)} \tag{3.16}$$

For fields discovered at time t_d, a plot of $G(t_d,\tau)$ versus τ represents the subsequent change in the estimated size of those fields relative to the initial estimate. An illustration was provided earlier in Figure 2.2. The rate of change of this function with respect to τ is given by:

$$G'_\tau(t_d,\tau) = \frac{\partial G(t_d,\tau)}{\partial \tau} \tag{3.17}$$

Since there is a limit to how much fields can grow, we would expect $G'_\tau(t_d,\tau) \to 0$ as $\tau \to \infty$. However, reserve growth may persist for very long periods of time. For example, cumulative proved discovery estimates for US fields appear to be still growing after an interval (τ) of 70 years (Nehring, 2006a; b; d).

Frequently, reserve growth is assumed to be independent of the time of discovery of a field (i.e. $G(t_d,\tau) = G(\tau)$ and $G'_\tau(t_d,\tau) = G'(\tau)$ for all t_d). But this implies that future reserve growth in recently discovered fields will be comparable to that observed in fields discovered many decades ago. Given the improvements in exploration and production technology that have occurred throughout the last century, this appears a rather questionable assumption. For example, it seems likely that modern seismic techniques will allow the size of newly discovered fields to be estimated more accurately than they were in the past. Similarly, since recovery factors have improved over time, the estimated reserves for a given estimate of the oil in place should also have increased. If this is the case, the potential for future reserve growth in recently discovered fields may be comparatively smaller.

3.2.5 The backdated discovery cycle

As the time of discovery (t_d) increases, the estimates of backdated cumulative discoveries ($B_{t_d,t}$) will also increase as a result of *new discoveries* (i.e. $B_{t_d+\lambda,t} \geq B_{t_d,t}$). The *rate of change of backdated cumulative discovery estimates with respect to the time of discovery* is given by:

$$B'_{t_d}(t_d,t) = \frac{\partial B(t_d,t)}{\partial t_d} \tag{3.18}$$

This is frequently termed the *backdated rate of discovery*, or more simply *backdated discovery*. A plot of $B'_{t_d}(t_d,t)$ versus t_d for a particular value of t represents a *backdated discovery cycle*. The estimated size of these discoveries will depend upon the time that the estimate was made (t) and hence upon the time interval between the discovery and the resource estimate ($\tau = t - t_d$). Compared to a non-backdated discovery cycle (*D'(t)*), this permits a more accurate estimate of the size of resources that were discovered at a particular time, with the degree of accuracy being proportional to τ.

As before, let t_{d_0} represent the date at which resources are first discovered in a region and let t_{d_f} represent the date at which the rate of change of backdated cumulative discovery estimates with respect to the time of the estimate falls to zero (i.e. $D'(t) = B'_t(t_d,t) = 0$ for all $t \geq t_{d_f}$). A plot of $B'_t(t_d,t_{d_f})$ against t_d represents the *full backdated discovery cycle* or the *ultimate discovery cycle*. This may be interpreted as a plot of the ultimately recoverable resources (URR) that were discovered in each time interval (dt_d) up until time t. (i.e. the ultimate amount of oil that fields discovered at time t_d will eventually produce).

The ultimately recoverable resource from fields that were discovered before time t_d ($B_{t_d,t_{d_f}}$) is then given by:

$$B_{t_d,t_{d_f}} = \int_{t_{d_0}}^{t_d} B'_{t_d}(t_d,t_{d_f})\,dt_d \tag{3.19}$$

Or, since $B'(t_d,t_{d_f}) = 0$ for $t_d < t_{d_0}$ and $t > t_{d_f}$, the ultimately recoverable resource from fields that were discovered before time t_d ($B_{t_d,\infty}$) may be written as:

$$B_{t_d,\infty} = \int_0^{t_d} B'_{t_d}(t_d,\infty)\,dt_d \tag{3.20}$$

The total cumulative discovery over the full discovery cycle ($B_{\infty,\infty}$) is then given by:

$$B_{\infty,\infty} = \int_0^{\infty} B_{t_d}'(t_d,\infty)\,dt_d \tag{3.21}$$

Where $B_{\infty,\infty} = D_\infty$ represents all the oil that will ultimately be discovered in the region. This is equal to the ultimately recoverable resource (URR) for the region.

In the empirical literature, unadjusted estimates of $B_{t_d,t}$ are sometimes referred to as the 'ultimate' resources contained in fields that were discovered before time t_d. Similarly, unadjusted estimates of $B'_{t_d,t}$ are sometimes referred to as the 'ultimate' resources that were found at time t_d. Both of these statements are incorrect since they neglect the potential for future reserve growth in these fields. In practice, $B_{t_d,t} \leq B_{t_d,\infty}$ and $B'_{t_d,t} \leq B'_{t_d,\infty}$. In other words, at any point in time (t), backdated cumulative discovery estimates will typically underestimate the ultimately recoverable resources found at or before t_d. For similar reasons, cumulative discovery estimates (D_t) will typically underestimate the cumulative resources discovered through to time t.

In both cases, the amount of underestimation will depend upon the length of time since the relevant fields were discovered ($\tau = t - t_d$). If this interval is relatively long, $B_{t_d,t}$ and $B'_{t_d,t}$ may provide relatively accurate estimates of $B_{t_d,\infty}$ and $B'_{t_d,\infty}$ since most of the reserve growth will have occurred. Conversely, if this interval is relatively short, they may provide relatively poor estimates since most of the reserve growth is still to occur. For any (t_d,t), estimates of $B_{t_d,t}$ represent the sum of reserve estimates from fields that were discovered at different times (i.e. at any time between t_{d_0} and t_d) and hence have experienced different amounts of reserve growth (i.e. different values of $\tau = t - t_d$) As a result, plots of $B(t_d)$ and $B'(t_d)$ for a given t can potentially be misleading since the sizes of fields discovered at different times (t_d) have not been estimated on a consistent basis. Despite its importance, this point does not appear to be widely recognised.

To provide a more accurate estimate of ultimately recoverable resources, the $B'(t_d, t)$ estimates may be adjusted to allow for future reserve growth using an estimated growth function (Equation 3.16). Assuming first, that the growth function is independent of the time of discovery ($G(\tau)$), and second that that a sufficiently long time series is available to allow $G(\infty)$ to be estimated, the relevant formula is:

$$B'(t_d, \infty) = B'(t_d, t) * \frac{G(\infty)}{G(\tau)} \qquad (3.22)$$

Where $\tau = t - t_d$ and $t \geq t_d$. Hubbert (1967) was one of the first authors to make such an adjustment. It is important to note that, in absence of backdated discovery estimates, estimates of *D'(t)* cannot be adjusted in this way since dates of discovery of the individual fields are not recorded. This suggests that such estimates may provide a less reliable means of estimating URR, unless the discovery process is well advanced ($D'(t) \rightarrow 0$) so that future reserve growth is expected to a small. Moreover, estimates of *D'(t)* will not provide a reliable guide to the 'actual' quantity of resources that was discovered at a particular time (t). Whether this is a drawback or not is a matter of debate.

3.2.6 Discovery as a function of effort

The previous sections have considered discovery as a function of time. This has disadvantages, however, as the rate of discovery will be influenced by a variety of economic and political factors which could invalidate the extrapolation of historical trends. For example, the rate of discovery may fall as a result of economic recession rather than through depletion of the underlying resource. An alternative approach is to consider cumulative discovery and the rate of discovery as a function of some measure of *effort (ε)* - for example, the number of exploratory wells drilled. In principle, these measures should be less sensitive to economic and political influences: for example, a recession could reduce exploratory activity as well as the number of new discoveries, with the result that the rate of discovery per unit of effort could remain relatively unchanged. However, there seems no reason to assume that such measures are "highly insensitive" to economic and political influences as Hubbert (1982) claimed.

Let $D(\varepsilon)$ represent the total amount of oil discovered for a cumulative level of effort (ε). The rate of change of cumulative discovery with respect to effort is then given by:

$$D'(\varepsilon) = \frac{dD(\varepsilon)}{d\varepsilon} \qquad (3.23)$$

As effort increases, the rate of addition to cumulative discoveries should fall as a result of physical depletion of the resource. However, a variety of other factors will influence the trend, including technologies that increase the success rate of exploratory drilling. As exploration and development exhausts the resources in a region, the rate of addition to cumulative discovery will fall to zero ($D'(\varepsilon) \to 0$) and cumulative discoveries will approach the URR for the region ($D_\varepsilon \to D_\infty$).

In a similar manner, *backdated cumulative discoveries* (*B)* may be expressed as a function of the cumulative level of effort through to the point of discovery (ε_d) and the *time* at which the estimate was made (*t*): $B(\varepsilon_d, t)$. Hence $B_{\varepsilon_d, t}$ represents the cumulative discoveries contained in fields that were discovered through to cumulative effort ε_d as estimated at time *t*.

The *rate of change of backdated cumulative discovery estimates with respect to cumulative effort* is then given by:

$$B'_{\varepsilon_d}(\varepsilon_d, t) = \frac{\partial B(\varepsilon_d, t)}{\partial \varepsilon_d} \qquad (3.24)$$

This may also be termed the *backdated rate of discovery with respect to effort.* A plot of $B'_{\varepsilon_d}(\varepsilon_d, t)$ versus ε_d for a particular value of *t* represents a *backdated discovery cycle with respect to effort*. The estimated size of $B'_{\varepsilon_d}(\varepsilon_d, t)$ will depend upon the time at which the estimate is made (*t*) and hence upon the interval between the discovery and the resource estimate ($\tau = t - t_d$).

As before, let t_{d_f} represent the date at which the rate of change of backdated cumulative discovery estimates with respect to the time of the estimate ($B'_t(\varepsilon_d, t)$ falls to zero. Then, a plot of $B'_{\varepsilon_d}(\varepsilon_d, t_{d_f})$ against ε_d represents the *full backdated discovery cycle with respect to effort* or the *ultimate discovery cycle with respect to effort*. This may be interpreted as a plot of the ultimately recoverable resources (URR) that were discovered in each effort interval ($d\varepsilon_d$) up until time t_{d_f}. (i.e. the ultimate amount of oil that will eventually be produced from the fields discovered at effort ε_d). A commonly used term for this is the *yield*. The full backdated discovery cycle with respect to effort $B'_{\varepsilon_d}(\varepsilon_d, t_{d_f})$ is then referred to as the *yield per effort (YPE)* curve (Cleveland, 1992b).

Note that the term *yield* is strictly only applicable if $t \geq t_{d_f}$ and $B'_t(\varepsilon_d, t) = 0$. If instead $t \leq t_{d_f}$ and $B'_t(\varepsilon_d, t) \geq 0$ the yield will be underestimated. To provide an accurate estimate of *yield*, the estimates of $B'_{\varepsilon_d}(\varepsilon_d, t)$ should be adjusted to allow for future reserve growth:

$$B'_{\varepsilon_d}(\varepsilon_d, \infty) = B'_{\varepsilon_d}(\varepsilon_d, t)\frac{G(\infty)}{G(\tau)} \qquad (3.25)$$

Where $\tau = t - t_d$ and t_d represents the time coinciding with cumulative effort ε_d. The ultimately recoverable resource over the full backdated discovery cycle with respect to effort is then given by:

$$B_{\infty,\infty} = \int_0^\infty B'_{\varepsilon_d}(\varepsilon_d, \infty) d\varepsilon_d \qquad (3.26)$$

Hubbert (1967) pioneered the analysis of yield per effort, using data on backdated cumulative discoveries in the US, which he adjusted for future reserve growth. His measure of effort was the cumulative *length* (in feet) of exploratory drilling. This was aggregated in units of 10^8 feet, a quantity subsequently termed a 'Hubbert Unit' (Cleveland and Kaufmann, 1991; Haun, 1981). Since this measure of effort is a product of the *number* and *depth* of exploratory wells, it captures both the vertical and horizontal dimension of oil exploration – including the technical advances that permitted drilling at greater depths. However, this information is not recorded for all regions of the world and is rarely available in the public domain. Other authors have used alternative measures of effort, including:

- the total number of exploratory wells drilled ('new field wildcats' or NFW) (Laherrère, 2002a; Ryan, 1973);
- the number of *successful* exploratory wells drilled (i.e. excluding 'dry holes') (Moore, 1962);
- the cumulative length of successful exploratory wells (Stitt, 1982);
- the cumulative length of all wells drilled (exploratory and development) (Cleveland, 1992b); and
- the cumulative number of discovered fields.

The first of these measures is most widely cited, since it has been popularised by Campbell and Laherrère (1995) and the relevant data is available in the IHS database. Campbell and Laherrère also followed Hubbert and other authors in using *exploratory* drilling activity as their main explanatory variable. However, this is potentially misleading since much of the increase in cumulative discoveries derives from *development* drilling activity at existing fields - which contributes to reserve growth.[23] Moreover, even when data is available, there may be ambiguities and inconsistencies in classifying different types of drilling activity - for example, when previously separate fields are merged into a single larger field (Drew, 1997). In these circumstances, it may be better to employ a measure of *total* drilling activity as the explanatory variable - as used, for example by Cleveland (1992b). There are also difficulties

[23] The IHS database contains information on the number of development wells drilled and hence allows a measure of total trading activity to be used as the explanatory variable.

with accounting for the delays between successful drilling activities and the subsequent additions to resource estimates (R_t) (Byrd, *et al.*, 1985).[24]

One issue that appears to be neglected in the literature is the allocation of exploratory activity between oil and gas resources. The data sources used by Campbell and others simply record *total* exploratory activity for relatively aggregate regions and do not distinguish between the search for oil and the search for gas. This may follow from the fact that individual fields frequently produce both. However, the opening up of a new, predominantly gas-producing region in a region may lead to a major increase in the exploratory activity for a country, with little or no increase in discovered oil resources (or vice versa). If so, this could seriously distort the time series for discovery over effort ($B'(\varepsilon_d, t)$) for oil. Note that the same problem does not occur with a time-series for discovery over time ($B'(t_d, t)$). Moreover, even if a time-series does allow the exploration of oil and gas to be distinguished, there may still be difficulties in allocating 'dry holes' between oil and gas fields and the results may be sensitive to the particular method that is used. Hence, there are some technical issues associated with the definition and measurement of exploratory effort, the importance of which is not always acknowledged.

3.2.7 Summary of explained and explanatory variables

Curve-fitting models utilise a number of different definitions for explained and explanatory variables. In particular, this involves choices between:

- production versus discovery measures;
- cumulative versus rate of change measures;
- *time* versus *effort* measures; and
- *current* versus *backdated* measures of discovery.

There are also additional complications, such as the appropriate definition and measurement of exploratory effort. In practice, the available choices will be greatly constrained by data availability.

This section has classified the different extrapolation methods in terms of their choice of explained and explanatory variables and clarified the mathematical relationship between these variables. The different options are summarised in Table 3.3.

While the analysis of production trends is relatively straightforward, the analysis of discovery trends (whether measured with respect to time or effort) is greatly complicated by reserve growth. As a result, measures of the rate of discovery do not necessarily correspond to the amount of new discoveries. While the use of backdated discovery estimates can help in this regard, this is only possible if the relevant information is available (i.e. the date of discovery of fields, together with the cumulative production and declared reserves of those fields). Even then, the estimates should ideally be adjusted to allow for future reserve growth. If, as is

[24] For example, the discovery well for Prudhoe Bay in Alaska was drilled in 1967, but the reserves were not added to official estimates until 1970 (Cleveland, 1992b).

frequently the case, such an adjustment is not made the associated extrapolation techniques may underestimate the URR.

The following sections described the curve fitting techniques in more detail. Section 3.3 describes the *production over time* techniques, while Sections 3.4 and 3.5 do the same for the *discovery over time* and *discovery over effort* techniques respectively. In each case, the aim is to describe the techniques, identify their historical origins and contemporary application, clarify some relevant methodological issues and highlight their strengths and weaknesses. While many issues are common to all the techniques, there are others which are specific to each. Discovery process models are discussed separately in Section 3.6.

Table 3.3 Classification of curve-fitting methods by explained and explanatory variables – notational summary

Group	**Method**		**Explained variable**	**Primary explanatory variable**
Production over time	Cumulative production projection		$Q(t)$	t
	Rate of production projection		$Q(t)$	t
	Rate of production decline curve		$Q'(t)$	$Q(t)$
Discovery over time	Cumulative discovery projection	Current Backdated	$D(t)$ $B(t_d, t)$	t t_d
	Rate of discovery projection	Current Backdated	$D'(t)$ $B'_{t_d}(t_d, t)$	t t_d
	Rate of discovery decline curve (time)	Current Backdated	$D'(t)$ $B'_{t_d}(t_d, t)$	$D(t)$ $B_{t_d}(t_d, t)$
Discovery over effort	Creaming curve	Current Backdated	$D(\varepsilon)$ $B_{\varepsilon_d}(\varepsilon_d, t)$	ε ε_d
	Yield per effort curve	Current Backdated	$D'(e)$ $B'_{\varepsilon_d}(\varepsilon_d, t)$	ε ε_d
	Rate of discovery decline curve (effort)	Current Backdated	$D'(t)$ $B'_{t_d}(t_d, t)$	$D(\varepsilon)$ $B_{t_d}(\varepsilon_d, \varepsilon)$

3.3 Production over time techniques

The simplest, although not necessarily the most reliable method of estimating URR relies upon time-series data on either *cumulative production* (*Q(t)*) or the *rate of production* (or more simply *production - Q'(t)*). Typically, a curve is fit to this data using non-linear regression techniques.[25] This curve may take a variety of functional forms with its shape being defined by three or more parameters, one of which corresponds to the *URR*.

A very similar approach is frequently used to forecast future production and to identify the date of peak production. With this approach, *assumed* values for the URR are used to constrain the shape of the curve and hence the estimated parameter values. In contrast, when using such an approach to *estimate* the URR, the curve-fitting is constrained only by the assumed functional form and the historical data. The following section summarises the origin of the production projection technique, the use of the standard 'logistic' model for cumulative production and the use of both alternative functional forms and multi-cycle models. Section 3.3.2 describes the closely related technique of production decline curves - which is frequently referred to as 'Hubbert Linearisation'.

3.3.1 Production projection

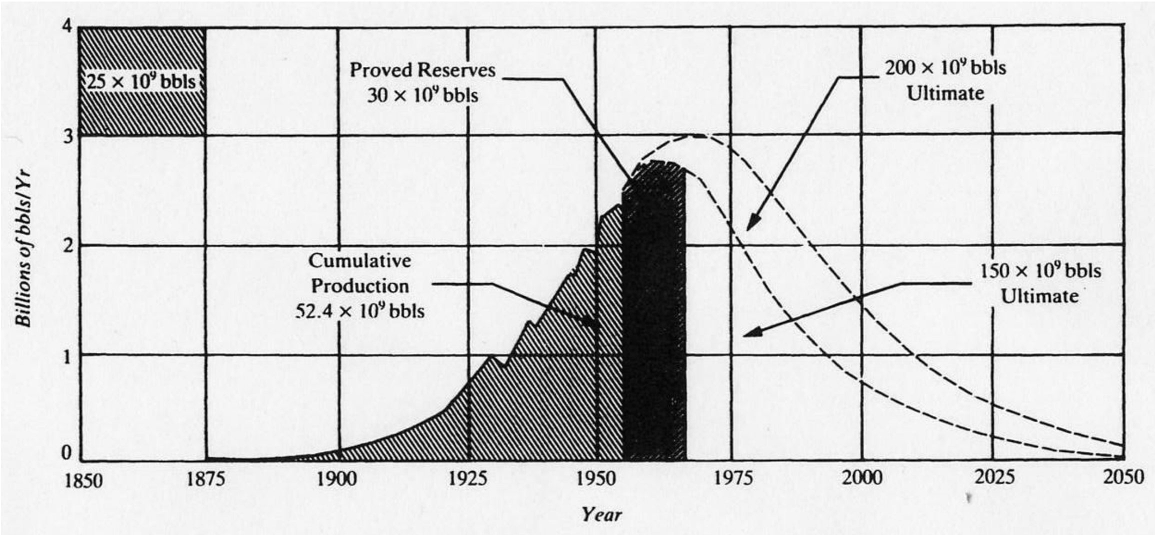
3.3.1.1 Origins of the technique

Production projection has its origins in a seminal paper by Hubbert (1956) in which he used assumed URR values in conjunction with time-series data on US oil production to forecast the future US production cycle ($Q'(t)$), including in particular the date of peak production. Hubbert's URR assumptions were based upon industry estimates and ranged from 150 to 200 billion barrels – which compared to cumulative production of 52.4 billion barrels through to 1956 and proved reserves of 30 billion barrels. Using simple graphical techniques, Hubbert fitted a curve to historical data on annual US oil production and projected this forward under the assumption that production must eventually decline exponentially and the area under the curve must equal the assumed URR. The result is illustrated Figure 3.1. Hubbert subsequently observed that, given these constraints, "…. it became impossible to draw this curve very differently from the way it is shown." (Hubbert, 1982).[26]

[25] These are discussed further in Section 5. A very accessible guide to such techniques is provided by Motulsky and Christopoulos (2004b). With modern computer technology and software packages, non-linear regression is relatively straightforward. However, this was not the case for much of Hubbert's lifetime. The earlier literature therefore uses simpler methods, including in particular the linear transformation of the relevant functional form followed by a linear regression (Deffeyes, 2003).

[26] Hubbert's approach was strongly influenced by the concept of a life cycle for mineral industries, first advanced by Hewett (1929). Hewett argued that the time path of production will follow a life cycle, involving initiation, rapid increase, levelling off, decline and eventual exhaustion. Hubbert called this "a truly great paper, one of the more important papers ever written by a member of the US Geological Survey."

Figure 3.1 Hubbert's 1956 projection of the forthcoming peak in US oil production



On the basis of this curve, Hubbert famously forecast that US (lower 48) oil production would reach a peak sometime between 1965 and 1971. Since this forecast ran counter to the prevailing optimism about the future of the US oil industry, it proved highly controversial (Bowden, 1985).[27] In 1956, it was estimated that less than half of the recoverable oil in the US had been discovered and less than one third had been produced. Also, annual discovery rates significantly exceeded the rate of production and both exploration and production technologies were improving rapidly. Hence, while Hubbert's URR assumptions were widely accepted, their implications for the future production cycle were poorly appreciated (Hubbert, 1959). Hubbert's paper provoked a debate about US oil resources that continued until the 1970s. When US production peaked in 1971 and began its long decline, many commentators considered Hubbert's approach to have been validated (Strahan, 2007b).

Shortly after the publication of Hubbert's paper, the industry consensus on URR estimates evaporated. Numerous commentators disputed Hubbert's approach and a series of more optimistic estimates of URR began to appear. Most famously, the USGS estimated a value of 580 billion barrels for the lower 48 states, based upon forecasts of future drilling activity (Zapp, 1961).[28] Such discrepancies motivated Hubbert to develop more formal methods to estimate URR, using historical data on oil production and discovery. Hubbert developed and applied several such techniques between 1962 and 1982, of which production projection was the first. All of these produced estimates for the US in the range 150-200 billion barrels.

3.3.1.2 The logistic model

Hubbert's first empirical estimate of URR was made by fitting a *logistic* curve to time-series data on cumulative production (*Q(t)*) and cumulative discoveries (*D(t)*) in the US (Hubbert, 1962).[29] The logistic curve was introduced by Verhulst in 1838 and was later popularised in mathematical biology by Lokta (1925). In the case of oil production, the use of logistic curve

---

[27] As an illustration, Hubbert's employer (Shell) required changes to be made to the paper, with specific predictions about the future of the oil industry being replaced with much vaguer statements (Bowden, 1985).
[28] The study concluded that: "….the size of the resource base would not limit domestic production capacity in the next 10-20 years at least and probably [not] for a much longer time"
[29] Hubbert (1962) was one of a number of reviews of natural resources policy commissioned by President Kennedy. Hubbert's primary interest was the determination of the future date of maximum oil production.

implies that cumulative production will initially grow exponentially, but as the URR is approached, production (*Q'(t)*) will fall and eventually decline to zero.

Exponential growth may be represented as follows:

$$\frac{dQ}{dt} = Q'(t) = aQ(t) \qquad (3.27)$$

Which may be solved to yield:

$$Q(t) = Q_0 e^{at} \qquad (3.28)$$

The logistic curve modifies exponential growth with a 'feedback' term that slows the rate of production (*Q'(t)*) as the URR is approached:

$$Q'(t) = aQ(t)\left(1 - \frac{Q(t)}{Q_\infty}\right) \qquad (3.29)$$

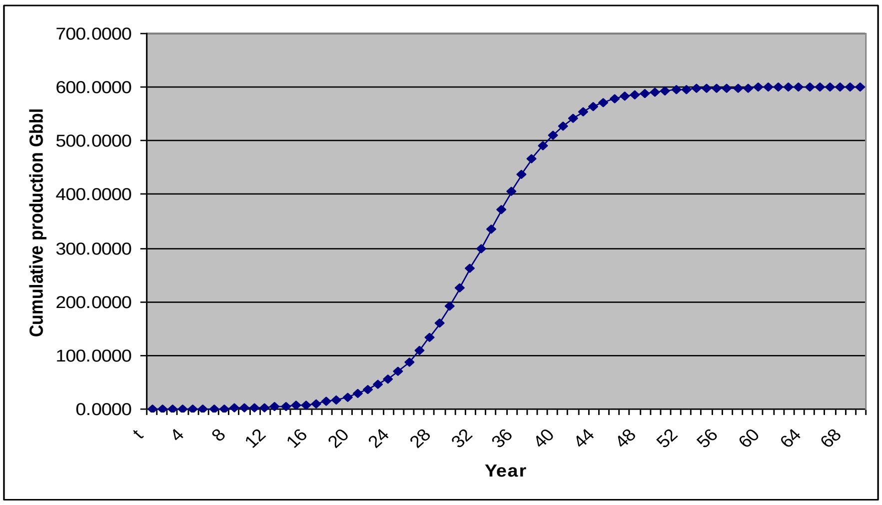
With this formulation, when $Q(t) \approx 0$, $Q'(t) \approx 1$ and as $Q(t) \to Q_\infty$, $Q'(t) \to 0$. This leads to a sigmoidal, or elongated S-shaped growth trajectory, tending to an asymptope as $t \to \infty$ which represents the URR (Meyer, *et al.*, 1999). Equation 3.29 can be solved analytically to yield:

$$Q(t) = \frac{Q_\infty}{1 + e^{-a(t-t_m)}} \qquad (3.30)$$

In this formulation[30] the logistic curve is defined by three parameters, Q_∞, *a* and t_m. The first (Q_∞) represents the ultimately recoverable resource, while the second (*a*) represents the 'steepness' of the cumulative production curve. Parameter *a* is sometimes replaced by $\ln(81)/C$, where *C* specifies the time required for cumulative production to grow from 10% to 90% of the URR (the 'characteristic duration'). The parameter t_m specifies the time when cumulative production reaches one half of the URR ($Q_\infty = 0.5 * Q_{t_m}$)), or the midpoint of the growth trajectory. At this point production is at a maximum - given by $(Q'(t_m) = Q_\infty * |a|)/4$. A key assumption of the logistic model is that the production cycle is *symmetric* about this midpoint. Figure 3.2 illustrates a logistic cumulative production cycle, with a characteristic duration of 17 years and an URR of 600 Gb.

[30] There are a variety of alternative formulations of the logistic function in the depletion literature (and in the growth literature more generally) although they all take the same general form. For example, Hubbert (1982) uses $Q(t) = Q_\infty / (1 + N_0 e^{-a(t-t_0)})$ where $N_0 = (Q_\infty - Q_0)/Q_0$ and Q_o represents the cumulative production in year t_o – the first year of the model fit.

Figure 3.2 Logistic model of cumulative production cycle



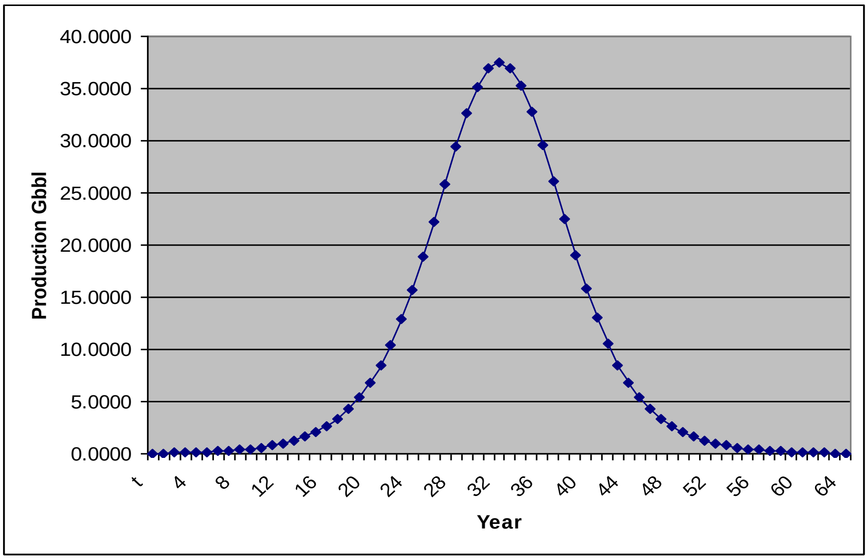
An expression for *production* over time ($Q'(t)$) can be obtained by differentiating Equation 3.30:

$$Q'(t) = \frac{aQ_\infty e^{-a(t-t_m)}}{\left(1+e^{-a(t-t_m)}\right)^2} \qquad (3.31)$$

The resulting 'bell shaped' production cycle is illustrated in Figure 3.3. The shape of this curve is determined by the same three parameters (Q_∞, a and t_m) as the cumulative production cycle, with t_m defining the time at which production reaches a peak. It is interesting to note that the (hand drawn) curve that appeared in Hubbert's 1956 paper has a 'wider' top than suggested by the above formula (Laherrère, 2000b). Also, while a bell shaped production cycle is commonly referred to as a 'Hubbert curve', Hubbert repeatedly stated that a production cycle need not take this form:

> "…The complete cycle curve has only the following essential properties: the production rate begins at zero, increases exponentially during the early period of development and then slows down, passes *one or more principal maxima*, and finally declines negative-exponentially to zero. *There is no necessity that the curve, P as a function of t, have a single maximum or that it be symmetrical*. In fact, the smaller the region the more irregular in shape is the curve likely to be...." (Hubberrt, 1982, emphasis added)

Figure 3.3 Logistic model of production cycle



Using nonlinear regression techniques, the ultimately recoverable resource (Q_∞) may be estimated by either fitting Equation 3.30 to historical data on cumulative production or by fitting Equation 3.31 to historical data on production.[31] In principle these approaches should give broadly equivalent results, but in practice the results can be significantly different (Carlson, 2007a). The usefulness of such techniques may be expected to depend in part upon the length of time-series available. Estimates of URR from production projections will be more reliable if production has passed its peak (i.e. ($t \geq t_m$) and can only be obtained if the *rate of increase of production* (i.e. $Q''(t)$) has past its peak. This corresponds to the point of inflection on the rising production trend ($Q'(t) > 0, Q''(t) > 0, Q'''(t) = 0$).

As illustrated in Figures 3.4 and 3.5, the US production cycle fits the logistic model relatively well, despite covering a period that includes two world wars, several recessions, two oil shocks and revolutionary changes in exploration and production technologies. This helps explain why the US experience is so widely quoted. The logistic model now suggests a URR for crude oil and natural gas liquids (NGLs) combined of 257 Gb, which compares to cumulative production through to 2007 of 227 Gb, of which approximately 197 Gb was crude oil. Hence, 25 years after Hubbert's last paper, cumulative production of crude oil in the US was 21% higher than his last estimate of URR. Hubbert’s logistic model provided no means of anticipating the subsequent development of the deep-water resources of the Gulf of Mexico, which now account for a large fraction of US production and URR.

In contrast to the US, the logistic model provides a relatively poorly approximation to the cumulative production cycle for other producing regions - perhaps in part because production derives from a smaller population of fields (Brandt, 2007).[32] This has led researchers to investigate the use of alternative functional forms to model cumulative production as well as multi-cycle models. These are discussed below.

[31] An alternative (rarely used) employs the second order differential – the rate of change of production (*Q’’(t)*).

[32] Hubbert (1962; 1982) emphasised that the technique works best for a large region such as the United States.

Figure 3.4 A fit of the logistic model to US cumulative production data (crude oil +NGLs)

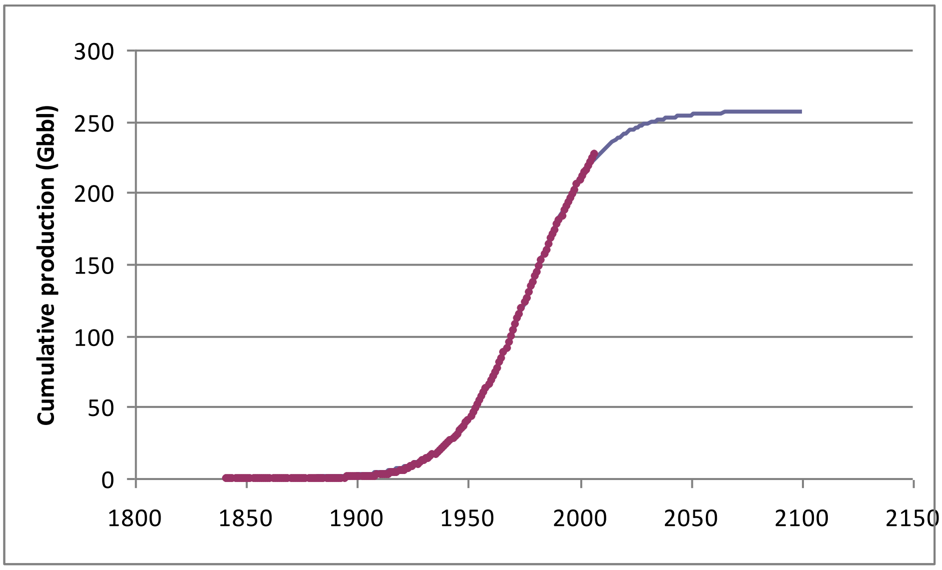
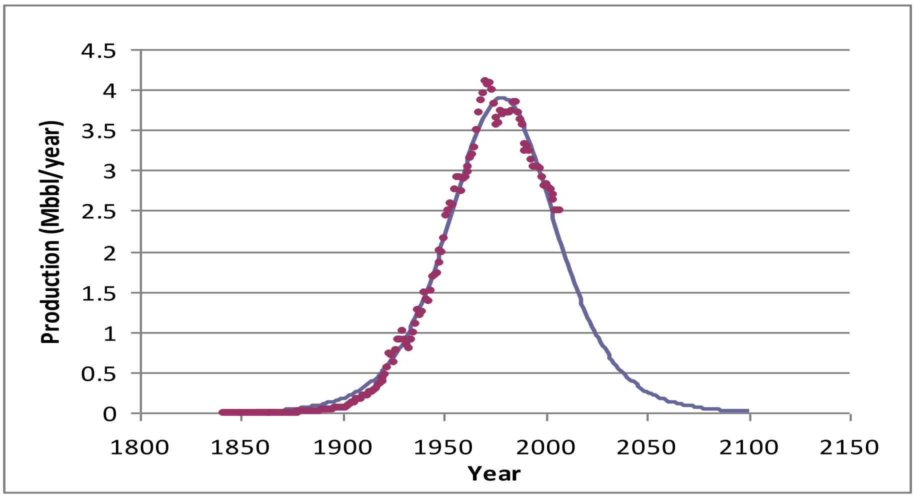


Figure 3.5 A fit of the logistic model to US production data (crude oil +NGLs)



3.3.1.3 Alternative functional forms

It seems likely that Hubbert chose a logistic equation to model cumulative production because it was analytically straightforward. In fact, the logistic model is one of a family of ‘sigmoidal’ (S-shaped) curves that are widely employed in biology and other disciplines to to simulate growth processes (Meade, 1984; Tsoularis and Wallace, 2002). Alternative growth models include the generalised logistic (Nelder, 1971), the Bass (Bass, 1969) and the bi-logistic (Meyer, 1994) as well as a variety of functions derived from probability theory, including the cumulative Cauchy, Weibull and lognormal distributions (Meade, 1984).[33] However, only a subset of these has been applied to the study of oil depletion.

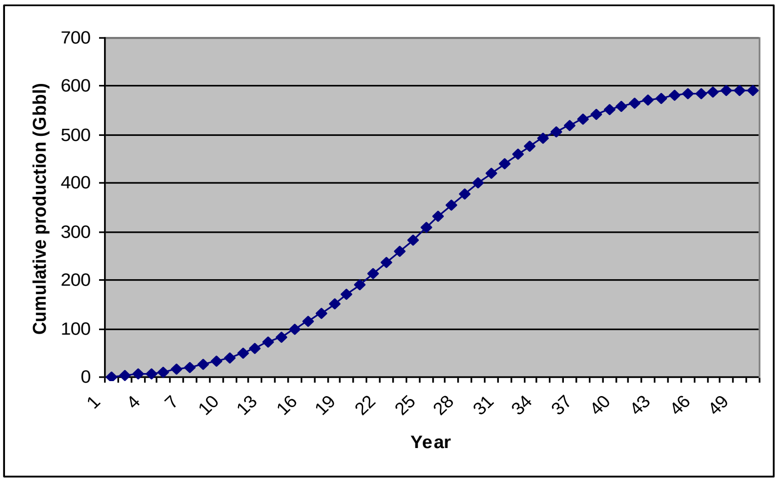
[33] All cumulative density functions are monotonically increasing between zero and unity and many approach unity asymtopically.

One commonly cited alternative, which also assumes a symmetric production cycle, is the cumulative normal (or cumulative Gaussian) distribution (Brandt, 2007; Deffeyes, 2003). For example, Bartlett (2000) fits an normal distribution to US production ($Q'(t)$) data to estimate a URR of 222 billion barrels. The corresponding equation for cumulative production ($Q(t)$) is:

$$Q(t) = \int_0^\infty \frac{Q_\infty}{\sigma(2\pi)^{0.5}} \exp\left[-\frac{(t-t_m)^2}{2\sigma^2} \right] dt \qquad (3.32)$$

Where t_m represents the time of peak production and σ is a 'width' parameter equivalent to the standard deviation. The resulting cumulative production curve is illustrated in Figure 3.6 for a region with t_m=25, $\sigma = 10$ and a URR of 600 Gb. For the same peak year (t_m) and URR, the corresponding production curve ($Q'(t)$) is slightly wider than the logistic around the midpoint and narrower on the flanks.

Figure 3.6 Cumulative normal model of cumulative production cycle



Both the cumulative normal and the logistic distribution are symmetric about t_m. But there are a variety of reasons for expecting the production cycle to be asymmetric. For example, a highly skewed field size distribution (Section 2.5) could lead to a rapid decline in production after the peak as the large fields that dominate production are depleted. Conversely, techniques such as enhanced oil recovery may act to slow the decline. A symmetrical production cycle also appears poorly supported by real-world experience. In the most systematic study to date, Brandt (2007) analysed 74 oil producing regions that were past their peak of production and found that the rate of production increase exceeded the rate of decline in over 90% of cases – in other words, most production cycles were asymmetric to the left.[34]

These observations have led a number of authors to employ *asymmetric* functional forms to model cumulative production cycles. One of the most widely cited is the Gompertz function, defined as follows:

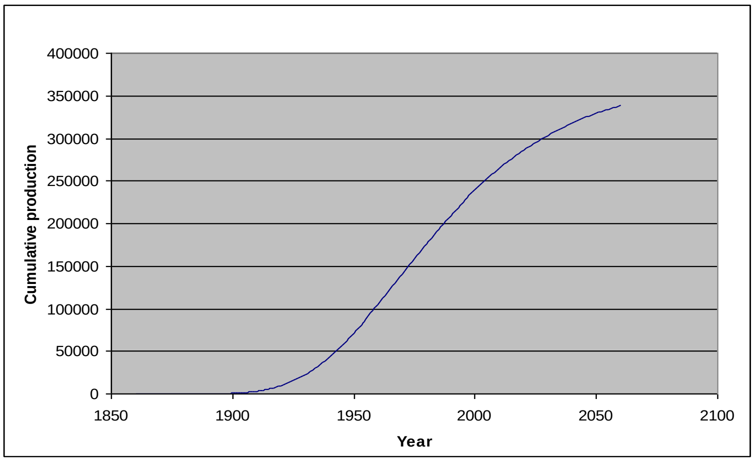
[34] Brandt (2007) found that the median rate of decline (2.6%) was approximately 5% less than the median rate of increase (7.8%) for 67 out of 74 post-peak regions fit with an exponential model. This analysis suggests that production profiles tend to be slightly asymmetric, with slower rates of decline than rates of increase. Since the mean rate of decline (4.1%) was significantly less than the production-weighted mean (1.9%), Brandt's results also suggest that decline is slower in larger regions.

$$Q(t) = Q_\infty e^{-ae^{-b(t-t_m)}} \tag{3.33}$$

This leads to a cumulative production cycle with a point of inflection around 35- 40% of the URR (Figure 3.7), implying a production cycle with a rate of decline that is slower than the rate of increase (as Brandt's work suggests). As with the logistic, the Gompertz function is defined by only three parameters and the degree of asymmetry is fixed. Moore (1962) fitted a Gompertz function to the same US production data as used by Hubbert and arrived at an URR estimate that was almost twice as large for a comparable goodness of fit. To Ryan (1966), this demonstrated the limitations of the technique:

> "The fact that the Gompertz curve, like the logistic, fits historical data quite well is scarcely surprising. Both the Gompertz and the logistic curves are monotonically increasing for relevant values of the parameters. Since each has three disposable constants, either one should give a reasonable fit to most monotonically non-decreasing functions such as cumulative oil production. The fact that such flexible curves appear to fit historical data well does not, therefore, provide a sound basis for long-term extrapolation. There are many different functional forms which would approximate the historical series reasonably well but which give widely varying estimates of ultimate recoveries." (Ryan, 1966)

Figure 3.7 Gompertz model of a cumulative production cycle



Other asymmetric functional forms employed in the oil depletion literature include:

- *Asymmetric normal*: Brandt (2007) used an asymmetric version of the normal curve in which the σ parameter for $t<t_m$ is different to that for $t>t_m$ and gradually shifts from one to the other in the region of t_m.
- *Asymmetric exponential*: Wood, *et al* (2003) assumed exponential growth until the ratio of URR to production fell to a specified threshold level, at which point production declined exponentially at a rate sufficient to keep this ratio constant.[35]
- *Asymmetric bell shaped:* Kaufmann and Shiers (2008) simulate a variety of future global production cycles using an iterative set of equations in which values of URR are assumed, together with the initial production growth and decline rates

[35] This assumption has little theoretical or empirical support and the model leads to implausibly sharp production peaks followed by extremely rapid declines in production.

The three studies quoted above focus on forecasting the peak year of production (t_m) using assumed values of URR. Provided that the rate of increase of production ($Q''(t)$) has passed its peak, the same functional forms may be used to estimate the URR. However, a variable degree of asymmetry implies a more complex model with additional parameters which may be difficult to justify on statistical grounds. This issue is discussed further in Section 4.

3.3.1.4 Multi-cycle models

Cumulative production cycles often have more than one point of inflection, corresponding to production cycles with more than one peak in production (Laherrère, 2000b). This behaviour may result from economic, technical or political disruptions, such as the reduction in UK oil production following the 1988 Piper Alpha disaster, or it may result from the opening up of a new oil producing region, such as Alaska in the US. This possibility was first recognised by Hubbert (1956), who noted that Illinois had experienced two discovery cycles as a consequence of changes in exploration technology (rather than the opening up of a new region), leading subsequently to two production cycles. Indeed, Laherrère (2004) has observed that *most* countries appear to have several cycles of exploration activity and then of production.

This suggests that the full production cycle could potentially be more accurately modelled using two or more curves. Ideally, each curve would represent the production of resources from a geologically homogeneous region with a corresponding URR. The cumulative production cycle for the aggregate region would then be formed from the sum of these individual curves:

$$Q(t) = \sum_{i=1,n} Q_i(t) \qquad (3.34)$$

Similarly, the aggregate URR would be derived from the sum of the URR estimates for the individual curves. Several authors have followed this approach, including Patzek (2008), Mohr and Evans (2007) and Imam, *et al* (2004). Meyer, *et al.* (1999) have developed a formal approach for decomposing a time-series into the sum of an arbitrary number of logistic curves ('loglet analysis') which could potentially be applied to oil depletion.[36] However, the use of multi-cycle models raises similar statistical issues to the use of more complex functional forms.

3.3.2 Production decline curves

As an alternative to modelling the growth in production over time, it is possible to model production as a function of cumulative production. This leads to an alternative approach to estimating URR that is closely related to the *decline curve analysis* used by reservoir engineers to model the production decline of individual wells or fields. This section first describes the 'aggregate' approach developed by Hubbert and then clarifies its relationship to the analysis of production decline in individual fields.

3.3.2.1 'Hubbert Linearisation'

Hubbert (1982) notes that:

---

[36] The application of this approach to the estimation of *URR* has so far been confined to 'peak oil' websites (e.g. http://www.theoildrum.com/story/2006/9/3/113719/7594). However, Kemp and Kasim (2005) have applied the technique to model production decline rates from individual fields.

"…..A difficulty in analysing either P or Q as a function of the time arises from the asymptopic approaches of these quantities to their respective limits as time increases without limit. On the other hand, Q has the definite finite limits of 0 and Q_∞. It is convenient, therefore, to consider the production rate *dQ/dt* as a function of Q rather than of time. In this system of the coordinates, *dQ/dt* is zero when Q=0 and when $Q = Q_\infty$. Between these limits *dQ/dt >0* and outside the limits, equal to zero."

The simplest functional form for the relationship between production (*Q'(t)*) and cumulative production (*Q(t)*) that will meet these boundary conditions is a second-degree (parabolic) equation:

$$Q'(t) = bQ(t) + cQ^2(t) \qquad (3.35)$$

By applying the boundary conditions, Hubbert arrives at the following equation for the relationship between *Q'(t)* and *Q(t)*:

$$Q'(t) = b\left[Q(t) - \frac{Q^2(t)}{Q_\infty}\right] \qquad (3.36)$$

As shown in Figure 3.8, this is the equation of a *parabola.* If this provides a poor approximation to the observed data, higher order terms may be required in the polynomial. However, if it provides a reasonable approximation, Equation 3.36 may be transformed as follows:

$$\frac{Q'(t)}{Q(t)} = b\left[1 - \frac{Q(t)}{Q_\infty}\right] \qquad (3.37)$$

Equation 3.37 states that the ratio of production to cumulative production is proportional to the fraction of resource remaining to be produced ($1 - Q(t)/Q_\infty$). This is the equation of a straight line, with a slope of ($-b/Q_\infty$) intersecting the vertical axis at b and the horizontal axis at Q_∞ (Figure 3.9). Hence, given historical production data for a region, the URR (Q_∞) may be estimated by plotting *Q'*(*t*)/*Q(t)* as a function of *Q(t)* and fitting a linear regression to estimate the parameters b and Q_∞. This straightforward technique has been popularised by Deffeyes (2003) and has become known as *Hubbert Linearisation (HL)*. Figure 3.10 illustrates the application of this approach to US production data. Early in the production cycle the data shows considerable scatter, but after cumulative production exceeds 50Gb (corresponding to the mid-1950s) it settles down into an approximate straight line. Fitting a linear regression to this data leads to an estimate of 260Gb for the URR, which is comparable to that estimated from the production projection (Section 3.3.2).

Figure 3.8 Production versus cumulative production as an idealised parabola

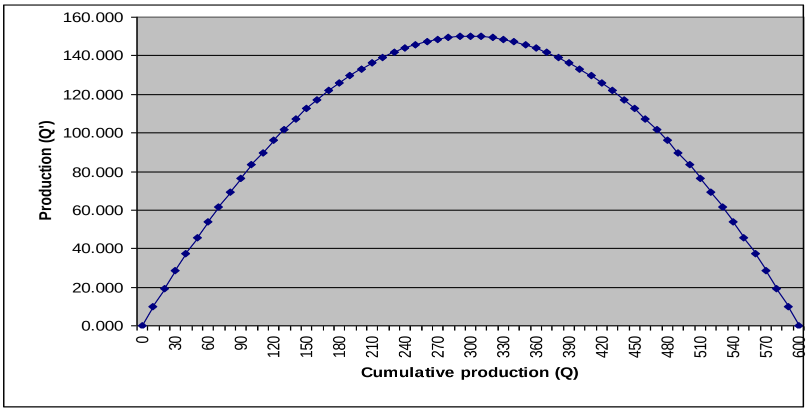


Figure 3.9 'Hubbert Linearisation' of parabolic relationship between production and cumulative production

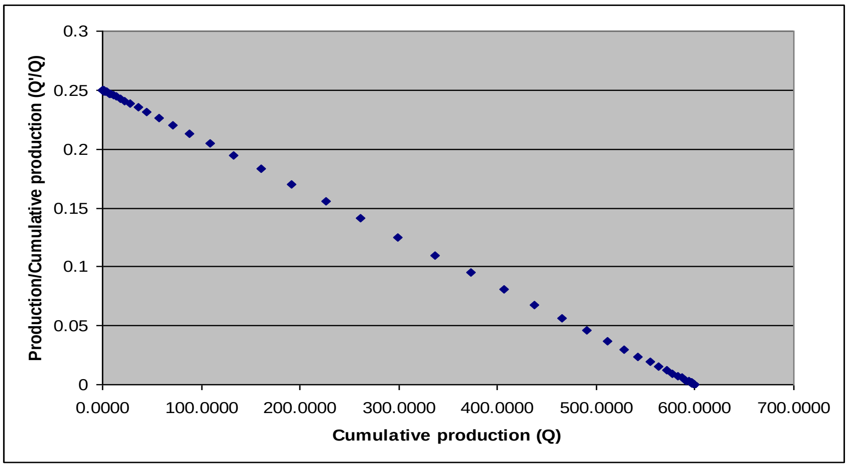
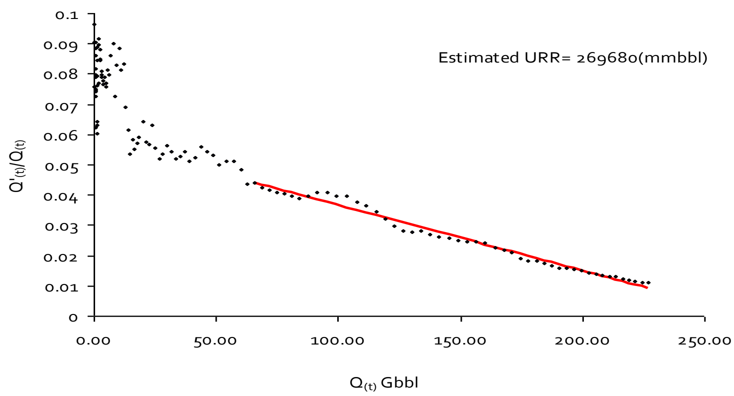


Figure 3.10 Hubbert Linearisation of US oil production



Hubbert first applied production projection in the early 1960s and only described the linearisation method much later. But in his comprehensive synthesis of extrapolation techniques, Hubbert *begins* with an assumed parabolic relationship between production

(*Q'(t)*) and cumulative production (*Q(t)*) (Equation 3.36) and uses this as a basis for deriving an equation for cumulative production over time (*Q(t)*) (Hubbert, 1982). By a process of substituting variables and integrating with respect to time, Hubbert eventually arrives at a version of the logistic equation:

$$Q(t) = \frac{Q_\infty}{1 + N_0 e^{-at}} \tag{3.38}$$

Where $N_0 = (Q_\infty - Q_0)/Q_0$ and Q_0 is the value of cumulative production for the first data point that is available. While Hubbert referred to the logistic equation as early as 1959, it was not until 1982 that he provided this formal derivation.[37] Hence, the logic by which the technique is presented in his 1982 paper may not necessarily reflect the sequence through which it was developed.

From Hubbert's work, it is clear that the assumption that $Q(t)$ takes a *logistic* form (Equation 3.38) is equivalent to the assumption of a *parabolic* relationship between $Q'(t)$ and $Q(t)$ (Equation 3.36) which in turn is equivalent to the assumption of a *linear* relationship between $Q'(t)/Q(t)$ and $Q(t)$ (Equation 3.37). Hence, if $Q(t)$ departs significantly from a logistic form, then a linear regression of $Q'(t)/Q(t)$ against $Q(t)$ is likely to provide a relatively poor approximation to the data. Deffeyes (2003) demonstrates that, if $Q(t)$ takes a cumulative normal form (which is very close to a logistic) the relationship between $Q'(t)/Q(t)$ and $Q(t)$ should be approximately linear through the latter part of the production cycle (i.e. once production is close to or past peak). But if $Q(t)$ is better approximated by other functional forms (e.g. the Cauchy or Gompertz), the 'Hubbert Linearisation' technique is unlikely to be reliable.

Hubbert Linearisation (HL) has proved very popular over the last few years, in part because production data is readily available and linear regression is relatively straightforward.[38] But in principle the technique is equivalent to non-linear regression of cumulative production against time assuming a logistic functional form. While analogous techniques are employed in population biology,[39] this does not remove concerns about the consistency of technique or its statistical robustness (Section 5).

3.3.2.2 Decline curve analysis

Although Hubbert did not make the connection, the linearisation technique is closely related to *decline curve analysis*, which has long been used by reservoir engineers to project the future production of individual oil wells and fields (Ahmed, 2006; Arps, 1945; 1956; Miller, *et al.*, 2009). As with production projection, decline curve analysis is based upon the assumption that past production trends and their controlling factors will continue into the future and can therefore be extrapolated using simple mathematical expressions. Decline curves project the future rate of production ($Q'(t)$) from a field or well once it has past its peak of production – which is normally relatively early in the field's life (Bentley, *et al.*, 2000).

---

[37] While Hubbert integrates Equation 3.35 to derive Equation 3.38, Deffeyes (2003) reverses the process and differentiates Equation 3.38 to derive Equation 3.35. The latter approach is rather easier to follow.

[38] Numerous examples are to be found on 'peak oil' web sites such as the Oil Drum.

[39] Deffeyes (2003) cites Smith (1963) as an early source.

Decline curves are normally characterised by three parameters: the initial production rate, the curvature of decline and the rate of decline. In a classic paper on the topic, Arps (1945) proposed three functional forms to model curvature, namely the exponential, harmonic and hyperbolic - with the first two being special cases of the third (Towler and Bansal, 1993). Subsequent authors have introduced additional functional forms, as well as developing more statistically robust approaches to estimation and prediction (Chang and Lin, 1999) and improving the theoretical basis of the technique (Li and Horne, 2007). However, it remains a largely ‘curve-fitting’ exercise in which the appropriate choice of functional form is likely to vary from one circumstance to another.

Exponential (i.e. constant percentage) decline is the simplest functional form and the one most widely used. It may be represented as:

$$Q'(t) = Q_0' e^{-bt} \tag{3.39}$$

Where Q_0' represents the level of production when it begins to decline (at t=0) and b is the decline rate. Integrating with respect to time and rearranging leads to:[40]

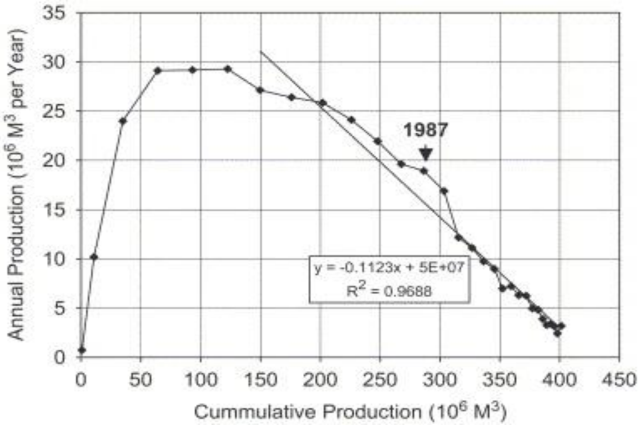
$$Q'(t) = b\left[Q_\infty - Q(t)\right] \tag{3.40}$$

Equation 3.40 states that the production is proportional to the amount of resource remaining to be produced ($Q_\infty - Q(t)$). This is the equation of a straight line, with a slope of $-b$ intersecting the vertical axis at b and the horizontal axis at Q_∞. Hence, if production from a field exhibits approximately exponential decline, the URR for that field may be estimated by plotting the rate of production against cumulative production, fitting a linear regression and extrapolating this regression until it crosses the *Q(t)* axis.

Figure 3.11 illustrates the application of this technique to the Forties field in the North Sea, leading to an estimated URR of approximately 420 million m^3. Figure 3.12 illustrates the corresponding production cycle (*Q’*(*t*)). A notable point is that the introduction of enhanced oil recovery techniques for this field in 1986 appears to have only temporarily increased production without having a significant impact on the URR. Similar patterns are observed in the Yates field in Texas and at Prudhoe Bay in Alaska, where EOR appears to have increased production at the expense of steeper decline rates in later years (Gowdy and Roxana, 2007). Whether this conclusion applies more generally is a topic of dispute.

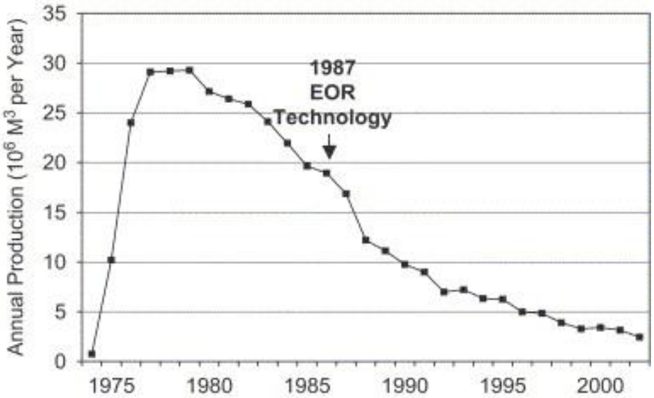
[40] Integrating with respect to time gives: $Q(t) = -\frac{Q_0'}{b} e^{-bt} + c$, or: $Q(t) = -\frac{1}{b} Q'(t) + c$. The constant c represents the *URR* - Q_∞.

Figure 3.11 Linearisation of exponential production declined for the UK Forties field



Source: Gowdy and Roxana (2007)

Figure 3.12 Production cycle of theUK Forties field



Source: Gowdy and Roxana (2007)

Laherrère (2001a; 2004) has applied this technique to large fields around the world to derive 'bottom-up' estimates of the URR of those fields. This forms part of the URR of the corresponding region, but an estimate of regional URR must also take into account the URR for smaller fields as well as those fields that have yet to begin production decline, those that have yet to be developed and those that have yet to be found. Exponential decline curves also form the foundation of bottom-up models of global oil supply, as developed by groups such as Energyfiles. But it is difficult for third parties to verify the corresponding URR estimates, since field-by-field production data is rarely available in the public domain. Also, a more robust approach to estimating URR would investigate alternative functional forms to model decline rates. While the harmonic model can be linearised by taking logs, the hyperbolic model requires the use of non-linear techniques. Since it is well established that the exponential model tends to underestimate the URR (while the harmonic model tends to overestimate it), the neglect of alternative models could lead to excessively conservative estimates of individual field (and hence regional) URR (Li and Horne, 2007).

Despite its close relationship to production projection, decline curve analysis appears to have neglected the logistic functional form. But in a comprehensive study of UK offshore oil fields, Kemp and Kasim (2005) found that the logistic model was preferred in the majority of the cases and that a better fit was obtained through a multi-logistic model, involving the sum of several cycles (Section 3.3.4). In this instance, estimates of URR based upon an exponential model would have been quite inaccurate. Kemp and Kasim also found that newer fields tended to have higher decline rates than older fields.

In summary, both Hubbert Linearisation and exponential decline curves take a non-linear relationship between cumulative production and time and convert this into a linear relationship between the rate of production and cumulative production. The former provides a straightforward means of estimating the URR for a region, while the latter provides a straightforward means of estimating the URR for individual well or field (Table 3.4). However, since both approaches assume a particular functional form for cumulative production they will only give consistent results if that form provides a reasonable fit to the historical data. Moreover, any forecasts using this approach will only be valid if this functional form continues to apply in the future.

Table 3.4 Comparison between Hubbert Linearisation and exponential decline curve

	Hubbert Linearisation	**Exponential decline curve**
Used for	Estimating the URR of aggregate regions	Estimating the URR of individual wells or fields
Variables	$Q'(t)/Q(t)$ versus $Q(t)$	$Q'(t)$ versus $Q(t)$
Assumptions	$Q(t)$ is logistic	$Q'(t)$ is negative exponential post peak
Only feasible if	Rate of increase in production ($Q''(t)$) is past peak	Production is past peak
More likely to be reliable if	Production is past peak	Decline is advanced and no enhanced recovery techniques are to be used

3.3.3 Summary

The production projection and production decline techniques are straightforward to apply and rely upon data that is readily available, relatively accurate and free from the complications of reserve growth. As a result, these techniques are very popular and may provide reliable estimates of URR in some circumstances. However, they are only useful for regions that are relatively advanced in their production cycle (and preferably past their production peak).

An important drawback of production projections is the lack of a robust basis for the appropriate choice of functional form. Different functional forms may often fit the production data equally well but yield very different estimates of URR. Multi-cycle models may often be more appropriate, but create the risk of ‘over-fitting’ and highlight the possibility of new production cycles occurring in the future.

While the production decline technique is computationally straightforward, it is equivalent to fitting a logistic curve to cumulative production and hence is no more reliable. Comparable techniques are often applied to the level of individual fields, but the neglect of alternative functional forms could lead the individual field URR to be underestimated. Finally, the techniques have several generic limitations, including: the limited theoretical basis; the neglect of economic, political and other variables that may modify the production cycle; the assumption that historically identified trends will continue to apply in the future; the tendency to apply to regions that are not geologically homogenous; and so on. These limitations apply to varying degrees to all extrapolation techniques and are discussed further below.

3.4 Discovery over time techniques

One drawback of using production projection as a basis for estimating URR is that the estimates are likely to be unreliable during the early stage of the production cycle. Indeed, estimates based upon the logistic function cannot be made if the rate of increase of production (i.e. $Q''(t)$) has yet to reach its peak. But since oil must be discovered before it is produced, the discovery cycle should be more advanced than the production cycle. Recognising this, Hubbert (1962; 1966; 1982) developed comparable curve-fitting techniques to estimate URR from time-series data on cumulative discovery and the rate of discovery. In much of this work, Hubbert's primary interest was the estimation of the future date of maximum oil production, which he assumed was preceded by the date of maximum oil discovery. Hence, the estimation of URR was often a secondary concern.

Hubbert's *discovery projection* and *discovery decline* techniques have since been adopted and developed by other authors, including in particular Laherrère (2003; 2004; 1999b; 2005). They have much in common with the production projection and production decline techniques described above and therefore raise a comparable set of issues and concerns. However, the use of discovery data introduces a number of additional complications, including the uncertainty in reserve estimates (especially for OPEC countries which hold the majority of the world's reserves), the relative suitability of 1P and 2P estimates, the choice between current and backdated measures of discovery, and the implications of 'smoothing' erratic discovery trends. The following discussion focuses primarily on these additional issues.

3.4.1 Discovery projection using current data

Hubbert's discovery projection technique is based upon the life-cycle model illustrated in Figures 3.13 and 3.14 (Hubbert, 1959). Hubbert assumed that both cumulative discovery - $D(t)$ - and cumulative production - $Q(t)$ - followed the same, broadly logistic functional form, with the former preceding the latter by some time interval Δt. Since the peak rate of discovery precedes the peak in production by Δt, identification of the former can form the basis for a prediction of the latter. Cumulative discoveries are calculated from the sum of cumulative production and declared reserves ($D(t) = Q(t) + R(t)$). When reserves reach their maximum value ($R'(t) = 0$), the rate of production (which is still increasing) is equal to the rate of discovery (which is decreasing) - $Q'(t) = D'(t)$.

Figure 3.13 Hubbert's idealised relationship between cumulative discoveries, cumulative production and proved reserves as a function of time

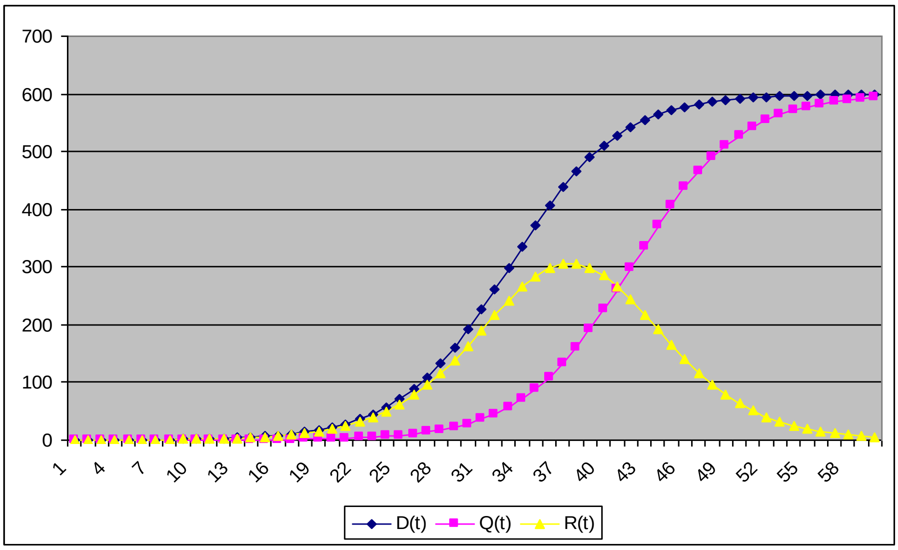
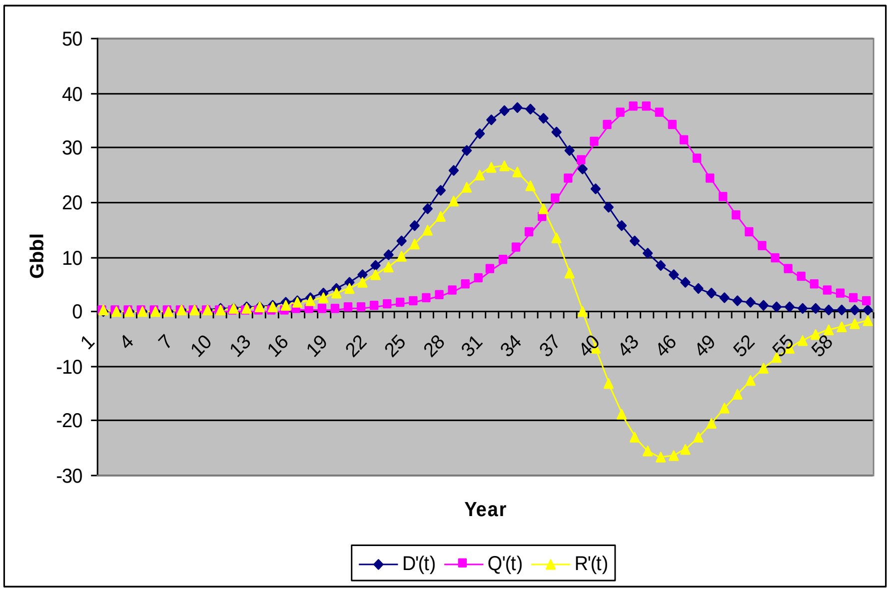


Figure 3.14 Hubbert's idealised relationship between rate of discovery, rate of production and reserve additions as a function of time



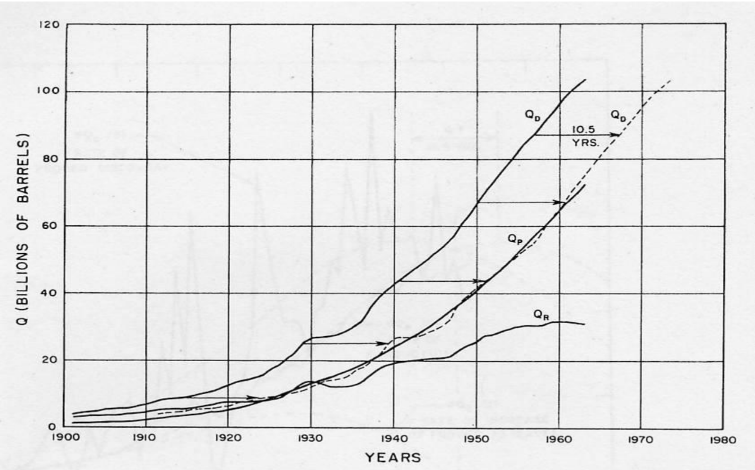
Using time-series data on cumulative production and proved reserves for the US, Hubbert (1962) estimated a time-series of cumulative proved discoveries. As shown in Figure 3.15, the plot of cumulative discoveries closely resembled the plot of cumulative production, but with the former preceding the latter by around 10.5 years. Hubbert took this data as broadly supporting his lifecycle model. By fitting a logistic curve[41] to the cumulative discovery data,

---

[41] Hubbert (1982) notes that: "….various forms of empirical equation were tested, but none gave satisfactory agreement with the data until finally the logistic equation was tried and found to fit the data with remarkable fidelity." But, Hubbert did not use nonlinear regression techniques to fit the equation and only introduced 'Hubbert Linearisation' some time later. Instead, he rearranged the logistic equation ($Q(t) = Q_\infty / (1 + N_0 e^{-a(t-t_0)})$) to give $\ln N = \ln N_0 - at$ where $N = (Q_\infty - Q)/Q$ and $N_0 = (Q_\infty - Q_0)/Q_0$ and obtained Q_∞ through an interactive graphical procedure.

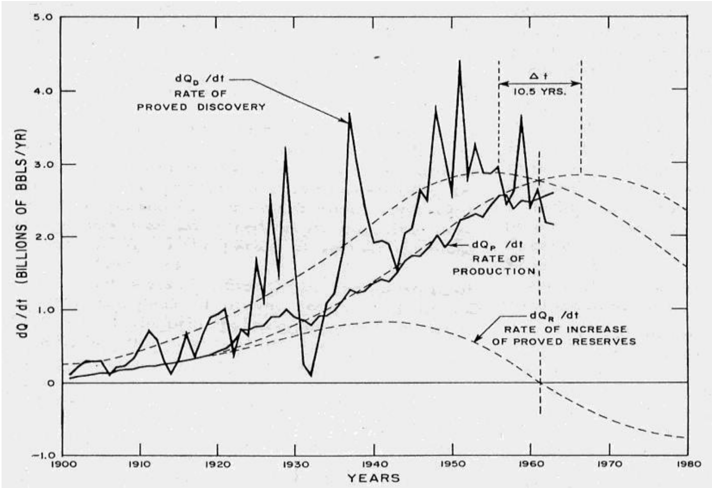
he estimated a URR for the lower 48 of 170 billion barrels and a peak year for discoveries of 1957. The estimates were subsequently confirmed by a linear regression of $D'(t)/D(t)$ against $D(t)$ (a discovery decline curve) (Hubbert, 1982). The results underpinned Hubbert's forecast of a peak in US proved reserves in 1962 and a peak in production around 1967. Hubbert updated his estimates in 1972 (after US production had peaked), when he again arrived at an estimate of 170 billion barrels for the lower 48 URR (Hubbert, 1974) and finally in 1982 when he lowered the estimate to 162 billion barrels (Hubbert, 1982). For comparison, the cumulative production of crude oil in the US through to 2007 was approximately 197Gb, but this includes Alaska and the deepwater Gulf of Mexico.

Figure 3.15 US cumulative proved discoveries, cumulative production and proved reserves from 1900 to 1962



Source: Hubbert (1966)

Figure 3.16 US rate of discovery, rate of production and rate of change of proved reserves from 1900 to 1962



Source: Hubbert (1966)

The use of discovery projection raises issues about the plausibility of the underlying assumptions, the appropriate choice of functional form and the appropriate treatment of multiple discovery cycles. These were discussed above and will be returned to in Section 4. In addition, users of the discovery projection technique frequently follow Hubbert in assuming that the discovery cycle is best modelled using the *same* functional form as the production cycle (e.g. a logistic) (Laherrère, 2000b). However, this is not a necessary assumption for estimating URR from a discovery projection and neither does it seem a plausible one (although it works fairly well for the US). The geological, economic, technological and political factors influencing discovery trends at different points in time are likely to be substantially different from those influencing production at a later point in time. In particular, the highly skewed field size distribution (Section 2.5) would be expected to lead to large peaks in discovery very early in the exploration history of a region. This would create an asymmetric (to the left) discovery cycle (possibly with large discontinuities representing the discovery of giant fields) which may not necessarily be reflected in the subsequent production cycle (Nehring, 2006b). For similar reasons, the time-lag between the peak in discovery and the subsequent peak in production (Δt) is unlikely to be predictable. Laherrère (2002a; 2002b; 2005) shows that the degree of correspondence between the discovery and production cycles can vary widely from one oil producing region to another. While the US discovery history is widely cited, it seems unlikely to be representative:

> "…..Because petroleum exploration in the US began very early, because the initial exploration and discoveries occurred in what has proved to be relatively minor basins, because early drilling technology was very limited in its drilling depth capabilities, and because discoveries in the major basins only hit their stride between 1910 in 1950, the US comes closest to a symmetric discovery curve of any major oil producing country or region." (Nehring, 2006b)

There are also concerns about the consistency and statistical robustness of discovery projections. For example, Ryan (1965; 1966) fitted logistic curves to US production and discovery data and found that they led to widely different estimates for URR. He also showed that much larger estimates of URR could be cited with equal justification and that the estimates increased rapidly with the addition of only a few more years of discovery data. In reply, Hubbert (1966) argued that the URR estimates based on discovery data should be considered more reliable since the discovery cycle was more advanced than the production cycle. Hubbert also claimed that much higher URR estimates could *not* be justified from the data and that the estimate could be verified by identifying the year of peak discovery (at which point cumulative discoveries should be one half of the URR).

Hubbert's replies are not wholly convincing and the points made by Ryan have been repeated by more recent critics. For example, Cavallo (2004) recreated Hubbert's original 1956 dataset and found that the R^2 for the best-fitting models changed only from 0.9946 to 0.9991 as the value of URR varied from 150 to 600 Gb. Similarly, Cleveland and Kaufmann (1991) fitted a logistic curve to US production data through to 1988 and found that the *adjusted* R^2 changed only from 0.9880 to 0.9909 as the value of URR varied from 160 to 250 billion barrels. Worse, they found the URR estimate from the cumulative discovery data to be highly unstable - thereby calling into question Hubbert's preference for discovery projection over production projection.

The estimation of URR from discovery data is also complicated by the highly erratic nature of the relevant time-series - illustrated most clearly in Figure 3.16. On several occasions, Hubbert addresses this by *smoothing* the discovery data by using an *N*-year running average. This procedure has also been followed by more recent authors, including Laherrère (2000b). But smoothing introduces difficulties, since the resulting data violates some core assumptions of statistical regression. In particular, the error terms from one time interval to another will no longer be independent and the variation of the error terms around the 'true' value will no longer be Gaussian (even assuming they were in the first place). As a result, the precision of the parameter estimates will be overstated and comparisons between different models will become invalid (Motulsky and Christopoulos, 2004b). These issues of consistency and statistical robustness will be explored further in Sections 4 and 5.

A point not picked up by earlier critics is Hubbert's reliance upon estimates of proved (1P) reserves to derive his time-series of cumulative discoveries. Proved reserves are acknowledged to be highly conservative estimates of remaining resources and subsequent authors such as Campbell (1997), Laherrère (2004) and Bentley *et al.* (2007) have advocated the use of proved and probable (2P) reserve estimates instead. While these are rarely available in the public domain (and were not available to Hubbert), they should provide a more realistic estimate of remaining resources. But since 2P reserve estimates are generally larger than 1P estimates, a discovery projection based upon the former would be expected to lead to a *higher* estimate of URR than one based upon the latter. The two sets of estimates would only be expected to converge when the discovery cycle was relatively advanced (at which point $R_t^{1P} \rightarrow R_t^{2P} \rightarrow 0$), but this was not the case for any of the time periods in which Hubbert was making his estimates. Hence, contemporary advocates of discovery projection techniques appear to making an argument that would call into question Hubbert's original results.

3.4.2 Discovery projection using backdated data

Discovery projection has been employed by Campbell and Laherrère (1995) to estimate URR for all oil producing regions around the world. But despite their debt to Hubbert, their method differs in two fundamental respects: first, they use 2P rather than 1P reserve estimates (derived from an industry database); and second, they use backdated measures of cumulative discoveries rather than current estimates (i.e. their measure of cumulative discoveries is $B^{2P}(t_d, t)$ while Hubbert's was $D^{1P}(t)$). Both have important implications for the results.

In an apparent dismissal of Hubbert's work, both Campbell and Laherrère argue that "…. proved reserves are useless for forecasting" (Laherrère, 2002b). In contrast, they claim that the 2P estimates contained in the Petroconsultants (now IHS) database provide an accurate estimate of the remaining recoverable resources.[42] Laherrère (2002b) argues that "…mean reserve growth will be close to zero, because some 2P values will grow while others will decrease, or disappear as in the case of fields that cannot be developed" (Laherrère, 2002b). In other words, the 2P estimates made at the time of field discovery should be relatively accurate. This implies that, at any point in time (*t*), the backdated cumulative discovery time-series (i.e. $B^{2P}(t_d,t)$)) should provide a fairly accurate estimate of the ultimately recoverable resources that have been discovered (i.e. $B^{2P}(t_d,\infty)$)). Although the estimated size of individual fields may change, the aggregate estimates should not be substantially different from those made at earlier times (i.e. $B^{2P}(t_d,t') \approx B^{2P}(t_d,t)$ where $t' > t$). Indeed, it is puzzling that Campbell and Laherrère advocate the use of backdated cumulative discovery estimates based on 2P data, but *also* claim that 2P estimates should not change much following field discovery. This seems contradictory: if 2P estimates are relatively stable, what is the advantage of using *backdated* discoveries?

The implication of Campbell and Laherrère's argument is that substantial revisions will *not* be made to cumulative 2P discovery estimates as a result of reserve growth. However, analyses of the IHS data set by Klett, *et al* (2005b) and Thompson *et al.* (2009b) demonstrate *substantial* growth in 2P estimates over time. Campbell and Laherrère have access to comparable data, but consider that much of this growth is due to unwarranted revisions. But if the apparent reserve growth is taken at face value, it implies that *both the 'height' and shape of the backdated cumulative discovery curves will change over time* (*t*). This is quite different from Hubbert's discovery projection analysis, where the cumulative discovery estimates for a particular point in time (*D*(*t*)) remain fixed. In the case of backdated estimates ($B(t_d,t)$), reserve growth leads to a change in the estimates for *earlier* years ($t_d \leq t$), while in the case of current estimates ($D(t)$), reserve growth simply contributes to the increase in cumulative discoveries in the *current* year (*t*).

Hubbert used current estimates of cumulative discoveries for his discovery projections, but was also one of the first to develop backdated estimates of cumulative discoveries (Hubbert, 1967). Hubbert's primary objective here was to estimate a reserve growth function ($G(\tau)$) for the US and thereby estimate the 'ultimate' amount of oil discovered in each year. However, he only used these backdated estimates when analysing discovery as a function of exploratory effort ($B(\varepsilon_d,t)$) and not when analysing discovery as a function of time ($B(t_d,t)$) - a choice that is not adequately explained (Hubbert, 1982). Hubbert found that the backdated discoveries per unit of exploratory effort ($B'_{\varepsilon_d}(\varepsilon_d,t)$) declined approximately exponentially, which implies that backdated cumulative discoveries as a function of effort $B(\varepsilon_d,t)$ can be approximated by the following functional form:

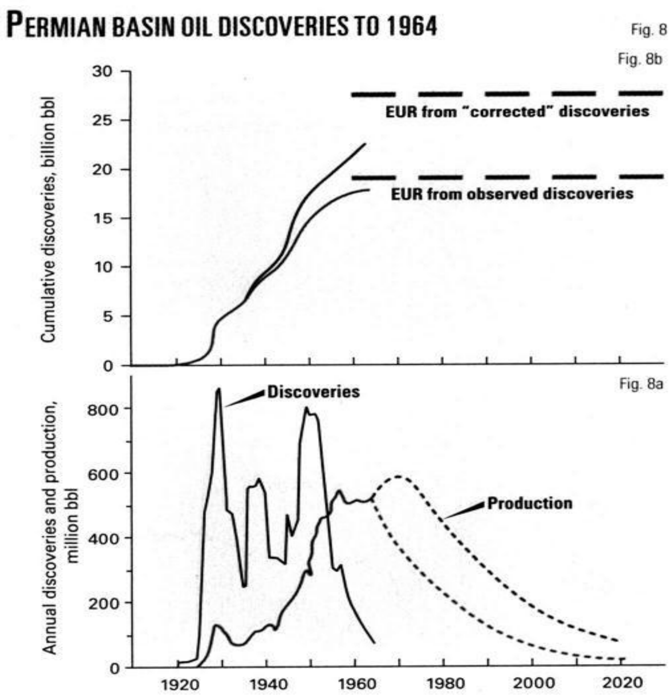
[42] However, Laherrère (2005) claims that the quality of the IHS estimates are deteriorating and highlights the large difference between recent estimates and those produced by a competitor, Wood Mackenzie.

$$B(\varepsilon_d, t) = B_{\infty,\infty}\left[1 - e^{-\beta\varepsilon_d}\right] \quad (3.41)$$

Where $B_{\infty,\infty}$ represents the URR to which the curve asymtopically trends. This functional form is consistent with the observation that most of the oil is contained within a small number of large fields that tend to be found relatively early in the discovery cycle (Section 2). Although exploration does not proceed at a constant rate, the plot of cumulative discoveries as a function of time $B(t_d, t)$ may be expected to have a broadly similar shape (i.e. exponential rather than logistic). In other words, *the shape of the discovery cycle based on backdated estimates may be expected to be different from that based upon current estimates*. The date of peak discovery may also be expected to be different, together with the time lag between peak discovery and peak production.

Nehring (2006a) employs discovery projection to estimate the URR from the Permian Basin and the San Joaquin Valley in the US. These regions have produced oil for more than 80 years and together account for around one quarter of the US URR. Unlike Hubbert, Nehring employs backdated discovery estimates and corrects these with a growth function to estimate the ultimate resources discovered in each time interval. The function is based on Hubbert (1967) and is only applied to the most recent 30 years of the discovery data.[43] Figure 3.17 shows cumulative discoveries in the Permian Basin through to 1964 and illustrates how the growth function leads to a higher estimate of the URR than would be obtained from the uncorrected data. This in turn leads to a later date for the forecast peak in production (assuming a symmetric production cycle)

Figure 3.17 Discovery projection for the Permian Basin using backdated discovery estimates through to 1964

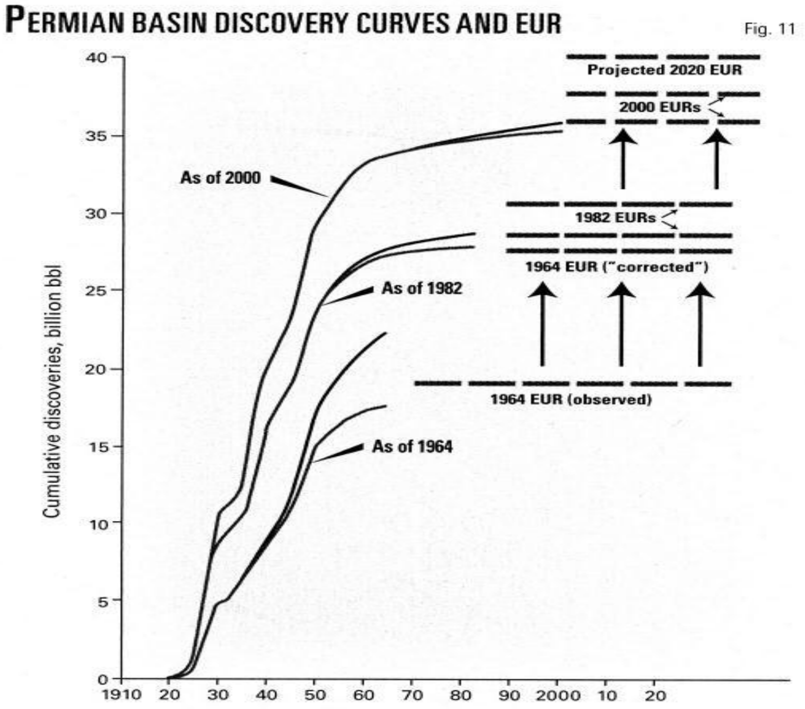


Source: Nehring (2006a)

[43] The growth function used by Hubbert declines to less than a 10% adjustment from known to ultimate for older discoveries.

Nehring then repeats this analysis using data through to 1982 and 2000 respectively (Figure 3.18). The results show clearly how the use of discovery projections can lead to misleading estimates of the URR. Since the estimated size of the discovered fields has increased over time (t) the cumulative discovery curves have increased in 'height' (i.e. $B(t_d, t') \geq B(t_d, t)$ for $t' > t$). This leads to a 36% increase in the estimated URR (from 27.5Gb as of 1964 to 37.5Gb as of 2000).

Figure 3.18 Discovery projection for the Permian Basin using backdated discovery estimates through to 1964



Source: Nehring (2006a)

While the purpose of 'correcting' the data is to avoid such underestimation, the results demonstrate that Hubbert's growth function seriously underestimates the reserve growth that actually occurred. A closer analysis reveals that the ultimate size of recent discoveries is estimated reasonably well, but the growth of older fields is greatly underestimated. As a result, the estimated peak date of discovery moves back in time while the corresponding estimate of the peak date of production moves forward in time - thereby undermining the notion of predictable time lag between the two. The peak in production in the Permian Basin actually occurred in 1974, but since the URR estimates have continuously increased, there has been a corresponding decrease in the estimated percentage of URR that had been produced by the time of peak. As Nehring comments:

> "…..The [discovery projection] method consistently underestimates future production because it consistently underestimates the ultimate recovery. It underestimates ultimate recovery because it is incapable of estimating the appreciation (growth) in the ultimate recovery that occurs in older fields...... the continuous upward movement in the cumulative discovery curve makes this curve useless as a tool for predicting the ultimate recovery. Estimates of ultimate recovery derive from cumulative discovery curves are only valid if one can guarantee that there

> will be no further increases in the ultimate recovery of discovered fields………no such guarantee can be made." (Nehring, 2006b)

Nehring further argues that this problem is also relevant to other methods of estimating URR, including discovery process modelling:

> "….Growth if the monkey wrench in the works of all such methods, particularly when it is unevenly distributed among fields and reservoirs (as it almost always is)." (Nehring, 2006b)

However, one should be careful about generalising these criticisms. First, Nehring's analysis is based upon US 1P estimates which may reasonably be expected to experience more reserve growth than the 2P estimates used by Campbell and Laherrère. Hence, if the latter had been used, the degree of underestimation of URR may have been less. Second, the observed reserve growth derives primarily from a small number of old, large fields which were subject to enhanced oil recovery techniques. The same opportunities may not be either available or appropriate for all fields and regions. Third, the analysis relies upon a reserve growth function that is nearly 40 years old and is only applied to the most recent 30 years data (Hubbert, 1967). Again, had a more representative function being used, derived using contemporary data and applied to the full time series, the degree of underestimation may well have been reduced.[44] Nevertheless, Nehring's analysis does raise some important questions regarding the reliability of extrapolation methods in general and the suitability of backdated discovery estimates in particular.

3.4.3 Summary

Discovery projection and decline curves have many similarities to production projection and decline curves. But since the discovery cycle is always more advanced than the production cycle, these methods should be applicable to a larger number of regions and could potentially provide more reliable estimates of regional URR.

These techniques raise similar concerns to those identified in Section 3.3, including the plausibility of the underlying assumptions, the appropriate choice of functional form (which could be different from that used for production projection), the appropriate level of aggregation and the treatment of multiple discovery cycles. In addition, the use of discovery data introduces a number of additional complications, including the uncertainty in reserve estimates, the relative suitability of 1P and 2P estimates, the implications of 'smoothing' erratic discovery trends and the treatment of reserve growth.

Reserve growth is of particular importance and suggests the need for backdated (B) rather than current (D) measures of discovery. However, the shape of the discovery cycle based on backdated estimates may be expected to be different from that based upon current estimates and both the 'height' and shape of the backdated discovery cycle will change over time. Hence, if reliable and consistent estimates of URR are to be obtained, the backdated discovery data needs to be adjusted to allow for future reserve growth. Given the paucity of data on reserve growth for different types, ages and sizes of field, the estimates of URR will be sensitive to the particular function employed.

[44] As shown in Section 2.4, there are plenty of more recent studies to choose from, which estimate reserve growth for periods up to 80 years.

3.5 Discovery over effort techniques

Regardless of whether current or backdated estimates are employed, the rate of discovery will be influenced by a variety of economic and political factors which could invalidate the extrapolation of historical trends. For example, the rate of discovery may fall as a result of economic recession rather than through the depletion of the underlying resource. Recognising this, Hubbert (1967) developed an alternative approach that examined cumulative discovery and/or the rate of discovery as a function of *exploratory effort* (ε).[45] In principle, this measure should be less sensitive to economic and political influences: for example, lower oil prices or political conflict may reduce exploratory activity as well as the number of new discoveries, with the result that the rate of discovery per unit of effort could remain relatively unchanged.

Variants of this approach have subsequently been employed by Campbell (1996) and Laherrère (2002b), who argue that it leads to more reliable estimates of URR than either production or discovery projection. In addition, this approach has much in common with statistical techniques for estimating URR that date back at least to the 1950s and go under the heading of discovery process modelling (Section 3.6). All these methods rely upon backdated measures of discovery, and (at least in principle) require some method for estimating future reserve growth. However, Hubbert's and Laherrère's curve-fitting techniques differ from discovery process modelling in at least three ways:

- The former are typically applied to relatively aggregate regions defined on political grounds (e.g. the UK) while the latter are applied to smaller regions defined on the grounds of geology and/or exploration history.
- The former can be used with an aggregate data from a region, while the latter requires data on individual fields.
- The former are straightforward and have relatively little theoretical support, while the latter are typically more complex with rather more basis in statistical theory.
- The former are most commonly used to provide single value estimates of URR while the latter typically provide probabilistic estimates.

But these are not absolute distinctions and the boundaries between the two approaches are somewhat blurred. The following section describes the analysis of cumulative discoveries as a function of function of exploratory effort (*creaming curves*), while Section 3.5.2 describes the analysis of discovery as a function of exploratory effort (*yield per effort curves*). Discovery process modelling is described in Section 3.6.

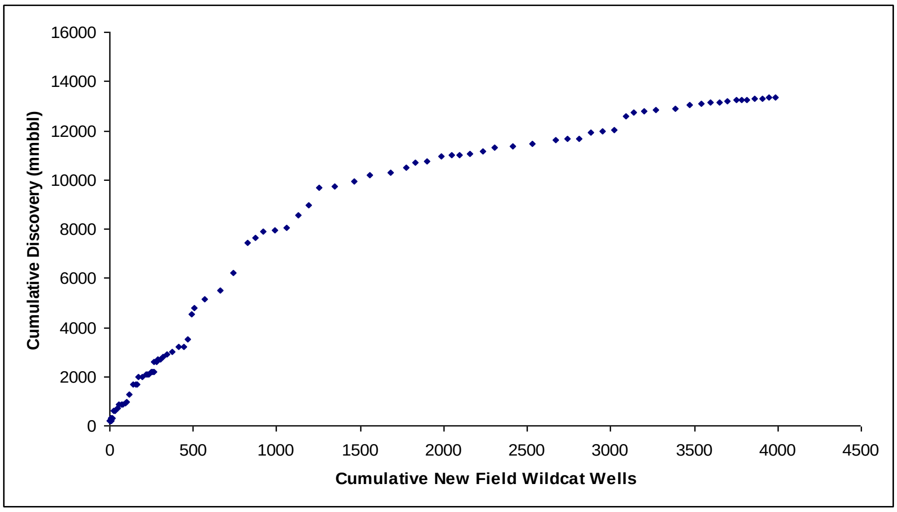
[45] Zapp (1962) was the first to use exploratory effort as an explanatory variable. However, Zapp assumed that the rate of discovery per unit of exploratory effort (yield) would remain unchanged, leading to an unrealistically large estimate for the US *URR* (590Gb). Zapp's approach was subsequently adopted by Hendricks (1965), who simply assumed that the yield would decline linearly. In contrast, Hubbert (1967) based his forecast of future yield upon a detailed analysis of past trends, which showed a negative exponential decline.

3.5.1 Creaming curves

3.5.1.1 The approach

A 'creaming curve' is a plot of backdated cumulative discoveries ($B(\varepsilon_d, t)$) against some measure of exploratory effort (ε_d). Creaming curves have been extensively employed by Laherrère (2002b), who measures exploratory effort through the cumulative number of exploratory wells drilled (termed 'new field wildcats', or NFWs).[46] Figure 3.19 shows a relatively 'well-behaved' creaming curve in which the 'yield' (i.e. $dB/d\varepsilon$) falls as the number of exploratory wells drilled increases, with the curve apparently tending towards an asymptope. The assumption underlying creaming curves is that the fall in the yield is due primarily to physical depletion of the resource.

Figure 3.19 Example of a creaming curve



The creaming curve hinges upon the notion of diminishing returns to exploratory effort. In principle, the amount of resources discovered for a given increment of exploratory effort should reflect the combined effect of two independent factors, namely: the success rate (the fraction of exploratory wells drilled that yield commercially viable quantities of oil) and the average size of the discovered fields (as measured by the estimated URR of each field) (Meisner and Demirmen, 1981).

Since there are a finite number of fields in each region, the probability of making another discovery should be inversely proportional to the number of fields already discovered. But at the same time, improvements in technology should allow more accurate identification of viable prospects, thereby increasing the success rate of exploratory drilling. Evidence suggests that the success rate of exploratory drilling in most regions declines only relatively gently, if at all, indicating that technical improvements partially or wholly offset the declining number of undiscovered fields (Forbes and Zampelli, 2000; Meisner and Demirmen, 1981).[47]

---

[46] In contrast, Hubbert (1967) used the cumulative length of exploratory drilling. Neither considered development drilling activity although this may make a major contribution to cumulative discoveries through the process of reserve growth. Similarly, neither considered the difficulties of classifying drilling activity as applying to either oil or gas resources.

[47] The IEA (2008) reports that, over the last 50 years, the global average success rate has increased from one in six exploratory wells to one in three. Similarly, Lynch (2002) reports that the average success rate in the US increased by 50% between 1992 and 2002 and Forbes and Zampelli (2000) report that the US offshore success rate doubled between 1978 and

But even if drilling were wholly random, large fields would tend to have a greater chance of being discovered than small ones since they generally occupy a larger surface area (Arps and Roberts, 1958; Meynerd and Shaman, 1975). Given the highly skewed field size distribution of most oil producing regions (Section 2.5), the largest fields should be found relatively early in the exploration history, with subsequent finds becoming progressively smaller. The creaming curve reflects the net effect of these two factors, with the declining field size tending to be more important.[48]

Laherrère (2002a; 2004; 2002b) has published creaming curves for a variety of regions around the world. Using backdated discovery estimates, he argues that the curves tend to rise steeply in the early stages of exploration, reflecting the discovery of small number of large fields. As exploration proceeds, the curves flatten as the discovered fields become progressively smaller. As an illustration, the (estimated) mean size of fields discovered in the Middle East before 1972 was 600 Gb, but this fell to 120 Gb over the subsequent 20 years (Laherrère, 2004). Although the exploration success rate actually increased after 1972 (from 24% to 33%), this was insufficient to compensate for the declining field size, with the result that the creaming curve flattened considerably.

As with discovery and production projection, the regional URR may be estimated by using non-linear regression to fit a curve to this data and identifying the value of the relevant parameter(s).[49] Alternatively, the URR may be estimated visually by identifying the asymptope towards which the curve is tending. In principle, the rate of discovery with respect to effort ($B'_{\varepsilon_d}(\varepsilon_d, t)$) could be plotted as a function of cumulative discovery with respect to effort ($B(\varepsilon_d, t)$) to give a *discovery decline curve to with respect to effort*. However, as with the *discovery decline curves with respect to time*, this approach does not seem to be widely used.

If backdated discovery estimates are employed, the creaming curve may rise very steeply in the early stages, reflecting the discovery of large fields. As a result, the logistic curve used for production and discovery projection may not provide a very good fit to the data. An alternative suggested by Hubbert is an exponential:

$$B(\varepsilon_d, t) = B_{\infty,\infty}\left[1 - e^{-\beta\varepsilon_d}\right] \qquad (3.42)$$

In various publications, Laherrère talks of fitting 'hyperbolas' to this data, but he never provides either the functional form for this hyperbola or statistical information on the goodness of fit. One possible functional form is the rectangular hyperbola, which may be defined as follows:

1995. In an econometric analysis, Forbes and Zampelli (2000) estimate that, over the period 1986-1995, technological progress increased the US offshore success rate by 8.3%/year.

[48] While a negative exponential decline implies that field are discovered in descending order of size, this need not imply that the industry has perfect information on field size and location. Instead, an approximately negative exponential decline may be generated from: a) a search strategy that is equal to or better than a random search; b) a skewed field size distribution; and c) a reasonably strong correlation between the aereal extent of the field and the volume of oil contained (Arps and Roberts, 1958; Cleveland and Kaufmann, 1991; Kaufman, 1975b).

[49] If the relationship is approximately exponential, an alternative is to take logs and estimate a linear regression.

$$B(\varepsilon_d, t) = \frac{(\varepsilon_d * B_{\infty,\infty})}{a + \varepsilon_d} \quad (3.43)$$

Section 4 describes our attempts to fit these functional forms to backdated cumulative discovery data from a variety of regions.

Although Laherrère (2004) states that the creaming curve was invented by Shell in the 1980s, variants of this approach have been used in the oil industry for much longer (Arps, *et al.*, 1971; Arps and Roberts, 1958; Harbaugh, *et al.*, 1995; Odell and Rosing, 1980b). Two employees of Shell published a paper on 'the creaming method' in 1981, but this describes a highly sophisticated (and not widely used) discovery process model that relies upon Monte Carlo simulation of trends in both success rates and average field sizes and assumes a lognormal field size distribution (Meisner and Demirmen, 1981). While Meisner and Demirmen model the same phenomenon as creaming curves, they do so in a completely different way using the data on the size of individual fields.

3.5.1.2 Limitations

Although the creaming curve method may have some advantages over production and discovery projection, it embodies a number of questionable assumptions, including:

- the data is sufficiently smooth to provide a good fit to the chosen functional form;
- the region is sufficiently homogeneous in geological terms for the field size distribution to contribute to the creaming phenomenon;
- exploration in the region has proceeded in a relatively orderly fashion; and
- the backdated discovery estimates (ideally adjusted for future reserve growth) provide a sufficiently reliable indication of the URR for the relevant fields.

Both theoretical reasoning and empirical evidence suggest that the data will frequently not provide a good fit to either exponential or hyperbolic functional forms. Given both the highly skewed field size distribution in most oil producing regions (including in some cases a King effect' where the largest field is very much larger than the rest), the data may exhibit very large 'jumps' in the early stages of exploratory effort (indicating the discovery of giant fields), followed by a long 'plateau' where the average size of discovered fields is very much smaller (Section 2.5). As a result, a smooth functional form may provide only a poor fit to the data. However, if the 'curve' is tending towards an asymptope, the URR may still be estimated visually.

Sneddon *et al.* (2003) show how exploration 'plays' (Box 2.2) with a long drilling history typically exhibit two or three plateaus, which could make it difficult to fit a single curve to the whole dataset. Sneddon *et al.* also provide some technical reasons for why this may be the case (related to the nature and accessibility of different oil-bearing structures) and illustrate this with examples from the Gulf of Mexico, Indonesia, Texas, and Norway.

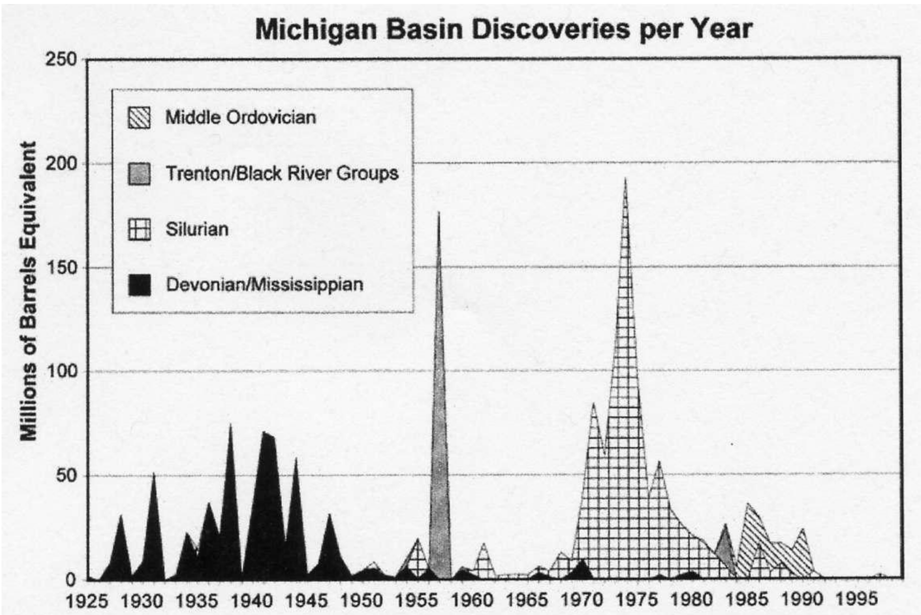
The creaming curve may potentially be 'smoothed' by using data from a larger geographical region, but this may not necessarily lead to a smooth curve since the highly skewed field size distribution applies at all levels of aggregation, including the world as a whole (Section 2.5). Moreover, the use of data from larger geographical regions introduces difficulties of its own. While diminishing returns to exploratory effort are widely observed at the level of the exploration play, the same may not necessarily be observed at larger geographical scales:

"…..The finding of larger fields early in exploration is not necessarily true at scales other than the play level because larger plays are not necessarily developed earlier. The factors that tend to make larger fields within a play be found earlier do not make the larger plays be developed first. On the contrary, plays tend to be developed in order of ease of exploration and development - often those with shallow oil reservoirs or easily detectable structural traps are developed first. Some large plays may not be developed until technological improvements can make them viable." (Charpentier, 2003)

Charpentier (2003) gives the example of the Michigan basin (Figure 3.20), where a combination of geological accessibility and improvements in exploration and production technology led to a multi-cycle exploration history in which the largest (Silurian) play was developed relatively late. If aggregate data were used, such a discovery history would most probably lead to a 'stepped' creaming curve, formed from four individual creaming curves reflecting the exploration history of the four individual plays. If the asymptope of the aggregate creaming curve was identified *before* all the plays had been opened up to exploration, the total basin URR would be underestimated.

Similar phenomena are reported by Wendebourg and Lamiraux (2002) in their study of petroleum resources in the Paris basin. This area experienced two major exploration cycles, corresponding to the opening up of two regions of sedimentary rocks. A creaming curve estimated using data through to 1986 leads to an estimate of 15MT for the URR of the Paris basin, but a similar curve estimated using data through to 1996 (after exploration had begun in the second region) leads to a much larger estimate of 46Mt.

Figure 3.20 Exploration history of the Michigan basin

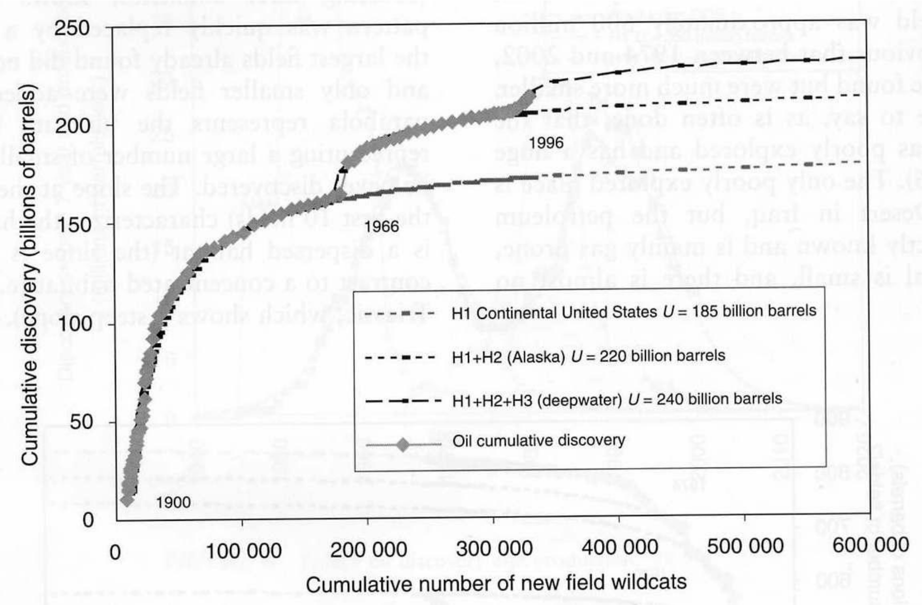


Source: Charpentier (2003)

The larger the geographical region for which the creaming curve is estimated, the more significant the problem of multiple exploration cycles is likely to become. This difficulty is recognised by Laherrère (2003; 2004), who typically models the discovery history of individual countries (or larger regions) as the sum of two or more creaming curves. Figure

3.21 gives the example of the US, which Laherrère models with three creaming curves, corresponding respectively to the continental region, Alaska and the Gulf of Mexico.[50]

Figure 3.21 Laherrère's creaming curve analysis of the United States



Source: Laherrère (2004)

The use of multiple curves to simulate the exploration history of multiple geological regions leads to three difficulties however. First, there is the risk of 'over-fitting', which means the use of an overcomplicated model with an excessive number of parameters to describe a small data set. The risk of over-fitting will increase with the number of curves that are used. While there is no general rule to determine precisely when a model is overfit, it could potentially destroy the ability of the model to generalize beyond the available data (see Section 4).[51]

Second, Laherrère never provides any statistical support for his choice of curves and in many cases the appropriate number of curves is not clearly apparent from the data. Indeed, the chosen number of curves appears to differ from one publication to another. For example, Laherrère (2003) models the oil and gas resources of the Middle East with two creaming curves but in a subsequent paper this has increased to four (Laherrère, 2004). While the choice could be based upon knowledge of the exploration history of a region (e.g. the level of exploratory effort at which different regions or depths were opened up to exploration), this information is rarely provided. Instead, the number and location of curves appears to be determined largely from the 'shape' of the data, without any statistical justification being provided and with any interpretation in terms of exploration history being made *ex post*.

Third, if multiple exploration cycles have occurred in the past, it seems reasonable to assume that *new* exploration cycles may occur in the future. The potential for these cannot be established from the statistical analysis of historical data, but only from a detailed evaluation

[50] Laherrère provides no information on statistical fit, but the data appears to give little very confidence on the shape of the third (Gulf of Mexico) creaming curve.

[51] The use of multiple curves, implying individual *URR*s, could also be problematic if (as is usually the case) the discovery history reflects the outcome of exploration proceeding in multiple regions simultaneously.

of the geological potential of the region, including an assessment of how various political, economic and technical constraints may have influenced exploration trends in the past. Failure to anticipate the possibility of new exploration cycles (and to estimate their size) will lead the creaming curve technique to *underestimate* the regional URR.

The probability of new exploration cycles may be expected to be proportional to the geographical size of the region and inversely proportional to its exploration maturity. Hence, for small, geologically-defined regions where exploration is well advanced, the probability of new cycles may be relatively low, while for large, politically-defined regions, which are largely unexplored (e.g. owing to the depth of drilling required, or geographical remoteness, or political restrictions) the probability of new cycles may be relatively high. Since both Laherrère and Campbell estimate creaming curves for large, politically defined regions, the reliability of their estimates of regional and global URR depend heavily on the assumption that any new exploration cycles will have only a small impact on aggregate resources (either because there will be few such cycles or because the discovered resources will be relatively small). While this judgement appears reasonable for many regions around the world, it remains problematic for key regions such as Iraq. Unfortunately, these are precisely the regions that have a disproportionate influence on global URR estimates and the associated forecasts of global oil production. This inability to anticipate new exploration cycles applies to all the extrapolation techniques discussed in this section and is of considerable importance. It is the key reason why advocates of discovery process modelling warn against applying such methods to large regions:

> "….In making forecasts we should be aware of the particular geographic area or geological domain for which we are making forecasts, by excluding from consideration those parts of the geographic area or geological domain that were not, and could not be, explored before. Hence, for example, deep sea prospects should never be lumped with land prospects in making a forecast, and similarly prospects located in a previous concession area should be kept separate from prospects in an area that was never opened to exploration." (Meisner and Demirmen, 1981)

Since the aggregate data used by Laherrère and others fails to make these distinctions, their methods may underestimate the regional URR. At the same time, more disaggregate analysis is likely to be resource intensive and require access to proprietary, field-level data. What matters, therefore, is the size of the error that may result from using the simpler techniques. Views differ on this issue, but the percentage error for *global* URR estimates should fall over time as more of the world's oil bearing regions become comprehensively explored.

Creaming curves have two other weaknesses that are also shared with the discovery projection techniques described earlier. First, the backdated discovery estimates should be adjusted to allow for future reserve growth. This adjustment was made by Hubbert (1967) who was using 1P reserve data for the US,[52] but is *not* made by Laherrère or Campbell who are using 2P reserve data for the world. To the extent that URR estimates based upon 2P data *do* grow over time (as Klett *et al.* (2005a) suggest), Laherrère's and Campbell's creaming curves are likely to underestimate the regional and global URR. While their creaming curves may trend towards asymptotes, the same curves estimated at a later point in time may trend towards *different* asymptopes since reserve growth increases the 'height' of the curves.

[52] The reserve growth function estimated by Hubbert (1967) suggests that the *URR* of discovered fields will be 5.8 times larger than the initial 1P reserve estimate for those fields and that half of this growth will occur in the first nine years following field discovery.

Illustrating this with data from the North Sea, Lynch (2002) argues that the neglect of reserve growth amounts to "....comparing old orchards with newly planted saplings and extrapolating to demonstrate declining tree size".

Hence, the use of the creaming curve technique *without* allowing for future reserve growth is highly questionable. But while accuracy should be improved by allowing for future reserve growth, the results will still be sensitive to the particular growth function that is employed. As Nehring (2006a; b; d) has shown, it is quite possible for such functions to underestimate URR by underestimating the amount of reserve growth in particular categories of fields (of course, the converse is also a possibility).

A final weakness of creaming curves is the uncertainty and inaccuracy of much of the relevant discovery data and the lack of a consistent exploration history in many regions. Davies (1981) highlights how datasets may be significantly biased by political restrictions in accessing regions, the economic incentives to under or overestimate discovery sizes, the dependence of recorded discoveries on contemporary oil prices and a variety of other factors. As a result, Davies (1981) argues that creaming curves are only useful for individual companies with access to all the relevant data relating to a relatively small and geologically homogenous region. These conditions rarely apply and the data will be particularly suspect for those regions, such as the Middle East which hold the largest resources.

3.5.2 Yield per effort curves

Many of the same difficulties are faced by a second, closely related technique for estimating URR, namely the *yield per effort* (YPE) curve. While a creaming curve is a plot of backdated cumulative discoveries ($B(\varepsilon_d, t)$) against exploratory effort (ε_d), a yield per effort curve is a plot of the backdated discovery rate ($B'(\varepsilon_d, t)$) against exploratory effort. Since both techniques utilise the same information, they are effectively equivalent. While an estimate of URR may be derived from the asymptope of the creaming curve, a corresponding estimate may be derived by integrating the yield per effort curve:

$$B_{\infty,\infty} = \int_0^{\infty} B'(\varepsilon_d, t) d\varepsilon_d \qquad (3.44)$$

The yield per effort curve was introduced by Hubbert (1967) who, after painstakingly constructing the relevant time series, analysed the yield in the lower 48 US states over the period 1860 to 1967. Hubbert's measure of discovery was based upon 1P reserve estimates while his measure of exploratory effort was the cumulative footage of exploratory drilling. The latter was combined into units of 10^8 feet - subsequently turned 'Hubbert Units' or HUs (Cleveland and Kaufmann, 1991). At the time of Hubbert's analysis, cumulative exploratory drilling in the US amounted to 15 HUs - with the first HU occurring over a period of 61 years and the last nine HUs requiring only two years each. Hubbert observed that the yield exhibited an approximately exponential decline, which he approximated with the following functional form:[53]

[53] Hubbert justified his use of a negative exponential function as follows: "…. the assumption of approximately negative exponential decline is justified by the fact that as more fields are discovered, the volumetric density of the remaining fields becomes progressively less, and the probability of discovery by any given amount of drilling must, in the long-run, decline continuously " (Hubbert, 1967)

$$B'(\varepsilon_d, t) = Ke^{-\beta\varepsilon_d} \tag{3.45}$$

Integrating between zero and ε_d yields an expression for backdated cumulative discoveries ($B(\varepsilon_d, t)$):

$$B(\varepsilon_d, t) = \frac{K}{\beta}(1 - e^{-\beta\varepsilon_d}) \tag{3.46}$$

Letting $\varepsilon_d \rightarrow \infty$ gives an expression for URR:

$$B_{\infty,\infty} = \frac{K}{\beta} \tag{3.47}$$

However, Hubbert did not use standard statistical techniques to estimate the 'best fit' exponential curve. Instead, he imposed two restrictions: a) the curve had to pass through the last observation for the rate of discovery; and b) the area under the curve had to equal the backdated cumulative (proved) discoveries through to the last observation. This led to an estimate of ~170Gb for the US URR which was consistent with his earlier estimates from production and discovery projection (Hubbert, 1962).

In a subsequent evaluation of this approach, Harris (1977) criticised Hubbert's method of curve-fitting. As well as violating standard statistical procedures, the method places excessive weight on the estimated size of newly discovered fields that have only been partially explored and tested (i.e. the last data point). Harris fit a curve to this data using standard OLS techniques and obtained an URR estimate of only 134 Gb, which was *less* than the cumulative discoveries through to 1966. Harris also used Hubbert's method to fit curves to *truncated* time-series (from 5 through to 14 HUs respectively) and found a strong trend towards *higher* estimates of URR as the amount of exploratory effort increased – suggesting that the method led to systematically biased estimates. Moreover, the only reason that Hubbert's estimate was consistent with his earlier work was that the discovery rate had recently increased - something which Hubbert considered to be both anomalous and temporary. This combination of poor fit to the data and inadequate theoretical support for the model was considered by Harris to be a serious weakness:

> "…. Use of a fitted exponential which is not in fact a good fit the data must invoke some scepticism about the reliability of the predictions by the model. If the selected model is strongly indicated by theory, it may be accepted by critics even when the fit to available data is not a good one. But when theory is not the basis for the model, the entire responsibility for generating confidence in the selected model rests upon the suitability of the model as a representation of the data." (Harris, 1977)

Echoing a point made above for creaming curves, Harris also argues that the exponential model is inappropriate since the relevant region (the continental US) is not geologically homogeneous. He uses a numerical example to show that an aggregate discovery trend will not necessarily be exponential, even if the trends for individual regions can be approximated by this form. As a result, the URR estimated from a curve fit to the aggregate data may be quite different from the sum of the estimates from two or more curves fitted to the regional data:

> "…. even if each mode of occurrence in each province were to possess the negative exponential functional relationship..., since the provinces were explored at different times and usually possessed oil reservoirs at more than one depth, only a unique set of circumstances would cause

> the data aggregated across all depths and provinces for the entire conterminous US to exhibit a negative exponential pattern. It must follow that it is also dangerous to suppose that future data must conform to the assumed negative exponential, or for even that matter to a decreasing trend."(Harris, 1977)

This implies that: a) a yield per effort or creaming curves should only really be fitted to relatively small and homogeneous regions; b) the URR for a larger region should be estimated from the sum of curves for smaller regions; c) oil and gas should be studied separately; and c) for large regions, the possibility of future exploration cycles should always be considered.

Harris also criticised Hubbert's method for neglecting variables other than physical depletion that influence the discovery history and which lead to departures from the negative exponential decline. These include the tendency to drill the more accessible formations first (which are not necessarily the largest), changes in exploration and drilling technology, and the varying influence of costs, prices and other incentives on drilling activity - including in particular licensing rules and government subsidy schemes. As the result of these factors, the exponential curve provides a particularly poor fit to the early years of US exploration activity, since the efficiency of the industry in finding large fields was lower than that predicted by random search (Menard and Sharman, 1975).[54]

By neglecting all these variables, Hubbert effectively assumed that the effects of physical depletion would outweigh them – at least over the time-scales that are relevant to the estimation of URR. However, this need not necessarily be the case, and even if it was the inclusion of additional explanatory variables could potentially improve the resulting estimates. This argument is explored further in Section 5 which provides a closer look at the statistical basis of curve fitting techniques.

3.5.3 Summary

Creaming curves and yield per effort curves are equivalent techniques that may offer significant advantages over production and discovery projection. However, they must still be used with care. Difficulties can arise with the accuracy of the relevant data, the required corrections for future reserve growth, the poor fit of particular functional forms, the existence of multiple exploration cycles and the failure to anticipate new exploration cycles in the future. These difficulties will be more significant when the technique is applied to large geographical regions that have neither a homogeneous geology nor a consistent exploration history. Overall, these difficulties appear more likely to lead to *underestimates* of the regional URR.

The accuracy of these methods may be improved through the explicit allowance for reserve growth and a focus on smaller geographical regions. But the information required to estimate and apply reserve growth functions do not be available and a more disaggregated analysis is likely to be resource intensive, as well as requiring access to confidential data. While a more aggregate analysis may potentially lead to biased estimates, the size of the error may vary widely from one region to another and may be expected to decline in the future as exploration

[54] Menard and Sherman (1975) use Monte Carlo simulation to show that if exploratory wells had been drilled at random in the US, the super giant East Texas oilfield would have been discovered at the very least by 1902. But in practice, East Texas was not discovered until 1930.

matures. For regions such as the continental United States, the error may already be relatively small.

3.6 Discovery process models

Discovery process models have many similarities with the ‘curve-fitting’ approaches described in the previous sections, but are typically more sophisticated.[55] Both approaches implicitly assume that the field size distribution is highly skewed, with most resources tending to be found in a small number of large fields. They also assume that large fields tend to be found first, leading to diminishing returns to exploratory effort and declining field sizes as exploration proceeds. Discovery process models also overlap with the field size distribution techniques discussed in Section 2, but do not necessarily assume a particular size distribution.

Perhaps the most important difference between discovery process models and curve-fitting is that the former require data on *individual fields*, while the latter use *aggregate* data for a region. Also, the former are typically used to study the discovery process in geologically homogeneous areas with a relatively consistent exploration history, while the latter are applied to more aggregate regions, frequently defined on political grounds (such as a country). While in many ways a superior technique, the extensive data requirements of discovery process models can make them both impractical for many researchers and problematic for estimating the URR of larger geographical regions.

Most discovery process models are based upon statistical analyses of the number and size of discovered fields as a function of either discovery sequence, time or exploratory effort. This may be combined with assumptions about the field size distribution and/or information about the location of fields. Most of these models simulate a probabilistic law governing the process of new field discovery and can be used to provide forecasts of the number, size and sequence of future discoveries together with the success rate of exploratory drilling. They may also be combined with economic models to estimate the anticipated returns to exploratory drilling and improved through the incorporation of economic variables influencing the success of exploratory drilling. Hence, the estimation of URR for a region is frequently a secondary concern.

At least ten different discovery process models exist, either as original formulations or significant modifications of such formulations. The following discussion is confined to the two approaches that appear to be most widely used and which originate from Arps and Roberts (1958) and Barouch and Kaufman (1975) respectively. Table 3.5 identifies the explained and explanatory variables for these models, while Table 3.6 summarises the mathematical notation used in the subsequent sections. As before, the ‘size’ of the field refers to an estimate of the URR from that field.

---

[55] According to Kaufman (1975b): "…. ‘Discovery process’ is a descriptive label for the sequence of information gathering activities (e.g. seismic surveys) and acts (drilling of exploratory wells) that culminates in the discovery of petroleum deposits. In building models of the discovery process, we will regard it as being effectively described by a small number of quantitative attributes (such as the number of exploratory wells drilled into a geological formation in a given area and the oil in place in the newly discovered pools and postulated relationships among them." Major references on discovery process modelling include Arps and Roberts (1958), Barouch and Kaufmann (1975), Kaufman (1975b; 1993) Schuenemeyer and Drew (1994) and Meisner and Demirmen (1981).

Table 3.5 Classification of discovery process models by explained and explanatory variables

Group	Method	Explained variable	Explanatory variable
Discovery process modelling	Arps-Roberts model	Number of fields in each size class	Effort
	Barouch-Kaufmann model	Probability of discovering a particular size class of field	Number and size of previously discovered fields

Table 3.6 Mathematical notation for discovery process models

Notation	Definition
t	Time
ε	Effort
i	Discovery number (i=1 represents the first field to be discovered)
k	Size class of field (k=1 represents the smallest size class)
$V_{k,t}$	Mean size of fields in size class k as estimated at time t
$V_{i,t}$	Size of ith field, as estimated at time t
$V_{i,\infty}$	URR of ith field
A_i	Surface area of ith field
A_k	Mean surface area of fields in size class k
$N_{k,t}$	Number of fields in size class k discovered up to time t
$N_{k,\varepsilon}$	Number of fields in size class k discovered up to exploratory effort ε
$N_{k,\infty}$	Ultimate number of fields in size class k
$M_{k,i}$	Number of discovered fields in size class k which precede the ith discovery.

3.6.1 Arps-Roberts model

In a classic paper, Arps and Roberts (1958) introduced discovery process modelling, provided one of the first pieces of evidence for a lognormal distribution of field sizes and highlighted the importance of economic truncation in determining the observed distribution of field sizes. Versions of their approach have subsequently enjoyed widespread use in resource appraisal, most notably by the USGS (Drew, *et al.*, 1995).

Arps and Roberts investigated the history of oil discovery in the east flank of the Denver-Julesburg basin in Colorado. This was an area of 5.7 million acres in which oil at first been found in 1930. Between 1949 and 1958 a total of 9504 wells had been drilled in this 'play', of which 5035 were exploratory wells ('wildcats'). This led to the discovery of 338 oil fields of widely varying sizes. The combination of a geologically homogeneous region, unrestricted exploration throughout the region and a relatively large sample size of wells and fields made the area an excellent candidate for statistical examination.

Arps and Roberts first investigated the relationship between the surface area of each field (A_i) and current estimate of the URR of these fields ($V_{i,t}$), as derived from decline curves and/or volumetric analysis (the potential for secondary recovery was ignored). They found that the estimated URR was approximately proportional to the 1.275 power of the surface area ($V_{i,t} = 530A_i^{1.275}$), suggesting that average recovery per unit area ($V_{i,t} / A_i$) improved as the fields grew larger. They then arranged the 338 fields into 15 size classes (k), based upon the estimated URR of each field - with each class representing an mean URR approximately twice that of the preceding class ($V_{k+1} \approx 2V_k$). A plot of the number of discovered fields within each size class against the natural log of the average URR within that size class was approximately normal - suggesting that the underlying field-size distribution was lognormal.

Arps and Roberts used the number of exploratory wells drilled as their measure of exploratory effort (ε) and postulated that the probability of finding another field of a particular size class for each additional exploratory well must be proportional to the product of the number of undiscovered fields of that size remaining and the average surface area of such fields. They proposed that the *success rate* (the number of discovered fields divided by the number of exploratory wells) for each size class (k) should decline over time in accordance with the following functional form:

$$\frac{dN(k,\varepsilon)}{d\varepsilon} = CA_k\left[N(k,\varepsilon) - N(k,\infty)\right] \qquad (3.48)$$

Where C is a constant representing *discovery efficiency* (with larger C implying more rapid discovery). Discovery efficiency depends upon the exploration method used, but Arps and Roberts assumed that it was independent of the field size class and did not change over time time. Integrating this equation leads to:

$$N(k,\varepsilon) = N(k,\infty)(1 - e^{\varepsilon CA_k}) \qquad (3.49)$$

This may be interpreted as a 'creaming curve' (although Arps-Roberts did not use that term), for the <u>number</u> of discoveries in a particular size class (k) as a function of the cumulative number of exploratory wells (ε). As ε increases, $N(k,\varepsilon)$ increases towards an asymtope which provides an estimate of the ultimate number of fields in this size class ($N_{k,\infty}$). Multiplying this number by the estimated average URR of fields in this size class ($V_{k,\infty}$) and summing over all size classes leads to an estimate of the URR for the region ($URR = \sum_k N_{k,\infty}V_{k,\infty}$).

Arps and Roberts reasoned that if drilling was entirely random and the total area covered by the discovered fields was small compared to the total area of the basin (B), then the constant C must be equal to *1/B*. in other words, the probability of hitting a field in size class *i* would be equal to the ratio of the total surface area of remaining fields of that size to the area of the unexplored part of the basin. If drilling was instead informed by geological assessments, C was likely to be greater than *1/B*. On the basis of the historical record of drilling success in the United States, they estimated a value of 2.75 for C which was assumed to be constant for all size classes and over time.

Given historical data on exploratory drilling activity and the estimated size of discovered fields within the region, the Arps-Roberts model provides a useful basis for estimating the size distribution of undiscovered fields, the future discovery rates for different sizes of field and the ultimately recoverable resource for the region. Although relatively simple in structure, it has produced apparently accurate forecasts for a number of exploration plays, basins and provinces. For example, subsequent analysis by Attanasi *et al.* (1981) found that the Arps-Roberts discovery predictions for the 1956-74 period were fairly accurate.

3.6.1.1 Use of the AR model by the USGS

Beginning in 1975, the US Geological Survey (USGS) has employed versions of the Arps-Roberts model in their assessment of the oil and gas resources of the United States (Drew and Schuenemeyer, 1993). Early applications of this technique to the Permian Basin and the Gulf of Mexico led to two important modifications (Drew, 1997). First, since discovery efficiency typically increases with field size, the USGS specified discovery efficiency as a function of the size class (C_k, with $C_k \geq 1.0$) and used non-linear regression to simultaneously estimate the relevant parameters.[56]

Second, the method tended to underestimate the number of undiscovered small fields since the observed field size distribution was subject to 'economic truncation' (Drew and Schuenemeyer, 1993). If future trends in technology and prices led to smaller fields becoming economically viable, this could lead to a significant underestimate of total undiscovered resources. The USGS corrected for this with a two-stage procedure. First, the Arps-Roberts model was used to estimate the number of fields in each size class *larger* than the mode. Second, the number of fields in each size class *smaller* than the mode was estimated by assuming a *log-geometric* field size distribution, in which the ultimate number of fields in each size class was a multiple of the next larger size class ($V_{k-1,\infty} = rV_{k,\infty}$). Estimates of the log-geometric multiplier (*r*) were derived in part from geological assessments and usually varied between 1.5 and 2.0 (Root and Attanasi, 1993).[57] While this procedure improved upon the unadjusted model, the results were sensitive to the assumed field size distribution and the estimated values for *r*.

The USGS use the modified Arps-Roberts model to forecast future discovery rates in the Gulf of Mexico (Drew, *et al.*, 1982a). A subsequent evaluation showed that discovery rates were consistently underestimated across all size classes (Drew and Schuenemeyer, 1992). But while the volume of oil discovered was underestimated by as much as 50%, the number of fields discovered was only underestimated by 9%. Closer examination revealed that the primary source of the underestimation was the neglect of future reserve growth in the discovered fields. On average, the estimated URR of discovered fields (based upon 1P estimates) nearly doubled between 1977 and 1988, with the result that many fields moved from one size class to another. Hence, as with the simpler yield per effort models, it is necessary to estimate future reserve growth if discovery process models are to provide accurate forecasts of future discovery rates. These corrections are now routinely incorporated into the USGS models, but the appropriate algorithm for different sizes and classes of field remains a topic of controversy (Drew, 1997).

3.6.2 Barouch-Kaufman model

Barouch and Kaufmann (1975) introduced a sophisticated model which simulated the discovery process as sampling without replacement from a population of oil fields. The

---

[56] $C_k=1$ implies that fields in this size class are discovered at more or less equal intervals throughout the discovery history. $C_k>1$ implies that the number of fields discovered declines as exploratory effort increases. Typically, C_k increases with *k*, where larger *k* corresponds to larger fields.

[57] If the ratio were less than 1.0, then the smaller size classes would have fewer fields than the larger size classes. Alternatively, if the ratio were greater than 2.0 by smaller classes would contain more oil than the larger size class. If the ratio is only a 1.0, the most of the oil is contained in the few large fields. If the ratio is near 2.0, then the oil is more uniformly distributed among all size classes.

original model assumed a lognormal field size distribution and used maximum likelihood techniques to estimate this continuous distribution given information about the number, size and sequence of discovered fields. Smith and colleagues (Smith, 1980; Smith and Paddock, 1984; Smith and Ward, 1981) developed a simpler alternative that group fields into discrete size categories and makes no assumptions about the field size distribution.

Following Barouch and Kaufman, Smith (1980) assumed that: a) the discovery of fields (or reservoirs) in a petroleum play can be modelled as sampling without replacement from an underlying population of fields; and b) the discovery of a particular field is random, with the probability of discovery being proportional to the size of that field divided by the sum of the sizes of all remaining undiscovered fields. The probability that the first discovery (D_1) will fall into size class *k* is then given by:

$$P(V_{1,\infty} \in k) = \frac{N_{k,\infty} * V_{k,\infty}}{\sum_{k=1,K} N_{k,\infty} * V_{k,\infty}} \qquad (3.50)$$

Let the cumulative number of discovered fields in size class *k* which precede the *i*th discovered field be represented by $M_{k,i}$. Hence, the probability that the i_{th} discovery will fall into size class *k*, conditional on the sizes of preceding discoveries, is given by:

$$P(V_{i,\infty} \in k) = \frac{(N_{k,\infty} - M_{k,i}) * V_{k,\infty}}{\sum_{k=1,K} (N_{k,\infty} - M_{k,i}) * V_{k,\infty}} \qquad (3.51)$$

If the number of fields in each size class ($N_{k,\infty}$) was known it would be possible to estimate the *likelihood* of any particular sequence of discoveries by taking the *product* of their conditional probabilities. Alternatively, given an observed sequence of discoveries, maximum likelihood techniques can be used to estimate the number of fields in each size class that maximises the likelihood of having observed the particular discovery sequence. Having obtained these estimates, it is possible to generate a sequence of predictive probability distributions for the sequence of future discoveries.

Smith (1980) used this model to estimate the URR and future discovery sequence of petroleum (i.e. oil and gas) resources in the North Sea. The data included 99 discoveries made between 1967 and 1976, which were grouped into seven size classes. The model was found to accurately reproduce the past discovery history and estimated a mean URR of 43Gb. Similarly, Barouch and Kaufman (1978) tested their model on the Leduc play of the western sedimentary basin of Canada. They used the first 15 discoveries to predict the sizes of the 16th through to the 55th discoveries. Since the 43rd discovery had been made, they were able to compare the predicted and actual values up to the 43rd. In total, the model predicted values that were within 7% of the actual values, although the errors were large for some individual discoveries. In another study, Power and Fuller (1992) compared the predictive accuracy of the Barouch and Kaufmann model to those of competing models using offshore data from Canada. The Barouch and Kaufmann model was generally found to perform better than competing models (including Hubbert's YPE curve) in forecasting future discovery rates.

Hence, there is some evidence that this approach performs well under certain conditions, although the limited number of studies we have been able to access prevents any general conclusions from being drawn. Table 3.7 provides a comparison of the Arps-Roberts and Barouch-Kaufman models.

Table 3.7 Comparison of the Arps-Roberts and Barouch-Kaufman models

Arps-Roberts	**Barouch-Kaufman**
Requires information about the area of the exploratory region	Does not require information about the area of the exploratory region
Computes results for changes in the exploratory effort	Does not require data on exploratory effort
Computes changes in the number of fields in discrete size categories	Computes the expected size of new discoveries in a particular order
Requires a time series of the number of fields in each size category – but does not require the order of discovery	Requires a time series on the order and the size of discoveries
Not necessary to estimate the population number of fields to be discovered	Necessary to estimate the population number of fields to be discovered ($N_{k,\infty}$)
Computationally straightforward	Computationally difficult involving maximum likelihood methods
Not based on probabilistic propositions	Derived from probabilistic propositions

Source: Herbert (1982)

3.6.3 Summary

Discovery process models have a stronger theoretical basis than simple curve-fitting and may potentially provide more reliable estimates of URR for the regions in which they can be applied. However, there has yet to be a comprehensive synthesis of research in this area and the existing literature appears both patchy and difficult to interpret. While numerous authors have developed variations on the basic themes – many of which are highly sophisticated[58] – the techniques appear to be little used and there is a lack of comparative studies on their relative performance. Also, the extensive data requirements and methodological sophistication would appear to present a barrier to their more widespread use.

Contrary to the claims of Kaufman (1975a)[59] and others, the difference between discovery process models and simple curve-fitting appears to be one of degree rather than kind. As a result, these models have many of the same limitations as a simple curve-fitting. For example, discovery process models works best when applied to maturely explored and geologically homogeneous regions with relatively open access to exploration. If applied to larger or poorly explored regions or to regions with restricted access to exploration, the results can be misleading (e.g. in failing to anticipate new discovery cycles).[60] Similarly, the models frequently assume that success rates decline continuously as a result of physical depletion, which implies that changes in prices, costs and technology do not have a significant influence. But this assumption is inconsistent with econometric evidence from a variety of regions (Forbes and Zampelli, 2000; Iledare and Pulsipher, 1999) and raises the

[58] For example, Meisner and Demirmen (1981); Rabinowitz (1991), Arps et al. (1971), Forman and Hinde (1985), Lee and Wang (1983; 1985; 1986) and Lee (2008).

[59] Kaufman (1975) claims that "….. there is an important difference between a model whose output is a logical consequence of relations among a set of primitive assumptions describing the process by which data are generated and one in which the 'law' governing the mathematical form of its output is the primitive assumption. An example of the former is the model of Barouch and the present author; the models of Hubbert and Moore are instances of the latter. The user will generally have more confidence in forecasts generated by a model that passes tests of the validity of primitive assumptions from which it is structured, independently of tests of the predictive quality of its output."

[60] In some cases, it may not be possible to use such models at all. For example, if the prime areas are leased and explored relatively late, the discovery rate for even the larger field size classes may not decline over time (Schuenemeyer and Drew, 2004).

possibility of historic or future shifts in discovery rates as a result of factors such as changes in tax policy (see Section 5). Similarly, the models will only generate reliable forecasts if the field size data is adjusted to allow for future reserve growth. But this adjustment is not always made and is difficult owing to the lack of relevant data and the considerable variability in the growth process between different regions, types of field and reserve reporting standards (Schuenemeyer and Drew, 2004).

Most importantly, both discovery process models and curve-fitting rely upon historical data that was generated under one set of economic, political and technical conditions and use this to forecast future discoveries that may take place under very different conditions. Hence, the forecasts will only be reliable if the geological determinants of future discoveries significantly outweigh the other influences. The accuracy may potentially be improved by the development of *hybrid* models that introduce additional economic and other variables (Walls, 1994), but relatively few examples of such models appear to be available. Section 5 examines these approaches in more detail.

3.7 Summary

This section has described the various historical extrapolation techniques that are used to estimate ultimately recoverable resources. In each case, it has identified the historical origins of the technique, highlighted some relevant strengths and weaknesses and identified the conditions under which the technique appears more likely to be reliable. The key conclusions are as follows:

- *Curve-fitting versus discovery process*: Extrapolation techniques fall into two groups - curve-fitting and discovery process models. While the former use aggregate data for a region, the latter require data on individual fields and hence may not always be feasible. Both groups assume that the field size distribution is highly skewed and that fields tend to be discovered and produced in declining order of size. These assumptions appear more likely to be accurate for geologically homogeneous regions where exploration has been relatively uninterrupted. Many of the difficulties with curve-fitting techniques arise from the fact that these conditions may not hold for the regions for which they are applied.
- *Classification*: Curve-fitting techniques may be classified according to their choice of explained and explanatory variables (Table 3.3). While the analysis of production is relatively straightforward, the analysis of discovery is greatly complicated by reserve growth at known fields. While the use of backdated discovery estimates can help in this regard, this is only possible if the relevant information is available. Even then, it is desirable to adjust the estimates to allow for future reserve growth. If such an adjustment is not made, curve-fitting to discovery data may lead the URR to be underestimated. But given the paucity of information on reserve growth, the estimates will be sensitive to the particular growth function employed.
- *Production over time:* Production projection is straightforward to apply and relies upon data that is readily available, relatively accurate and free from the complications of reserve growth. But the technique is only useful for regions that are relatively advanced in their production cycle. There is no robust basis for choosing a particular functional form and different forms are often found to fit the data equally well but yield very different estimates of URR. Multi-cycle models may often be more appropriate for aggregate regions, but they create the risk of over-fitting and highlight the possibility of new discovery and production cycles occurring in the future. The results will only be reliable

if any future cycles have only a small impact on aggregate resources - either because there are few such cycles or because the resources are relatively small. The popular 'Hubbert Linearisation' technique is equivalent to fitting a logistic curve to cumulative production and hence is both less flexible than production projection (since only one functional form is employed) and no more reliable.

- *Discovery over time*: Discovery projection and decline techniques have many similarities to those for production, but since discovery is more advanced these techniques should be applicable to a larger number of regions. The techniques raise similar concerns to those identified above but also introduce additional complications such as the uncertainty in reserve estimates, the relative suitability of 1P and 2P estimates and the implications of 'smoothing' erratic discovery trends. The existence of reserve growth suggests the need for backdated rather than current measures of discovery, but the shape of the discovery cycle based on backdated estimates will be different from that based upon current estimates and both the 'height' and shape of the backdated discovery cycle will change over time. Hence, if reliable and consistent estimates of URR are to be obtained, backdated discovery data should be adjusted to allow for future reserve growth.

- *Discovery over effort*: Creaming curves and yield per effort curves offer advantages over production and discovery projection, but must still be used with care. Difficulties can again arise with the accuracy of the relevant data, the required corrections for future reserve growth, the poor fit of particular functional forms, the existence of multiple exploration cycles and the failure to anticipate new exploration cycles in the future. Overall, these difficulties appear more likely to lead to *underestimates* of the regional URR.

- *Discovery process:* Discovery process models have a stronger theoretical basis than simple curve-fitting and may potentially provide more reliable estimates of URR for the regions in which they can be applied. However, these techniques appear to be largely confined to North America, perhaps as a result of their extensive data requirements. Also there is a lack of comparative studies on the relative performance of different models and they have many of the same limitations as a simple curve-fitting. As with curve-fitting, the neglect of economic and other variables suggests that the forecasts will only be reliable if the geological determinants of future discoveries significantly outweigh the other influences - at least over the time-scales that are relevant to the estimation of URR. This need not necessarily be the case, and even if it was the inclusion of additional explanatory variables could potentially improve the results.

- *Implications*: All methods of estimating URR have their limitations, so the above criticisms need not imply that other approaches are any more reliable. The uncertainties associated with these techniques must always to be acknowledged and the results expressed in probabilistic form where possible. Accuracy may potentially be improved through the incorporation of additional explanatory variables, explicit allowance for reserve growth and a focus on smaller geographical regions. But a more disaggregated analysis is likely to be resource intensive, as well as requiring access to confidential data. While a more aggregate analysis may potentially lead to biased estimates, the size of the error may vary widely from one region to another and may be expected to decline in the future as exploration matures. For regions such as the continental United States, the error may already be relatively small. Estimates of the *global* URR may be derived from the *sum* of estimates derived from applying curve-fitting techniques to individual regions. These estimates should become more accurate over time as more of the world's oil bearing regions become comprehensively explored.

4 Consistency of curve fitting techniques

4.1 Introduction

This section uses illustrative examples to investigate the *consistency* of URR estimates from different curve fitting techniques (i.e. the extent to which one estimate differs from another). Using discovery and production data from a number of regions, it assesses:

- *Consistency over time*: whether estimates for a region that are made with one technique using data through to year X are consistent with estimates made by the same technique using data through to a later year Y.
- *Consistency between functional forms*: whether estimates for a region that are made assuming one functional form are consistent with the estimates made by assuming a different functional form that has a comparable goodness of fit;
- *Consistency over the number of curves*: whether estimates for a region that are made by fitting a single curve to the whole data set are consistent with the estimates made by fitting two curves sequentially to individual components of the dataset; and
- *Consistency between techniques*: whether estimates for a region that are made using one technique are consistent with those made by another technique.

Tests for consistency were conducted with the following three curve-fitting techniques, representing each of the groups identified in Section 3:

- *Hubbert Linearisation* - a production over time technique;
- *Discovery Projection* - a discovery over time technique; and
- *Creaming Curves* - a discovery over effort technique

These techniques are widely used for deriving estimates of regional URR, so the consistency of results derived from them is of considerable interest. Table 4.1 indicates the tests that were conducted for each of the three techniques. Whether two URR estimates are judged to be ‘consistent’ with one another will depend upon the level of accuracy that is expected from the technique. In general, we would expect more accurate estimates to be obtained for those regions that are at a later stage of their production and/or discovery cycle. For illustrative purposes, in what follows *we judge two estimates to be consistent if they differ by less than 20% of the cumulative discoveries in the region through to 2007*.

Table 4.1 Consistency tests on curve fitting techniques

Test	Hubbert Linearisation	Discovery projection	Creaming Curves
Consistency over time	Yes	Yes	-
Consistency between functional forms	-	Yes	Yes
Consistency over the number of curves	-	-	Yes
Consistency between techniques	Yes	Yes	Yes

The relevant data was obtained from the 2007 edition of the Petroleum Economics and Policy Solutions (PEPS) database provided by IHS Energy. This contains time-series data on petroleum production, remaining resources, discoveries and exploratory drilling for all oil producing countries going back to the 19th century. This is based upon a much more detailed field-level database and is the continuation of statistics maintained by Petroconsultants Inc before their purchase by IHS Energy in 1998. A very similar database forms the basis for the work of Colin Campbell and Jean Laherrère. The information in the IHS database derives from a variety of sources including published information and expert assessments. It is widely recognised as an authoritative and valuable source of information on the global oil and gas industry.

The IHS database classifies petroleum into gas and liquids. The latter includes crude oil, natural gas liquids (NGLs), condensate, heavy oils (less than 10 degrees API) and oil sands, but the relative proportion of each is not identified.[61] The estimates of remaining resources are based upon 2P (proven plus probable) reserves and hence differ from those published in public domain sources such as the BP Statistical Review. The measure of exploratory activity is the total number of new field wildcat (NFW) wells. Importantly, no distinction is made between the search for oil and the search for gas resources.

In order to examine the reliability of URR estimates generated through the curve-fitting techniques, we analysed data from **ten** oil producing regions – labelled A to J. Most of these were individual countries, while some represented groups of countries. The regions were chosen to represent a broad range, both geographically and with regard to the relative maturity of their production and/or discovery cycle – with the majority being at a relatively mature stage. From inspection, we estimate that all but Region B have passed their peak of discovery, while five regions (A, D, F, H and J) have passed their peak of production. In principle, we would expect the curve fitting techniques to provide more accurate and consistent results for those regions where the production/discovery cycle is more advanced.

For reasons of confidentiality, the individual regions are not identified below and the associated discovery and production figures are annoyomised. Where comparison of URR estimates is made, the values are given as a percentage of either the cumulative discovery in the region through to 2007 (D_{2007}) or the cumulative production through to 2007 (Q_{2007}). The focus throughout is on the consistency of the methodological techniques, rather than the absolute size of the relevant estimates. In principle, estimates of URR should be greater than Q_{2007} and D_{2007} but closer to the latter than the former. Also, smaller estimates of URR

[61] Biofuels, GTLs and CTLs are excluded.

relative to Q_{2007} (D_{2007}) may suggest that the region is more advanced in its production (discovery) cycle. However, these estimates may be inaccurate and the examples below include instances where the 'best' estimate of URR is less than the cumulative discoveries through to 2007.

The models are fit using both linear and nonlinear regression techniques and their relative 'goodness of fit' is compared using the coefficient of determination, or R^2. This measure has a number of limitations, as discussed in Box 4.1. However, it is useful as a first order indication of how well a particular model fits the data. The relative statistical issues associated with curve fitting are explored further in Section 5.

The following sections use the mathematical notation introduced earlier in Section 3.2. This is summarised again in Table 4.2.

Table 4.2 Notation for explained and explanatory variables.

Notation	**Definition**
t	Time
ε	Effort
t_d	Cumulative time for discovery
ε_d	Cumulative effort for discovery
$Q(t)$	Cumulative production
$Q'(t)$	Rate of change of cumulative production (rate of production)
$R(t)$	Reported reserves
$D(t)$ or $D(\varepsilon)$	Cumulative discovery
$D'(t)$ or $D'(\varepsilon)$	Rate of change of cumulative discovery (rate of discovery)
$B(t_d,t)$ or $B(\varepsilon_d,t)$	Backdated cumulative discovery
$B'(t_d,t)$ or $B'(\varepsilon_d,t)$	Rate of change of backdated cumulative discovery
$Q(\infty)$ or $D(\infty)$ or $B(\infty,\infty)$	Ultimately recoverable resource

Box 4.1 Measures of goodness of fit

The *goodness of fit* of a statistical model describes how well it fits a set of observations. In regression analysis, a common measure of goodness of fit is the R^2 value, which defines the proportion of variability in the data that is accounted for by the model. The R^2 is calculated from: $R^2 = 1 - \frac{SS_{err}}{SS_{tot}}$ where SS_{err} is the sum of squared errors (the distances of the points from the best-fit curve) and and SS_{tot} is the total sum of squares (the distances of the points from the mean of all Y values). R^2 is a fraction between zero and one, with higher values indicating a better fit. The name R^2 is a missnomer as it is not the square of anything and could be negative if the curve fits the data worse than the mean.

One drawback with R^2 is that it always increases when additional explanatory variables are added to a model. Hence, an alternative *adjusted* R^2 is frequently used, which penalizes the statistic as extra variables are included (nested models). The *adjusted* R^2 is defined as $\overline{R^2} = 1 - \frac{SS_{err}(T-k)}{SS_{tot}/(T-1)}$, where T is the sample size and k is the number of variables in the model. Note that the adjusted R^2 can never be higher than the R^2 and could principle be less than zero.

Neither R^2 or $\overline{R^2}$ should be used as the main criterion for whether a model fit is reasonable, since they can be quite high and yet lead to wide confidence intervals for the *forecast* ('out of model') values of the explained variable. It is also possible to have best-fit values for most of the data, but not over the most recent years which could be more important when analysing trends in a variable over time. For instance, two models may have the same R^2, but one may fit the data very well at the beginning of the time period but not so well towards the end. Conversely a second model may fail to provide a good fit at the start, but fit the data much better towards the end. This important fact is not captured by either R^2 or $\overline{R^2}$.

The R^2 and $\overline{R^2}$ statistics can be useful when comparing *nested* models: that is, where one model is a particular case of a second model (e.g. containing only a subset of the variables). However, these statistics are not appropriate for comparing *non-nested* models, where such a relationship does not apply. An example would be a comparison of a logistic with a Gompertz functional form. There are a variety of ways of comparing such models, but they all include the error sum of squares (SS_{err}) multiplied by a penalty factor that depends upon the complexity of the model (Ramanathan, 2002). A more complex model will reduce the SS_{err} but raise the penalty. A popular approach for time-series analysis is Akaike's Information Criterion (AIC), defined as: $AIC = \left(\frac{SS_{err}}{T}\right)e^{2k/T}$. Alternative options include the Schwartz Criterion (SC) and Ameniya's Prediction Criterion (PC) (Kennedy, 2003). None of these methods are definitive, since it is possible to find a model that is superior under one criterion and inferior under another. Note further that these statistics should not be used to compare models with different explained variables (e.g. a discovery projection versus a production projection).

4.2 Hubbert linearisation

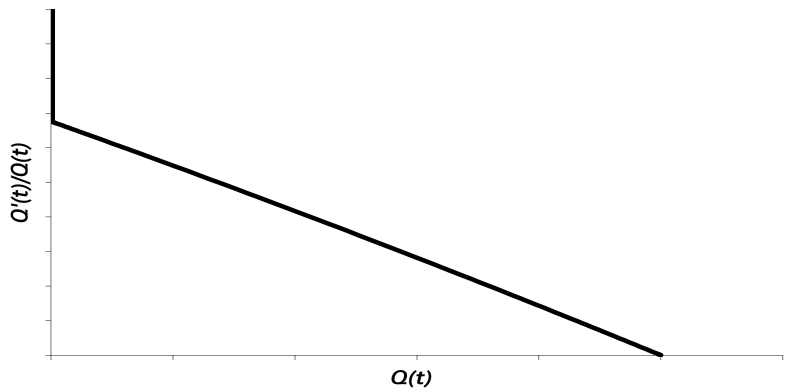
4.2.1 Background and approach

'Hubbert Linearisation' (HL) is a popular curve-fitting technique owing to its methodological simplicity and the relative availability of the required data. It involves:

- plotting the ratio of annual to cumulative production ($Q'(t)/Q(t)$) as a function of cumulative production (*Q(t)*);
- taking a linear regression; and
- estimating the URR from the intercept of the regression line with the *Q(t)* axis.

As shown in Section 3.3.2, the relationship between *Q*'(*t*)/*Q(t)* and *Q(t)* will only take a strictly *linear* form if cumulative production (*Q*(*t*)) takes a *logistic* form. To illustrate this, Figure 4.1 illustrates a Hubbert Linearisation for a hypothetical region in which the cumulative production grows logistically. This suggests that regions where the growth in cumulative production is approximatly logistic should linearise reasonably consistently, but departures from logistic growth will lead to departures from a linear trend in the HL transform. In addition, Hubbert stated that this approach was only likely to provide a reliable estimate of URR if over one third of the resources had already been produced. However, since the URR is not known, it is impossible to assess whether this criterion is met (Pesaran and Samiei, 1995).

Figure 4.1 Hubbert Linearisation of logistic growth in cumulative production

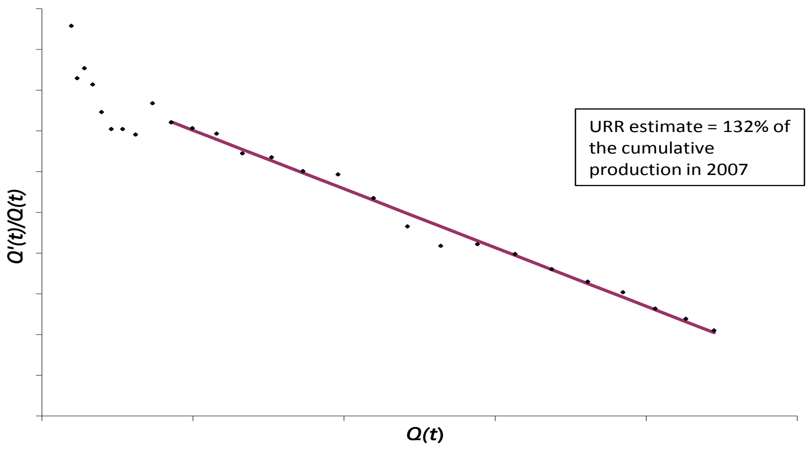


The data for the ten regions were first plotted in the prescribed fashion. Groups of data points that appeared to show an approximate linear relationship were then isolated and a linear regression performed using Microsoft® Excel. No attempt was made to fit a linear regression to the entire dataset, since the data invariably showed considerable scatter for the early years of production This is perhaps understandable, given that the numerator of the 'explained' variable (*Q'(t)/Q(t)*) will be relatively large during the early period compared to the denominator. Hence, small departures from a logistic model of cumulative production may lead to proportionately large departures from the implied ratio of *Q'(t)* and *Q(t)* in the early stages of the production cycle. In contrast, as cumulative production increases, the numerator of the explained variable will become a progressively smaller fraction of the denominator, with the result that departures from the logistic model should have a smaller effect on the magnitude of this ratio. This highlights an important weakness of the HL approach: since the 'explained' and explanatory variables are not independent, the errors cannot be normally distributed throughout the dataset. This issue is discussed further in Section 5.

4.2.2 Results - consistency over time

Figure 4.2 shows the results of our analysis for Region A. This example behaves in a similar manner to those presented by Hubbert (1982) and Deffeyes (2003).[62] Data points for earlier years show considerable scatter, but then the data 'settles' to an approximate linear relationship which can be modelled and extrapolated using linear regression. This leads to an estimate of URR which is 32% larger than the cumulative production through to 2007 – suggesting that the region is well advanced in its production cycle.

Figure 4.2 Hubbert Linearisation of production data for Region A



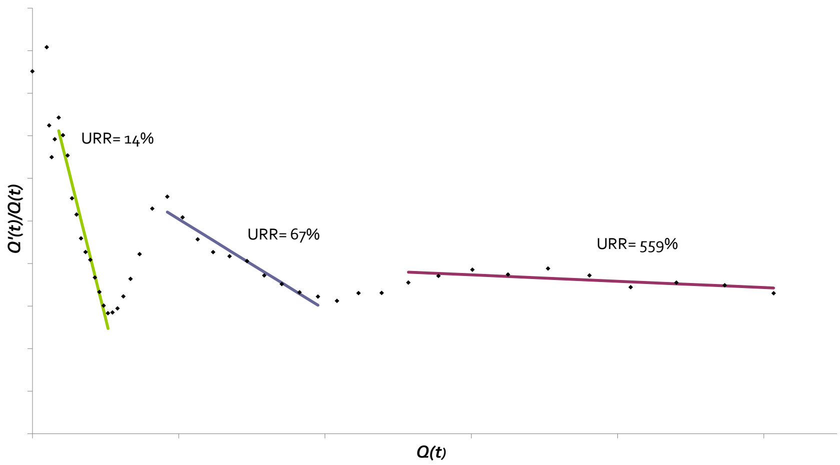
From this example and the references cited above, it could be assumed that the linear relationship will be maintained until production ceases, implying that the estimate of URR is consistent and accurate. However, the data only behaved in an approximately consistent fashion for four out of the ten regions examined here. In the remainder, the data failed to 'settle' into a single linear relationship even for regions where the production cycle appeared to be relatively advanced. An example is provided in Figure 4.3, where the data for Region B is modelled with three separate linear regressions. If similar analyses had been conducted at an earlier stage in the production cycle for this region, these would have led to *underestimates* of the URR.[63] The linear trends identified during the earlier stages of the production cycle were not subsequently maintained, but the HL method provides no way of anticipating this.

---

[62] Numerous examples can also be found on peak oil web sites such as the Oil Drum.

[63] For a similar example, see the following post on the Oil Drum by Robert Rapier: Predicting the Past: the Hubbert Linearisation.

Figure 4.3: Hubbert Linearisation of production data for Region B



One possible explanation for the behaviour illustrated in Figure 4.3 is that Region B has experienced several *cycles* of exploration and production. These may occur, for example, because different geographical areas were opened up to exploration at different times, or because technological developments allowed access to deeper or less accessible resources. In particular, several countries have both onshore and offshore oil-producing regions and the former have typically been developed many years before the latter. Combining both of these within a single analysis could be misleading – although it may also be unavoidable if production data is only available at the aggregate, country level. Fortunately, the IHS database reports onshore and offshore production separately. To explore whether separating the two improves the results, Figure 4.4 presents the results of a HL of offshore production in Region B, while Figure 4.5 does the same for onshore.

Figure 4.4 Hubbert Linearisation of offshore production in Region B.

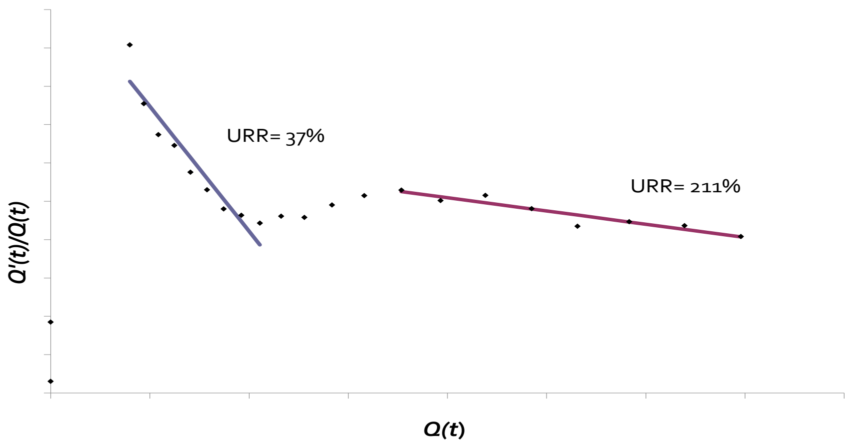
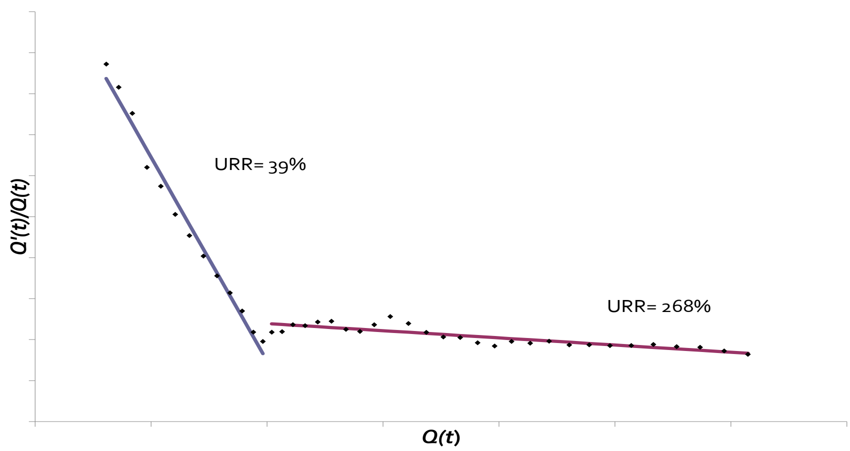


Figure 4.5 Hubbert Linearisation of onshore production in Region B



The results demonstrate that separating offshore from onshore production does not resolve the problems in using HL to estimate the URR for this region. Both sets of results show the same 'breaks' in the data series that were observed with the aggregate data analysed in Figure 4.3. If the HL had been estimated at an earlier stage of the production cycle, it would have led to a significant underestimate of the URR (e.g. an estimate that was less than 15% of the 'actual' URR).

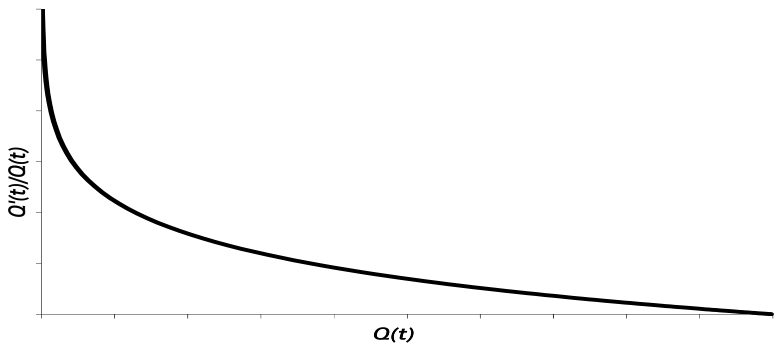
In Section 3.3.1, it was observed that the logistic model provided a poor fit to production trends in many of the world's oil producing regions. In the most systematic study to date, Brandt (2007) analysed 74 regions that were past their peak of production and found that the rate of production increase exceeded the rate of decline in over 90% of cases.[64] This implies that an *asymmetric to the left* model of cumulative production may often be more appropriate. One approach, first introduced by Moore (1962), is the *Gompertz* function, defined as follows (Gompertz, 1825):[65]

[64] The production weighted mean rate of exponential decline was found to be approximately 2% while the production weighted mean rate of increase was approximately 6%.
[65] Both the logistic and Gompertz functions have their origin in the study of population growth (Ausloos and Dirickx, 2005)

$$Q(t) = Q_\infty e^{-ae^{-b(t-t_m)}} \quad (4.1)$$

This leads to a cumulative production cycle with a point of inflection around 35- 40% of the URR, implying a production cycle with a rate of decline that is slower than the rate of increase. Figure 4.6 shows the corresponding Hubbert Linearisation. It is clear that if cumulative production grew in this manner, then the use of HL would lead to a consistent underestimation of the regional URR, with the percentage error falling over time.

Figure 4.6 Hubbert Linearisation of Gompertz growth in cumulative production



The data from several of the regions examined exhibited similar behaviour to that illustrated in Figure 4.6 - consistent with Brandt's observation that the production cycle is frequently asymmetric to the left. For example, Figure 4.7 presents the HL plot for offshore production in Region C. In regions such as this, an HL would again give unreliable results.

Figure 4.7 Hubbert Linearisation of offshore production data for Region C.

The overall results from the consistency over time tests for Hubbert Linearisation are summarised in Table 4.3 and illustrated in Figure 4.8. This demonstrates that only four of the regions (A, D, F and I) exhibited consistency over time with the aggregate data, while none of the regions exhibited consistency over time for both onshore and offshore data. These judgements are based upon the URR estimates from the linear regressions but for many of the regions the data only settled into an approximately linear relationship relatively late into the production cycle. While the regions that gave consistent results appeared to be at a relatively mature stage in their production cycle, other regions (e.g. Region J) that appeared to be at a comparable stage gave inconsistent results. Moreover, given the frequency of 'breaks' in the relationships for the 'mature' regions, it is quite possible that comparable breaks will subsequently be seen in the data for less mature regions. In other words, analyses that appear to provide consistent results using data through to 2007 may not continue to provide consistent results in the future.

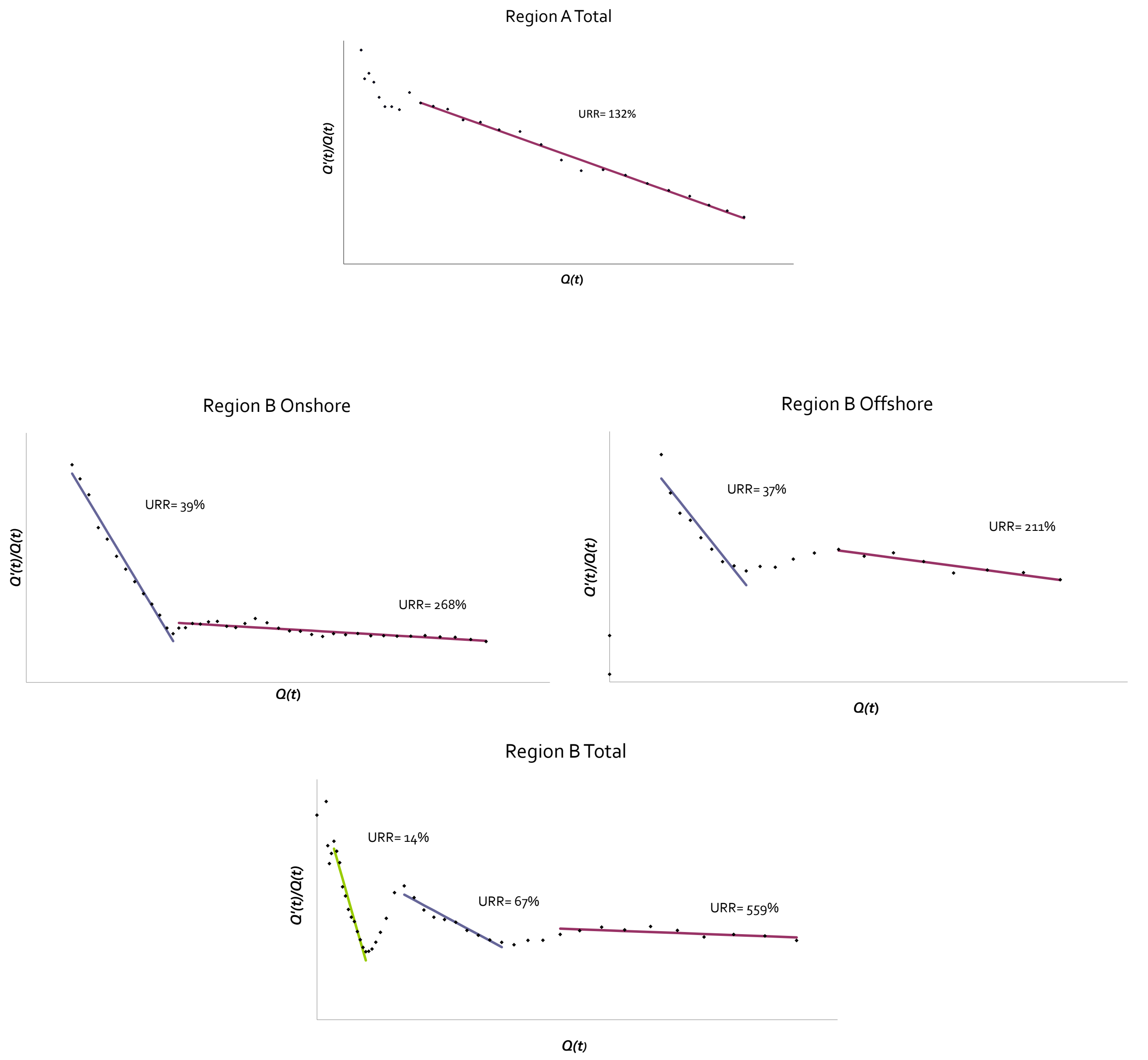
It is also useful to compare the URR estimates from the aggregate data with the sum of the estimates from the onshore and offshore data (Table 4.3). This shows that five regions (C, D, F, G and H) gave consistent estimates (i.e. differing by less than 20% of Q_{2007}).

The primary explanations for the lack of consistency appear to be that: a) the regional data is highly aggregate and in many cases includes several discrete oil producing regions that were developed at different times; and b) the cumulative production profile of individual regions is only poorly approximated by the logistic model. Further research is required to identify whether our sample is representative of oil producing regions as a whole, or of regions at different levels of aggregation. However, the results raise concerns about the usefulness of the HL technique. In particular, it appears more likely to *underestimate* the regional URR than to overestimate it, thereby contributing to overly pessimistic forecasts of future oil production. The following section examines the discovery projection technique to see whether the same difficulties apply.

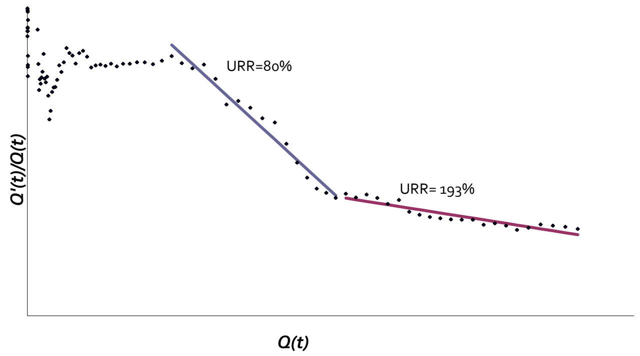
Table 4.3 Summary of consistency tests of Hubbert Linearisation technique

Region	Stage of production cycle (pre or post peak)	HL consistent over time for aggregate data?	HL consistent over time for onshore data?	HL consistent over time for offshore data?	URR estimate from aggregate data (% of Q_{2007})	Sum of URR estimates from on and offshore data (% of Q_{2007})	Difference (% of Q_{2007})
A	Post	Yes	-	-	132	132	-
B	Pre	No	No	No	559	262	297
C	Pre	No	No	Yes	210	192	18
D	Post	Yes	Yes	No	119	132	13
E	Pre	No	No	No	403	495	92
F	Post	Yes	Yes	No	124	119	5
G	Pre	No	No	No	258	252	6
H	Post	No	No	Yes	124	108	15
I	Pre	Yes	Yes	No	175	201	26
J	Post	No	Yes	No	124	170	46

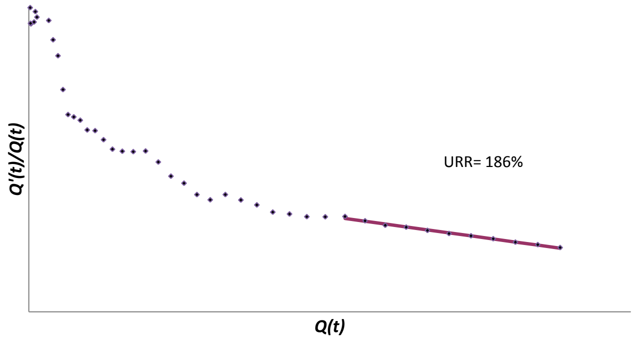
Figure 4.8 Summary of consistency over time tests for Hubbert Linearisation technique



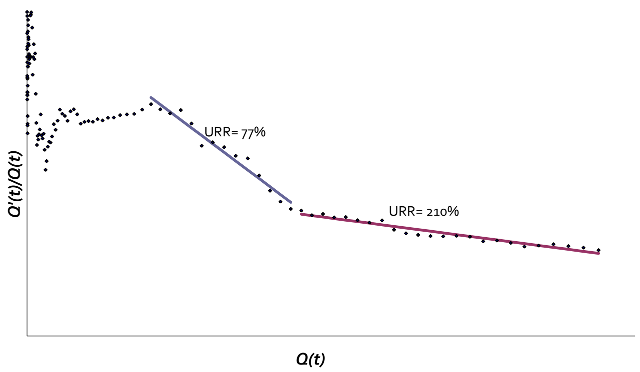
Region C Onshore



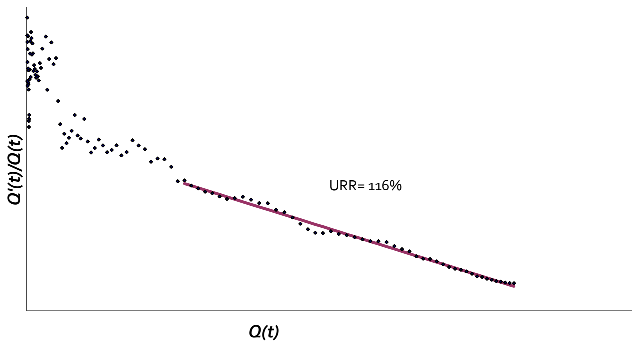
Region C Offshore



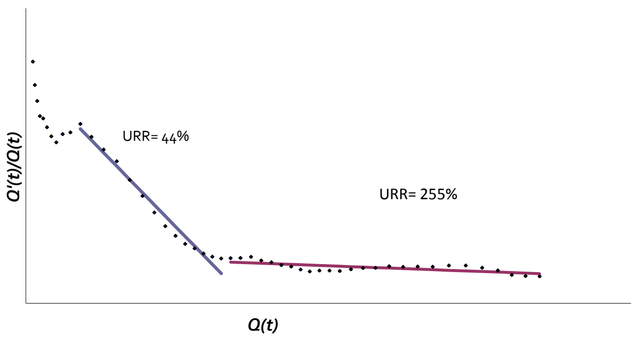
Region C Total



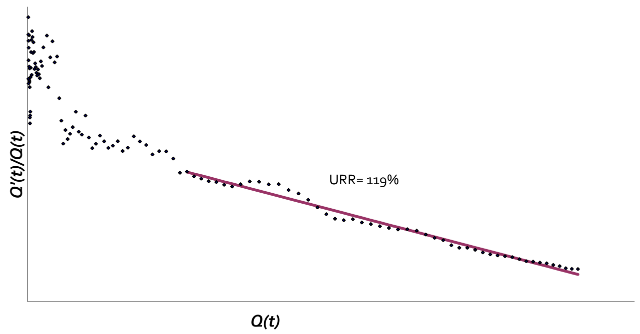
Region D Onshore



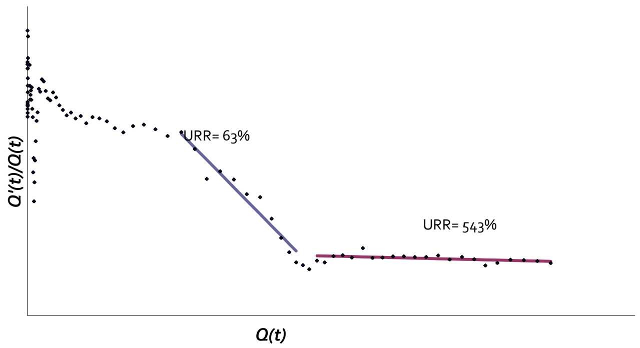
Region D Offshore



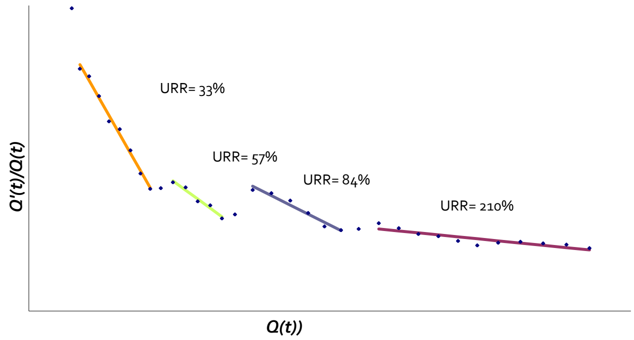
Region D Total



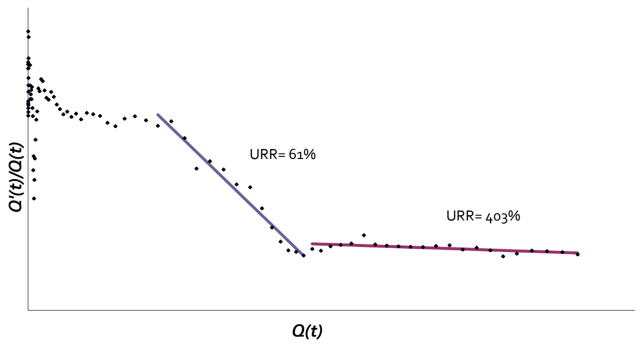
Region E Onshore



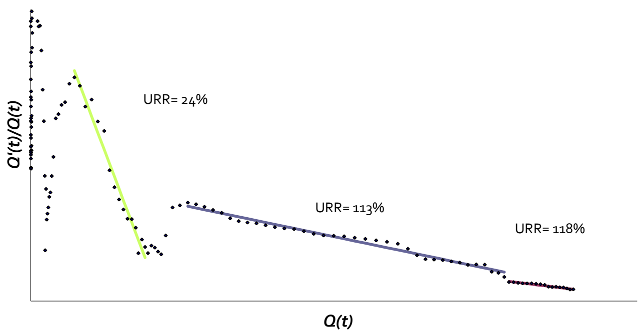
Region E Offshore



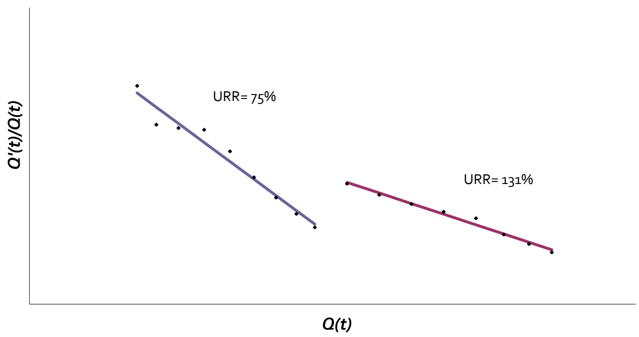
Region E Total



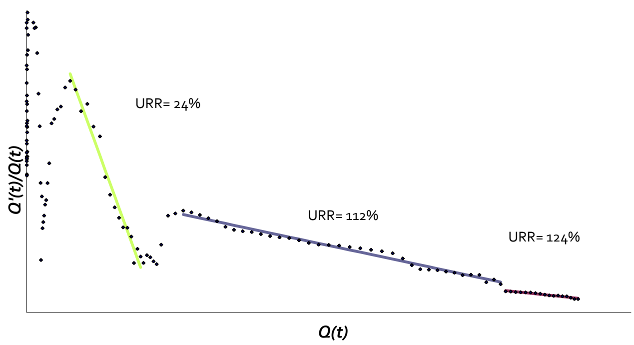
Region F Onshore

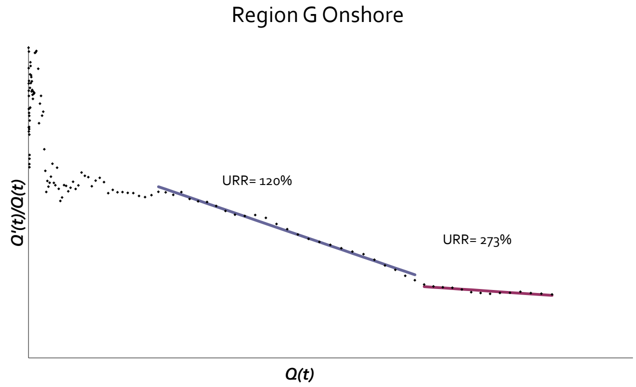


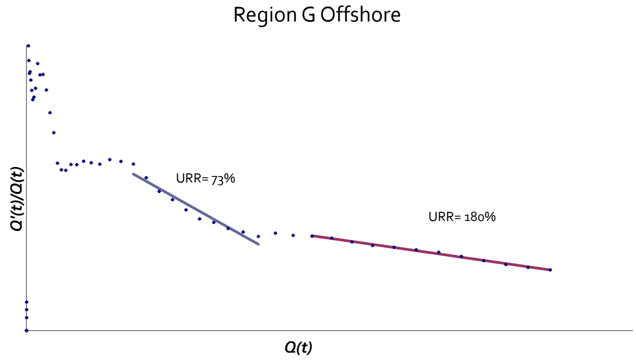
Region F Offshore

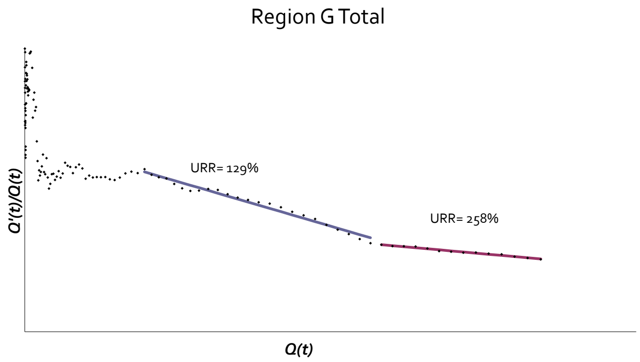


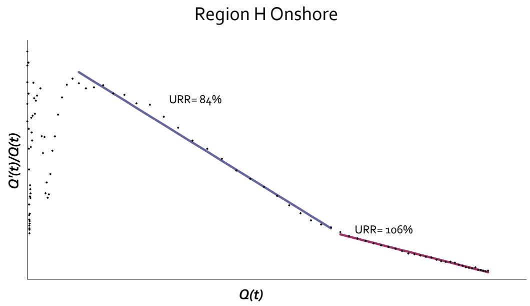
Region F Total

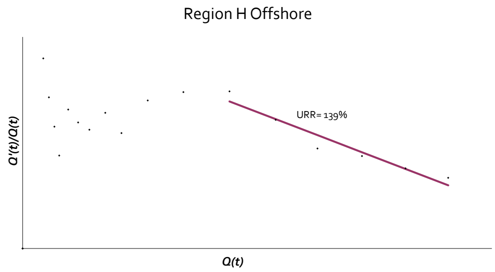


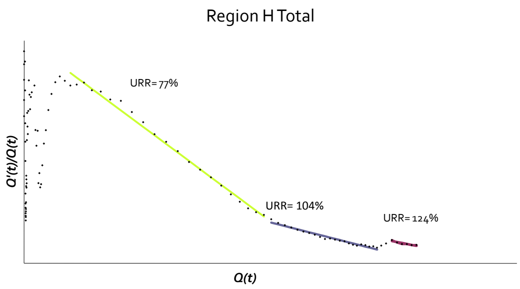

Region G Onshore
URR= 120%
URR= 273%
Q'(t)/Q(t)
Q(t)

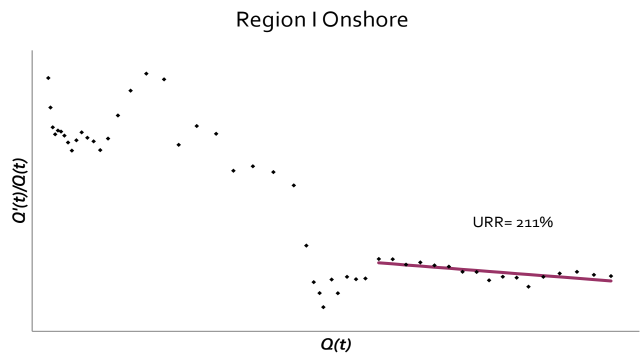

Region G Offshore
URR= 73%
URR= 180%
Q'(t)/Q(t)
Q(t)

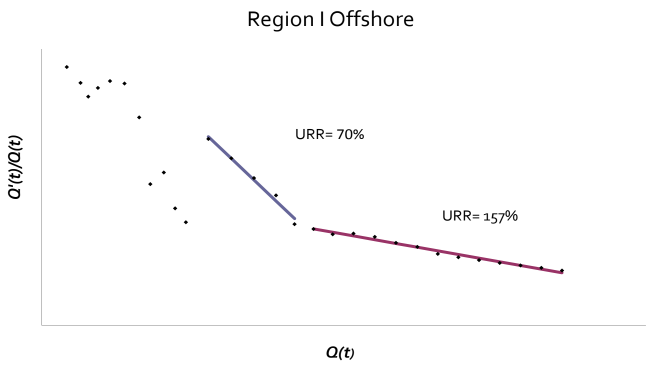

Region G Total
URR= 129%
URR= 258%
Q'(t)/Q(t)
Q(t)

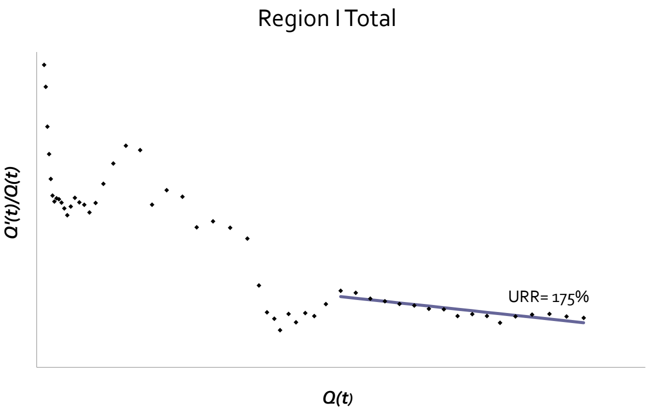

Region H Onshore
URR= 84%
URR= 106%
Q'(t)/Q(t)
Q(t)

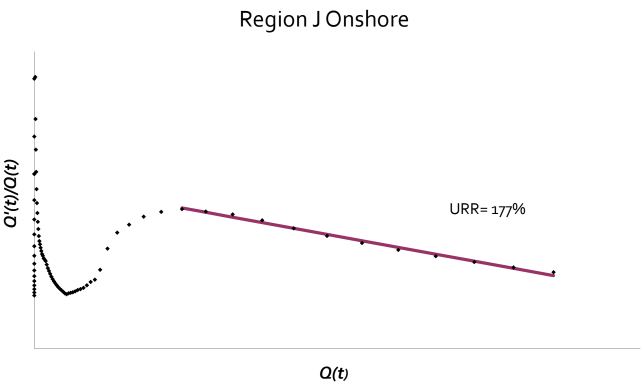

Region H Offshore
URR= 139%
Q'(t)/Q(t)
Q(t)

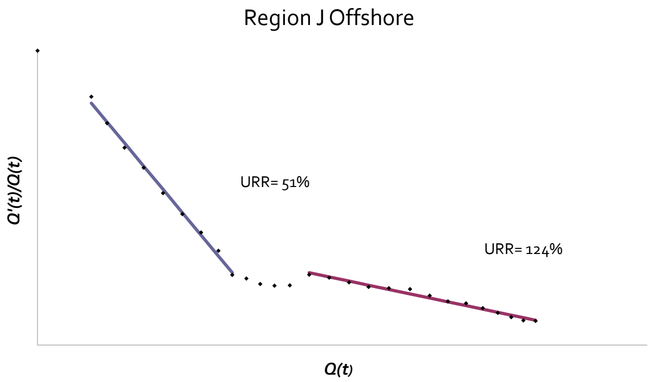

Region H Total
URR= 77%
URR= 104%
URR= 124%
Q'(t)/Q(t)
Q(t)

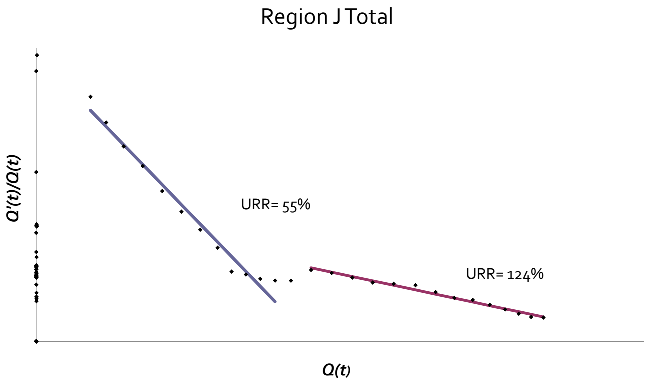

Region I Onshore
Q'(t)/Q(t)
URR= 211%
Q(t)


Region I Offshore
Q'(t)/Q(t)
URR= 70%
URR= 157%
Q(t)


Region I Total
Q'(t)/Q(t)
URR= 175%
Q(t)


Region J Onshore
Q'(t)/Q(t)
URR= 177%
Q(t)


Region J Offshore
Q'(t)/Q(t)
URR= 51%
URR= 124%
Q(t)


Region J Total
Q'(t)/Q(t)
URR= 55%
URR= 124%
Q(t)

4.3 Discovery projection

4.3.1 Background and approach

Discovery projection involves:

- Plotting cumulative discovery estimates as a function of the date of discovery (t_d);
- using non-linear regression techniques to fit a particular functional form to this data; and
- estimating the URR from the value of the relevant parameter(s) – which corresponds to the asymptote of the curve.

We use *backdated* cumulative discovery estimates ($B(t_d, t)$) in what follows, since this is what is available from the IHS database. We do *not* correct for future reserve growth since such adjustments are not made by authors such as Campbell and Laherrère. In other words, *we are testing the consistency of this method as used by the most prominent authors in this field*. Nevertheless, it is *possible* to adjust for future reserve growth with the IHS database (see Box 4.2) and it is likely that this would make the methods more consistent as well as more accurate. However, the results would necessarily be sensitive to the particular growth function that is employed.

Compared to the Hubbert Linearisation of production data, discovery projection should have two advantages. First, since the discovery cycle is more advanced than the production cycle, the technique should be applicable to a greater number of oil-producing regions and the corresponding estimates of URR should be more reliable. Second, discovery projection does not impose the restriction that the discovery cycle be approximately logistic, since a variety of functional forms can be employed. The main drawback is that the relevant discovery data is generally less accessible and should in principle be corrected to allow for future reserve growth. Also, nonlinear regression is less straightforward than linear regression, although modern software packages make the technique much easier to apply than was the case in the past. Some background on nonlinear regression is provided in Box 4.3.

We applied discovery projection techniques to the same ten regions as analysed in Section 4.2. In each case, we fitted one or more functional forms to the aggregate data using the nonlinear regression function in SPSS©. For each region, we first compared the consistency of the URR estimates obtained using different functional forms (Section 4.3.2) and then compared the consistency using different lengths of data series (Section 1.1.1). For reasons of time, the analysis was confined to data for the *aggregate* region, without distinguishing between onshore and offshore resources. Although this is also the approach used by Campbell and Heapes (2008) among others, combining onshore and offshore regions within a single analysis could adversely affect the consistency of the results.

Box 4.2 Adjusting IHS aggregate discovery estimates to allow for future reserve growth

The IHS database contains estimates of $B'_{t_d}(t_d,t)$ - i.e. the amount of oil discovered at time t_d as estimated in the current year t (where $t \geq t_d$). These are *backdated* estimates and hence tend to be larger than the corresponding estimates made in the year of discovery ($B'_{t_d}(t_d,t_d)$) as a consequence of reserve growth in the intervening period ($\tau = t - t_d$.). The growth in the estimates for a given discovery year (t_d) is given by: $G(\tau) = B'_{t_d}(t_d, t_d + \tau) / B'_{t_d}(t_d,t_d)$. Estimation of the 'growth function' ($G(\tau)$) requires data on $B'_{t_d}(t_d,t_d)$ and hence requires access to the IHS databases from *all* previous years (i.e. each t_d). Estimates of growth functions are also available from the technical literature. However, these mostly relate to cumulative discovery estimates based upon 1P reserves for US fields, while the IHS database contains cumulative discovery estimates based on 2P reserves for non-US fields. However, Klett *et al.* (2005a) find that growth functions derived from the former appear equally applicable to the latter.

Estimates of the size of discovered resources will continue to grow into the future, with the ultimately recoverable resources discovered in each year being given by $B'_{t_d}(t_d,\infty)$. If data on $B'_{t_d}(t_d,t_d)$ was available, then $B'_{t_d}(t_d,\infty)$ could be estimated by using a suitable growth function drawn from the technical literature: $B'_{t_d}(t_d,\infty) = G(\infty) * B'_{t_d}(t_d,t_d)$. Taking the estimated size of the ultimate resources discovered in each year ($B'_{t_d}(t_d,\infty)$) and dividing this by the database estimates of the size of those resources ($B'_{t_d}(t_d,t)$) gives an estimate of the *amount of growth that remains to be realized* for each age of field, assuming a particular growth function ($G(\tau)$):

$$\frac{B'_{t_d}(t_d,\infty)}{B'_{t_d}(t_d,t)} = \frac{G(\infty) * B'_{t_d}(t_d,t_d)}{G(\tau) * B'_{t_d}(t_d,t_d)} = \frac{G(\infty)}{G(\tau)}$$

An estimate of the ultimate resources discovered in each year can therefore be obtained from:

$$B'_{t_d}(t_d,\infty) = \frac{G(\infty)}{G(\tau)} B'_{t_d}(t_d,t)$$

Hence, given a suitable growth function, it is possible to correct the annual discovery estimates in the IHS database to allow for future reserve growth (i.e. convert $B'_{t_d}(t_d,t)$ into $B'_{t_d}(t_d,\infty)$). Corrected estimates of the cumulative discoveries through to year t may then be obtained from:

$$B(t_d,\infty) = \int_0^{t_d} B'_{t_d}(t_d,\infty) dt_d$$

Box 4.3 Nonlinear regression

Nonlinear regression is similar to linear regression in that is based upon minimising the sum of squared errors (i.e. the distances between the data points and the estimated curve). A number of algorithms are available, but the most commonly used method was developed by Levenberg and Marquardt. The starting point is an assumed nonlinear functional form which should in principle be based upon a theoretical model. Unlike with linear regression, the user must enter initial values for the relevant parameters, ensuring that these are a reasonable approximation of the best fit model. The computer then systematically adjusts these parameters through a series of iterations in order to achieve the minimum sum of squares. Depending upon the data set and the model, the results may be sensitive to the chosen initial values. As with linear regression, the technique relies upon a number of assumptions including independent and normally distributed error terms with constant variance. If these conditions are not met – which may well be the case – the standard errors of the parameter estimates become invalid. A more comprehensive explanation of nonlinear regression techniques can be found in Bates and Watts (2007).

Source: Motulsky and Christopoulos (2004b)

4.3.2 Results – consistency over functional form

To perform discovery projection, it is necessary to choose an appropriate functional form. Rather than being informed by theory, the choice is typically made on the basis of convenience and goodness of fit and in some cases a number of different functional forms would appear to fit the data equally well. In Section 3.3, it was argued that the appropriate functional form for backdated estimates of cumulative discoveries ($B(t_d,t)$) was likely to be different to that for current estimates of cumulative discoveries (*D(t)*) and that an exponential curve was more likely to be suitable for the former than a logistic curve. This is because an exponential curve can better reflect the discovery of the largest fields early in a region's exploration history. However, *this assumption was not borne out for the ten regions examined here.* Indeed, an exponential curve was found to provide a poor fit to the data in all cases while a logistic curve frequently provided a better fit. Hence, in what follows we confine attention to the fitting of *logistic* and *Gompertz* functional forms to the backdated discovery data. These were chosen because they are widely used and appeared to fit the data equally as well as other sigmoidal functional forms (e.g. the cumulative normal).

The reason for the poor fit of the exponential functional form is unclear. One possibility is that many of the regions were first explored many decades ago when exploration technology was relatively immature. As a result, the fields may not have been found in approximate declining order of size as theory suggests.[66] In regions that have been explored more recently, the time series of cumulative discoveries may be expected to be approximated more closely by an exponential form. To illustrate this, Figure 4.9 compares a 'pioneer' region in which exploration was begun many decades ago, with a 'young' region in which exploration began relatively recently using modern technology. The backdated discovery cycle for the former is approximately sigmoidal in shape, while that for the latter is approximately exponential. It is also possible that the backdated discovery cycle for more aggregate regions (such as the world as a whole) would be more sigmoidal as the slow learning phases of pioneer oil regions

[66] This is the case in the US for example, where for the first fifty years of exploration, the efficiency of the industry in finding large fields was lower than that predicted by a random search (Menard and Sharman, 1975).

would be included in the data. But this again serves to highlight the essentially arbitrary nature of the choice of functional form.

Figure 4.9 Comparison of backdated cumulative discovery trends in 'Pioneer' and 'Young' regions

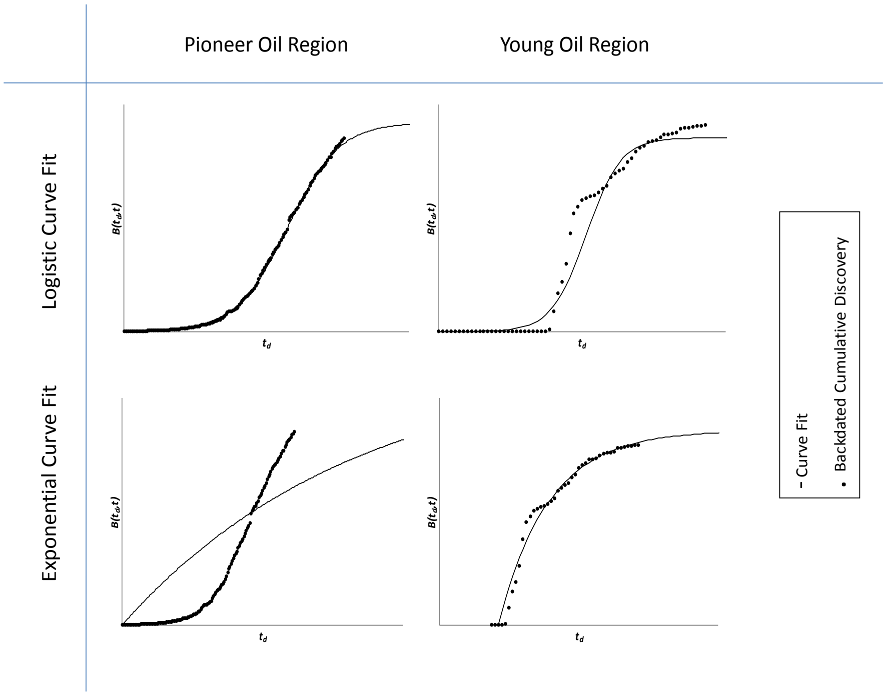


Figure 4.10 presents the results of fitting a logistic model to the cumulative discoveries in Region D. As with most of the regions examined here, this appears to have passed its peak of discovery. An R^2 of 0.999 indicates that the logistic curve provides a very good fit to the data (Box 4.1), but this does not discount the possibility that another model may describe the data more accurately (Motulsky and Christopoulos, 2004b). To illustrate this, Figure 4.11 presents the results of fitting a Gompertz model which is *also* found to give an R^2 of 0.999. However, the corresponding estimate for URR is 33% larger. While different (non-nested) models cannot be compared using R^2 (Box 4.1), the example serves to illustrate that URR estimates can be sensitive to the choice of functional form. Given the good fit of both models and the lack of any theoretical grounds for choosing between them, this raises concerns about the reliability of the discovery projection technique for estimating URR.

Figure 4.10: Logistic discovery projection for Region D

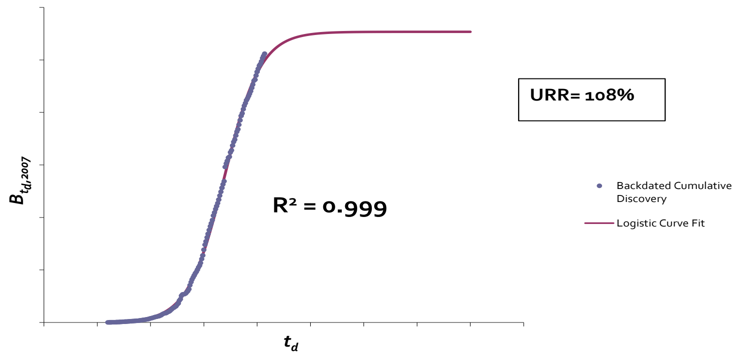
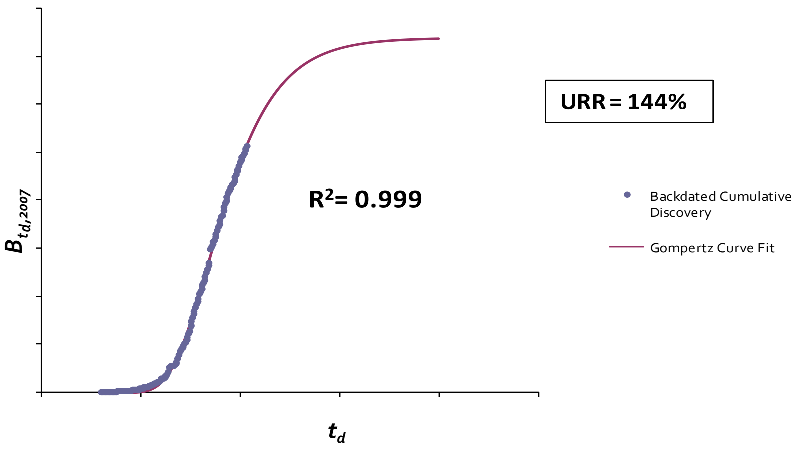


Figure 4.11: Gompertz discovery projection for Region D



The results of this test for all the regions examined are summarised in Table 4.4 and Figure 4.12. Of the ten regions examined, the R^2 value was higher using the Gompertz model in six cases and equal (to three decimal places) in one case (Region D). All the R^2 values were above 0.97 and the mean difference between them was 0.001. In contrast, the mean difference in the URR estimates from the two models, expressed as a percentage of the cumulative discoveries through to 2007, was 59%. The largest difference between the two models was 362% of the cumulative discoveries through to 2007 (Region B), while the smallest was only 1% (Region I). These differences are only related in part to the relative maturity of the discovery cycle. The Gompertz model produced a higher estimate of the URR in all cases, suggesting that the choice of functional form can bias the results.

For four out of the ten regions (B, D, E, F), the difference in URR estimates was more than 25% of the cumulative discoveries through to 2007, while for five of the ten regions (A, C, H, I, J) it was less than 10%. This suggests that the consistency between functional forms with discovery projection can vary widely from one dataset to another. However, there does not appear to be a strong correlation between the magnitude of the R^2 statistics and the consistency between the URR estimates from the two functional forms (e.g. compare regions *B* and *I*) and inconsistent results were also obtained for regions that appeared to be at a relatively mature stage of their discovery cycle (e.g. Region D).

For five of the regions (A, C, H, I and J), the logistic model gave a best estimate of the URR that was less than the current value of cumulative discoveries. For three of these regions, the Gompertz model did the same. These results are clearly in error, but the confidence intervals on these estimates include more plausible values. It is notable that the five regions that produced 'implausible' best estimates with the logistic model are the same regions that produced relatively consistent estimates between the two models (i.e. a difference of less than 10%).

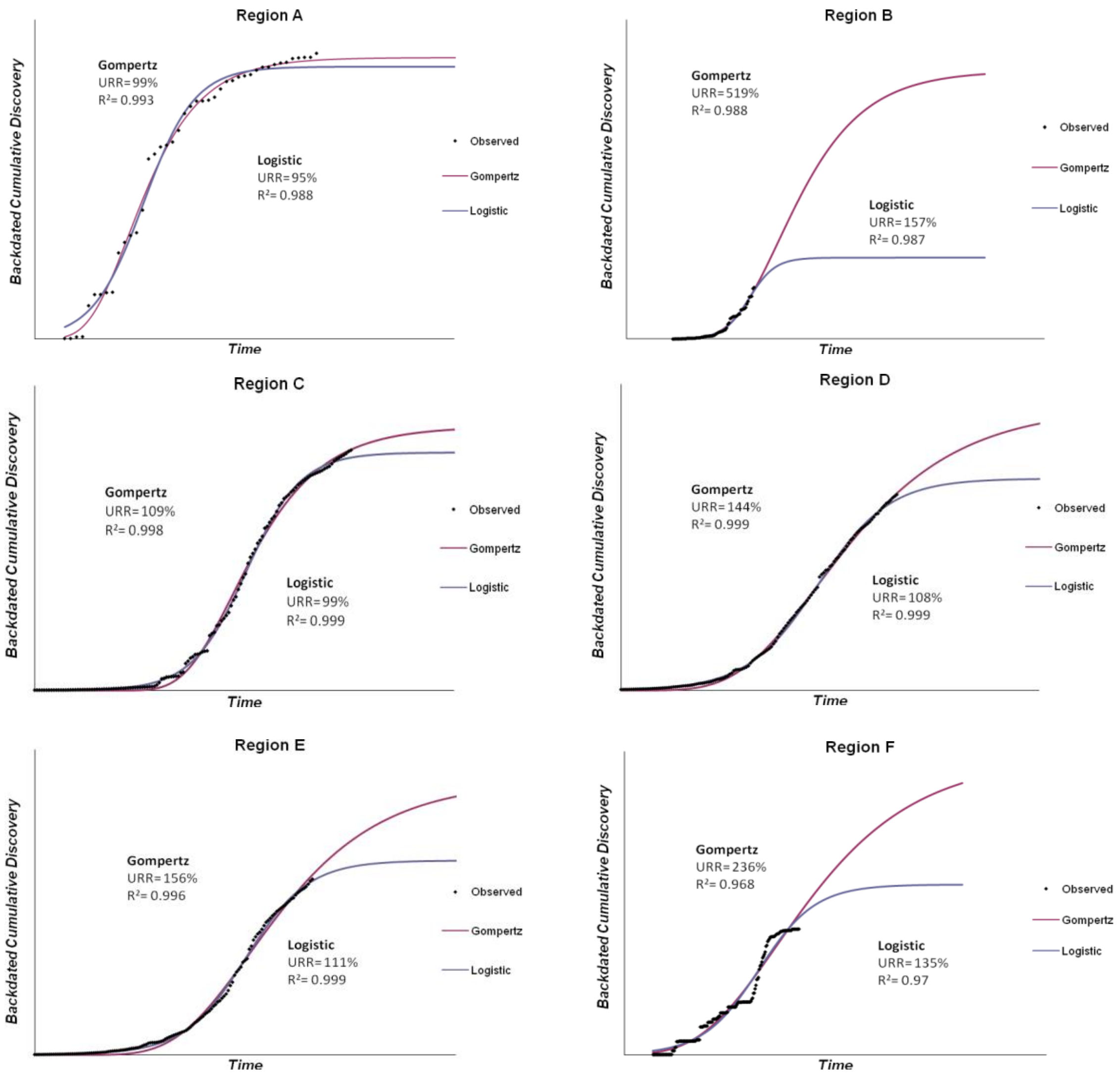
In principle a large estimate of URR relative to D_{2007} suggests that a region is at a relatively early stage in its discovery cycle (i.e. less of the URR has been discovered). This appears to be the case for two regions that gave inconsistent estimates of URR (i.e. B and E) but the same may not apply to regions D and F. Similarly, a small estimate of URR relative to D_{2007} suggests that a region is at a relatively late stage in its discovery cycle. This appears to be the case for the five regions that gave relatively consistent estimates of URR (i.e. A, C, H, I and J). These hypotheses are supported by inspecting the discovery projections (Figure 4.12) which suggests that cumulative discoveries in Regions A, H, I and J are approaching an asymptote.

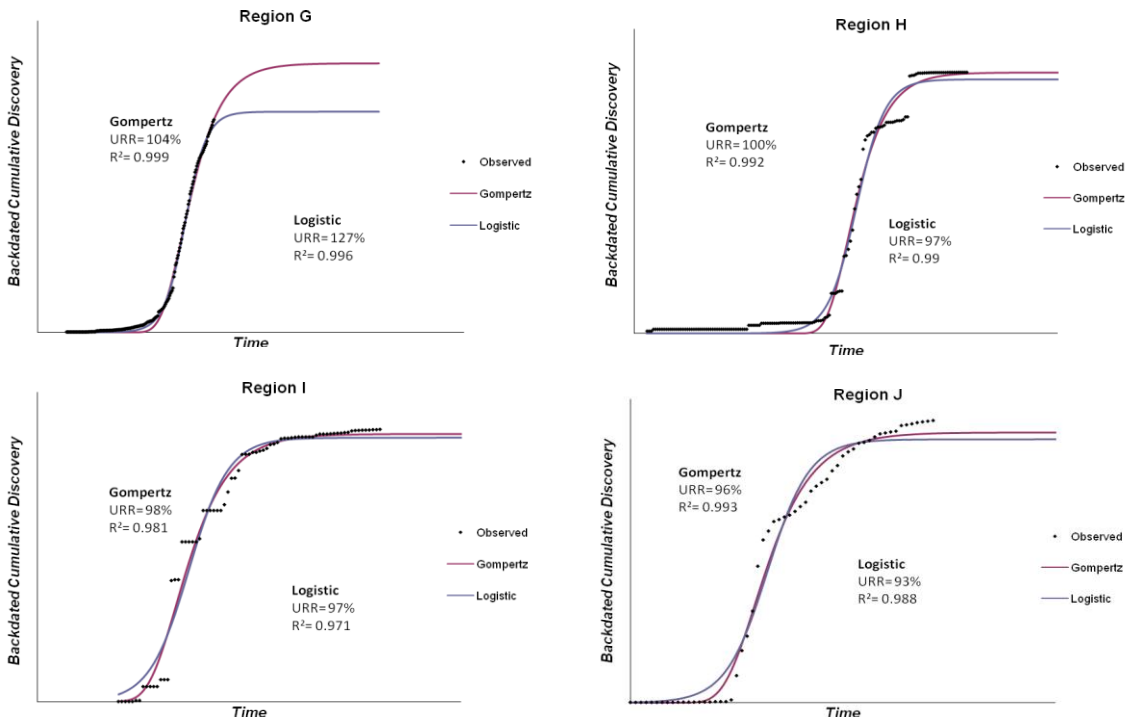
The obvious implication is that the discovery projection technique is more reliable in those regions where the discovery cycle is mature. In contrast, the discovery projection for Region B shows how the corresponding estimates for relatively immature regions can be highly uncertain. Given the confidence intervals on the estimates, an 'implausible' best estimate for the URR (i.e. less than D_{2007}) may be interpreted as suggesting that cumulative discoveries in a region are approaching their maximum. However, this neither rules out the possibility of future discovery cycles in these regions nor takes into account future reserve growth at known fields.

Table 4.4 Summary of consistency over functional form tests for discovery projection

		Logistic		**Gompertz**		**Comparison**	
Region	**Pre/post discovery peak**	**R^2**	**URR (% of D_{2007})**	**R^2**	**URR (% of D_{2007})**	**Logistic minus Gompertz R^2**	**Logistic minus Gompertz URR (% of D_{2007})**
A	Post	0.988	95	0.993	99	-0.005	-4
B	Pre	0.987	157	0.988	519	-0.001	-362
C	Post	0.999	99	0.998	109	0.001	-10
D	Post	0.999	108	0.999	144	0.000	-36
E	Post	0.999	111	0.996	156	0.003	-45
F	Post	0.97	135	0.968	236	0.002	-101
G	Post	0.999	104	0.996	127	0.003	-23
H	Post	0.99	97	0.992	100	-0.002	-3
I	Post	0.971	97	0.981	98	-0.010	-1
J	Post	0.988	93	0.993	96	-0.005	-3
Mean						*-0.001*	*59*

Figure 4.12 Summary of consistency over functional form tests for discovery projection





4.3.3 Results - consistency over time

The consistency over time of the discovery projection technique was investigated by systematically shortening the time series of cumulative discoveries ($B(t_d,t)$) and recording the corresponding change in the estimates of URR. The shortening related to the time period through to the last recorded discovery (t_d) rather than the time at which the estimate was made (t), which remained at 2007. So, for example, we estimated successive discovery projections using the data set $B(t_d,2007)$ where the maximum value of t_d was successively reduced from 2007 to 1985.

Each data set ($B(t_d,2007)$) represents the cumulative discoveries in each year through to time t_d *as estimated in 2007*. It is important to note that this is not equivalent to the cumulative discoveries in each year through to time t_d *as estimated at time* t_d ($B(t_d,t_d)$). This is because there will have been reserve growth in the intervening period ($\tau = 2007 - t_d$). In general, we would expect $B_{t_d,2007} \geq B_{t_d,t_d}$, with the amount of change being proportional to the time interval since the last discovery (τ). To apply discovery projection to the B_{t_d,t_d} estimates would require either: a) access to the IHS databases from all the earlier years (i.e. each t_d); or b) using $B(t_d,2007)$ and an assumed growth function ($G(\tau)$) to estimate $B(t_d,t_d)$ following a procedure similar to that described in Box 4.2.

This means that the regional URR estimates that we attribute to a time series ending at time t_d (i.e. based on $B(t_d,2007)$) may *differ* from the corresponding estimate that would have been obtained had the projection been made using only the data that was available through to time t_d (i.e. based on $B(t_d,t_d)$). It is likely that the use of the latter datasets would lead to URR estimates that were *less* consistent over time than those explored here. In principle, the two approaches would be more likely to give comparable results if the cumulative discovery estimates were corrected to allow for future reserve growth. But since we are testing the consistency of the methods used by the key authors in the field (who do not correct for reserve growth), such an adjustment is not made.

We again employed both logistic and Gompertz functional forms. As an illustration, Figure 4.13 and Figure 4.14 present the results for Region E which is well advanced in its discovery cycle and appears to have passed its discovery peak some decades ago. It is clear that the estimates generated by both models become more consistent as the length of time series increases and also that the confidence intervals converge. However, in contrast to the tests on Hubbert Linearisation, the URR estimates from both the models tend to *fall* as the length of the time-series increases – suggesting that a discovery projection using a shorter time series for this region could lead to an *overestimate* of the URR.

Figure 4.13: Region E – sensitivity of URR estimates from logistic discovery projection to the time through to discovery (t_d)

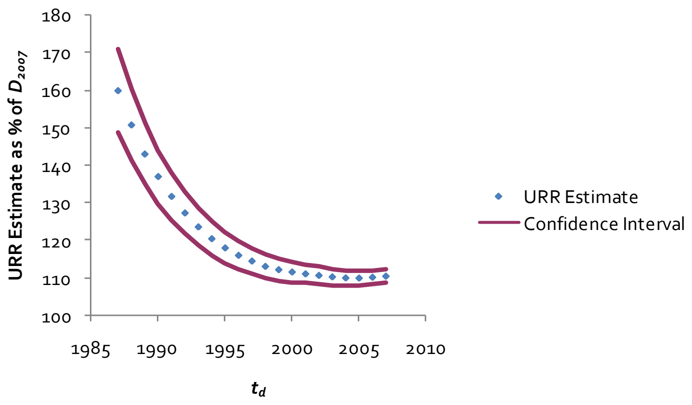


Figure 4.14: Region E – sensitivity of URR estimates from Gompertz discovery projection to the time through to discovery (t_d)

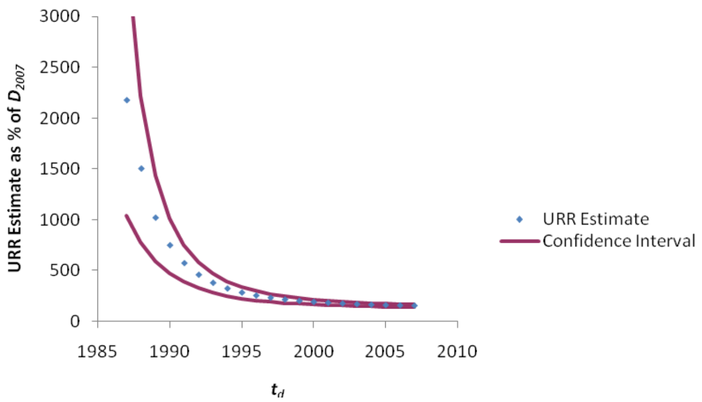


Figure 4.15 and Figure 4.16 present the results of the same process applied to Region B. This seems to be at a relatively early stage in its discovery cycle and does not appear to have passed its discovery peak. In both cases, the *URR* estimates initially decline as the length of time series increases and the confidence interval converges. But then the magnitude of the

estimates begins to increase again and the confidence interval diverges – particularly in the case of the Gompertz model. It is also notable that the estimates from the Gompertz model are generally larger than those from the logistic model – and more than three times larger when the full time-series is employed. This inconsistency results from the relative immaturity of this producing region.

Figure 4.15: Region B – sensitivity of URR estimates from logistic discovery projection to the time through to discovery (t_d)

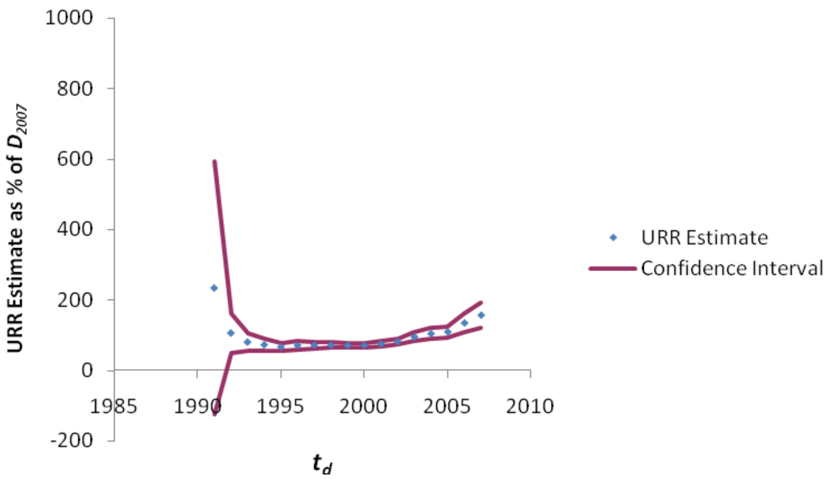
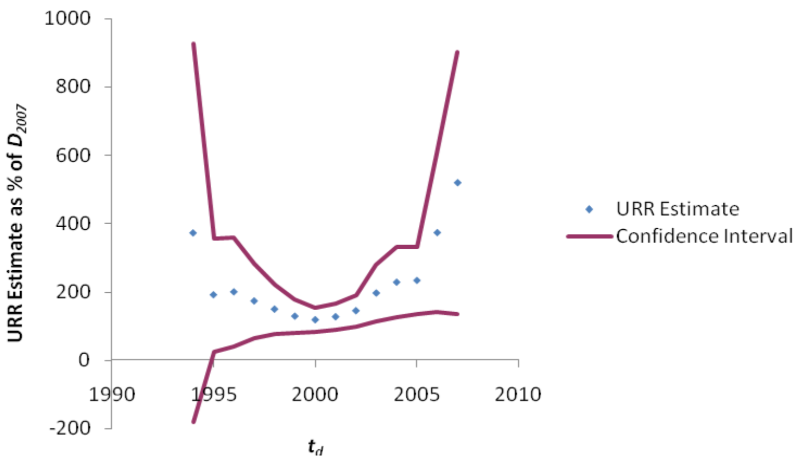


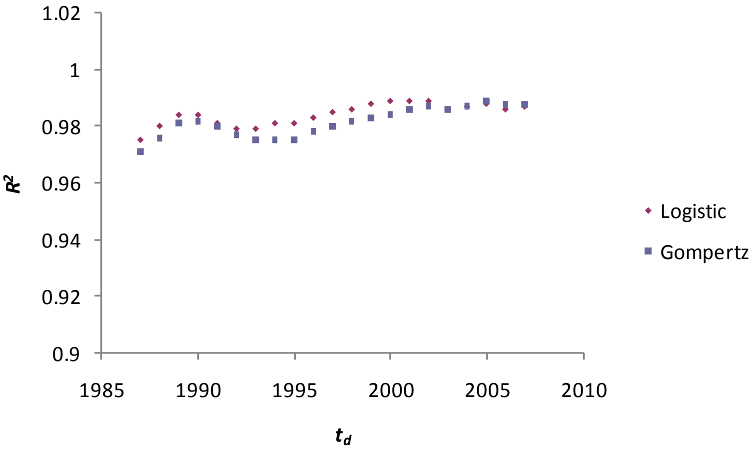
Figure 4.16: Region B – sensitivity of URR estimates from Gompertz discovery projection to the time through to discovery (t_d)



It is also useful to investigate the changes in the R^2 values for the model fits to Regions E and B (Figure 4.17 and Figure 4.18). This shows that the R^2 value does not change significantly in either case and, though the value is more erratic for Region B, it is generally high in all cases. Furthermore, although the model fits to Region B have an R^2 exceeding 0.97, this represents the lowest R^2 value seen over the ten regions examined (again because of the relative immaturity of this region). This again demonstrates that the R^2 value is of little help in selecting the appropriate functional form for discovery projection. But in the absence of any theoretical guidance on the appropriate functional form, and given the wide variation in URR estimates obtained using different functional forms, this again gives little confidence in the results.

Figure 4.17:Change in R^2 estimates for discovery projections in Region E using different lengths of time series

Figure 4.18 Change in R^2 estimates for discovery projections in Region F using different lengths of time series



The results of the consistency over time tests test for all the regions examined are summarised in Table 4.5 and Figure 4.19. The main points are as follows:

- The URR estimates for three out of the ten regions (A, H and I) were relatively consistent over time in both the logistic and Gompertz models, with the difference between the first and last estimate being less than 10%. All three regions appear to be at a relatively late stage in their discovery cycle.
- The URR estimates for four out of the ten regions (B, E, F and J) were inconsistent over time in both the logistic and Gompertz models, with the difference between the first and last estimate being at least 20% (and typically much more). Two of these regions (B and E) appear to be at a relatively early stage of their discovery cycle.

- The URR estimates for the remaining three regions (C, D and G) were relatively consistent over time in the logistic model, but inconsistent in the Gompertz model.
- With the logistic model, the URR estimates increased with the length of the time series in six regions (A, C, D, H, J and I) and decreased in three regions (E, F and G). The corresponding figures for the Gompertz model were four (A, H, I and J) and five (C, D, E, F and G), with two of the regions exhibiting inconsistent behaviour between the two models. Only Region B exhibited first a convergence and then a divergence in URR estimates. In general, these results do not suggest any systematic tendency to either underestimate or overestimate the URR using this technique.
- The regions that exhibited consistency over time in their URR estimates did not necessarily have a 'better' fit in terms of R^2 than those that exhibited inconsistency.

Again, further research is required to identify whether our sample is representative of oil producing regions as a whole, or of regions at different levels of aggregation. The results suggest that discovery projection may lead to more reliable results than the Hubbert Linearisation production data for regions that are at a relatively mature stage in their discovery cycle. However, the technique performs poorly for regions at an earlier stage in their discovery cycle and the degree of inconsistency, both between functional forms and over time, remains relatively high for many of the regions examined here. In contrast to Hubbert Linearisation results, there does not appear to be a systematic tendency to over or underestimate URR in the regions examined here, although the choice of functional form can bias the results. These results have been derived using aggregate data and it is likely that the consistency would be improved by distinguishing between onshore and offshore regions and by conducting the analysis at lower levels of spatial aggregation.

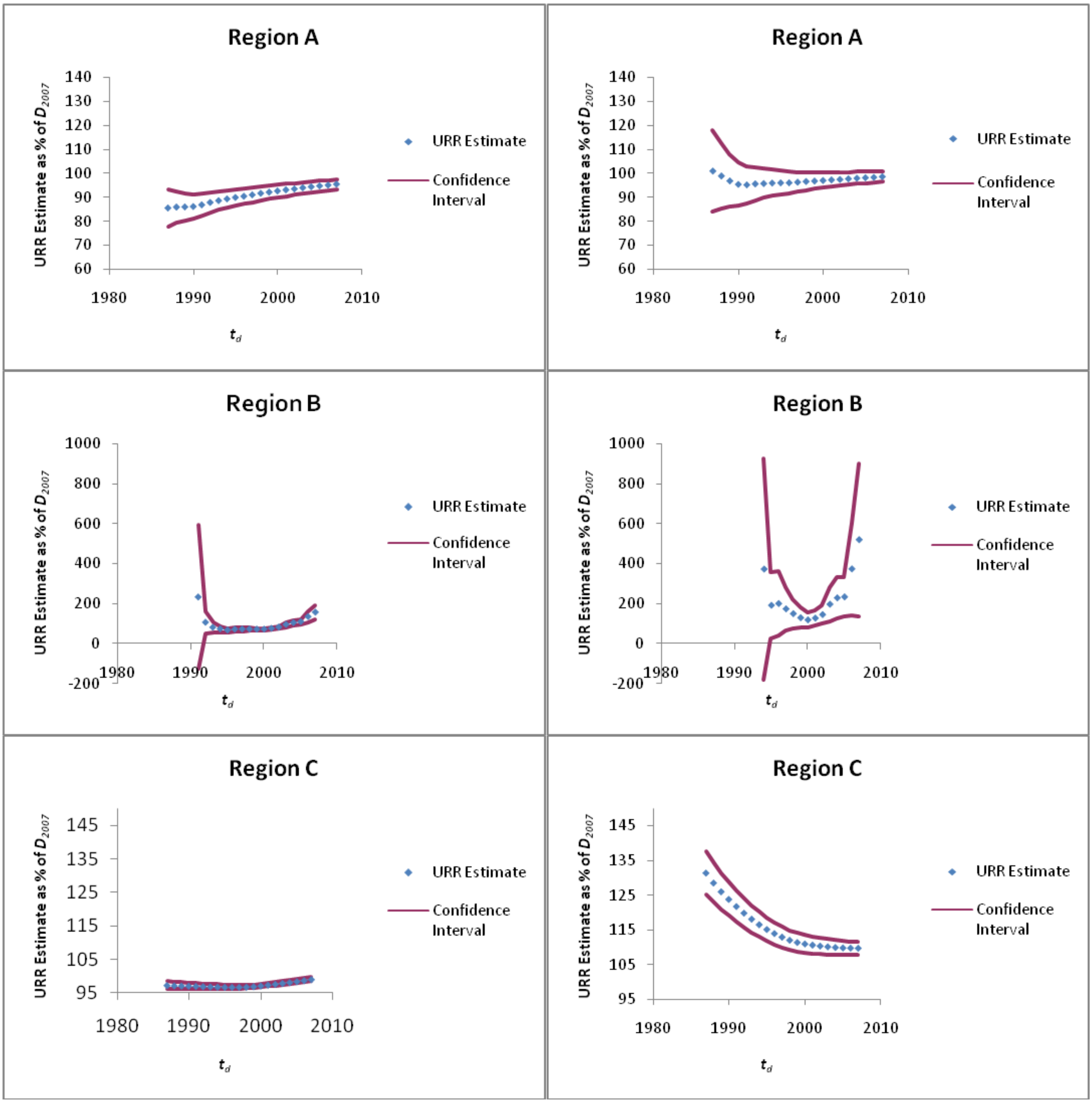
Authors such as Laherrère (2004) claim that some of the drawbacks of discovery projection can be overcome through the use of creaming curves. Section 4.4 investigates whether this is the case for the ten regions examined here.

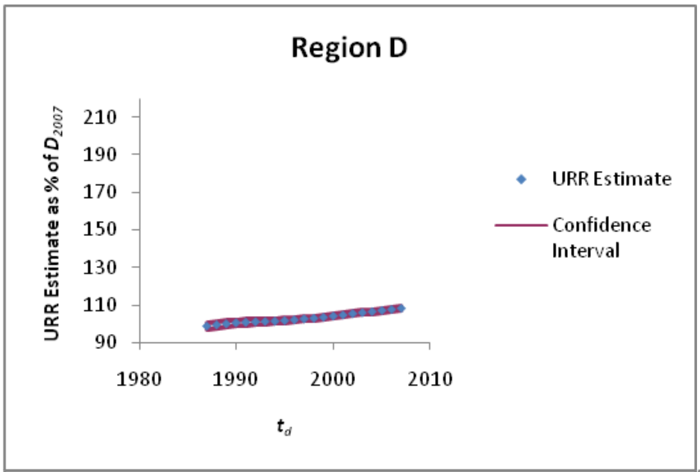
Table 4.5 Summary of consistency over time tests for discovery projection

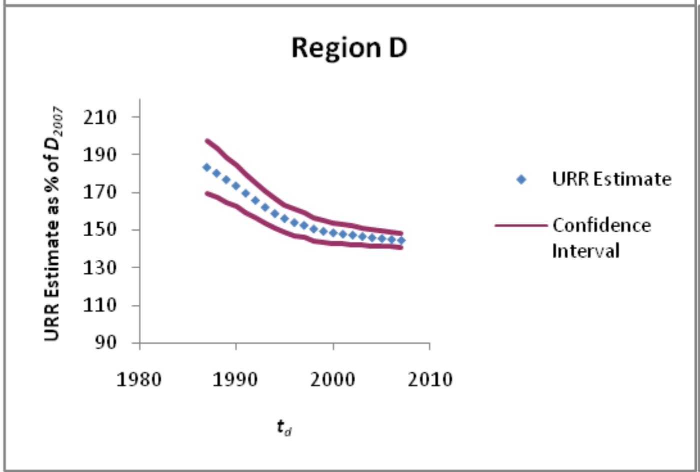
Region			**Logistic**		**Gompertz**	
	Pre/post discovery peak	**Final Gompertz URR estimate as % of D_{2007}**	**First minus last estimate as % of D_{2007}**	**Consistency**	**First minus last estimate as % of D_{2007}**	**Consistency**
A	Post	99	10	Good	2	Good
B	Pre	519	-77	Poor	147	Poor
C	Post	109	2	Good	-22	Poor
D	Post	144	10	Good	-39	Poor
E	Post	156	-49	Poor	-2019	Poor
F	Post	236	-179	Poor	-872	Poor
G	Post	127	-6	Good	-194	Poor
H	Post	100	6	Good	4	Good
I	Post	98	3	Good	2	Good
J	Post	96	21	Poor	21	Poor

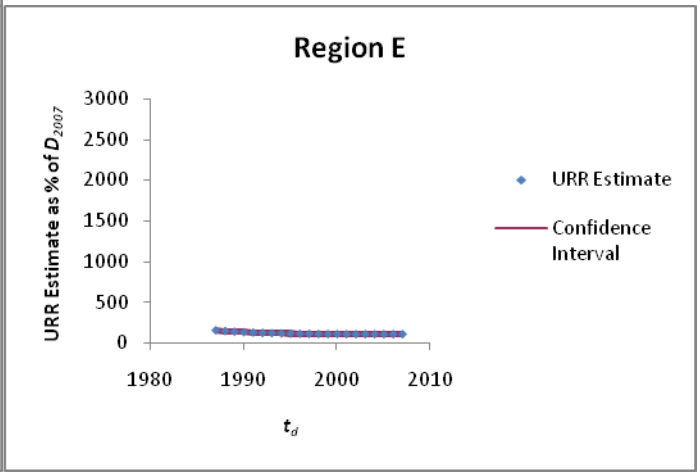
Figure 4.19 Summary of consistency over time for discovery projection

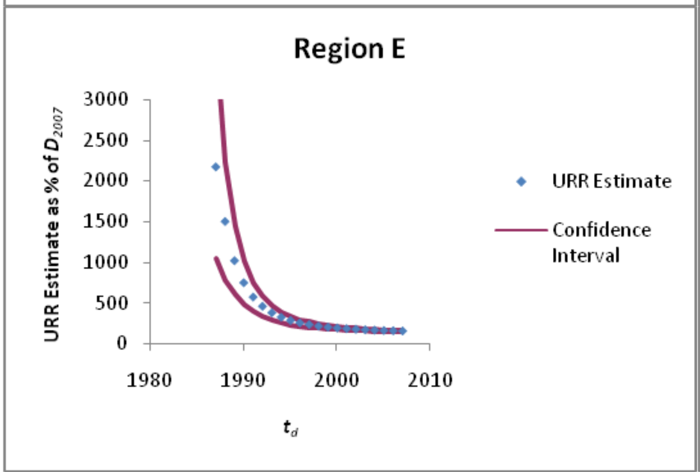
Logistic **Gompertz**

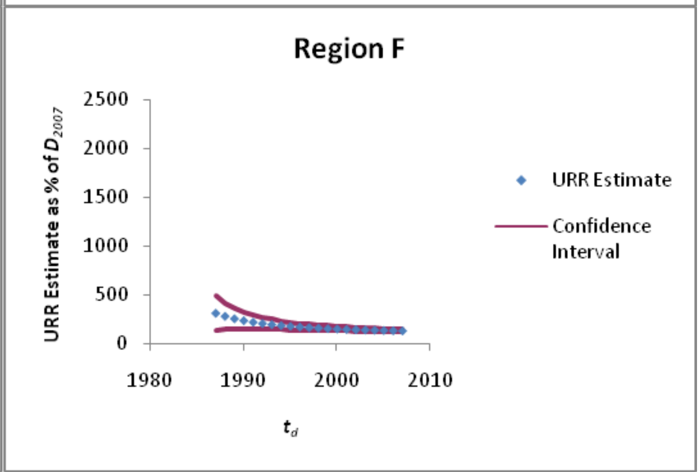


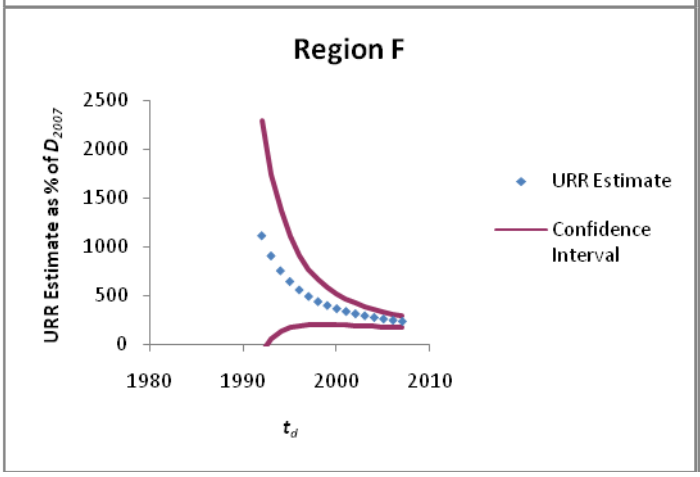

Region D
URR Estimate as % of D_{2007}
210
190
170
150
130
110
90
1980
1990
2000
2010
t_d
URR Estimate
Confidence Interval

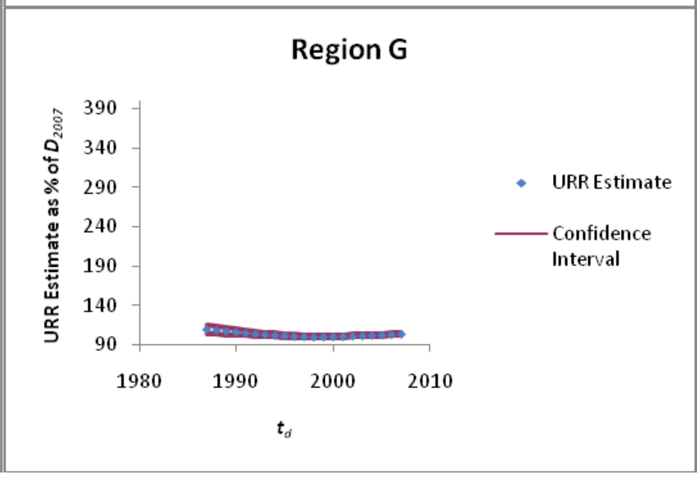

Region D
URR Estimate as % of D_{2007}
210
190
170
150
130
110
90
1980
1990
2000
2010
t_d
URR Estimate
Confidence Interval

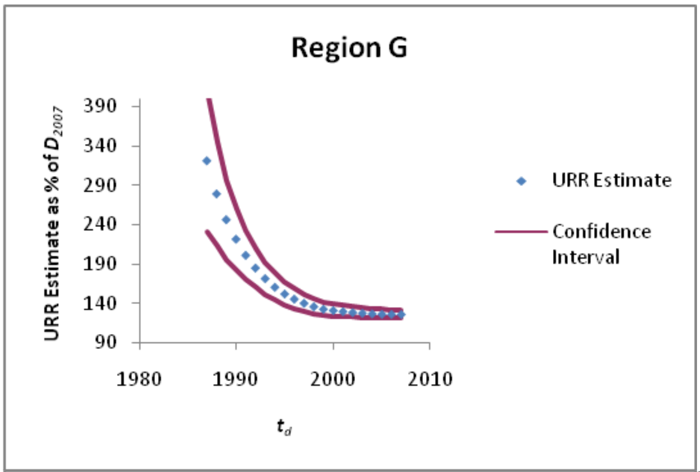

Region E
URR Estimate as % of D_{2007}
3000
2500
2000
1500
1000
500
0
1980
1990
2000
2010
t_d
URR Estimate
Confidence Interval

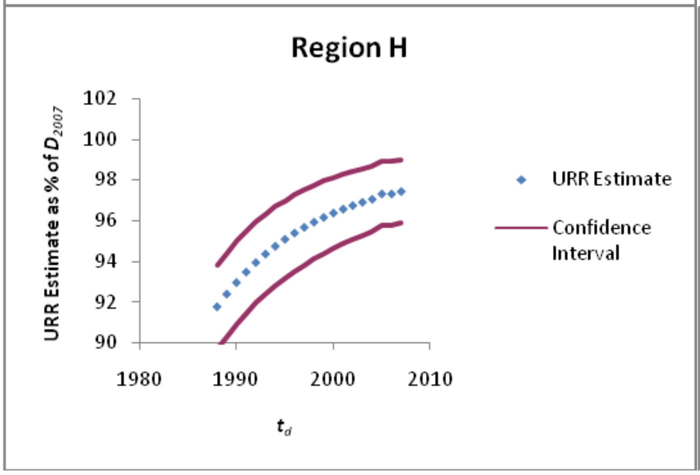

Region E
URR Estimate as % of D_{2007}
3000
2500
2000
1500
1000
500
0
1980
1990
2000
2010
t_d
URR Estimate
Confidence Interval

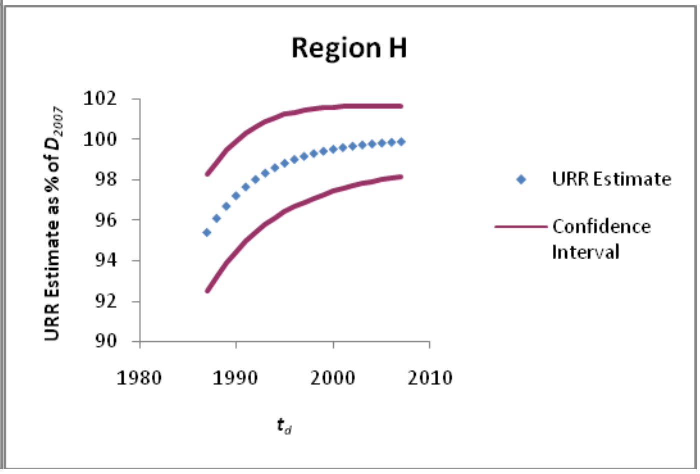

Region F
URR Estimate as % of D_{2007}
2500
2000
1500
1000
500
0
1980
1990
2000
2010
t_d
URR Estimate
Confidence Interval

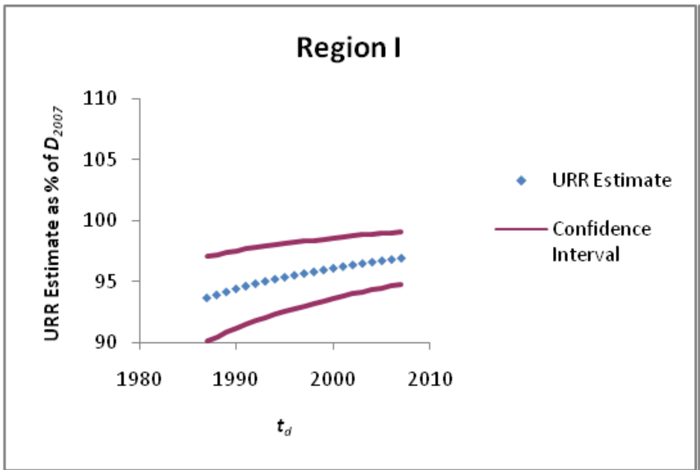

Region F
URR Estimate as % of D_{2007}
2500
2000
1500
1000
500
0
1980
1990
2000
2010
t_d
URR Estimate
Confidence Interval

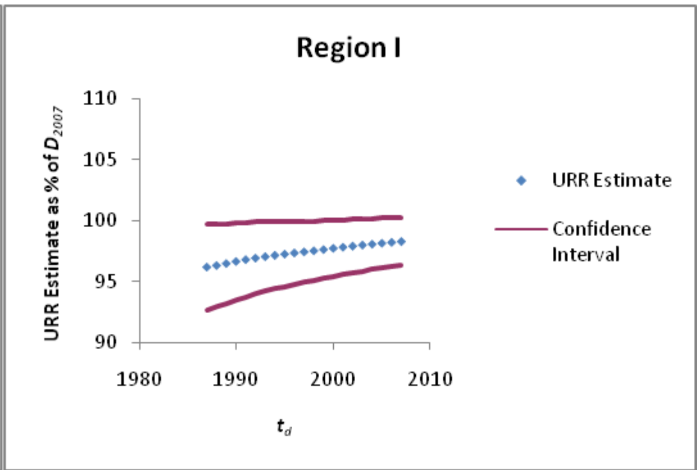

Region G
URR Estimate as % of D_{2007}
390
340
290
240
190
140
90
1980
1990
2000
2010
t_d
URR Estimate
Confidence Interval

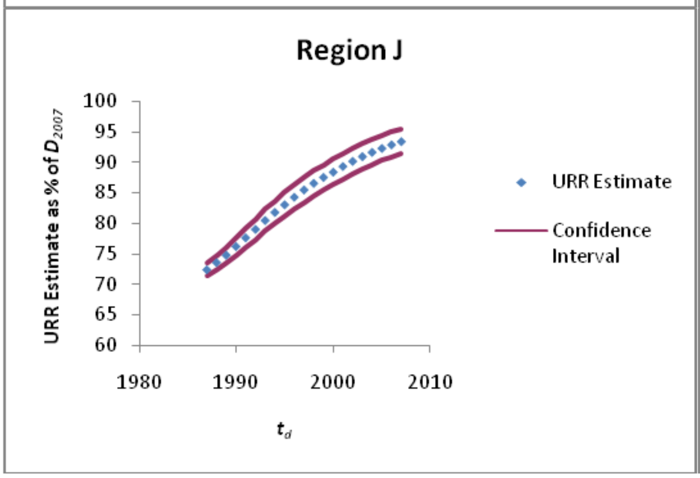

Region G
URR Estimate as % of D_{2007}
390
340
290
240
190
140
90
1980
1990
2000
2010
t_d
URR Estimate
Confidence Interval

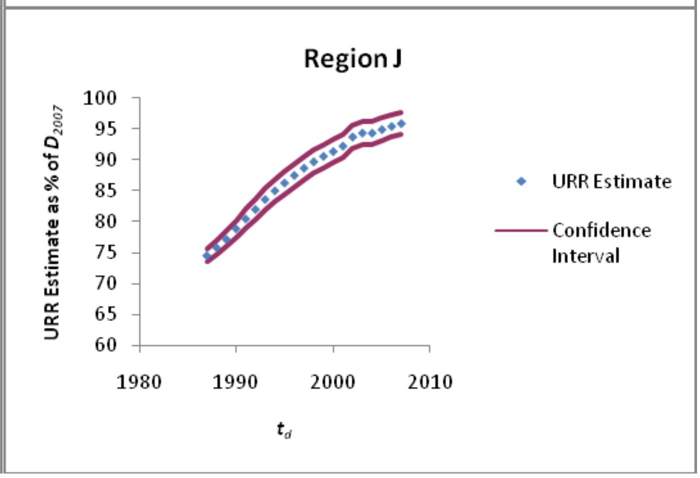

Region H
URR Estimate as % of D_{2007}
102
100
98
96
94
92
90
1980
1990
2000
2010
t_d
URR Estimate
Confidence Interval


Region H
URR Estimate as % of D_{2007}
102
100
98
96
94
92
90
1980
1990
2000
2010
t_d
URR Estimate
Confidence Interval


Region I
URR Estimate as % of D_{2007}
110
105
100
95
90
1980
1990
2000
2010
t_d
URR Estimate
Confidence Interval


Region I
URR Estimate as % of D_{2007}
110
105
100
95
90
1980
1990
2000
2010
t_d
URR Estimate
Confidence Interval


Region J
URR Estimate as % of D_{2007}
100
95
90
85
80
75
70
65
60
1980
1990
2000
2010
t_d
URR Estimate
Confidence Interval


Region J
URR Estimate as % of D_{2007}
100
95
90
85
80
75
70
65
60
1980
1990
2000
2010
t_d
URR Estimate
Confidence Interval

4.4 Creaming curves

4.4.1 Background and approach

The creaming curve technique involves:

- plotting cumulative discovery estimates as a function of exploratory effort (ε);
- using non-linear regression to fit a particular functional form to this data; and
- estimating the URR from the value of the relevant parameter(s) – which corresponds to the asymptote of the curve.

Again, we use *backdated* cumulative discovery estimates ($B(\varepsilon_d, t)$) that are not corrected for future reserve growth. The measure of exploratory effort (ε) is the cumulative number of exploratory, or 'new field wildcat' (NFW) wells that have been drilled in the region.[67] Compared to discovery projection, creaming curves should be less sensitive to various economic and political factors that could affect both the rate ($d\varepsilon / dt$) and success ($B'_{\varepsilon_d}(\varepsilon_d, t)$) or 'yield per effort' of exploratory activity. However, as Cleveland and Kaufmann (1991) have shown, exploratory activity is far from independent of such influences.

We applied the creaming curve technique to the same ten regions as analysed in the previous sections. In each case, we used fitted one or more functional forms to the data using the nonlinear regression function in SPSS©. We first examined the consistency of the URR estimates obtained using different functional forms (Section 4.4.2) and then the consistency of the estimates obtained using both single and multiple curves (Section 4.4.3). Again, the analysis is confined to aggregate data for reasons of time, although this could reduce the accuracy of the technique.

4.4.2 Results - consistency over functional form

Remarkably, while authors such as Campbell and Laherrère publish numerous examples of creaming curves, they never specify the specific functional form employed (Campbell and Laherrere, 1998; Laherrere, 2003). In many instances the curve is referred to as a 'hyperbola', but this term is somewhat ambiguous. In what follows, we use the rectangular hyperbola, which is frequently used in the modelling of biological interactions[68]:

---

[67] The IHS database also contains information on the number of: a) *appraisal* wells at existing fields, including new-pool wildcats, deeper-pool wildcats and shallow-pool wildcats; and b) *development* wells at existing fields, which are used to produce from, inject into, monitor or dispose of liquids from reservoirs. It appears sensible to exclude appraisal and development wells from the measure of exploratory effort, since they refer to drilling activity at *known* fields rather than exploration for *new* fields. But since appraisal and development activity contributes to reserve growth at known fields, it will necessarily affect the 'explained' variable of cumulative discoveries. This suggests that it may be interesting to explore the relationship between *total* drilling activity and cumulative discoveries as well as that between *exploratory* drilling activity and cumulative discoveries. However, only the latter approach is normally used.

[68] The "rectangular hyperbola" is one of the three conic section curves (others being parabola and ellipse). The curve consists of two asymptotic arms which meet in a symmetrical centre. Hyperbola are used in the modelling of many things including biological processes, mirroring the biological analogy seen in the logistic equation. The Michaelis-Menten equation is a rectangular hyperbola used to model enzyme kinetics (Motulsky and Christopoulos 2004).

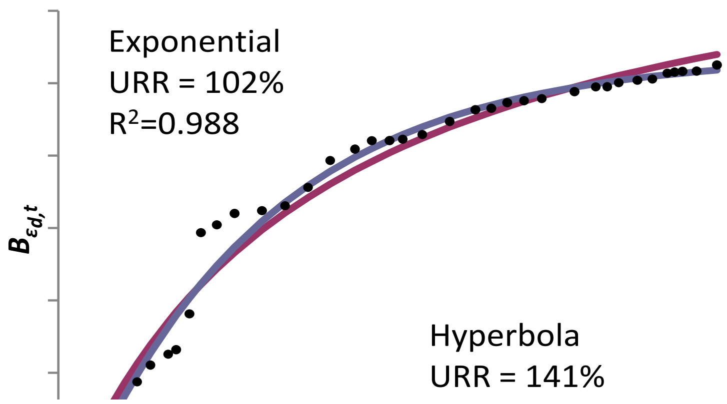
$$B(\varepsilon_d, t) = \frac{(\varepsilon_d * B_{\infty,\infty})}{a + \varepsilon_d} \tag{4.2}$$

However, as with discovery projection, there are no theoretical reasons for choosing this particular functional form and it is legitimate to investigate others that may provide a comparable fit. The primary requirements are that the curve should rise immediately from zero and exhibit asymptopic behaviour. The following exponential function also meets these criteria:

$$B(\varepsilon_d, t) = B_{\infty,\infty}[1 - a\exp(-b\varepsilon_d)] \tag{4.3}$$

As an illustration, Figure 4.20 presents the results of fitting both a rectangular hyperbola and an exponential function to the data for Region A. This region exhibits the 'classic' creaming curve shape, in that exploratory activity initially leads to the discovery of several large fields, but the returns to exploratory effort decrease rapidly as exploration proceeds. Both the exponential and hyperbolic curves provide a good fit to the data for this region, with the former having a marginally greater R^2. But as was seen with discovery projection, the two models provide significantly different estimates of URR, with the estimate from the exponential model being 25% smaller than that from the hyperbolic model. Indeed, the exponential model suggests that 98% of the URR for this region has already been discovered. Since neither of these models have any theoretical justification, the appropriate choice between them is unclear.

Figure 4.20: Hyperbolic and exponential creaming curves for Region A



Only six of the ten regions examined exhibited asymptopic behaviour comparable to that in Figure 4.20. For example, Figure 4.21 presents the results for Region B where the rate of discovery with respect to effort appears to be *increasing* over time – suggesting either increasing success rates or the discovery of increasingly large fields (or both). While a logistic model could be fit to this data, the accuracy of the URR estimate would be highly questionable. This example demonstrates that the creaming curve technique may only be applicable under certain circumstances. Region B is an unusual example, however, since it is at a relatively early stage in its discovery cycle (Figure 4.22).

Figure 4.21 Backdated discoveries as a function of exploratory effort in Region B

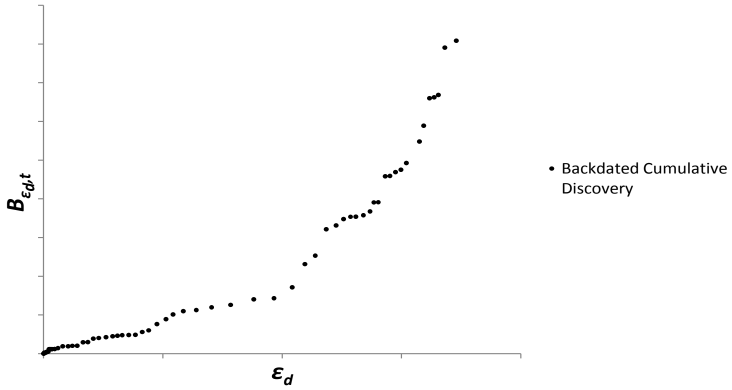
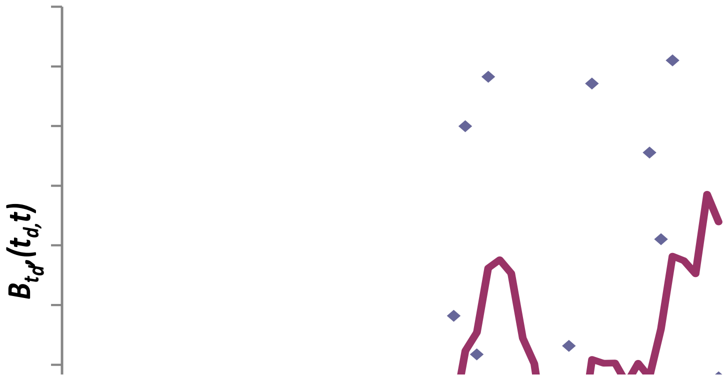


Figure 4.22 Rate of discovery over time for Region B with smoothed 5 year average



There are other examples, however, which suggest that the creaming curve technique may also be inappropriate for regions which are at a relatively advanced stage of their discovery cycle. As an illustration, Figure 4.23 presents data for the highly aggregated Region C. The data here is approximately linear which is supported by the high R^2 for the linear regression. As in Figure 4.21, the results for Region C do not exhibit the expected asymptotic characteristics. But unlike Region B, Region C appears to be well past its peak in oil discovery (Figure 4.24).

Figure 4.23: Hyperbolic and linear creaming curves for Region L

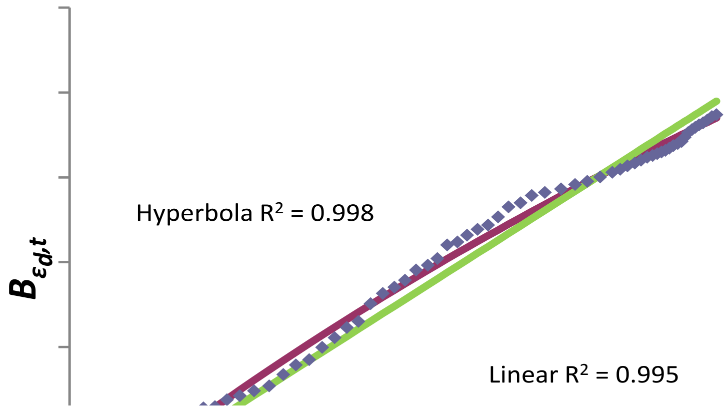
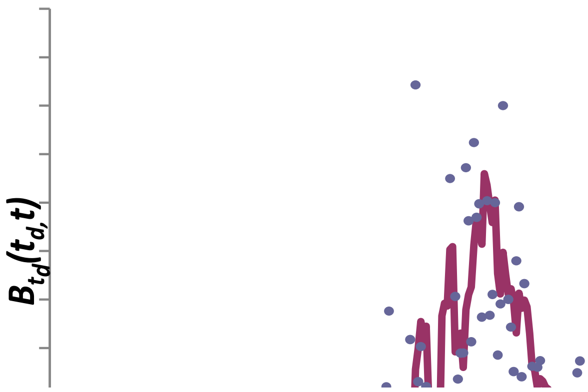


Figure 4.24: Rate of discovery over time for Region L with smoothed 5 year average



It is unclear why the creaming curves for regions C, D and G failed to exhibit asymptotic behaviour. One possibility is the limitations of new field wildcat's (NFWs) as a measure of exploratory effort. Historically, the knowledge of the underlying geology in an oil play was very limited until the first exploratory wells were drilled (NFW's). Under these circumstances, all other variables being equal, the number of NFWs drilled may have provided a reasonably consistent metric for the amount of effort being expended on exploration. But modern seismic technologies permit a much better understanding of the underlying geology and may reduce the need for exploratory drilling. This hypothesis is supported by the observation that the success rate of exploratory drilling is increasing in many regions (Forbes and Zampelli, 2000; IEA, 2008; Managi, *et al.*, 2005). We would still expect a trend towards declining field sizes, but the time-series may also be complicated by the fact that modern techniques allow field sizes to be estimated more accurately.

Three other factors could also be relevant:

- The IHS database does not distinguish between exploratory drilling for oil and that for natural gas. Hence, temporal variations in the relative proportion of exploratory effort devoted to each, together with differences in the yield per effort between oil and gas resources, could have a significant effect on the aggregate creaming curve.
- The analysis in this section uses aggregate, country level data and does not distinguish between onshore and offshore regions. This could complicate the time-series, both

because these regions were explored at different times and also because the YPE may be significantly different for offshore and onshore exploration.[69]

- Even when offshore and onshore regions are separated in the data, they may each encompass several sub-regions distinguished by either geology or exploration history, whose exploration proceeds at different rates at different times with different trends in yield per effort.

Each of these factors could contribute to a straightening of the creaming curve (Figure 4.23), or may make a logistic functional form fit better than an exponential. But whatever the explanation, our results suggest that the 'classic' creaming curve illustrated in Figure 4.20 may not necessarily be representative at the aggregate, country level.

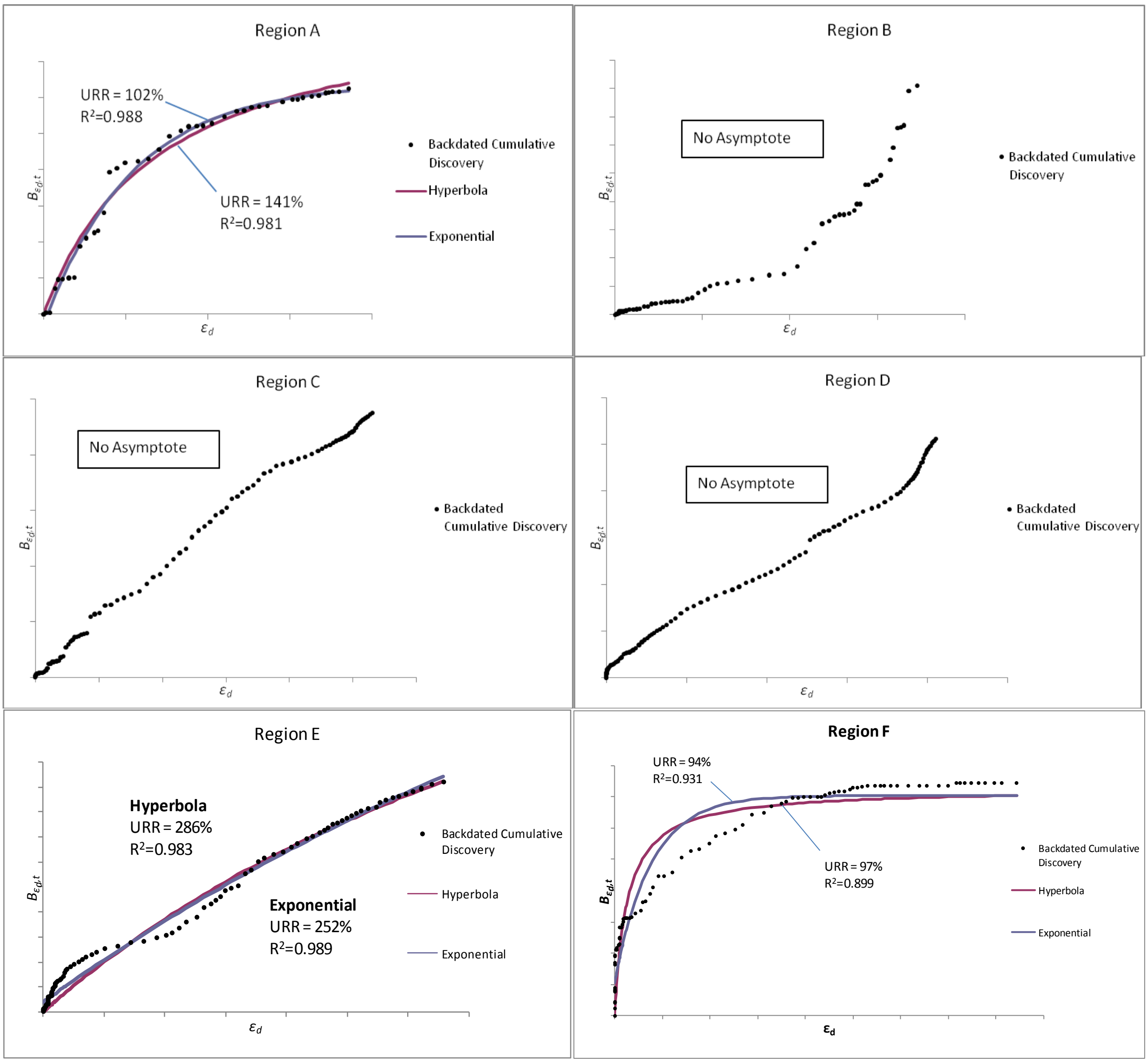
The full results of our consistency over functional form test for creaming curves are summarised in Table 4.6 and Figure 4.25. For the six regions that exhibited asymptopic behaviour, the mean difference in URR estimates from the exponential and hyperbolic functional forms was 27% of D_{2007} while the mean difference in the R^2 was only 0.017. For regions A, E and H the difference between the estimates from the two functional forms was very large, while for regions F, I and J the difference was relatively small. In all cases, the estimate from the exponential functional form was either smaller or equal to that from the hyperbola - again illustrating how the choice of functional form could bias the estimates, despite the minor differences in goodness of fit. As with discovery projection, several of the estimates (F, I and J) were less than the cumulative discoveries through to 2007 which suggests that there is limited scope for new discoveries in these regions - as indicated by the asymptopic behaviour of the creaming curves (Figure 4.25). But again, this neither rules out the possibility of future discovery cycles nor takes into account future reserve growth. While Region I is notable for the large fraction of URR contained in the earliest discovered fields, this pattern is unrepresentative of the regions examined here.

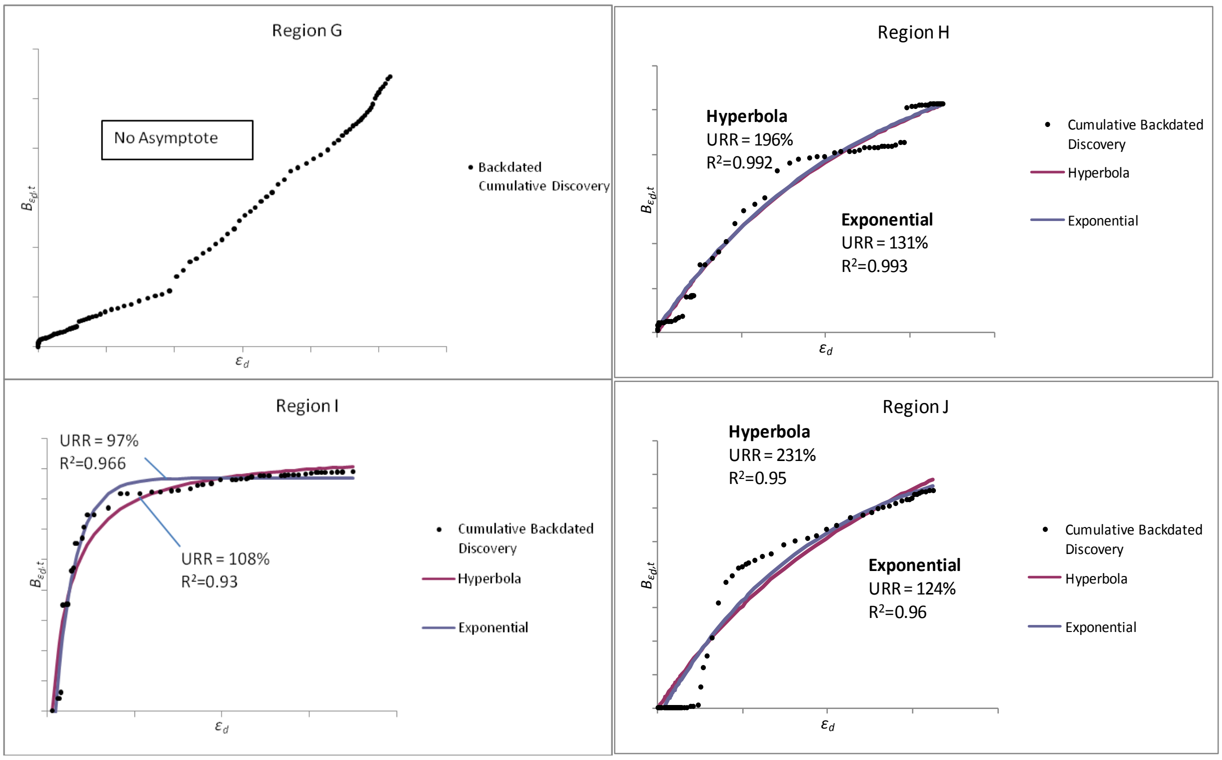
[69] For example, Cleveland and Kaufmann (1997) found that offshore exploratory effort for natural gas in the US had a YPE that was 2 to 20 times greater than that for onshore exploration.

Table 4.6 Summary of consistency over functional form tests for creaming curves

Region	Asymptotic?	Exponential URR (% of D_{2007})	Exponential R^2	Hyperbola URR (% of D_{2007})	Hyperbola R^2	Comparison Diff in R^2	Comparison Diff in URR as % of D_{2007}
A	Yes	102	0.988	141	0.981	0.007	39
B	No	-	-	-	-	-	-
C	No	-	-	-	-	-	-
D	No	-	-	-	-	-	-
E	Yes	252	0.989	286	0.983	0.006	42
F	Yes	94	0.931	97	0.899	0.032	3
G	No	-	-	-	-	-	-
H	Yes	131	0.993	196	0.992	0.001	65
I	Yes	97	0.966	108	0.93	0.036	11
J	Yes	124	0.96	231	0.95	0.019	14

Figure 4.25 Summary of consistency over functional form tests for creaming curves





4.4.3 Results - consistency over the number of curves

Of the six regions exhibiting asymptopic behaviour, only two (A and I) exhibited a 'smooth' curve, while three (E, H and J) appeared to be better approximated by *two* curves. Such situations are frequently encountered in the literature, where it is argued that the individual curves represent discrete areas (distinguished by spatial location, depth, availability for exploration or some other factor) that were explored approximately *sequentially* over the history of the region (Campbell and Heapes, 2008; Laherrère, 2004). In practice, it appears more likely that the individual regions were explored *in parallel*, but the techniques employed by Campbell and Laherrère do not allow this to be accurately simulated. The use of multiple curves can be viewed as an inevitable complication of applying the technique to an aggregate region that is not geologically homogeneous and lacks a consistent exploration history. However, when creaming curves are estimated for individual countries, such a situation may be the norm.

We tested this proposition by fitting *two* curves to regions E, H and J and comparing the resulting URR estimates with those obtained from a single curve. Under the assumption that the returns to exploratory effort have substantially reduced in one region before a second region is opened up for exploration, the appropriate 'breakpoints' can be identified visually and the individual curves can be fit sequentially to different time periods of data. This appears to be the approach taken by Laherrère (2004) and others, but is nevertheless rather crude. More sophisticated approaches are available, that could potentially allow the simulation of multiple areas being explored simultaneously (Meyer, *et al.*, 1999). However, these approaches have not been applied to cumulative discovery data.

Figure 4.26 presents the result of fitting a single rectangular hyperbola to the data from Region E. This gives an R^2 of 0.996, compared to 0.983 for the single curve model. While the two-curve model leads to an estimated regional URR that is nearly twice the cumulative discoveries through to 2007, the single curve model gives an estimate that is half as large again. Hence, the choice between the two models has a significant influence on the overall results.

Figure 4.26: Creaming curve data for Region E fitted with a single hyperbola

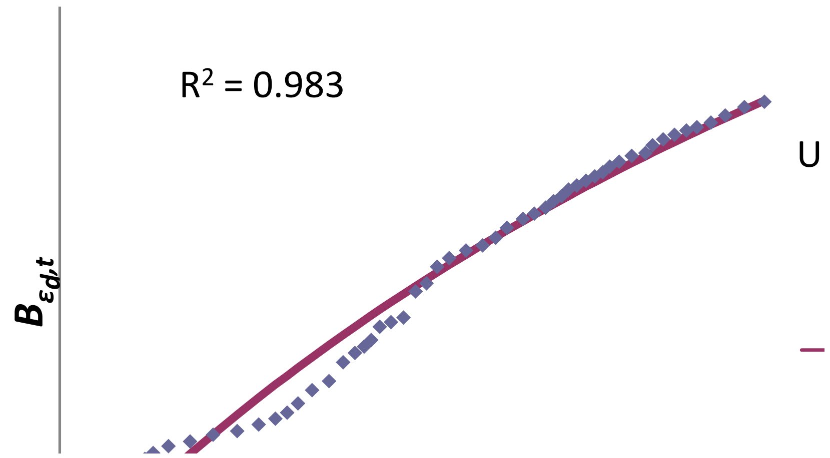
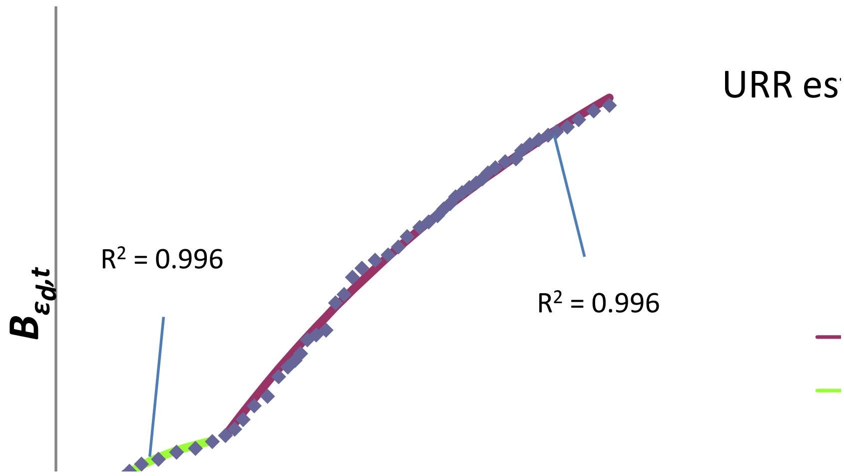
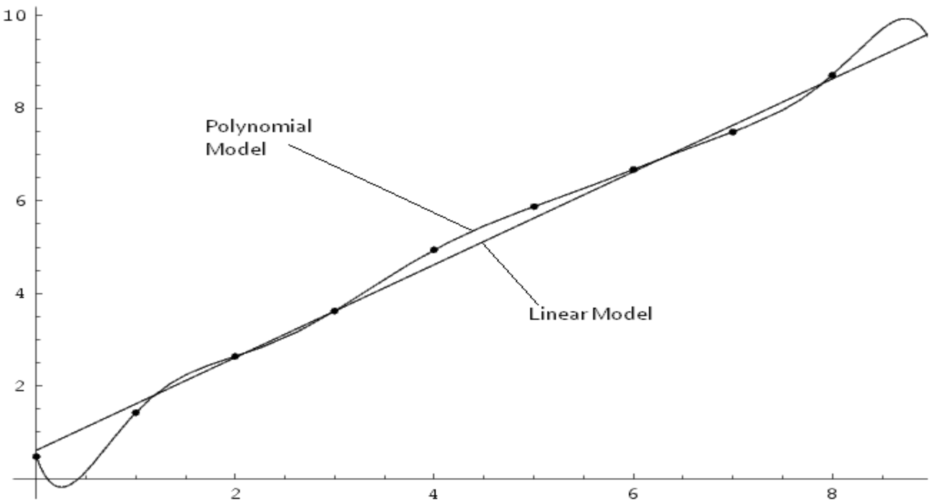


Figure 4.27: Creaming curve data for Region E fitted with two sequential hyperbola



One of the consequences of fitting multiple curves is the division of the dataset into smaller groupings. This creates the risk of *overfitting*, a statistical term which refers to the use of an overcomplicated model to describe too small a data set. In principle, a complicated model with many parameters could pass through every point in the dataset. But if this model is used to extrapolate forward, it is likely to be less accurate than a model containing fewer parameters as a result of increasing deviations at the extremes of the model. To illustrate this, Figure 4.28 compares an overfit polynomial model with a simple linear model. Note the large deviations at the extremes of the polynomial as a result of the overfitting.

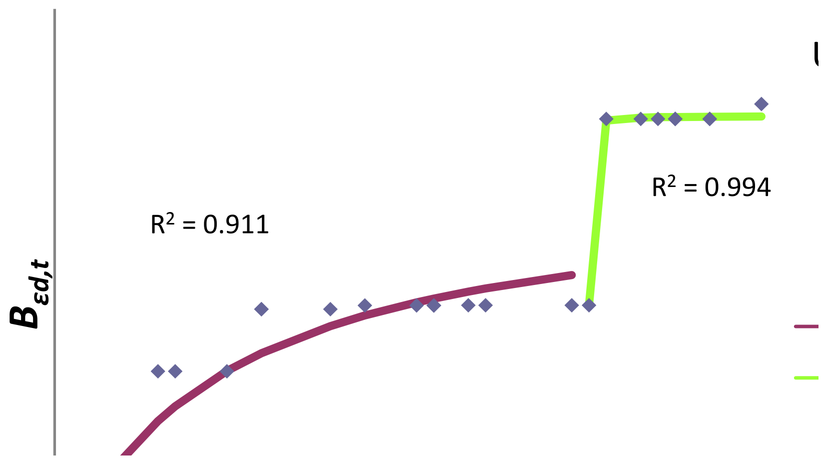
Figure 4.28: Example of the difference between simple linear model and overfitted polynomial



Source: 'Overfitting', Wikipedia, 2009

While there are no strict rules for deciding when a model is overfit, a useful rule of thumb is to ensure there are at least ten data points per parameter (Motulsky and Christopoulos, 2004a). Though the functional forms for creaming curves are relatively simple (with only two or three parameters), the use of multiple curves can easily lead to ratios that are less than this. As an illustration, Figure 4.29 shows how two rectangular hyperbola may be fit to the data for Region H. With two parameters, a rectangular hyperbola should be fit to at least 20 data points, but the second curve in this example is fit to only seven data points (a ratio of 3.5 to 1). Though this may seem a slightly extreme example, very similar curves are published by Campbell and Heapes (2008) and may also be found in the peer-reviewed literature (Laherrère, 2002a).

Figure 4.29: Creaming curve data for Region H fitted with two sequential hyperbola



The results of our consistency over the number of curves tests are summarised in Table 4.7. For the three regions where both one and two curves were fit to the data, the mean difference in the R^2 was 0.017 (with two curves providing a better fit in each case), while the mean difference in URR estimates was 103% of D_{2007} (with two curves providing a smaller estimate in each case). The choice between single or multiple curves can therefore have a significant influence on the URR estimates, but without a detailed knowledge of the exploration history of the region, it is difficult to justify one choice over the other. While authors such Laherrère (2004) make reference to either the geological characteristics or exploration history of region, this appears to be largely an ex-post rationale for a choice of curves that is driven primarily by the ‘look’ of the data.

Table 4.7 Summary of consistency over the number of curves tests for creaming curves

	Single curve		**Two curves**		**Comparison**	
Region	**R^2**	**URR**	**R^2**	**URR**	**Two curve minus single curve R^2**	**Two curve minus single curve URR (% of D_{2007})**
E	0.983	286	0.996	196	0.013	90
H	0.992	196	0.996	100	0.004	96
J	0.950	231	0.983	109	0.033	122
Mean					*0.017*	*102.7*

Note: All estimates derived assuming a rectangular hyperbola

4.5 Comparison of techniques

As a final consistency check, we compare the URR estimates obtained from each of the three extrapolation techniques. In each case, we chose the ‘best fit’ estimate on the basis of the R^2 value, although we recognise the limitations of this statistic when comparing non-nested models (Box 4.1). The results are summarised in Table 4.8. The main points are as follows:

- *Hubbert Linearisation versus discovery projection*: Four of the regions (C, H, I and J) provided relatively consistent URR estimates, while the remainder (A, B, D, E, F and G) provided inconsistent estimates.
- *Discovery projection versus creaming curves*: Of the six regions for which we were able to fit a creaming curve, only one (I) provided a URR estimate consistent with that from the corresponding discovery projection.
- *Hubbert Linearisation versus creaming curve*: Of the six regions for which we were able to fit a creaming curve, three (A, F and I) provided URR estimates consistent with those from the corresponding Hubbert Linearisations.
- *Overall*: Only one region (I) provided URR estimates that were consistent between all three techniques. Moreover, inconsistent URR estimates were frequently obtained for regions that appeared to be at a relatively advanced stage in their discovery and production cycle (e.g. E and A).

Table 4.8 Summary of consistency between techniques tests

Region	1. Hubbert linearisation 'best fit'	2. Discovery projection 'best fit'	3. Creaming curve 'best fit'	1 – 2 % diff	2 – 3 % diff	1 – 3 % diff	Consistency?
A	94	66	102	28	-36	-8	Poor
B	140	519	-	-379	-	-	Poor
C	100	99	-	1	-	-	Poor
D	105	144	-	-39	-	-	Poor
E	397	111	252	286	-141	145	Poor
F	91	135	94	-44	41	-3	Poor
G	160	104	-	56	-	-	Poor
H	96	100	131	-4	-31	-35	Poor
I	105	98	97	7	1	8	Good
J	95	93	124	2	-31	-29	Poor
Mean	*138.3*	*146.9*	*133.3*	*-8.6*	*-32.8*	*13.0*	*Poor*

4.6 Summary and implications

This section has used illustrative data from ten oil-producing regions to investigate the *consistency* of URR estimates from different curve fitting techniques (i.e. the extent to which one estimate differs from another). Any judgment of consistency will depend upon the accuracy expected from the relevant techniques which in turn will depend upon the maturity of the relevant discovery and/or production cycle. For illustrative purposes, we have judged two estimates to be consistent if they differ by less than 20% of the cumulative production (Q_{2007}) or cumulative discoveries (D_{2007}) in a region through to 2007. On this basis, Table 4.9 summarises the overall results. A more or less stringent definition of consistency would not significantly change these results, since most estimates were found to be either broadly consistent or substantially different.

The main findings are as follows.

- The results raise concerns about the reliability of URR estimates from curve-fitting techniques - at least when (as is usually the case) they are applied at the country or regional level. In particular, we note that: a) in only one of the regions examined were the estimates consistent between all three curve-fitting techniques; b) in most cases, inconsistent results were obtained more often than consistent results; and c) the degree of inconsistency was frequently very large. Generally, the estimates were more likely to be consistent for those regions at a later stage of their discovery and/or production cycle, but inconsistent results were frequently obtained for mature regions as well.
- Different functional forms were often found to fit the data equally well (on the basis of R^2) but to provide substantially different estimates of URR. Convergence only occurs at a relatively late stage in the discovery or production cycle when an asymptote is clearly apparent. Given the lack of theoretical guidance on the appropriate choice of functional form, this reduces the confidence in the results. It also suggests that estimates of URR for many of the world's oil producing regions are likely to remain uncertain.
- Particular concern must be expressed regarding the use of the 'Hubbert Linearisation' technique. In addition to important statistical limitations (e.g. the error terms cannot be normally distributed), the frequency of 'breaks' in the linear relationship suggests a

systematic tendency to underestimate the URR. Generally, it would be preferable to fit a non-linear curve to either production data or cumulative production data.

- Discovery projection may provide more reliable results for regions that are at a relatively mature stage in their discovery cycle. However, the technique performs poorly for regions at an earlier stage in their discovery cycle and the degree of inconsistency, both between functional forms and over time, remains relatively high. There does not appear to be a systematic tendency to over or underestimate URR with this technique, but the choice of functional form can bias the results. Contrary to expectations, a logistic functional form was found to be more appropriate than an exponential in all the regions examined here.
- The results do not support the claim that creaming curves are generally more reliable than discovery projection. Notably, the creaming curves for four of the regions did not exhibit asymptotic behaviour, although in two of these cases the corresponding discovery projection was asymptopic. Three regions could be fit with either one or two creaming curves, but the corresponding URR estimates were significantly different (i.e. more than 100% of D_{2007}). Without a detailed knowledge of the exploration history of a region, it is difficult to justify one choice over the other. Also, the use of multiple curves creates the risk of 'overfitting'.
- The lack of consistency between URR estimates may result in part from: a) applying the technique to large regions that are not geologically homogeneous and lack a consistent exploration history; b) not a distinguishing between onshore and offshore regions c) not correcting the discovery data to allow for future reserve growth; and d) (for creaming curves) not distinguishing between the exploration for oil and the exploration for gas. Further research should therefore use the lowest possible level of spatial aggregation, distinguish between onshore and offshore regions and adjust for future reserve growth using functions derive from the technical literature. Disaggregation of exploratory activity would also be desirable, although this is not possible with the data source used here. The reliability of the techniques also needs to be investigated in a much more systematic manner. Nevertheless, these results are sufficient to demonstrate the limitations of curve-fitting technique as applied by advocates such as Campbell. The associated URR estimates and oil supply forecasts should therefore be treated with considerable caution.

Table 4.9 Results of consistency tests - summary

	Maturity of region		**Hubbert Linearisation**	**Discovery projection**			**Creaming curve**	**Comparison between techniques**	**Overall judgment**
Region	**Post discovery peak**	**Post production peak**	**Consistency over time**	**Consistency over functional form**	**Consistency over time (logistic)**	**Consistency over time (Gompertz)**	**Consistency over functional form**		
A	Yes	Yes	Good	Good	Good	Good			Good
B									
C	Yes			Good	Good			Good	Good
D	Yes	Yes	Good		Good				
E	Yes								
F	Yes	Yes	Good				Good		
G	Yes				Good				
H	Yes			Good	Good	Good	Good		Good
I	Yes	Yes	Good	Good	Good	Good	Good	Good	Good
J	Yes	Yes		Good					

5 Statistical robustness of curve-fitting techniques

5.1 Introduction

The literature on curve-fitting techniques to estimate URR has generally paid little attention to the statistical issues involved. Hubbert's methods were relatively crude and most subsequent authors have neither significantly developed these techniques nor fully recognised their weaknesses. However, there are some important exceptions, including Wiorkowski (1981), Kaufmann (1991), Cleveland and Kaufmann (1997) and Pesaran and Samiei (1995).

Most authors have assumed that the 'shape' of the production or discovery cycle can be estimated from the historical data and that this shape will not be significantly affected by any future changes in prices, technology and other relevant variables. As a result, there has been a tendency to neglect these variables, despite the potential errors that may result. Moreover, some of the literature makes some very elementary errors, such as failing to specify either the relevant functional forms or the confidence intervals on the parameter estimates. In contrast, the statistical techniques used in both econometric models of oil supply and discovery process modelling are very sophisticated. This suggests that there could be considerable scope for improving the application of curve-fitting techniques: for example, by including additional variables such as oil prices in the model specification (Kaufmann, 1991); by allowing for structural breaks in the time-series (Reynolds, 2002); and by addressing some of the most important statistical weaknesses, such as serial correlation in the error terms (Pesaran and Samiei, 1995).

The following sections briefly indicate how this could be done. Section 5.2 introduces three statistical issues relevant to curve-fitting techniques, namely model specification and comparison, missing variables and serial correlation. Section 5.3 uses a case study of global production to illustrate these issues and compares results from four different model specifications. Section 5.4 examines missing variables in more detail and uses examples from the literature to show how econometric and curve-fitting techniques may potentially be reconciled. Finally, Section 5.5 summarises the main conclusions and implications.

5.2 Overview of statistical issues

5.2.1 Specification of time-series models

All curve-fitting techniques rely upon *time-series* data on oil production (*$Q(t)$ or $Q'(t)$*) or oil discovery (*$B(t)$ or $B'(t)$*), where the data is typically available on an annual basis. A wide range of statistical techniques are available to analyse such time-series, either for the purpose of understanding the underlying mechanisms that generated the data or to make forecasts. Ordinary least squares (OLS) regression forms the basis of many of these techniques and under certain assumptions this can provide parameter estimates (including URR) that are: a) *unbiased* (i.e. the expected value of the estimates equals the true value); b) *consistent* (i.e. the estimates approach the true value as the sample size increases); and c) *efficient* (i.e. the estimates have minimum variance) (Wooldridge, 2003). However, the required assumptions are frequently violated with time series data, creating the need for re-specification of the model and/or more sophisticated estimation techniques (Wooldridge, 2003).

Curve-fitting relies upon an assumed functional relationship between a time-series of the explained variable and time-series of one or more explanatory variables. Such an approach is a standard feature of econometrics, but what distinguishes curve-fitting is the absence of a theoretical framework that sufficient to justify the assumed relationship. Typically, the explanatory variable is simply time (t) or exploratory effort (ε), but in principle other explanatory variables (x_{nt}) could also be included. Taking production projection as an example, an assumed relationship between cumulative production and time may be written as:

$$Q_t = f(x_{1t}, x_{2t}, x_{3t}.........) + e_t \qquad (5.1)$$

Where each x_{nt} is an explanatory variable defined for each interval of time ($t=1,....T$) and e_t. represents an 'error' term indicating a departure between the assumed functional form ($f(x_{1t}, x_{2t}, x_{3t}.........)$) and the actual value of cumulative production (Q_t). For OLS to provide unbiased, consistent and efficient estimates, this error must be a normally distributed random variable with constant variance.

At a minimum, cumulative production could simply be a function of time itself (t):

$$Q_t = \alpha + \beta t + e_t \qquad (5.2)$$

Where $t=1,....T$. This represents a *global* trend which is assumed to hold at all points in time with the parameters remaining constant throughout – as indicated by line *a* in Figure 5.1. A more complex alternative is a local trend, as indicated by line *b* in Figure 5.1.

Figure 5.1 Gobal (a) and local (b) trends in time-series data

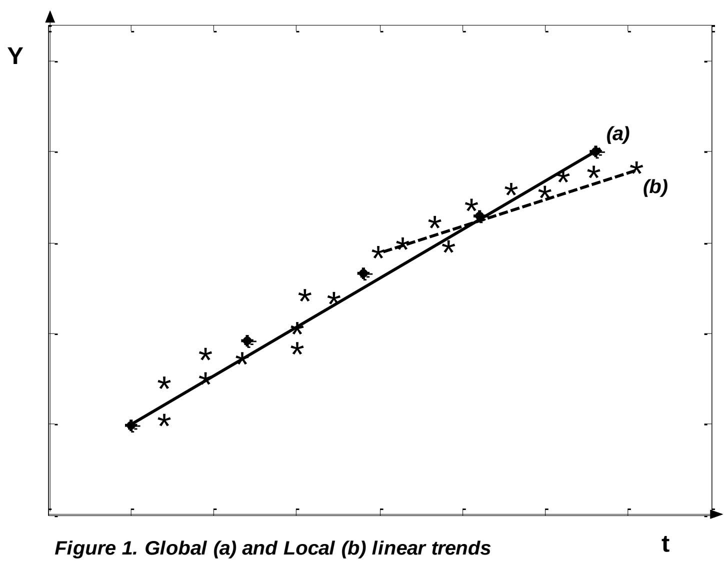


Figure 1. Global (a) and Local (b) linear trends

The specification could be made more complex by including additional explanatory variables (x_{nt}) and/or 'lagged' values those variables (e.g. x_{nt-1}, x_{nt-2}). For example, cumulative production could also be made a function of current oil prices (p_t) together with the oil prices in the previous time period (p_{t-1}):

$$Q_t = \alpha + \beta_1 t + \beta_2 p_t + \beta_3 p_{t-1} + e_t \tag{5.3}$$

Alternatively, the current value of the explained variable (Q_t) may depend upon previous values of that variable:

$$Q_t = \alpha + \beta_1 t + \beta_2 Q_{t-1} + e_t \tag{5.4}$$

A model in which the explained variable depends *only* on previous values of that variable is termed an *autoregressive model* (*AR*). An example is: $Q_t = \alpha + \beta_1 Q_{t-1} + e_t$ If e_t is a normally distributed random variable, this equation implies that Q_t is some percentage (β_1) of its value in period *t-1* plus a random shock. This is termed an autoregressive model of order one (*AR(1)*) since there is only one 'lag' of the explained variable. An *AR(2)* model is give by: $Q_t = \alpha + \beta_1 Q_{t-1} + \beta_2 Q_{t-2} + e_t$. For stability, the coefficients on the lag terms must be less than unity. If $\beta_n = 1.0$, the model is said to be integrated of order one (*I(1)*). Statistical procedures are available to test for *I(1)* variables and the standard approach to estimating such models when $\beta_1 = 1.0$ is to take the *first difference*: $\Delta Q_t = Q_t - Q_{t-1}$. The aim of differencing is to provide a *stationary* dependent variable, which means that the mean and variance are constant over time.

The assumed functional form ($f(x_{1t}, x_{2t}, x_{3t}.........)$) may either be *linear*, such as a polynomial, or *nonlinear*, such as an exponential. Note that the term linear refers to the parameters rather than the explanatory variables. So for example, if β is a parameter and *t* is an explanatory variable, then a model containing the term βt^2 is *linear* while a model containing the term t^β is *nonlinear*. In many cases, a model can be made linear in parameters through a transformation. For example, if cumulative production follows an exponential decay ($Q_t = \alpha \exp(-\beta t)$), the equation can be made linear by taking logarithms ($\ln Q_t = \ln \alpha - \beta t$). However, it is not always appropriate to transform the data in this way and sometimes such a transformation is not possible (Myers, *et al.*, 2002). In these circumstances, the equation may be estimated through nonlinear regression techniques.

Two key aspects of statistical analysis are: a) assessing the *goodness of fit* of a model (i.e. how close the predicted values are to the true values); and b) estimating the *precision* of the estimated parameters (i.e. their standard errors or confidence intervals). A discussion of various measures of goodness of fit was provided earlier in Box 4.1. The usual approach is to fit several different models and choose the one that gives the 'best' fit under one or more statistical measures. In addition, if the model is being used for forecasting, it is important to estimate confidence intervals for the forecast, based upon the standard errors of the parameter estimates.

Another criterion for judging a model is the accuracy of its 'out of sample' forecasts, which must be distinguished from the "goodness of fit" within the sample. To do this, set of data points from the end of the time-series is typically withheld during model estimation. These data points can then be compared with forecast values and the difference measured by the minimum mean squared error (*MMSE*) or some related measure.

An important weakness of the curve-fitting literature is the lack of attention to model specification and the limited discussion of goodness of fit, the precision of parameter estimates (including URR) and the confidence intervals of the relevant forecasts. For example, in their comprehensive 'atlas' of oil depletion, Campbell and Heapes (2008) provide no information on these issues at all.

5.2.2 Missing variables in model specification

In general, simpler models (i.e. fewer x_{nt}) are preferred because the inclusion of too many variables reduces the precision of the parameter estimates (β_{nt}). Also, a simpler model is easier to understand and adding variables reduces the degrees of freedom and hence the 'power' of hypothesis tests (Ramanathan, 2002). However, one of the conditions for OLS to give unbiased, consistent and efficient estimates is that *all* of the statistically significant variables (x_{nt}) are included in the model (including the relevant lagged values such as x_{nt-s}). If a statistically significant variable is omitted from the model, the result can be *omitted variable bias* which has the following consequences:

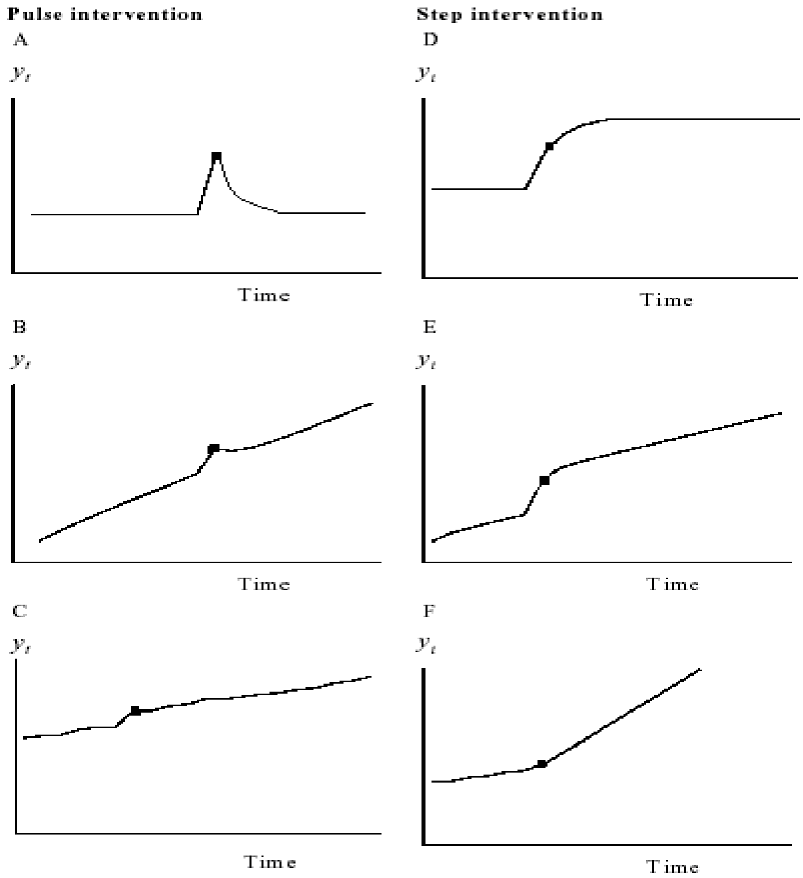
- The estimated values of all the regression coefficients will be biased unless the omitted variable is uncorrelated with every included variable.
- Even if this condition is met, the estimated constant term (α) will be biased and hence forecasts will be biased.
- The estimated variance of the regression coefficients will be biased and hence hypothesis tests will be invalid.

Omission of relevant variables can therefore have serious consequences. Since curve-fitting techniques typically only include a single explanatory variable (i.e. either time (t) or exploratory effort (ε)) the resulting estimates of URR will only be unbiased if the omitted variables are uncorrelated with time or exploratory effort or have little or no influence on production and/or discovery trends. This seems unlikely in principle and a number of authors have shown that it does not hold in practice (Kaufmann, 1991; Kaufmann and Cleveland, 2001).

One important category of missing variables results from specific, one-off events, such as a major fall in price or policy intervention. Relevant examples include the oil crises of 1973 and 1979, hurricanes in the Gulf of Mexico, the Piper Alpha disaster in the North Sea and the opening of a new region for exploration. These changes may be either abrupt or gradual and may lead to either a temporary or a permanent change in the relevant time-series. In Figure 5.2 for example, A and B represent an abrupt and temporary change, D and E represent an abrupt and permanent change and C and F represent a gradual and permanent change. While simple curve-fitting cannot accommodate such changes, specific econometric techniques are available to detect and model them (Perron, 1997).

The consequences of omitting relevant variables and the potential benefits of including them will be discussed further in Section 5.3.4.

Figure 5.2 Structural breaks in time series



5.2.3 Serial correlation in the error terms

A second (and related) condition for OLS to give unbiased, consistent and efficient estimates is that the error terms be *uncorrelated* (i.e. $Corr(e_t, e_s | \bar{x}) = 0$ for $t \neq s$). When this condition does not hold the errors are said to suffer from *serial correlation*, or *autocorrelation,* because they are correlated across time. For example, consider the case of errors from adjacent time periods. Suppose that when $e_{t-1}>0$ then, on average, the error in the next time period, e_t, is also positive. Then $Corr(e_{t-1}, e_t) > 0$. In the simple model of Equation (5.2), 'first order' serial correlation may be written as:

$$Q_t = \alpha + \beta t + e_t + \rho e_{t-1} \qquad (5.5)$$

Where ρ $(-1 < \rho < 1)$ is the first order autocorrelation coefficient. Higher order serial correlation is also possible. A common way of checking for the presence of serial correlation is to plot the *autocorrelation function* (ACF) which indicates the correlation between Q_t and Q_{t-k} as a function of k, and to look for any patterns. Also useful is the *partial autocorrelation function* (PACF) which indicates the correlation between Q_t and Q_{t-k} as a function of k after removing the effect of the intermediate Q's.[70] There are also formal statistical techniques to

[70] A large proportion of the correlation between Q_t and Q_{t-k} may be due to the correlation they had with the intervening lags – Q_{t-1}, Q_{t-2}, Q_{t-3} and so on. The partial correlation removes the influence of these intervening variables.

identify the presence of serial correlation, the most common of which is the 'Durbin Watson statistic.[71]

Serial correlation commonly occurs as a result of omitted variables (Pindyck and Rubinfeld, 1998). For example, in fitting a logistic model to global production data, the predicted values are consistently too high after the 1973 oil crisis because the logistic model does not include variables that allow the modelling of an oil embargo. Serial correlation may also be caused by measurement errors and incorrect specification of the functional form. In principle, if the appropriate number of lags of all the relevant variables is included within an appropriate functional form, there should be no serial correlation.

If the serial correlation results from omitted variables, the OLS estimates and the forecasts based upon them could be biased. If it derives from some other source, the estimates and forecasts will be unbiased and consistent, but will be *inefficient,* implying that the precision of URR estimates will be overstated. It is commonly the case that the explanatory variables grow over time and the serial correlation is positive. Under these conditions, serial correlation will lead to underestimates of the variance of the parameters (including URR) and overestimates of the R^2 for the model. This means that the goodness of fit of the model will be exaggerated, together with the accuracy of the parameter estimates.

If serial correlation is found, the model should first be re-specified to assess whether the cause is missing variables, incorrect functional form or inappropriate lag structure. If this possibility has (as far as possible) been eliminated, there are a number of techniques to estimate models in the presence of serial correlation which typically involve estimating the relevant autocorrelation coefficients (ρ) and using this to transform the variables. However, these tests and procedures are very rarely used in the simple curve-fitting models used to estimate URR. This is an important weakness since in practice the errors are very likely to be serially correlated (Considine and Dalton, 2008; Kaufmann, 1991; Pesaran and Samiei, 1995)

5.2.4 Forecasting

Curve-fitting techniques typically take historical data on cumulative production or discoveries and forecast *future* production or discoveries under the questionable assumption that the relationships identified in the past (e.g. logistic growth) will continue into the future. This can be understood, therefore as a particular form of *forecasting*.

The statistical literature recognises two general approaches to forecasting. *Econometric* forecasting relates an explained variable to one or more explanatory variables and therefore seeks to understand and model the physical and behavioural factors that influence the explained variable. In contrast, *time-series* forecasting attempts to predict the future values of the explained variable solely on the basis of the past values of that variable. The time-series approach has generally been found to be superior to the econometric approach for short-term forecasting, but less so for long-term forecasting. However, the boundary between these two approaches is blurred (Ramanathan, 2002).

[71] A Durbin-Watson statistic of 2 indicates that the data are not serially correlated, while values below 1 indicate significant positive serial correlation and values greater than 4 indicate significant negative serial correlation. However, this test is invalid if lagged dependent variables are present and it often gives inconclusive results. Hence, other tests are frequently preferred (Wooldridge, 2003).

Both econometric and time-series forecasting provide an alternative to simple curve-fitting as a basis for forecasting future production or discovery. However, pure time-series models do not permit simple 'what if' questions to be asked and since they do not explicitly model depletion, they are much less suitable for estimating URR. The basis of time-series forecasting is the autoregressive moving average (ARMA) model, which is defined as follows:

$$Q_t = \beta_1 Q_{t-1} + \beta_2 Q_{t-2} + \beta_3 Q_{t-3} + ...\beta_p Q_{t-p} + \\ v_t + \chi_1 v_{t-1} + \chi_2 v_{t-2} + \chi_3 v_{t-3} + ...\chi_q v_{t-q} \quad (5.6)$$

The ARMA model expresses the current value of the explained variable (Q_t) as a linear combination of an *autoregressive* model of order p (AR(p)) and a *moving average* model of order q (MA(q)), where the latter is a linear combination of random variables (v_{t-n}). The combination is typically referred to as a *ARMA(p,q)* model and is widely used.

ARMA models are only suitable if the time-series of the dependent variable (Q_t) is *stationary*, meaning that it has a constant mean and a variance that does not change over time.[72] Trending variables such as *Q(t)* are not stationary, but it is possible to convert most non-stationary time-series to a stationary form through the process of *differencing*. For example, consider a linear trend of the form $Q_t = \alpha + \beta t$. The *first difference* of Q_t is given by $\Delta Q_t = Q_t - Q_{t-1} = \beta$ which is constant and hence stationary. Hence, a linear trend can be removed by differencing once. Similarly, a quadratic trend can be removed by differencing twice. If a non-stationary time-series can be converted to a stationary one by differencing d times the series is set to be *integrated of order d* and is written *I(d)*. The differenced stationary series can then be modelled as an *ARMA(p,q)*. In this case, the process that generates the series is called an *autoregressive integrated moving average* and the models are termed ARIMA models, denoted as *ARIMA(p,q,d)* (Ramanathan, 2002).

Differencing may not be helpful in econometric analysis, since we are often interested in the determinants of the level form of the dependent variable rather than the first differenced form. Fortunately, a group of specialised econometric techniques based upon the so-called *cointegration* between different variables can allow such relationships to be investigated even when these variables are *I(1)* (Engle and Granger, 1997). These techniques are now standard feature of time-series econometrics, but there are very few applications to the estimation of URR.[73]

5.2.5 Summary

In summary, the use of curve-fitting techniques to estimate URR raises a number of standard but important statistical issues that are inadequately addressed in much of the literature. The most important of these are:

- the different options for model specification and the need to test different specifications;

---

[72] With a stationary series, the correlation between the variable at time *t* and that at time *s* depends only on the distance between the two time periods (*t-s*).

[73] A notable exception is Kaufmann and Cleveland (2001).

- the different approaches to comparing model specifications and the inadequacy of R^2 as a criterion for choosing between different models (Box 4.1);
- the possibility of missing variable bias and a consequent risk of bias in the estimates of URR; and
- the risk of serial correlation in the error terms and the consequent risk of understating the standard error of the URR estimates and overestimating the goodness of fit of the model if an appropriate estimation method is not used.

The following section illustrates these issues with the help of a case study.

5.3 Illustration - global production projection

Some of the issues introduced above may be illustrated by fitting the following four models to data on global oil production:

- Cumulative production projection assuming a logistic curve.
- Production projection assuming a first derivative of a logistic curve.
- Production projection assuming a first derivative of a logistic curve and including a lagged dependent variable.
- An ARIMA model of global production

5.3.1 Model 1: global cumulative production projection

The logistic model for cumulative production is as follows:

$$Q_t = \frac{Q_\infty}{1+\alpha e^{-\beta t}} + e_t \qquad (5.7)$$

Where the subscript t refers to annual data and Q_∞ represents the URR. This model was fit to the global cumulative production data for conventional oil and NGLs using the nonlinear regression facility of SPSS. Table 5.1 shows the parameter estimates and the R^2, with the estimates for URR being expressed as both a percentage of cumulative production through to 2007 (Q_{2007}) and a percentage of cumulative discoveries through to 2007 (D_{2007}). The R^2 is very high and the standard errors small, but this likely to be a consequence of serial correlation. It is notable that the estimated URR is 39% *less* than the cumulative discoveries through to 2007 and only 29% more than the cumulative production through to 2007 (although production has yet to peak). Despite the small standard error, these results are clearly implausible and illustrate the unreliability of production projection for regions have yet to reach their peak of production. Brecha (2008) and Bartlett (2000) find similar results in their projections of world oil production.

Table 5.1 Parameter estimates and goodness of fit for Model 1

Parameters	Estimate	Std. Error	R^2
Q_∞ (% of Q_{2007})	128.5	1.5	0.999
Q_∞ (% of D_{2007})	61.2	0.7	
α	24972.30	1882.95	
β	0.073	0.001	

Figure 5.3 shows the observed values of cumulative production and the fitted logistic curve. Visually, the fit appears to be relatively good but a closer examination of post 2002 data (Figure 5.4) shows that the difference between observed and fitted values is increasing with time, with the fitted cumulative production trending below the actual cumulative production.

Figure 5.3 Model 1 curve fit to cumulative production

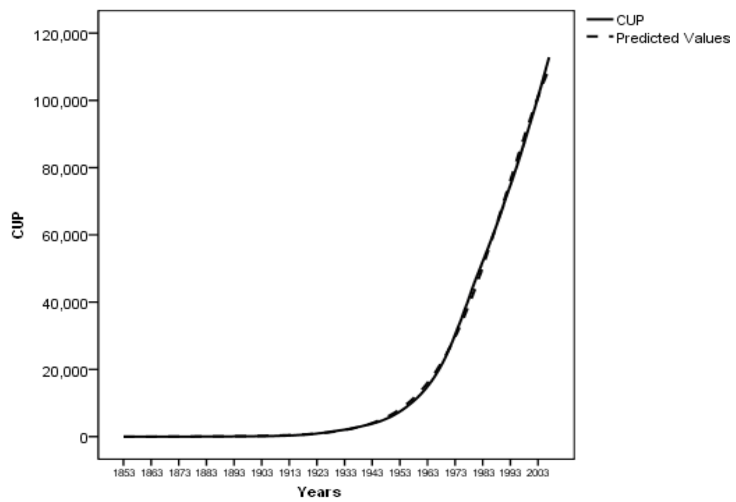


Figure 2. Plot of CUP and predicted values from model (5.1.1)

Figure 5.4 Model 1 curve fit to cumulative production post 2002

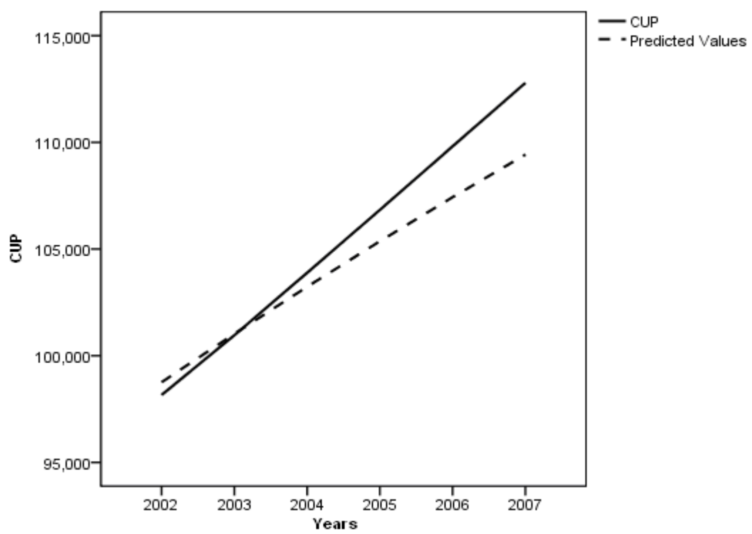


Figure 3. Plot of CUP and predicted values from model (5.1.1) after 2002

5.3.2 Model 2: global production projection

An alternative to Model 1 is to fit the first derivative of the logistic function to the data on annual production (Q'_t). The relevant functional form is:

$$Q'_t = \frac{Q_\infty \alpha e^{-\beta t}}{\left(1+\alpha e^{-\beta t}\right)^2} + e_t \qquad (5.8)$$

Table 5.2 shows the parameter estimates and the model R^2 while Figure 5.5 plots the observed and fitted values for production. Again, the R^2 is high and the standard errors small. The mean URR estimate from Model 2 is 28% greater than that from Model 1, but is still only 79% of cumulative discoveries through to 2007 and only 65% more than cumulative production through to 2007 (implying that more than half of the URR has been produced, despite production having yet to reach a peak). The reason for this highly conservative estimate is clear from Figure 5.5, which shows the fitted production trending downwards while actual production continues to increase (i.e. the ‘best fit’ curve is past peak while actual production is not).

Table 5.2 Parameter estimates and goodness of fit for Model 2

Parameters	Estimate	Std. Error	R^2
Q_∞ (% of Q_{2007})	165.0	5.3	0.978
Q_∞ (% of D_{2007})	78.5	2.5	
α	5491.29		
β	0.059		

Figure 5.5 Model 2 curve fit to rate of production

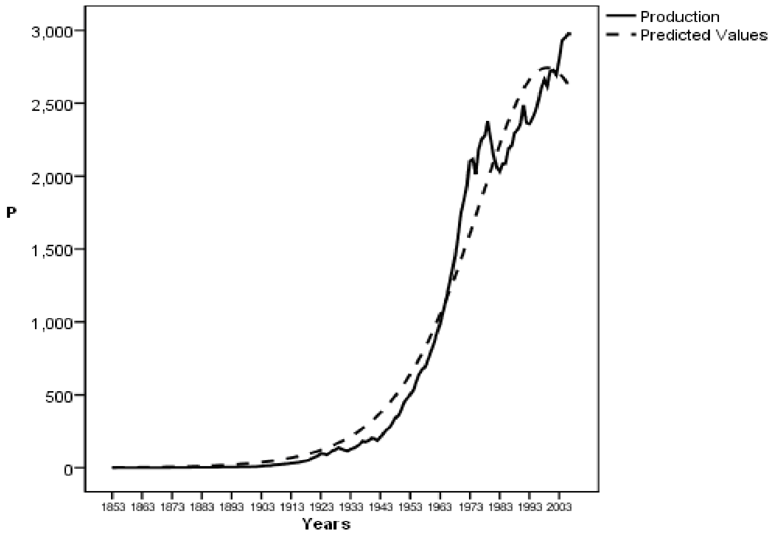
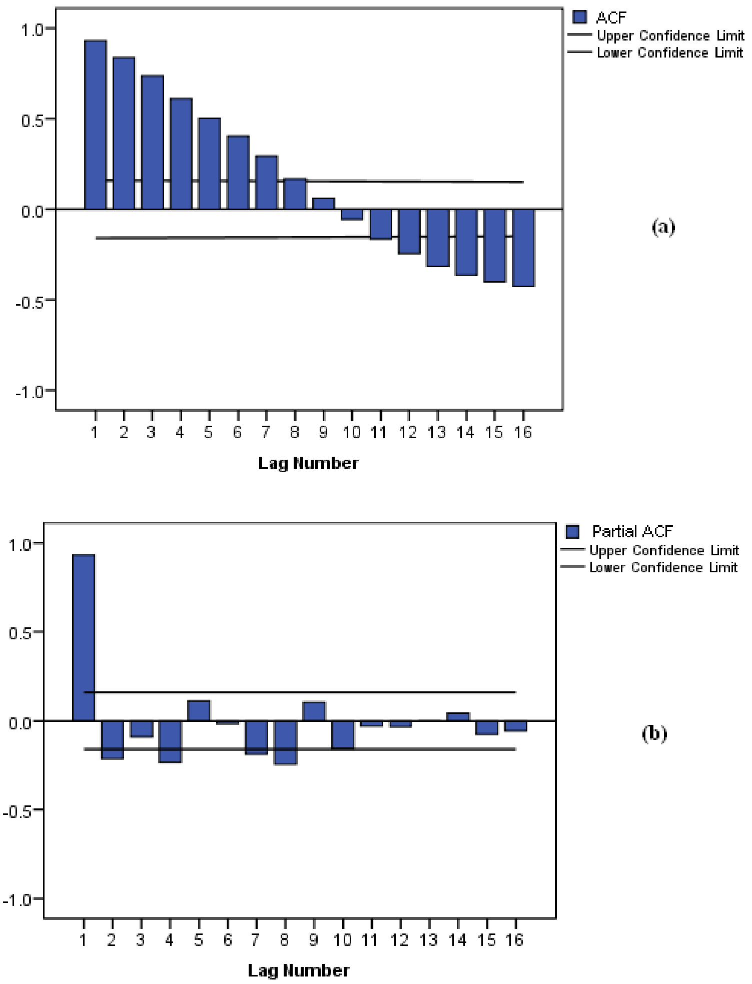


Figure 5. Plot of Production and predicted values from model (5.1.3)

Both of these models lead to implausible estimates of the global URR despite the high R^2 and small standard errors of the parameter estimates. This suggests the presence of serial correlation which for Model 2 is confirmed by the *autocorrelation function* (ACF) and *partial autocorrelation function* (PACF) (Figure 5.6). Both of these indicate significant serial correlation, as indicated by bars outside the confidence limits. As a result, the standard error of the URR estimate will be incorrect and the estimate itself may be biased.

Figure 5.6 Autocorrelation function (ACF) and partial autocorrelation function (PACF) for Model 2



5.3.3 Model 3: global production projection with lagged dependent variable

One possible response to these difficulties is to re-estimate the model using a different length of time series: for example, by removing some of the earlier data points. However, this is unsatisfactory since it means that the parameter estimates (including URR) will depend upon the particular sample chosen.

An alternative approach is to re-specify Model 2 to include a one-period lag of the dependent variable (Q'_{t-1}) (Pesaran and Samiei, 1995). However, since the parameters of the logistic equation will differ from those in Equation (5.8) the parameter corresponding to Q_∞ can no longer be interpreted as the URR. To illustrate this, we write this new equation as follows.

$$Q'_t = \frac{\hat{Q}_\infty \hat{\alpha} e^{-\hat{\beta}t}}{\left(1+\hat{\alpha}e^{-\hat{\beta}t}\right)^2} + \phi Q'_{t-1} + e_t \tag{5.9}$$

Where the hats ($\hat{x}$) are used to indicate that the parameters are different from those in Equation (5.8). Tests of this model showed a very slow convergence and large standard errors. As an alternative, we try the following specification (Nickerson and Madsen, 2004):

$$Q'_t = \frac{\hat{Q}_\infty \hat{\alpha} e^{-\hat{\beta} t}}{\left(1+\hat{\alpha} e^{-\hat{\beta} t}\right)^2} + \phi \left[Q'_{t-1} - \frac{\hat{Q}_\infty \hat{\alpha} e^{-\hat{\beta} t}}{\left(1+\hat{\alpha} e^{-\hat{\beta} t}\right)^2} \right] + e_t \tag{5.10}$$

The difference between this and Equation (5.9) is the second term on the right hand side, in which the effect of Q'_{t-1} on Q'_t is corrected by the trend $\frac{\hat{Q}_\infty \hat{\alpha} e^{-\hat{\beta} t}}{\left(1+\hat{\alpha} e^{-\hat{\beta} t}\right)^2}$. The R^2 and parameter estimates for Model 3 are summarised in Table 5.3, while Figure 5.7 shows the corresponding curve fit. It is clear that inclusion of the lagged dependent variable (Q'_{t-1}) allows Model 3 to provide a much better fit to the production data than Model 2.[74] Also, the serial correlation largely disappears (Figure 5.8).

Unfortunately, the introduction of lagged values into the model means that the URR (Q_∞) does not enter Equation (5.10) explicitly. In a comparable model for cumulative production in the US, Pesaran and Samiei (1995) are able to calculate Q_∞ algebraically from the new model specification. Unfortunately, the same is not possible for our model of global production, in part because production has yet to peak.[75] Hence, while the introduction of a lagged dependent variable allows a much better fit to historical production data, as well as removing the serial correlation, it may not necessarily help in the calculation of the URR.

Table 5.3 Parameter estimates and goodness of fit for Model 3

Parameters	Estimate	Std. Error	R^2
$\hat{Q}_\infty$	619.737	566.077	0.992
α	3367530.891	24580000	
β	0.130	0.058	
ϕ	1.074	0.031	

[74] Note that the standard error corresponding to α is very large which might indicate over-parameterization. One possibility is to try a smaller number of parameters through a simpler functional form.

[75] As an alternative, we could adopt the greatly simplifying assumption that the peak of production occured in 2007. By integrating the area under the *Q'(t)* curve up to 2007 and assuming that production has a symmetrical distribution around the peak, we can estimate a value for Q_∞ which is equal to 208.3% of cumulative production (Q_{2007}) and 99.1% of cumulative discoveries (Q_{2007}).

Figure 5.7 Model 3 curve fit to rate of production

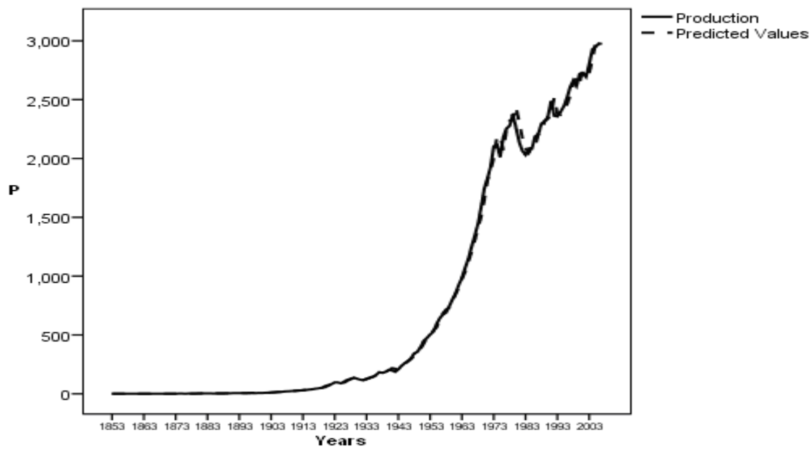
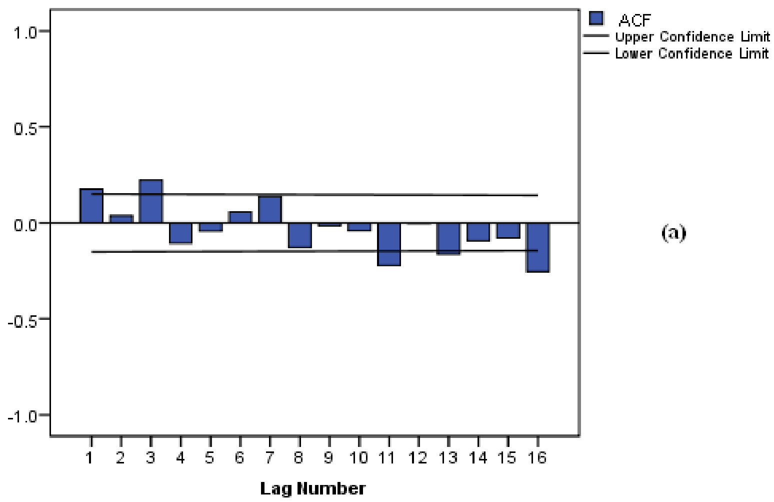


Figure 7. Plot of Production and predicted values from model (5.1.4)

Figure 5.8 Autocorrelation function for Model 3



5.3.4 Model 4: ARIMA model of global production:

If our interest is confined solely to forecasting, an alternative is to dispense with an assumed functional form altogether and to model production using a pure time series approach. Serial correlation may be removed altogether by using the *second difference* of production ($\Delta^2 Q'_t = Q'_t - 2Q'_{t-1} + Q'_{t-2}$) as the dependent variable. This type of differentiation ensures that the dependent variable is stationary (i.e. has constant mean and variance). An *ARIMA(1,2,2)* model for production (see Section 5.2.4) may then be defined as follows:

$$\Delta^2 Q'_t = \mu + \Phi \Delta^2 Q'_{t-1} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \tag{5.11}$$

This model was fit using a specialised time-series program which provides confidence intervals on the fitted values. The results are illustrated in Figure 5.9. The model appears to perform as well as Model 3, but a formal comparison is difficult since the non-linear regression function used to estimate Model 3 does not provide confidence intervals. Inspection of the autocorrelation functions for Model 4 (Figure 5.10) suggests no residual serial correlation.

However, while both the Model 3 and Model 4 provide an excellent fit to historical data, neither is of much little help in estimating the URR. A better alternative is to address the omitted variable bias *directly* by including variables such as oil prices within the model specification. The following section review several studies of US oil production that take this approach.

Figure 5.9 Model 4 – time series model of rate of production

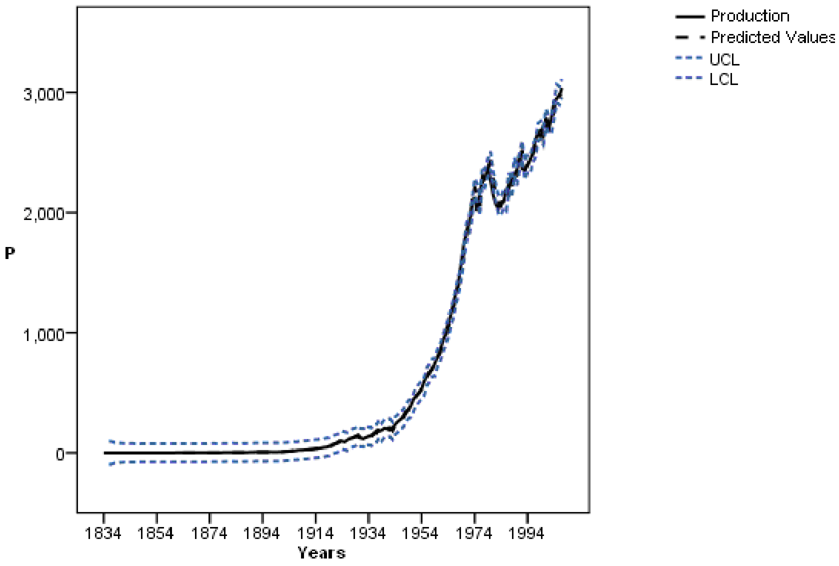
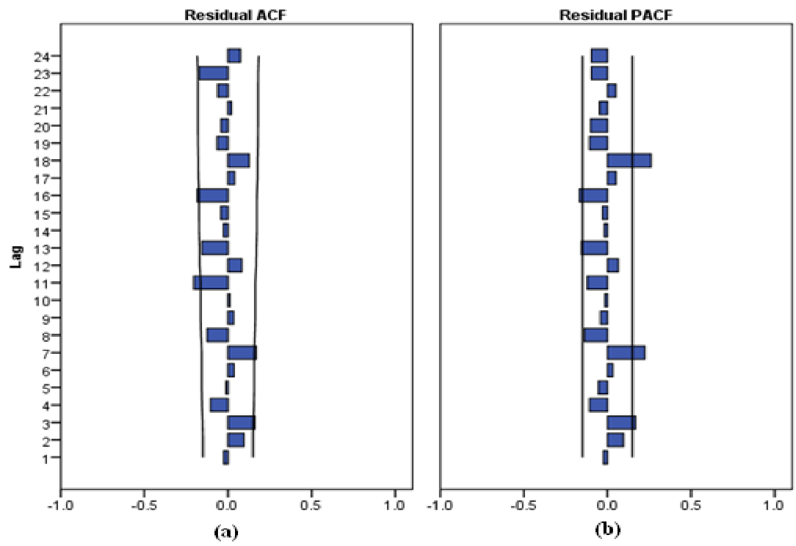


Figure 9. Plot of Production, predicted values and their confidence limits from model (5.1.5)

Figure 5.10 Model 4: (a) Autocorrelation function (ACF) (b) Partial autocorrelation function (PACF)



5.4 Reconciling econometrics and curve-fitting

Standard econometric models of oil supply use historical data to estimate relationships between exploration, discoveries and/or production and economic variables such as oil prices (Walls, 1992). Such models are likely to be biased because they ignore the geological determinants of discoveries and production, such as the skewed field size distribution and the tendency to discover the large fields first. In contrast, simple curve-fitting models reflect these geological determinants by assuming a particular functional form, but may be biased because they ignore political and economic influences. For example, low oil prices and political constraints may restrict production when resources are abundant, while high prices may stabilise or increase production when depletion is advanced. These biases are typically indicated by the serial correlation of the error term (Kaufmann, 1991).

This suggests that a promising approach would be to combine an assumed functional form to simulate the geological determinants of oil production or discovery with an econometric specification to simulate the economic and political determinants (Walls, 1994). The appropriate choice of explanatory variables will vary with both the nature of the model (e.g. production versus discovery projection) and the particular region under examination (e.g. the relevance of various political constraints) and will only be feasible if the relevant data is available (frequently it isn't). To date, relatively few studies have taken this approach and the majority have been confined to the United States. The following sections summarise these studies and highlight the lessons that can be learned.

5.4.1 A two-stage production projection

Kaufmann's (1991) study of oil production in the Lower 48 US states provides an interesting reconciliation of the two approaches. He begins by fitting a logistic curve to cumulative production in a similar manner to Model 1 above. He then takes the first difference of the fitted values ($\hat{Q}_t$) to obtain an estimate of the annual rate of production ($\hat{Q}'_t = \hat{Q}_t - \hat{Q}_{t-1}$) and calculates the normalised difference between this and the actual rate of production: $R_t = (Q'_t - \hat{Q}'_t)/\hat{Q}'_t$. Kaufmann then uses this residual (R_t) as the explained variable in linear regression in which economic and political factors serve as explanatory variables. The objective is to assess the extent to which such factors cause production to *deviate* from the logistic trend. The chosen variables are: short-run oil prices (a running average of real oil prices lagged one and two years); long-run oil prices (a running average of real oil prices lagged and three, four and five years); the price of oil relative to natural gas; the fraction of Texan production capacity that was allowed to operate by the Texas Railroad Commission (TRC); and a variable indicating whether the production curve should be asymmetric. The rate of production ($\tilde{Q}'_t$) is then calculated from $\tilde{Q}'_t = \hat{Q}'_t(1+\hat{R}_t)$, where $\hat{Q}'_t$ is the rate of production predicted by the first difference of the logistic equation and the $\hat{R}_t$ is the residual predicted by the econometric analysis.

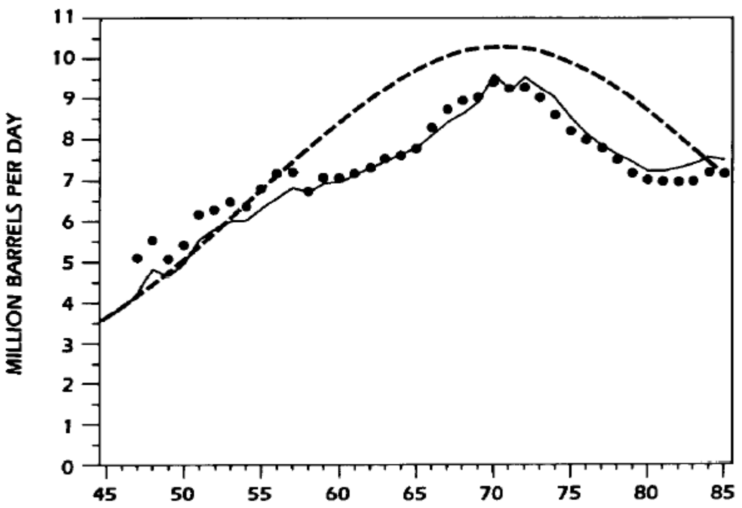
As illustrated in Figure 5.11, Kaufmann's two-stage model accounts for most of the variation in annual oil production in the Lower 48 states between 1947 and 1985. In contrast, the simple logistic curve overestimates production throughout much of this period, in part because it neglects the effect of both low oil prices between 1947 and 1973 and the 'prorationing' decisions of the TRC which shut in more than 50% of Texan capacity between 1957 and 1968.[76] Kaufmann found all of the explanatory variables to be statistically significant, with the projected rate of production falling faster after the peak than it rose prior to the peak. The URR is estimated from the sum of the URR generated from the first stage logistic model plus the sum of the deviations from the logistic curve over the full production cycle - which in turn, requires assumptions about the future values of the relevant economic

---

[76] Between 1947 and 1985, 38% of the oil produced in the lower 48 states came from Texas.

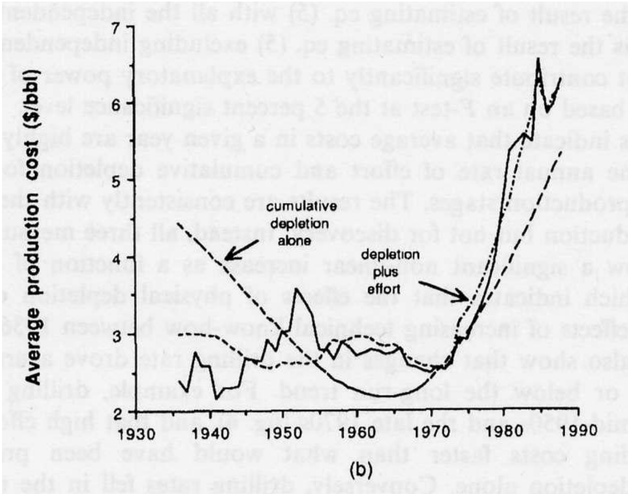
and political variables. For example, with pro-rationing eliminated and relative prices held at 1985 levels, Kaufmann estimates a URR for the Lower 48 of 190 Gb which is within the range of Hubbert's estimates.

Figure 5.11 Kaufmann's econometric model (solid line) of US lower 48 oil production (dots) as compared to logistic model (dashed line)



Three points about Kaufmann's approach are worth noting. First, economic factors lead to deviations from the logistic model, but in the long-run physical depletion dominates and production returns to low levels whatever the prevailing level of prices. Second, Kaufmann justifies the symmetrical logistic curve on the grounds that it imposes very few assumptions and criticises the asymmetric Gompertz curve because it implicitly assumes that oil prices will rise in the later phase of production, contributing to a slower production decline. Third, Kaufmann observes that the long-run average cost curve for US oil production appears to be U-shaped, which provides further support for the logistic model. This argument is based upon a study by Cleveland (1991), but has its origins in the work of Slade (1982). Cleveland finds that US production costs fell between 1936 and the mid-1960s since the cost reducing effect of technical change overshadowed the cost increasing effect of resource depletion. After that point, resource depletion began to dominate and costs rose rapidly (see Figure 5.12). The production curve peaked in 1970 while the long-run cost curve reached its trough in the mid-1960s. This seems unlikely to be a coincidence since the curves are estimated with different data (Kaufmann, 1991). However, evidence from other regions suggests that the long-run cost curve for oil production is not always U-shaped (Managi, *et al.*, 2004).

Figure 5.12 The long run average cost of oil production in the lower 48 US states



Source: Cleveland (1991)

5.4.2 Production projection using cointegration techniques

Kaufmann and Cleveland (2001) subsequently repeated their analysis of oil production in the Lower 48 using modern 'cointegration' techniques. The explanatory variables were the same as in Kaufmann (1991), but allow for asymmetric responses to price increases and decreases (e.g. the response to a price rise may not be undone by a subsequent price decrease) (Gately, 1992; Grubb, 1995). Also, by including average production costs as an explanatory variable, they no longer need to assume a particular functional form for the production cycle. In effect, the *empirically estimated* average cost curve (Figure 5.12) replaces the *exogenous* assumption of a bell-shaped production curve. The cost curve represent the net effect of resource depletion and technical change, with rising costs after 1970 indicating accelerating depletion.

Using a 'vector error correction model' (VECM) Kaufmann and Cleveland are able to account for most of the variation in oil production between 1938 in 1991. They conclude that economic variables have a critical influence on the shape of the production cycle and therefore on the date of peak production:

> "….the accuracy of Hubbert's original forecast for oil production in the lower 48 states is fortuitous. The cointegration analysis indicates that oil production in the lower 48 states shares stochastic trends with the decomposed price series, average costs, and pro-rationing decisions by the TRC. These stochastic trends are not present in the deterministic bell-shaped curve, so the first difference of the bell-shaped curve drifts away from the annual change in oil production for extended periods……Our results indicate that Hubbert was able to predict a peak in US production accurately because real oil prices, average real cost of production, and decisions by the TRC coevolved in a way that traced what appears to be a symmetric bell-shaped curve for production over time. A different evolutionary path for any of these variables could have produced a pattern of production that was significantly different from a bell shaped curve……In effect, Hubbert got lucky." (Kaufmann and Cleveland, 2001)

While an important innovation, this method may not be applicable for other regions owing to the lack of data on production costs. Moreover, if this method is to be used estimate the regional URR, some assumption is required about future trends in production costs. These could be developed in a variety of ways, but if this involves curve-fitting to historical data and extrapolation, the procedure begins to resemble standard curve-fitting. Also, the conclusion that 'Hubbert got lucky' refers primarily to his predicted date of peak production.[77] What is much less clear is whether these variables will make a significant difference to the final estimate of URR.

5.4.3 Production projection with variable URR

Pesaran[78] and Samiei (1995) provide an insightful account of the statistical issues associated with estimating URR and illustrate these by fitting a number of models to US production data. They argue that Kaufmann's (1991) two-stage model is biased because the estimation of URR in the first stage does not take into account the effect of economic factors which only enter the analysis in the second stage. A more appropriate method would be to estimate the long-term trend and short-term effects simultaneously which could be achieved by modifying the logistic model to allow for the dependence of URR on a number of economic and other variables. That is: $Q_\infty^t = A + \mathbf{BX}_t$, where $\mathbf{X}_t$ is a vector of relevant variables, **B** a vector of coefficents and Q_∞^t is the size of the URR given the economic, technological and other relevant factors that prevail at time *t*.

With this model, the URR is only fixed if economic and other factors have no impact on the size of the resource (i.e. all the elements of **B** are zero). Pesaran and Samiei use this formulation of URR within a model of US production (Q'_t), with the variables incorporated within $\mathbf{X}_t$ being the same as those employed by Kaufmann (1991). To eliminate problems of serial correlation, Pesaran and Samiei formulate a model that includes both the first difference of cumulative production and first and second order lags of the rate of production:

$$Q'_t = \chi Q'_{t-1} + \delta Q'_{t-2} + \left[\frac{A + \mathbf{BX}_t}{1 + \alpha e^{-\beta t}} - \frac{A + \mathbf{BX}_{t-1}}{1 + \alpha e^{-\beta t - 1}} \right] + e_t \qquad (5.12)$$

This model is found to explain over 98% of the variation in annual production over the period 1948-1990. Using 1990 values of the exogenous variables, it leads to an estimate of 209 Gb for the Lower 48 URR, with a 95% probability of lying within the interval 188-229 Gb. Importantly, this estimate is *higher* than obtained from other formulations that do not take economic and other factors into account. Pesaran and Samiei also re-estimated the model over the period 1926-85 and used it to generate forecasts of annual production over the period 1986-90. This provided estimates that were significantly more accurate than Hubbert's basic model.

5.4.4 Hybrid modelling of yield per effort

Very similar approaches can be used to modify and improve any of the curve-fitting techniques described in Section 3, although the relevant variables will vary in each case. A good example is Cleveland and Kaufmann (1991), who begin with Hubbert's basic exponential model of yield per effort (YPE) in the US:

---

[77] Cavello (2004; 2005a; b) reaches a very similar conclusion through a more qualitative argument.

[78] M.H. Pesaran is a Professor of Economics and Fellow of Trinity College Cambridge and an internationally recognised econometrician.

$$B'(\varepsilon_d, t) = Ke^{-\beta\varepsilon_d} \qquad (5.13)$$

As described in Section 3.5.2, this model provided a relatively poor fit to the historical data on yield per effort and tended to underestimate the US URR. Cleveland and Kaufmann modify this model by introducing two additional variables:

- *Drilling rates*: Each year, the industry ranks prospective sites in order of expected profitability. A low rate of drilling in any year implies that only those sites that are most likely to be profitable have been drilled ('highgrading'), while a high rate of drilling implies the exploration of increasingly marginal sites. As a result, there should be an inverse relationship between yield per effort and the annual rate of drilling.[79]
- *Oil prices:* At any level of drilling, some discoveries are not reported because they are uneconomic to develop (i.e. a 'dry' hole is not necessarily dry) (Schuenemeyer and Drew, 1983). The relevant cut-off point should be determined in part by real oil prices, with higher prices leading to more discoveries being reported. As a result, there should be a positive relationship between yield per effort and oil prices.

The revised equation is as follows:

$$B'(\varepsilon_d, t) = Ke^{-\beta\varepsilon_d}e^{+\chi p}e^{-\delta r} \qquad (5.14)$$

Where p is the real price of oil, r is the rate of exploratory drilling and χ and δ are coefficients. This equation is found to provide a much better fit for YPE in the Lower 48 than Hubbert's simple exponential decline. While YPE has declined since the 1930s, there have been periods of relative stability or increases in YPE associated with changes in the rates of drilling and/or oil prices. If not properly accounted for, these short run changes can temporarily mask the decline associated with depletion, which in turn may lead to misleading forecasts of future discoveries. Cleveland and Kaufmann's results show, however, that physical depletion dominates in the long-term, implying that policies to accelerate exploratory drilling in the US are unlikely to yield significant amounts of oil (Kaufmann and Cleveland, 1991).

Cleveland and Kaufmann (1997) estimate a very similar model for non-associated natural gas in the US. Historical discoveries are corrected for future reserve growth and allowance is made for shifts between offshore and onshore regions, since the former has a much larger YPE. Again, the regression model provided a much closer fit to the historical data that a simple exponential decline, but the long-run trend is only temporarily reversed by changes in energy prices.

5.4.5 Modelling technical change

Although technical change is not explicitly included in the above models, it is wrong to conclude that it is ignored.[80] In the case of the YPE models, cumulative exploratory effort

---

[79] This may be reinforced by the entry and exit from the industry of marginal, less efficient operators during periods of high and low drilling effort respectively.

[80] For example, Lynch (1999) argues that stable URR estimates make the "unrealistic assumption that technological progress will effectively cease". But the data used to estimate the curve-fitting models include the effects of *both* resource depletion and technical change. If the same model is used for forecasting, the implicit assumption is that the *relative* effect of resource depletion and technical change will be unchanged.

(ε_d) is used as a proxy for the net effect of technical change and resource depletion, with the relative size of these effects determining the sign of the relevant coefficient (β). Forecasts using these models do not ignore future technical change, but simply assume that the net effect of technical change and depletion will remain broadly unchanged. Nevertheless, given the revolutionary advances in exploration and production technology over the last 50 years, the explicit modelling of technical change could be valuable.

One possibility is to simply use time as a proxy for technical change – which effectively assumes that it occurs a constant rate. This is the approach taken by Iledare and Pulsipher (1999) in their study of the YPE for oil in onshore Louisiana (a mature province). They estimate that technical change increased the discovery rate in Louisiana by 7.5% per year over the period 1977 to 1994, but this was insufficient to offset the negative effect of resource depletion (measured by cumulative drilling), which reduced the annual discovery rate by an average of 12% per year. Hence, as with the Cleveland and Kaufmann models, depletion effects dominated over other variables.

A more sophisticated approach is taken by Managi, *et al.* (2005) in their study of YPE in the Gulf of Mexico over the period 1947-98.[81] Contrary to the standard assumptions of curve-fitting techniques, the YPE for this region appears to exhibit a U-shaped trend, with technical change more than offsetting the effect of resource depletion since the early 1970s (Figure 5.13). The trend appears to be linked to technical improvements that permitted exploration in deeper waters, thereby expanding the productive area and allowing access to larger fields. Drawing upon earlier work by Cuddington and Moss (2001), Managi, *et al.* develop an novel measure for this technical change which comprises an index of the annual number of innovations adopted by the offshore industry weighted by estimates of the relative importance of those innovations (Managi, *et al.*, 2004; NPC, 1995). The weights can increase over time to simulate both the slow diffusion of innovations and their continuing improvement.[82]

Figure 5.13 Yield per effort for oil exploration in the Gulf of Mexico 1947-98

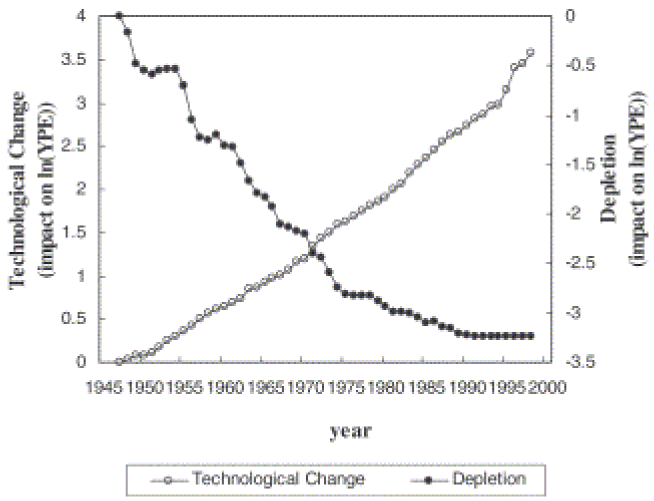
Source: Managi, *et al* (2005)

[81] This study is based upon Managi's PhD thesis (Managi, 2002).
[82] Lynch (2002) notes that it can take at least five years for significant innovations to become widely adopted, even in the US offshore industry, and the impact of factors such as YPE might not show up in the data for another five years.

Managi, *et al.* (2005) specify a exponential model for the YPE in the Gulf of Mexico ($B'(\varepsilon_d, t) = K\prod_i e^{\beta_i x_{it}}$) in which the explanatory variables (x_{it}) include annual and cumulative exploratory activity, oil prices and the average water depth of drilling. Depletion is modelled by cumulative discoveries while technical change is modelled by the above index. The estimation process corrects for serial correlation in the error terms. The estimated individual and net effects of technical change and depletion are illustrated in Figure 5.14 and Figure 5.15 respectively. This shows that the pace of technical change has increased since 1975, greatly expanding the area of exploration and leading to a YPE in 2000 that is comparable to that achieved 50 years previously.[83] While the geological diversity of the Gulf of Mexico may contribute to this result, a more likely reason is that exploration has been geographically restricted in the past, owing to the technical difficulties of deep-water drilling. As a result, fields may not have been found in the approximate declining order of size that is normally assumed.

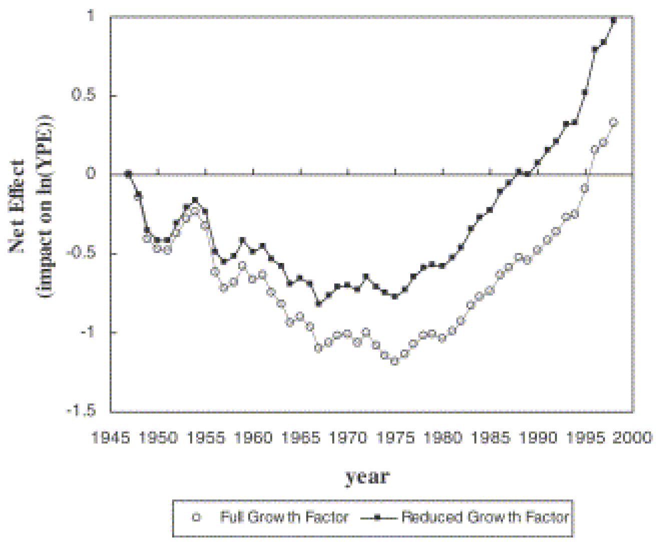
Since the URR does not appear explicitly in the equation for YPE, it can only be estimated by extrapolating the curve into the future and taking the integral. This in turn requires some explicit assumptions for the explanatory variables, including in particular the future rate of technical change. While depletion must eventually outpace technical change, the historical record gives little indication of when this will occur. As a result, the exploration history of this region currently provides an inadequate basis for the estimation of the URR.

Figure 5.14 Individual effect of technical change and depletion on yield per effort for oil exploration in the Gulf of Mexico 1947-98



[83] This is consistent with Forbes and Zampelli (2000), who found that technical change increased the exploratory success rate in the offshore US at an average annual rate of 8.3% over the period 1986-95.

Figure 5.15 Net effect of technical change and depletion on yield per effort for oil exploration in the Gulf of Mexico 1947-98



Source: Managi, *et al* (2005)

5.4.6 The challenge of hybrid modelling

This review suggests that a reconciliation of econometric and curve-fitting techniques can potentially overcome problems of missing variables and serial correlation. Such 'hybrid' models can potentially reduce the biases associated with curve-fitting techniques and allow the dependence of URR on prices and other variables to be explored. However, both curve-fitting and econometric techniques have their limitations and these are not necessarily overcome by combining them within a single model (Lynch, 2002). For example, the required data is frequently either lacking or unreliable and is rarely available at the level of disaggregation required. Since it is generally impractical to include more than a subset of the variables that could affect production and/or discovery trends (e.g. tax rules, leasing decisions, geographical restrictions on exploration, production/import/export quotas, relative fuel prices etc), these models may still be vulnerable to missing variable bias. While it would be useful to include technical change as an explanatory variable, it is difficult to find measures that adequately reflect the slow process of technical diffusion. Problems such as these may partly explain why there are so few 'hybrid' studies and why the available studies are largely confined to the United States where the relevant data is more readily available

5.5 Summary and implications

This section has summarised the main statistical issues raised by curve-fitting techniques, highlighted the potential consequences for estimates of URR and provided some illustrations of how these issues may be addressed. The main conclusions are as follows:

- The literature on curve-fitting techniques to estimate URR has paid insufficient attention to the statistical issues involved. Most authors have assumed that the 'shape' of the production or discovery cycle can be estimated from the historical data and that this shape will not be significantly affected by any future changes in prices, technology and other relevant variables. As a result, there has been a tendency to neglect these variables, despite the potential errors that may result.

- The omission of variables such as oil prices and technical change does not necessarily mean that they are ignored. Instead, it amounts to assuming that the relative effect of these variables as compared to the effect of resource depletion has been relatively stable in the past and will remain broadly unchanged in the future. In practice, this assumption appears unlikely to be justified.
- As a consequence of this neglect, many applications of curve-fitting techniques are likely to suffer from missing variable bias and/or serial correlation of the error terms. This could lead to biased estimates of model parameters (including URR), underestimates of the associated standard errors and overestimates of the model goodness of fit (i.e. the R^2). Examples of this have been provided in both this section and Section 4.
- These problems may potentially be addressed by including one or more lags of the dependent variable within the model specification. However, while this can provide a much better fit to the historical data, the re-specified model may not necessarily lend itself to the estimation of URR.
- A more promising approach is to include some of the economic and political determinants of discovery and/or production within the model specification. The appropriate choice of explanatory variables will vary with both the nature of the model and the particular region under examination and will only be feasible if the relevant data is available. Relatively few studies have taken this approach to date and the majority of these have been confined to the United States.
- The studies reviewed here include a two-stage production projection, a production projection involving cointegration techniques, a production projection in which URR is a made a function of economic and other variables and YPE models incorporating energy prices, annual drilling rates and other factors. All of these provided a much better fit to the relevant historical data than the basic curve-fitting model and largely remove the serial correlation. The study by Kaufmann and Cleveland (2001) is especially notable as their use of data on average production costs removes the need for an assumed functional form. However, the required data on production or discovery costs is unlikely to be available for other oil-producing regions.
- ‘Hybrid models’ such as these may be more suitable for short-term supply forecasting than for estimating the regional URR. The latter requires assumptions about the future values of the relevant explanatory variables and hence is subject to considerable uncertainty. Despite their better fit to historical data, it is not obvious that hybrid models lead to substantially different estimates of the regional URR than simple curve-fitting. However, they do allow the dependence of URR on energy prices and other factors to be directly explored.
- Hybrid models cannot resolve all of the problems associated with curve fitting techniques. In particular, such models may still lead to misleading conclusions if applied to regions that lack either geological homogeneity or a consistent exploration history.

6 Global estimates of ultimately recoverable resources and their importance for future oil supply

6.1 Introduction

For almost ninety years, analysts of global oil supply have produced estimates of the global ultimately recoverable resource (URR) of 'conventional' oil. These estimates frequently differ in their definition of conventional oil, as well as their methods, their assumptions and their results. Some have become trusted estimates, used to calibrate forecasts of future oil production, while others have proved highly controversial, inciting debate and criticism. While the estimates have grown over time as more of the world's oil regions have been explored and developed, a consensus on the global URR has yet to emerge.

As with many aspects of oil supply, the comparison of published estimates of global URR is complicated by several issues. The estimates may include different petroleum liquids, be developed using different methods (or combinations of methods), relate to time periods, use different data sources and/or rely upon different economic and technical assumptions that may not be transparent. For example, reserve growth may or may not be accounted for, and where it is accounted for, different assumptions about the rate of growth may be employed. As a result, there is a considerable risk of 'comparing apples and oranges' with such estimates.

Perhaps the most prominent and authoritative estimates of global URR are from the US Geological Service (USGS). The most recent of these was published in 2000, following 100 person-years of effort by a team of 41 geoscientists over a period of five years (Ahlbrandt, 2002; USGS, 2000b). These estimates are significantly larger than previous USGS estimates and several commentators have disputed their validity (Laherrère, 2001b). Nevertheless, they underlie projections of global oil supply and associated carbon emissions by bodies such as the International Energy Agency (IEA, 2008), the Energy Information Administration (EIA, 2008) and the Intergovernmental Panel on Climate Change (IPCC, 2007).

This section provides an overview and evaluation of global URR estimates and assesses their implications for future global oil supply. The structure is as follows. Section 6.2 summarises and compares some global URR estimates that have been made in the past, illustrates how these have grown over time and looks in more detail at three of the more prominent estimates. Section 6.3 summarises the methods and results of the USGS World Petroleum Assessment 2000 and evaluates whether the experience since 2000 is consistent with these projections. Section 6.4 summarises how these estimates have been updated by the IEA World Energy Outlook 2008 together with a recent study by the Colorado School of Mines. Section 6.5 identifies the implied range of uncertainty over this variable and the implications for future global supply. Section 6.6 concludes.

6.2 A brief history of global estimates of ultimately recoverable resources

Eugene Stebinger, once chief of Foreign Minerals at the USGS, was the first geologist to estimate the global URR of conventional oil (White, 1920). This early estimate totalled only 43 Gb of oil, which compares to cumulative production through to 2007 of 1100 Gb (IEA, 2008).[84] White (1920) described Stebinger's estimate as conservative and suggested that a further 20 Gb were probably available. By 1942, Stebinger had increased his estimate of global URR by a factor of 13, to some 600 Gb (Pratt, 1942). Since this time, around 100 estimates of global URR have been published from a variety of sources, including several repeated estimates by the same institutions or individuals. These estimates are summarised in Table 6.4. When comparing these estimates, it is important to remember that their assumptions and methods may differ, together with the purpose for which they were derived. Of particular importance is the differing coverage of petroleum liquids, which may or may not include lease condensate, natural gas liquids (NGLs), polar and deepwater oil and various types of heavy oil, including oil sands. Unfortunately, the coverage is not always clear and even where it is clear, different definitions and assumptions may apply (e.g. what is meant by 'heavy oil').

Most recent estimates of global URR have clustered in the range 2000 Gb to 3000 Gb, with the largest being Miller (1992) who estimates a URR of >4000 Gb for conventional oil. This compares with cumulative production of petroleum liquids through to 2007 of 1128 Gb and cumulative proved discoveries (i.e. cumulative production plus 1P reserves) of 2366 Gb (IEA, 2008) which is higher than many of the URR estimates! However, the latter figure includes NGLs, heavy oil and oil sands, while many of the URR estimates relate to a much narrower definition of 'conventional' oil (i.e. we are comparing apples and oranges). In addition, authors such as Campbell (2002) argue that the figure for cumulative proved discoveries is incorrect because several countries reserve estimates are overstated.

Figure 6.1 plots the URR estimates and fits linear regressions to indicate the approximate trends over three different time periods, namely: 1942 to 2007; 1970 to 2007 and 1987 to 2007. All suggest a gradual increase in URR estimates over time, which may be expected given the increase in geological knowledge and the improvements in exploration and production technology that have occurred. The regression over the full period suggests an increase of around 17 Gb/year in the estimates of URR, while the regression over the period 1970-2007 suggest that the rate of increase may have slowed (to 5 Gb/year). However, the post 1987 regression suggest an increase of 47 Gb/year. A similar upward trend can be seen in the repeated estimates from individual sources (Andrews and Udall, 2003). However, it would be wrong to attribute too much significance to such trends, given the considerable difficulties in comparing one estimate with another. In particular, several of the most recent estimates include NGLs while most of the earlier estimates do not. Since the USGS estimates that NGLs account for around 13% of the remaining resources of conventional liquids (Ahlbrandt, 2002), their exclusion or underestimation in earlier studies will have contributed to the observed upward trend. While such a trend cannot continue indefinitely, there does not seem to be any strong evidence of a levelling off over the last 20 years.

---

[84] This figure includes conventional oil, condensate, NGLs, heavy oil and oil sands but excludes CTLs, GTLs and biofuels.

Table 6.4 Historical estimates of the global ultimately recoverable resource of conventional oil

Year of Publication	Author and/or Organisation	Estimated URR (Gb)	Notes
1942	Pratt and Weeks	650	
1946	Duce (Aramco)	500	100 in the United States, 400 abroad.
1946	Pogue (Chase Manhattan Bank)	605	49.2 cumulative production, plus 65.8 proved reserves, plus 490 future discoveries.
1948	Weeks (Standard Oil Co., New Jersey)	610	487 future discoveries.
1949	Levorsen (Stanford)	1635	65 cumulative production, plus 65 discovered reserves, plus 1500 undiscovered reserves. Based on estimates by Pogue (1946), and Weeks (1948) for onshore, and Pratt (1947) for offshore.
1950	Levorsen (Stanford)	1635	
1950a	Weeks (Standard Oil Co., New Jersey)	1010	Discussion to Levorsen (1950).
1950b	Weeks (Standard Oil Co., New Jersey)	1100	610 onshore, 400 offshore shelves.
1953	MacNaughton, personal communication	1000	
1956	Hubbert (Shell)	1250	
1958	Weeks (Standard Oil Co., New Jersey)	3000	1500 primary recovery, plus 1500 secondary recovery; includes natural-gas liquids.
1959	Weeks (Standard Oil Co., New Jersey)	3500	2000 primary recovery, plus 1500 secondary recovery; includes natural-gas liquids.
1961	Weeks (Weeks Petroleum Corp.)	3500	2000 primary recovery, plus 1500 secondary recovery; includes natural-gas liquids.
1962	Hubbert (Shell)	1250	
1963	Weeks (Weeks Petroleum Corp.)	2000	
1965	Hendricks (U.S. Geological Survey)	1984-2480	6162-6200 oil in place, 40% recovery.
1967	Ryman (Standard Oil Co., New Jersey)	2090	According to Hubbert (1969).
1967	Royal Dutch Shell	1800	
1968	Weeks (Weeks Petroleum Corp.)	3550	2200 primary recovery, plus 1350 secondary recovery.
1969	Hubbert (U.S. Geological Survey)	1350-2100	
1970	Weeks (Weeks Petroleum Corp.)	3550	2200 primary recovery, plus 1350 secondary recovery.
1970	Moody (Mobil)	1800	
1971a, b	Warman (British Petroleum)	2000	
1971	Weeks (Weeks Petroleum Corp.)	3650	2290 primary recovery, plus 1360 secondary recovery.
1972	ESSO	2100	Oil increasingly scarce by 2000
1972	Bauquis *et al* (IFP)	1900	

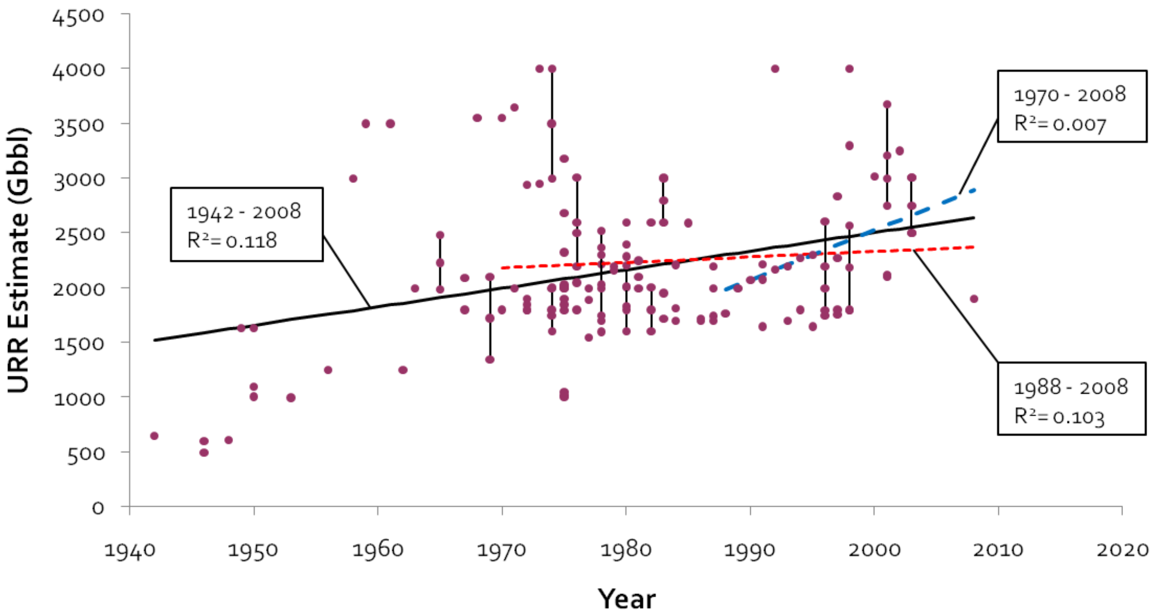
1972	Warman (British Petroleum)	1800	
1972	Linden (Institute of Gas Technology)	2945	
1972	Moody and Emmerich (Mobil)	1800-1900	1000 discovered, 800-900 yet to be discovered.
1973	Schweinfurth (USGS)	2950	
1973	Odell (Erasmus)	4000	
1974	Bonillas (SOCAL)	2000	
1974	Howitt (BP)	1750	
1974	Kirby and Adams (British Petroleum)	1600-2000	
1974	Parent and Linden (Institute of Gas Technology)	3000-4000	
1975	MacKay (Bank of Montreal, Calgary) and North (Carleton University, Ottawa)	1000-1050	
1975	Weeks (Weeks Petroleum Corp.)	3180	1900 onshore, 1280 offshore.
1975a,b	Moody (consultant) and Esser (Mobil)	2000-2030	
1975	Moody (consultant) and Geiger (Mobil)	2000	280 (90%), 2000 (50%), and 2200 (10%)
1975	Linden and Parent (Institute of Gas Technology)	2685	Estimated total remaining recoverable.
1975	Moody (consultant)	1800-1900	1000 discovered, 800-900 yet to be discovered.
1975	National Academy of Sciences	2326	
1976	Grossling (U.S. Geological Survey)	2200-3000	
1976	Folinsbee	1800	
1976	American Petroleum Institute	2050	
1976	Barthel *et al.* (West Germany Geological Survey)	2500	
1977	Nelson (SOCAL)	2000	
1977	Parent and Linden (Institute of Gas Technology)	2000	Includes natural-gas liquids.
1977	World Energy Conference	1889	
1977	Klemme (Weeks Petroleum Corp.)	1550	350 cumulative production, 600 proven reserves, 600 undiscovered.
1978	De Bruyne (Shell)	1600	
1978	Klemme	1750	
1978	Desprairies (Institut Francais du Petrole)	2220-2520	Delphi pool: 1620-2200-2870.
1978	Moody (consultant)	2030	
1978	Nehring (Rand Corp.)	1700-2300	
1979	Meyerhoff	2200	
1979a,b	Wood (Cities Service)	2163	1038 already discovered, 1125 to be discovered; 1500 (95), 3300 (50), 3100 (10).
1980	Halbouty and Moody (consultant)	2288	
1980	Schubert (World Energy Conference)	2600	
1980	Nehring (Rand Corp.)	1600-2000	Includes natural-gas liquids.

1980	Desprairies and Tissot (Institut Francais du Petrole)	1830-2200	
1980	Roorda (Shell)	2400	
1981	Strickland (Conoco)	2100	
1981	Colitti (AGIP)	2100	
1981	Halbouty	2250	
1981	Hubbert and Root (U.S. Geological Survey)	2000	
1982	Nehring (Rand Corp.)	1600-2000	Includes natural-gas liquids.
1982	Bois (Institut Francais du Petrole)	2600	
1983	Odell and Rosing	3000	
1983	Masters *et al.* (U.S. Geological Survey)	1718	
1983	Riva (Library of Congress)	1953	
1984	Martin (BP)	1700	
1984	Burollet (Total)	2213	524 billion cumulative production, plus 2313 reserves and to be discovered.
1984	Masters *et al.* (U.S. Geological Survey)	1818	
1985	Tanzil (Consultant)	2594	
1986	Masters *et al.* (U.S. Geological Survey)	1718	
1986	Ivanhoe (consultant)	1700	
1987	Jenkins (BP)	1700	
1987	Masters *et al.* (U.S. Geological Survey)	1744	
1987	Pecqueur (Elf Aquitaine)	2200	Includes enhanced oil recovery.
1987	Roadifer (Mobil)	2000	
1988	Riva (Library of Congress)	1765	
1989a, b	Bookout (Shell)	2000	
1990	Masters *et al.* (U.S. Geological Survey)	2074	
1991	Masters *et al.* (U.S. Geological Survey)	2079	
1991	Campbell (consultant)	1650	
1991	Riva (Library of Congress)	2215	
1992	Masters *et al.* (U.S. Geological Survey)	2171	
1992	Miller (British Petroleum)	>4000	
1993	Laherrère	1700	
1993	Townes (independent petroleum geologist)	2600-3000	
1993	Miremadi and Ismail (OPEC)	2200	
1994	Masters *et al.* (U.S. Geological Survey)	2272	
1994	Laherrère (Petroconsultants)	1800	
1995	Riva	2300	

1995	Campbell (consultant)	1650	
1996	Ivanhoe	2000	
1996	MacKenzie (World Resources Institute)	1800-2600	
1996	Campbell (consultant)	1750	
1997	Campbell (consultant)	1800	
1997	Edwards (University of Colorado)	2836	
1997	Masters *et al.* (U.S. Geological Survey)	2272	
1997	Al-Jarri and Startzman (Texas A&M)	1760	
1998	Campbell and Laherrère (consultants)	1800	
1998	Hiller (Hanover, Germany)	1800-2570	
1998	Linden (Illinois Institute of Technology)	4000	
1998	Schollnberger (Amoco)	3300	
2000	U.S. Geological Survey 2000	3021	Oil available by 2025
2001	Deffeyes (Princeton)	2100-2120	
2001	Odell	3000	
2001	Edwards (University of Colorado)	2750-3670	
2002	Edwards (University of Colorado)	3251	
2003	Nehring (personal estimate)	2500-3000	All liquids; 3500 “aggressive”
2004	Odell	3000	
2008	Campbell and Heapes	1900	

Source: Salvador (2005), Andrews and Udall (2003)

Figure 6.1: Comparison of global URR estimates over the last 70 years



The following sections examine three of these estimates in more detail.

6.2.1 Campbell and Laherrère

Campbell and Laherrère have produced several very conservative estimates of the global *URR* over the last 20 years - most of which are less than half of the most optimistic estimates. These estimates have underpinned their forecasts of an early peak in global oil production – which have repeatedly proved incorrect (Lynch, 1998).

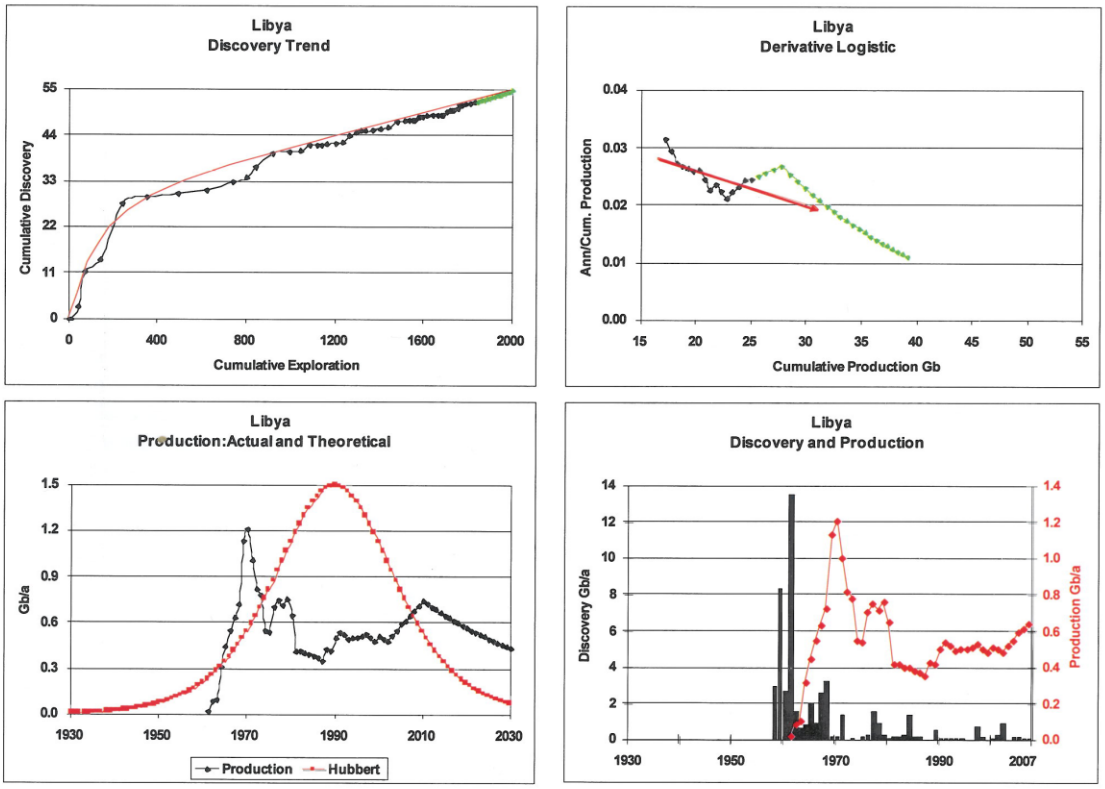
Campbell and Laherrère obtain their global estimates by aggregating estimates for individual countries and regions. These are obtained from a combination of extrapolation techniques, including production decline curves and creaming curves. They emphasise the limitations of official reserve estimates, including inaccurate reporting by individual countries, inflation of OPEC reserves as a consequence of quota negotiations (by 287 Gb during the 1980s), the unlikely reporting of unchanged reserves from year to year (by 59 countries in 1997) and variations in reserve definitions from one country to another. To avoid these difficulties, Campbell and Laherrère rely upon 2P reserve estimates derived originally from the Petroconsultants database (subsequently subsumed into the IHS PEPS database). They argue that these estimates provide a more reliable indication of remaining resources than do official 1P estimates from sources such as the BP Statistical Review. Though they acknowledge reserve growth, Campbell and Laherrère make no allowance for this in their estimates and also make a number of judgemental adjustments, most notably in significantly downgrading the reserve estimates for OPEC countries. Most importantly, Campbell uses a highly restrictive definition of conventional oil:

> "….. what is here termed *Regular Conventional Oil* is defined to exclude oil from coal and shales; Bitumen; Extra Heavy Oil (<10º API), Heavy Oil (10-17.5º API); Deepwater Oil and Gas (>500m); Polar Oil and Gas; Natural Gas Liquids from gas plants." (Campbell and Heapes, 2008)

Campbell's earliest estimate of a global URR of 1650 Gb was made in 1991 (Campbell, 1991), before he joined Petroconsultants and gained access to more reliable data. Based on this data, Campbell and Laherrère's published an influential report in 1997 which estimated the global URR to be 1800 Gb. Campbell's most recent estimate was for 1900 Gb, (Campbell and Heapes, 2008) indicating a 250 Gb increase over 17 years. This is still less than two thirds of the USGS 2000 estimate (Section 6.3) and 469 Gb less than cumulative 2P discoveries (IEA, 2008), but these figures relate to a broader definition of conventional oil that includes NGLs.

The process Campbell uses to derive these estimates does not appear very robust. As an illustration, Figure 6.2 presents his extrapolation techniques as applied to Libya. It is notable that the extrapolation of the Hubbert Linearisation (in green) does not follow the actual data points (in black) in any obvious way and the first derivative of the logistic curve does not bear any relation to the actual production data. It is unclear from the surrounding text how these charts contribute to the estimate of URR for Libya. Similar problems are apparent in many of the regions examined.

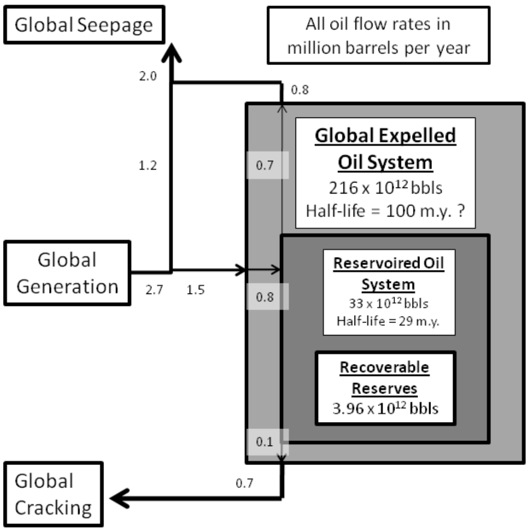
Figure 6.2: Presentation of extrapolation methods for Libya oil data as presented in Campbell and Heapes (2008).



6.2.2 Miller (1992)

As a counterpoint to the relatively small URR estimates of Campbell and Laherrère, one of the largest modern estimates of the global URR was published by Miller in 1992. This estimate excludes heavy oil, tar sands and shale oil. Miller proposes an alternative to the traditional methods of URR estimation, which he refers to as the "global oil system" (Figure 6.3). Using this model, URR is estimated by attempting to quantify the total oil in place (referred to as the "reservoired" oil system) by calculating the balance between 'oil generation' and 'oil seepage'. The recoverable resource is estimated as a proportion of the oil in place, using assumptions about recovery factors. As can be seen in Figure 6.3, Miller concludes that a URR of 3960 Gb is plausible but notes that this depends upon uncertain assumptions about the rate of seepage and the state of equilibrium between oil generation and natural loss within the system.

Figure 6.3 The global oil system model as presented by Miller (1992)



Source: Miller (1992)
Notes: Arrows represent flow rates, expressed in million barrels per year.

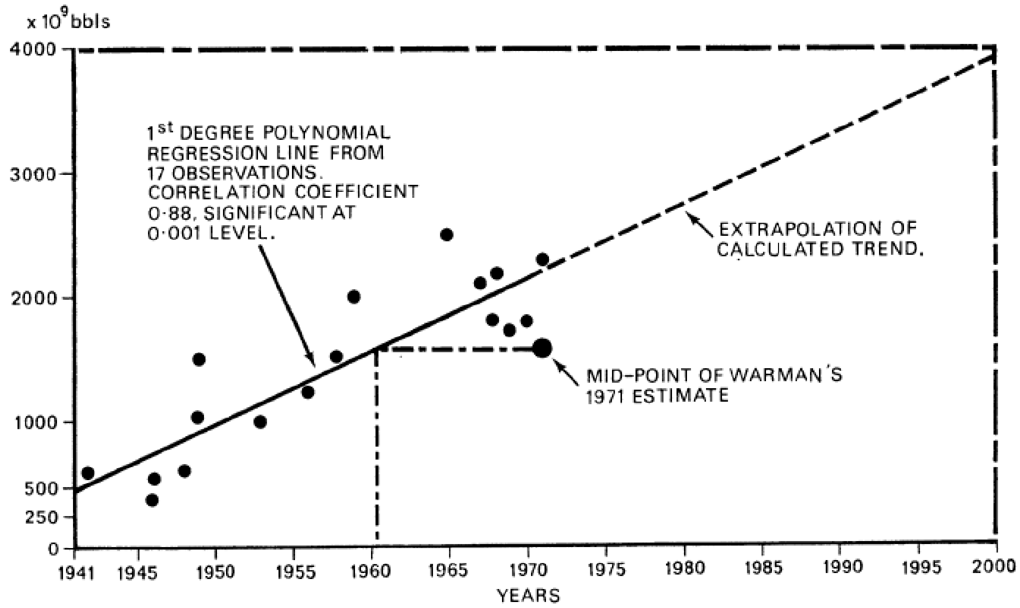
- *Global Seepage;* oil which oxidises upon seeping to surface or oxidised within reservoir (mb/y)
- *Global Generation:* Volume of oil generated within the source rock (mb/y)
- *Global Cracking:* oil lost due to thermal cracking (mb/y)
- *Global Expelled Oil System:* total volume of oil expelled by source rocks, with fixed volume, input and output at any specific point in time (b)
- *Reservoired Oil System: v*olume trapped in reservoirs, including all original oil-in-place, both known and yet-to-find (b)
- *Recoverable Reserves:* estimate of URR based on recovery factor applied to Reservoired Oil System (b)

6.2.3 Odell

In contrast to Campbell and Laherrère, Odell has consistently provided relatively optimistic forecasts of regional and global oil supply, together with correspondingly optimistic estimates of the global URR (Odell and Rosing, 1980b). Much of Odell's optimism regarding oil supply comes from his estimates of the non-conventional URR and his estimates of the conventional URR have appeared more realistic over time in the light of other studies.

In 1973, Odell forecast that estimate of conventional oil URR would reach 4000 Gb by 2000, on the basis of a linear extrapolation of historical trends (Figure 6.4) (Odell, 1973c). The assumption that estimates of global URR will continue to increase is perhaps unrealistic and 4000 Bb is greater than most contemporary estimates (although see Section 6.4). Odell and Rosing (1980a) subsequently considered estimates of the global URR ranging from 2000 Gb to 11000 Gb as inputs to their model of future oil supply. Odell identifies 3000 Gb as a realistic estimate and has used this in subsequent modelling (Odell, 2004). Throughout his work, Odell does not make any primary estimates of global URR, but simply draws upon other published estimates.

Figure 6.4: Odell's estimates of world ultimate reserves of crude oil from conventional sources (with extrapolation to the year 2000)



Source: Odell (1973c)

6.3 The USGS World Petroleum Assessment 2000

6.3.1 Methods

As indicated above, the most comprehensive and influential estimates of global URR have been produced by the US Geological Survey (USGS) which has published five assessments of the global URR for petroleum (conventional oil, NGLs and natural gas) since 1980 (USGS, 2000b). Each of these assessments used a combination of methods to estimate the URR of geologically homogeneous regions, which were then aggregated to the level of the world as a whole. The assessment methods have changed significantly since 1980 and have greatly increased in complexity. The most recent and comprehensive assessment was completed in 2000 and runs to 30,000 pages, available only on CD-ROM (USGS, 2000b).

The 2000 assessment considered petroleum resources that had *the potential to be added to reserves* between 1995 and 2025 using existing technology.[85] This involves assumptions about technical and economic viability and implies that the results could both *underestimate* the global URR (since some resources may only be technically and economically accessible in the longer term) and *overestimate* resource availability up to 2030 (since political and other constraints may prevent resources from being accessed and exploited). However, the estimate will be referred to as the global URR in what follows.

The USGS divided the world into eight oil producing regions, namely:

- Former Soviet Union

[85] Previous USGS assessments did not specify a particular time span.

- Middle East and North Africa
- Asia Pacific
- Europe
- North America
- Central and South America
- Sub-Saharan Africa and Antarctica
- South Asia

Seven of these regions were assessed in depth using a common methodology (termed the 'Seventh Approximation'), while resource estimates for the United States were taken from previous studies by the USGS (1995) and the Minerals Management Service (1996). The world was divided into 937 petroleum provinces,[86] 406 of which were *known* to contain petroleum resources (354 outside the US and 52 within the US) while an additional 5 were considered *likely* to contain petroleum although no discoveries had yet been made.[87] Formal assessments were made of 128 non-US provinces located in 96 countries and two jointly held areas. These were subdivided into 76 'priority' provinces containing 95% of the world's discovered petroleum and 52 'boutique' provinces which were more prospective (USGS, 2000b). The distribution of resources between these provinces resembled the distribution of resources between different fields in a region, in that most of the petroleum was estimated to be contained within a small number of provinces. By implication, 538 provinces were excluded altogether from the assessment, presumably because they were considered unlikely to contribute to global oil supply over the 30 year time horizon of the study. The implications of this are discussed further in Section 6.4.2.

The 128 provinces for which formal assessments were conducted (exclusive of the US) were divided into 159 Total Petroleum Systems (TPS) and 270 Assessment Units (AU). Assessments were made of 149 TPS and 246 AUs that were judged to be 'significant' on a world scale.[88] The TPS was a new geological concept, introduced for the first time in the 2000 WPA, and comprising 'all genetically related petroleum generated by a pod of mature source rocks' (USGS, 2000b).[89] Each TPS was divided into one or more AUs, which were designed to be sufficiently homogeneous in terms of geology, field size distribution, exploration considerations, accessibility and risk to be examined using a single resource assessment methodology. The study excluded unconventional resources and imposed a minimum field size, which ranged from 1 to 20 million boe depending upon the individual AUs. For example, a larger minimum field size was assumed for offshore regions.

[86] The 937 provinces were defined to encompass all the world's major land areas and adjoining water to depths of a least 2000 m. A province is defined as an area having dimensions of perhaps hundreds to thousands of kilometres encompassing a natural geological entity (e.g. sedimentary basin, thrust belt, delta) or some combination of contiguous geological entities.

[87] Only three of the extra five provinces are actually included in the USGS 2000 study, namely North Barents, Provence, and the East Greenland Rift.

[88] In addition, 24 AUs were identified as containing unconventional resources, but these were not quantitatively assessed (USGS, 2000b).

[89] More specifically: "….an entity encompassing genetically related petroleum that occurs in seeps, shows and accumulations (discovered or undiscovered) that has been generated by a pod or by closely related pods of mature source rock, together with the essential mappable geological elements (source, reservoir, seal and overburden rock) that control fundamental processes of generation, migration, entrapment and preservation of petroleum." (USGS, 2000b).

The assessment of undiscovered resources was based upon a mixture of geological assessments and discovery process modelling. The choice depended upon the level of exploration maturity of the region, with the results of several methods generally being combined. These provided estimates of the minimum, mode and maximum number of undiscovered fields, together with their minimum, median and maximum size. These numbers were used as inputs to a Monte Carlo simulation which derived probabilistic estimates of the size of undiscovered resources under the assumption of a shifted lognormal field size distribution (Charpentier, 2005). Probablistic assumptions about 'co-product' ratios were then used to estimate the volume of oil in gas fields (i.e. NGLs) and vice versa.[90]

Since the study uses a baseline of 1st January 1996, the estimates of cumulative discoveries (cumulative production plus 2P reserves) were already five years out of date by the time the study was published. Similarly, the estimates of 'undiscovered' resources refer to resources that had the potential to be added to reserves between 1996 and 2025. Again, a portion of this had already been discovered by the time the study was published.

6.3.2 Results

Table 6.5 and Figure 6.5 present the USGS mean estimates of the global URR for petroleum liquids while Table 6.6 shows how this is split between oil and NGLs, together with the corresponding range of uncertainty. Table 6.7 shows the estimated regional breakdown of undiscovered resources. The following points may be highlighted:

- The mean estimate for the global URR for petroleum liquids was **3345 Gb**, of which 90.3% (3021 Gb) was for conventional oil and the remainder NGLs. This represented a 47% increase on the previous USGS estimate of 2273 Gb.[91]
- This large increase derived in part from the inclusion of reserve growth for the first time and also from a significant increase in the estimated size of NGL resources. The latter were estimated to comprise 9.7% of the URR for liquids, 12.1% of the remaining resources and 22.0% of undiscovered resources, compared to only 1% of cumulative production through to 1996.
- The mean estimate for undiscovered liquid resources was **939 Gb**, or 28.1% of the estimated URR and 35.6% of the remaining resources. This is 48% larger than the mode estimate in the previous USGS assessment (471 Gb).
- One third of the estimated undiscovered resources were estimated to be located in the Middle East, followed by the FSU (18.2%) and North America (17.2%).
- Assuming a constant discovery rate, the mean estimates imply that an average of 31 Gb should be found each year between 1996 and 2025. This compares to an average of 14 Gb in the previous ten years (1986-1995) and 22 Gb in the previous twenty years (1976-1995). In other words, these potential reserve additions could only be achieved through a major turnaround in global exploration success, which has been declining fairly continuously since the mid 1960s (Figure 6.6). Whether this decline is a result of physical depletion, restricted access to the most promising areas or price-induced reductions in

[90] A more comprehensive summary of the methodology used can be found in Appendix 1 of Klett *et al.* (2005a) or Chapter AM of USGS (2000).
[91] Earlier USGS estimates were: 1719 Gb (1981), 1744 Gb (1985) and 2171 Gb (1990) respectively (USGS, 2000b).

exploration activity is disputed (Bentley, 2002; Mills, 2008). In practice, each is likely to have played a role.

- The mean estimate for reserve growth was **730 Gb**, or 21.8% of the estimated URR and 27.8% of the remaining resources. Hence, reserve growth at existing fields was expected to contribute almost as much to future resource additions as new discoveries.
- Only 21.4% (717 Gb) of the estimated URR had been produced through to January 1996. However, with cumulative production through to December 2007 of 1128 Gb, this figure has now increased to 33.7%, with the growth in annual production averaging 1.5%/year. Put another way, about a third of the oil that has ever been produced has been produced in the last fourteen years (1994-2007) and one quarter in the last ten years.
- If consumption continues to grow at an average of 1.5%/year, the midpoint of the mean estimate of URR (1672 Gb) would be reached around 2024 (with production at 103 mb/d). The midpoint would be reached later if demand grew more slowly or earlier if it grew more rapidly. It could also be reached earlier if new discoveries and/or reserve growth deliver less than the mean estimate of reserve additions. More importantly, various factors may prevent reserve growth and new discoveries from contributing to reserve additions at the *rate* that is required.

Table 6.5 USGS WPA 2000: mean estimates of global URR for petroleum liquids (Gb)

	US total liquids	**World (non-US) oil**	**World (non-US) NGLs**	**World Total**
Cumulative production	171	539	7	**717**
Remaining 2P reserves	32	859	68	**959**
Reserve growth	76	612	42	**730**
Undiscovered resources	83	649	207	**939**
URR	**362**	**2659**	**324**	**3345**
Remaining resources	*191*	*2120*	*317*	*2628*

Source: USGS (2000)

*Note*s: All figures refer to January 1996.

Figure 6.5 USGS 2000: components of the estimated global URR for conventional oil

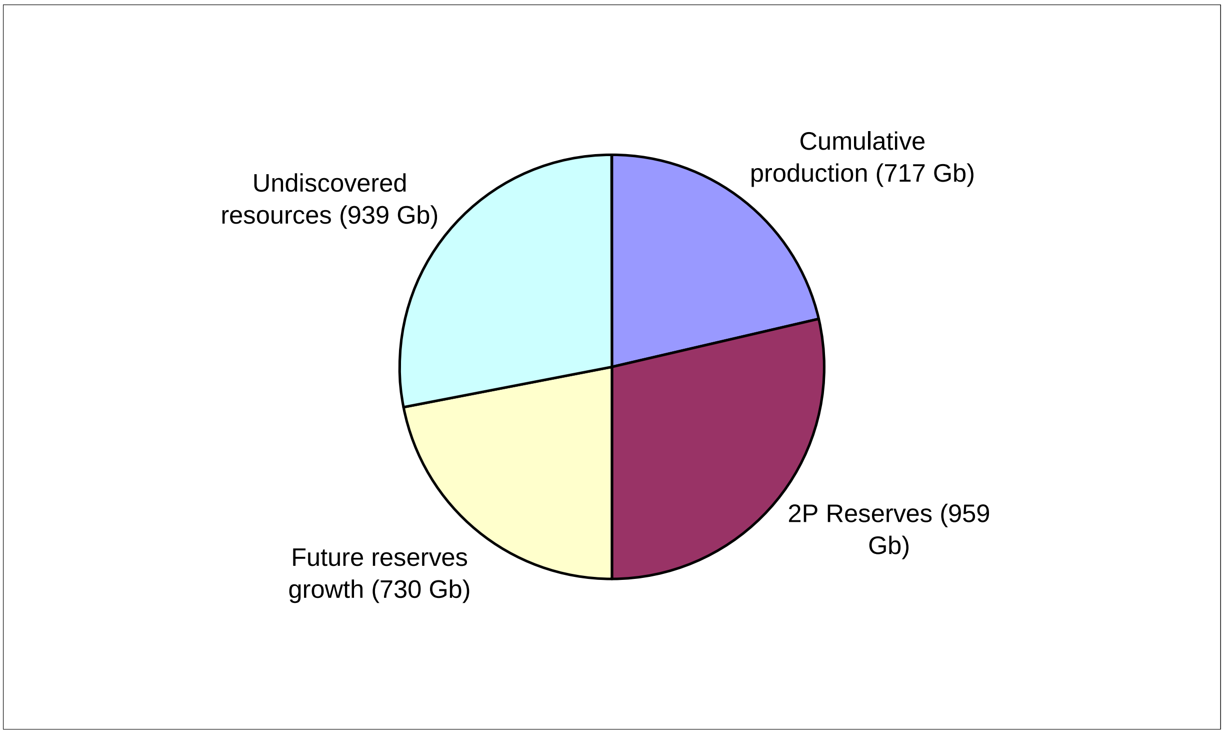


Figure 6.6 Comparing historical trends in backdated 2P discoveries with those implied by the USGS 2000 for the period 1995-2025

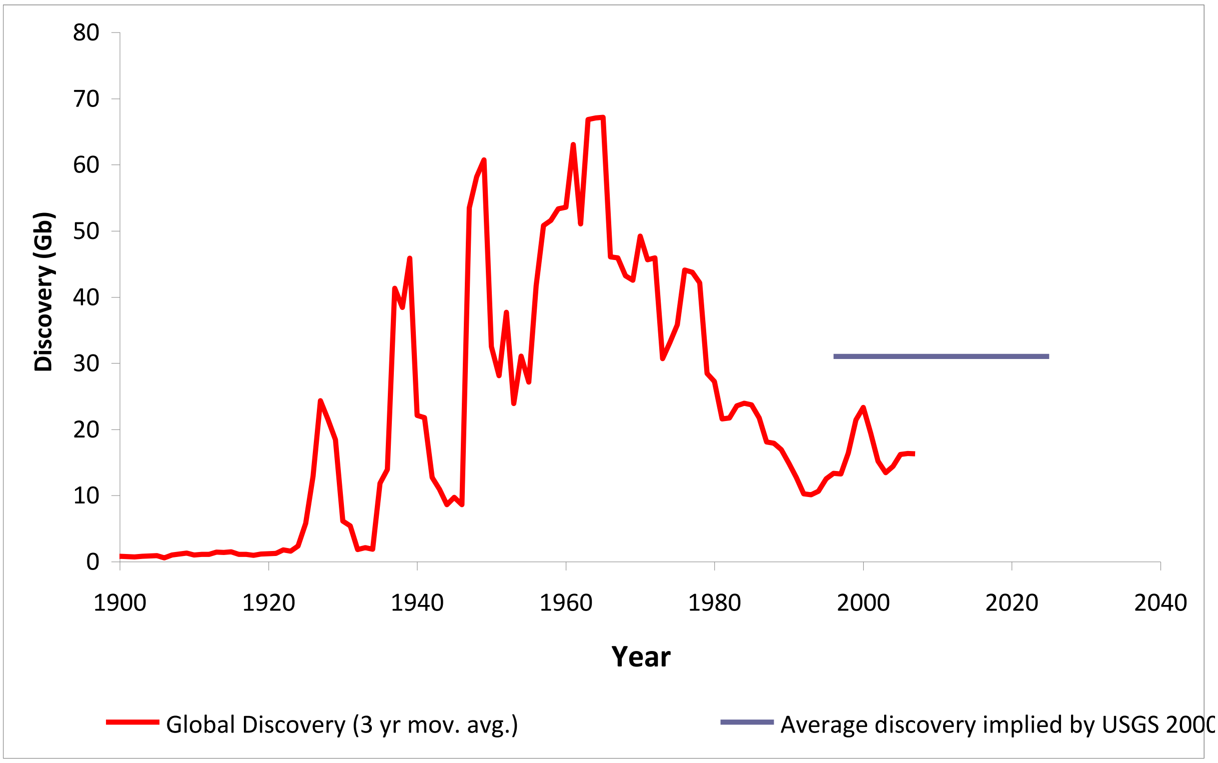


Table 6.6 USGS World Petroleum Assessment 2000: summary of global URR estimates for petroleum liquids

	Oil				NGLs			
	F95	**F50**	**F5**	**Mean**	**F95**	**F50**	**F5**	**Mean**
Rest of World								
Undiscovered	334	607	1107	649	95	189	378	207
Reserve growth	192	612	1031	612	13	42	71	42
Remaining reserves				859				68
Cumulative Production				539				7
Total				**2659**				**324**
United States								
Undiscovered	66		104	83	Combined with oil			
Reserve growth				76	Combined with oil			
Remaining reserves				32	Combined with oil			
Cumulative Production				171	Combined with oil			
Total				**362**	Combined with oil			
World Total				**3021**				

Source: USGS (2000)
Notes: All figures refer to January 1996. Fx implies an estimated x% probability of the resource exceeding the indicated size.

Table 6.7 USGS WPA 2000: Mean estimates of undiscovered resources by region

Region	Oil	NGLs	Liquids	% of total
Former Soviet union	116	54.8	170.8	18.2%
Middle East & North Africa	229.9	81.7	311.6	33.2%
Asia-Pacific	29.8	15.4	45.2	4.8%
Europe	22.3	13.7	36	3.8%
North America	153.5	7.9	161.4	17.2%
Central & South America	105.1	20.2	125.3	13.3%
Sub-Saharan Africa & Antarctica	71.5	10.8	82.3	8.8%
South Asia	3.6	2.6	6.2	0.7%
Total	**731.7**	**207.1**	**938.8**	**100%**

Source: USGS (2000)

*Note*s: All figures refer to January 1996.

The most controversial aspect of the USGS WPA 2000 was the assumptions about reserve growth (Laherrère, 2001b). While this process had been well studied in the US, it had been excluded from previous global assessments, owing to insufficient data. However, this neglect was becoming increasingly inappropriate, given that reserve growth appeared to be accounting for an increasing proportion of global reserve additions. Using the Petroconsultants database, the USGS (2000) found that the estimated URR for 186 giant fields outside the US had increased by 26% between 1981 and 1996. This was *greater* than would have been predicted by the reserve growth functions estimated from US oil fields, despite the Petroconsultants database containing 2P reserve data while the US function was estimated from 1P data. Hence, the neglect of non-US reserve growth no longer seemed viable.

In contrast to the comprehensive and sophisticated assessment of undiscovered resources, the USGS methodology for estimating reserve growth was remarkably crude. Since the available data was considered inadequate - in terms of completeness, quality and internal consistency - to accurately estimate reserve growth functions for regions outside the US, the USGS chose instead to apply a single US reserve growth function to *all* oil and gas fields throughout the world.[92] This gave point estimates for the ‘grown size’ (by 2025) of each field that had been discovered before 1996. To reflect the uncertainty in these estimates, the USGS assumed a symmetrical triangular probability distribution around this estimate, with a minimum value of zero.[93]

Much debate has arisen over the validity of applying US growth functions to global oil reserves in this way. For example, Laherrère (2001b) has argued that much of the US reserve growth is a reporting phenomenon, linked to the SEC rules regarding disclosure of highly conservative 1P reserve estimates, so it is inappropriate to apply this to 2P data. Similarly, an

[92] The function was a weighted average of the oil and gas field functions used in the 1995 US national assessment (Attanasi, *et al.*, 1999; Gautier, *et al.*, 1995).

[93] Arguably, a minimum value of less than zero should have been used to reflect the possibility that 2P estimates could reduce overtime.

increasing proportion of global supply is projected to come from offshore fields, but reserve growth should be less for these fields owing to the more limited opportunities for additional drilling. However, the USGS was quite open about the limitations of its approach:

> "…..The forecast of world potential reserve growth described here is considered to be preliminary.... the present study is an attempt to provide a numerical hypotheses for world potential reserve growth that is valuable in itself and will perhaps acted as a stimulus for discussion and research aimed at reducing the uncertainty of world reserve growth estimates." (USGS, 2000b)

The USGS notes that its approach may *overestimate* reserve growth if the criteria for reporting non-US reserves was less restrictive than in the US (as they are), if reserves are overstated in some countries (as seems likely), or if non-US fields have more development prior to the release of initial field size estimates (leading to more accurate initial reserve estimates and reducing the potential for future growth). At the same time, the approach could *underestimate* reserve growth over the next 30 years if non-US fields benefit from better technology than that which determined the historical US growth function, or if these fields have not been developed as fast as US fields of the same age. In practice, the relative importance of these different factors may be expected to vary widely from one region to another and one type of field to another. But in the absence of good data, there is room for a range of views on the net effect at the global level.

6.3.3 Evaluation

Since we are now 40% of the way through USGS assessment period (1995-2025), some evaluation of the 'accuracy' of the assessment can be made. But it is important to recognise that the study did not predict what would *actually* be found in 30 years, but instead estimated what could *potentially* be found using existing technology. In other words, various political, economic and investment constraints could (and almost certainly have) prevented reserve additions.

The USGS evaluated their assessment using IHS data through to December 2003 (i.e. 27% of the assessment period) (Klett, *et al.*, 2005a). They found that only 69 Gb of oil had been discovered in the 128 assessed provinces (i.e. those outside the US), or less than 11% of the mean estimate of undiscovered resources (649 Gb) for those provinces. Assuming a constant discovery rate, a total of 173b should have been discovered by 2003 or 27% of the undiscovered resource. In other words, real-world oil discoveries outside the US were less than half of what was 'expected' over this period. Klett *et al*. (2005a) highlight a number of possible reasons for this, including limited access to resources in Iraq, Iran[94] and Libya, political and economic instability in Russia and the Central Asian republics and low oil prices during the late 1990s leading to low rates of exploratory drilling. The IEA (2008) reports a fall in the average *number* of fields discovered per year since 1996 as well as the average *size* of those fields which suggests that there has been some reduction in exploratory activity.

Although not mentioned by Klett *et al.*, an additional reason for the apparently low discovery rate may be that the 69 Gb figure represents the 2003 estimate of the URR of the fields discovered in the previous eight years and therefore has *not* been adjusted for future reserve growth (i.e it is $B'_{t_d,2007}$ and not $B'_{t_d,2025}$). If this adjustment was made, the 'actual'

[94] Although, Iran had the largest amount of discoveries between 1993 and 2002.

discoveries may be much closer to the 'forecast' discoveries. For example, the 'modified Arrington' reserve growth function depicted in Figure 2.2 projects a quadrupling of the initial estimate of URR in 20 years, with the majority of the increase occurring in the first ten years (Verma, 2005).

In contrast to new discoveries, reserve growth appeared to be tracking the USGS projections relatively well. From analysis of the IHS database, Klett, *et al.* (2005a) conclude that a total of 171 Gb had been added through reserve growth by 2003, which is 28% of what was expected over the full 30 years and more than twice the resource additions through new discoveries. This suggests that (contrary to be expectations of many critics), the US reserve growth function works relatively well when applied to 2P data - at least at the global level. However, some of the apparent reserve growth in the IHS PEPS database may result from revisions to the reserve estimates of various OPEC countries which remain a subject of controversy (Thompson, *et al.*, 2009b). For example, Stark and Chew (2005) found a global total of 465 Gb of reserve growth between 1995 and 2003, of which 175 Gb was attributed to 'classic' reserve growth and the remainder to 'new and revised data'. This distinction suggests that much of the apparent reserve growth could derive from factors such the inclusion of previously omitted fields in the industry databases and from revised estimates of fields where the data was poor. The biggest growth in absolute terms derived from Middle East fields where the reserves data is particularly uncertain.

The most uncertain estimates in the USGS 2000 assessment were for relatively unexplored regions of the world, such as the East Greenland Rift. Here, the USGS estimated a 95% probability of oil resources exceeding zero, a 5% probability of more than 111.8 Gb and a mean estimate of 47 Gb. In a recent reappraisal of this area, the USGS (2007) downgraded the mean estimate to only 8.9 Gb (although the estimate for NGLs was doubled from 4 to 8Gb). While this downgrading is regionally significant, it only amounts to 1.4% of the 2000 estimate of remaining resources (2628 Gb). Also, given the formidable difficulties of accessing resources in East Greenland, this region seems unlikely to make a significant contribution to global oil production before 2025.

As the USGS acknowledges, the estimates for relatively unexplored regions must remain highly uncertain. While very few regions are wholly unexplored, there are still large areas that remain poorly explored, including parts of central and southern Africa, the offshore regions of Kenya and Namibia, large parts of Libya and the Middle East, and offshore regions of Argentina, Colombia, Peru, Venezuela and Mexico. Many of these were excluded altogether from the USGS assessment, while exploration in many of the included regions was restricted by various political, economic and technical constraints. If these continue, the potential reserve additions identified by the USGS will not be realised before 2025, even if the resource is actually there. This in turn could lead to supply difficulties occurring at an earlier date. For example, suppose that the anticipated reserve growth is fully realised by 2025, but only half the estimated undiscovered resources are developed. If consumption grows at 1.5%/year, the midpoint would be reached as early as 2017 (Strahan, 2007a).

6.4 Recent modifications to the USGS estimates

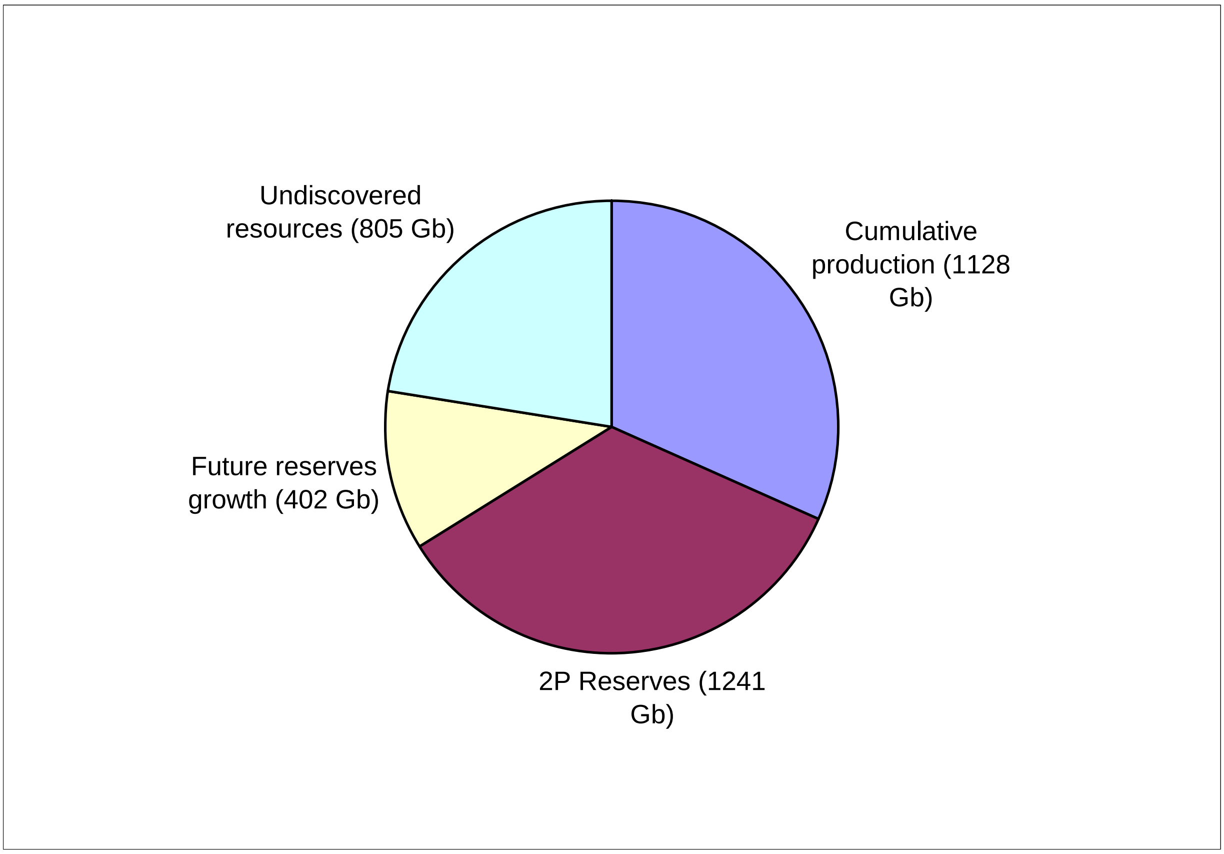
6.4.1 The IEA World Energy Outlook 2008

An updated assessment of the global URR was presented in the IEA World Energy Outlook 2008 (IEA, 2008). This took the USGS 2000 study as the primary data source, but updated this with information from the IHS database, a recent evaluation of the USGS 2000 assessment (Klett, *et al.*, 2007), additional information from the USGS and the IEA's own databases and analyses. The results for total petroleum liquids are summarised in Table 6.8 and Figure 6.7 (note that the IEA do not provide a breakdown between oil and NGLs). The most notable point is that the IEA estimate the global URR for conventional petroleum liquids to be **3577 Gb** which is 6.9% *larger* than the earlier USGS estimate. Furthermore, they estimate a 90% probability of the URR exceeding 2400 Gb and a 5% probability of it exceeding 4495 Gb. The mean estimate of remaining resources is 2448 Gb, or 68% of URR. Remaining 2P reserves are estimated to have increased by 29% since 1996 which indicates that reserve growth and new discoveries have more than offset the 411 Gb of production over this period. The remaining contribution from reserve growth is estimated at 402 Gb (16% of remaining resources), while that from undiscovered resources is estimated at 805 Gb (33% of remaining resources). The estimates of reserve growth and undiscovered resources are lower than in the USGS 2000 study, in part because a proportion of both have already been converted into production and/or 2P reserves. It is also notable that 86% of the remaining resources are estimated to lie outside the OECD.

Table 6.8 IEA 2008 WEO: mean estimates of global URR for petroleum liquids (Gb)

	OECD	**Non-OECD**	**World**	**% diff from USGS 2000**	**OECD as % of total**
Cumulative production	363	765	**1128**	57.3%	32.2%
Remaining 2P reserves	95	1147	**1241**	29.4%	7.7%
Reserve growth	27	375	**402**	-44.9%	6.7%
Undiscovered resources	185	620	**805**	-14.3%	23.0%
URR	**670**	**2907**	**3577**	**6.9%**	**18.7%**
Remaining resources	*307*	*2142*	***2448***	*-52.8%*	*24.7%*

Figure 6.7 IEA 2008: components of the estimated global URR for conventional oil



The IEA estimates (Table 6.8) exclude 'conventional oil produced by unconventional means', but the precise definition of this category is unclear. The estimate would appear to exclude the Artic region, for example, which remains both largely unexplored and extremely difficult to exploit. Much of the Artic was excluded from the USGS 2000 assessment, presumably because it was considered unlikely to contribute to global supply before 2025. But in an updated assessment for this region published in 2008, the USGS provided a mean estimate of 134 Gb of undiscovered petroleum liquids (90 Gb oil and 44 Gb NGLs) (USGS, 2008). This would increase the global URR for petroleum liquids by 3.7% and the remaining resources by 5.5%.

The IEA (2008) also provide a long-term oil supply cost curve which appears to exclude NGLs but include 'conventional oil produced by unconventional means'. This curve includes the following estimates of the size of remaining resources:

Conventional oil:	2100Gb[95]
Enhanced oil recovery (EOR):	400 - 500 Gb
Deepwater and ultra deepwater:	160 Gb
Artic:	90 Gb

[95] Comparison with Table 6.5 implies a mean estimate of 348 Gb for the remaining resources of NGLs. This is larger than the corresponding estimate in the USGS 2000 study (317 Gb), despite significant production of NGLs in the intervening 12 years.

Combining these estimates and adding the implied contribution from NGLs (348 Gb) leads to a *higher* estimate of the remaining resources of petroleum liquids - 3148 Gb. This in turn implies a URR of **4276 Gb** which is significantly larger than the figure in Table 6.8 but still excludes heavy oil and oil sands. Hence, the URR implied by the IEA supply curve is larger than the URR given in the table and represents one of the largest estimates seem to date. It is notable, however, that this estimate is very sensitive to the assumptions about EOR which the IEA states could potentially increase recovery factors from their current average of 35% to around 50%. Although this would provide an additional 1200 Gb of recoverable resources, the IEA estimates that it would take 'much more than two decades' to achieve. The estimated contribution from EOR in the supply curve data appears to imply a increase in the average recovery rate to around 40%, but in its supply forecasts to 2030, the IEA anticipates a cumulative contribution of only 24 Gb from EOR before 2030 which is less than 12% of the estimated potential. Similarly, it projects that only 14% (114 Gb) of the undiscovered resources will be found before 2030.

6.4.2 Colorado School of Mines

A comparably optimistic assessment of the global URR is provided by Aguilera, *et al.* (2009). This estimate is also based upon the USGS 2000 study but (unlike the IEA) does not update the figures to allow for production, discoveries and reserve growth since 1996. Instead, it increases the USGS estimates in two interesting and somewhat unconventional ways.

In the first stage, Aguilera *et al.* argue that the USGS underestimates total petroleum resources because they only assessed 409 provinces out of the global total of 937.[96] Moreover, only 200 of these were assessed in detail. Table RH-1 in the USGS study ranks the 409 provinces in descending order of size, which shows that the provinces that were not assessed in detail typically contain less than <0.1% of cumulative non-US discoveries (USGS, 2000b).

The 528 provinces that were excluded from the study remain unexplored but could potentially contain petroleum resources. This includes regions such as the Artic, Antarctic and much of sub-Saharan Africa. The main reason for excluding these was that they were considered unlikely to contribute to global supply over the 30 year time horizon of the USGS study. This could be because they were relatively small or because they were relatively inaccessible and hence difficult to exploit (e.g. the Artic). By including estimates of the resources contained in these provinces, Aguilera *et al.* may obtain a better estimate of the 'long-term' global URR. However, if these resources are difficult to access and exploit, they may not necessarily contribute to oil supply in the short to medium term and hence may have no influence on the date of the global peak. In this context, the following two quotes should be noted:

> " the USGS assessment is not exhaustive, because it does not cover all sedimentary basins of the world. Relatively small volumes of oil or gas have been found in an additional 279 provinces, and significant accumulations may occur in these or other basins that were not assessed. The estimates are therefore conservative." (Ahlbrandt and McCabe, 2002)

[96] This includes the 406 provinces that were known to contain petroleum, but only three of the five that were considered likely to contain petroleum. Of these, only 200 were assessed in detail.

> "We believe that the USGS (2000) estimates are conservative for a variety of reasons, chief among which are that the USGS assessment did not encompass all geologically conceivable small sources of conventionally reservoired crude oil and was limited to the assessment of reserves that would be added within a 30 year time frame because, in part, technological changes beyond 30 years are difficult, if not impossible, to conceptualize and quantify." (US Energy Information Administration)

To estimate the size of the resources in the unassessed provinces, Aguilera *et al* fit a 'variable size distribution' (VSD) model to known resources contained within the 200 assessed provinces. This approach is analogous to the field size distribution models for estimating URR that were reviewed in Section 2.5. But rather than assuming a particular (e.g. lognormal) size distribution, the model "…. allows the data to determine the size distribution relationship rather than specifying this relationship ex ante" (Aguilera, *et al.*, 2009). Aguilera, *et al* then extrapolate this curve to estimate that 593 Gb of petroleum liquids are contained in the remaining 735 provinces – which is approximately one fifth of the remaining resources identified by the original USGS study.

In the second stage, Aguilera *et al.* assume that reserve growth *also* applies to undiscovered resources and can be estimated using the reserve growth functions used previously by the USGS. As a result, they *adjust* these estimates upwards by as much as 50% for both of assessed and unassessed provinces. This leads to an additional 1025 Gb of petroleum liquids.

The net result of these two adjustments is an estimate of 3516 Gb for the remaining resources or petroleum liquids and an estimate of **4233 Gb** for the URR.[97] This figure is comparable to that implied by the IEA's supply curve (see above), but derived through an entirely different route. Since Aguilera, *et al* make no assumptions about EOR, a combination of their approach with the IEA's assumptions could potentially lead to an even more optimistic estimate for the global URR.

Aguilera *et al.*'s approach is questionable in a number of respects. First, it may be unreasonable to assume that all of the 937 provinces contain recoverable liquids and many of those that do (e.g. in the Artic regions) are likely to remain relatively inaccessible in the medium-term. Second, the unassessed provinces are significantly smaller (in resource size) than the assessed provinces, raising questions about both the economic viability of extraction and the net energy yield. Third, the application of reserve growth multipliers to the mean estimates of undiscovered resources seems difficult to justify. These multipliers were derived from studies of existing (frequently very old) fields, they relate to estimates derived from exploratory drilling and production experience and they reflect factors such as conservative initial reporting and improvements in recovery technology over the past 50 years. In contrast, the estimates of undiscovered resources are based largely on geological information, already contain wide confidence intervals to reflect uncertainty and are implicitly based upon assumptions about recovery factors that reflect modern technology. It therefore seems something of a leap of faith to multiply these estimates by such a large factor and as far as we are aware, no other researchers have done so.

While the studies by Aguilera *et al* and the IEA have 'pushed the envelope' of global URR estimates, the relevance of these estimates to forecasts of global oil supply needs to be questioned. This is because the relevant question is not simply whether the resource is there,

[97] Since Aguilera, *et al* are working with the original USGS data (which applies to 1996), the *URR* is estimated here by adding Aguilera, *et al's* estimate of remaining resources to the USGS figure for cumulative production through to 1996.

but whether it can be accessed sufficiently quickly to contribute to global oil supply in the medium-term. Multiple factors will influence this, including the physical accessibility of the relevant regions, the technical difficulties of extracting the resource, the implied investment requirements, the degree to which the resource can be accessed by independent oil companies and so on. Hence, even if the more optimistic URR estimates are correct, this may not make any difference to the date of the global oil supply peak. The following section explores this relationship between stocks (global URR) and flows (global supply) in more detail.

6.5 The implications of global URR estimates for future global supply

The above review demonstrates that contemporary estimates of the global URR for conventional oil fall within the range **2000-4300 Gb**, while the corresponding estimates of the quantity of remaining resources fall within the range **870 to 3170 Gb**. In other words, the highest estimate of remaining resources is four times greater than the lowest estimate. While the lower end of this range arguably results from an excessively narrow definition of conventional oil, the upper end arguably results from excessively optimistic assumptions about reserve growth, undiscovered resources and/or the future potential of enhanced oil recovery. But while excluding both could narrow the range, the degree of uncertainty is likely to remain very high for the foreseeable future. This in turn, leads to a corresponding uncertainty in the projections of future global oil supply. In particular, precise forecasts of the date of peak production appear wholly unwarranted when there so much uncertainty over this key parameter.

It is useful to explore the implications of this uncertainty with the help of a simple logistic model of global liquids production – where liquids is defined here to include crude oil, NGLs, condensate, heavy oils (<10^{o} API) and oil sands. As illustrated in Section 5.3, the logistic model provides a relatively poor fit to global production trends for these liquids, in part because it fails to account for the effects of the oil shocks in the 1970s. The 'best-fit' logistic model for cumulative production leads to a global URR estimate for these liquids of 1440 Gb which is less than cumulative global discoveries through to 2007 and only 29% more than the cumulative production. Similarly, a 'best fit' model of the rate of production gives a global URR estimate of 1860 Gb, which is only 65% more than cumulative global production and lower than the most pessimistic URR estimates considered here – despite a more inclusive definition of liquids. However, instead of using non-linear regression to *estimate* the URR, we can instead *assume* a value for the URR and estimate 'best-fit' values for the other two parameters – including the date of peak production (t_m). By assuming a range of values for the global URR, we can investigate the sensitivity of the data of the peak to the size of the global resource.[98]

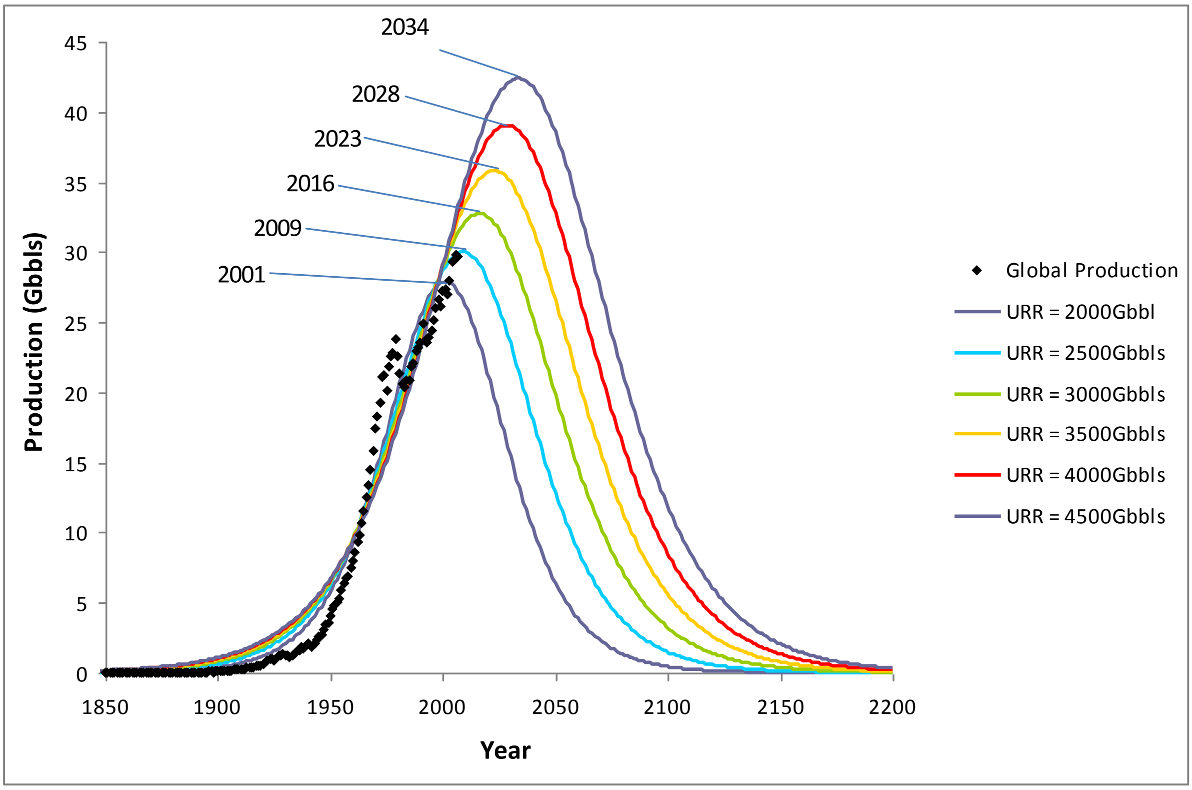
Figure 6.8 and Figure 6.9 show the results of such an exercise in which a first differential of the logistic curve has been fit to annual global production data using assumptions for the global URR ranging from 2000 Gb to 4500 Gb. For an estimate of 2500 Gb, the model gives peak production in 2009 at a level of 30 Gb/year (82 Mb/day), while for an estimate of 4500 Gb, the model gives peak production in 2032 at a level of 42 Gb/year (115 Mb/day). Hence, a

[98] Earlier investigations along these lines have been conducted by Bartlett (2000), Carlson (2007a; b) and Brecha (2008; Brecha, *et al.*, 2007) among others.

125% increase in the size of the URR (or a 260% increase in the size of the remaining resource), delays the date of peak production by only 23 years. Put another way, *increasing the global* URR *by one billion barrels delays the date of peak production by only 4.7 days*.[99] To delay the date of peak production by one year would require the addition of some 78 billion barrels to the global URR, which is two and half times greater than global production in 2007 and almost seven times greater than global discoveries in that year.[100] To put this in perspective, it implies that the discovery of new resources equivalent to the UK portion of the North Sea would delay the date of global peak production by only six months – even assuming that those resources could be developed and produced within the required timeframe. Similarly, the discovery of resources equivalent to those of the entire United States would delay the date of the global peak by less than four years.

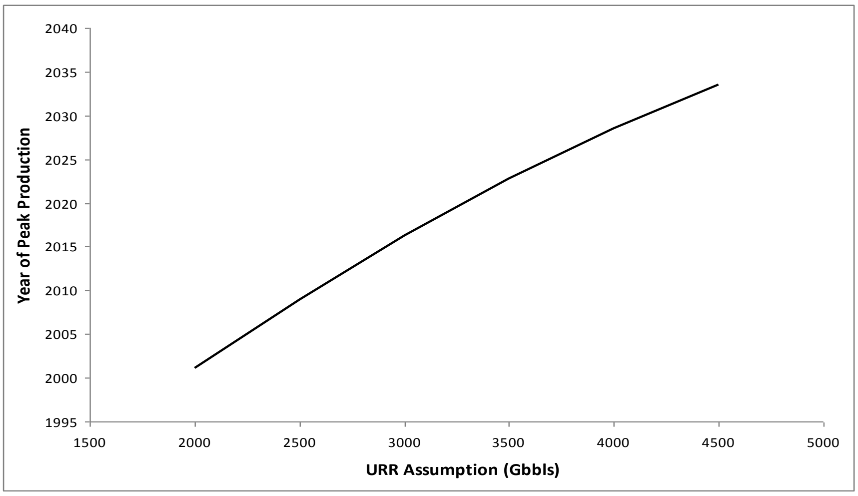
These sobering figures merely reflect the overwhelming power of exponential growth.[101] As a consequence, the range of uncertainty over the date of peak production must be significantly less than the range of uncertainty over the size of the resource. As Brandt (2007) notes: "…..Hubbert-like theories based on good estimates of ultimate recovery cannot be wrong by decades, regardless of the details".

Figure 6.8 The peaking of global conventional oil production under different assumptions about the global URR - simple logistic model



[99] Using production data through to 1995 and a slightly less inclusive definition of liquids, Bartlett estimated a corresponding figure of 5.5 days (Bartlett, 2000).
[100] This may be an overstatement, since the discovery figure has not been corrected for future reserve growth.
[101] Bartlett (1969) comments that 'The greatest shortcoming of the human race is our inability to understand the exponential function.'

Figure 6.9 Sensitivity of the date of global peak production of conventional oil to different assumptions about the global URR – simple logistic model



A fair criticism of the above analysis is that the global production cycle is unlikely to be symmetric. For example, it is possible that production could decline rapidly as the giant fields are depleted and production shifts towards much smaller fields. Equally, it is possible that production could decline more slowly as price signals provide incentives for demand reduction and enhanced oil recovery.[102] Hence, uncertainties about the size of the global URR are compounded by uncertainties about the future shape of the global production cycle.

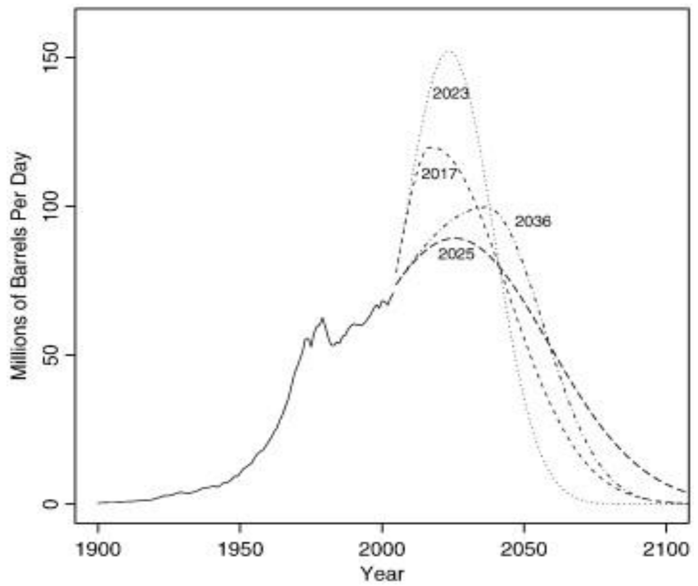
In an insightful paper, Kaufmann and Shiers (2008) address these uncertainties through a form of sensitivity testing. They develop a bell-shaped model for global production that is solved iteratively as a set of three equations:

$$Q'(t) = \left\{ \begin{array}{ll} Q'_{t-1}\left(1 + r_{inc} - t\left(\dfrac{r_{inc}}{t_m + 1}\right)\right) & \text{for } t \le t_m \\ Q'_{t-1}\left(1 + r_{dec} - t\left(\dfrac{r_{dec}}{t_m + 1}\right)\right) & \text{for } t > t_m \end{array} \right\}. \tag{6.1}$$

Where t_m is the date of peak production, r_{inc} is the rate of production growth immediately before the peak and r_{dec} is the rate of production decline immediately after the peak. The equations are constrained by the requirement that cumulative production is less than the assumed URR. By combining, four assumptions about the size of the global URR with four assumptions about r_{inc} and four assumptions about r_{dec}, Kaufmann and Shiers are able to generate 64 scenarios for the future global production cycle that include varying degrees of asymmetry (to both the left and the right) and varying degrees of optimism about the size of the resource. As an illustration, Figure 6.10, shows four possible scenarios for an assumed URR of 3000 Gb.

[102] For example, Brandt's (2007) empirical study of 74 oil-producing regions found that the production cycle was most commonly asymmetric, with a rate of decline typically being less than the rate of increase.

Figure 6.10 Illustrative scenarios for future global oil production with a URR of 3 Gb



Source: Kaufmann and Shiers (2008)

Note: a) dashed line: r_{inc}=0.08, r_{dec}=0.02; b) dotted line: r_{inc}=0.08, r_{dec}=0.08; c) long dashed line: r_{inc}=0.02, r_{dec}=0.02; d) dashed-dotted line: r_{inc}=0.02, r_{dec}=0.08. All scenarios assume a URR of 3000 Gb.

Despite this wide range of scenarios, Kaufmann and Shiers find that large differences in the assumed URR lead to relatively small differences in the date of peak production. In particular, *for 53 of the 64 scenarios, the date of peak production is found to lie between 2009 and 2031*. Holding r_{inc} and r_{dec} constant, the date of peak production is found to increase by 20-25 years as the assumed URR is increased from 2 Gb to 4.5 Gb. Kaufmann and Shiers also show that, for any given value of URR, changing the initial growth and decline rates has relatively little effect on the date of peak production. Scenarios with two peaks were also considered, but the URR constraint implies either that the annual rate of production for the second peak is lower than that for the first peak, or that production declines very rapidly after the second peak. As a result, the single peak scenarios were considered to be 'best case'.

The results imply that delaying the peak in global oil production beyond 2030 requires both optimistic assumptions about the size of the global URR (i.e. 4 Gb or more), together with a relatively steep post-peak decline rate (e.g. an initial decline rate of 8%/year, or an average post-peak decline rate of ~4%/year). This form of asymmetry appears relatively unlikely and could also have serious implications once the peak is passed. For most combinations of the three parameters, the annual production of conventional oil is found to fall by 10 Mb/day within 5 to 10 years after the peak. To put this in perspective, this is equivalent to one eighth of current production or equal to the current output of Saudi Arabia. Later peaks generally imply faster rates of decline which in turn implies that substitutes must be developed faster and/or demand must fall more rapidly. While the implications of this are beyond the scope of the present study, the challenge appears formidable (Hirsch, 2008; Hirsch, *et al.*, 2005).

One weakness of the Kaufmann and Shiers study is the failure to consider demand-side shocks. The smallest value they considered for r_{inc} was 2%/year, but in practice the supply of liquids fell in 2008 as a result of the economic recession. Supply is anticipated to fall further in 2009 and it is not yet clear when it will recover. The result may be to delay the data of global peak production by several years in a similar manner to the oil shocks of the 1970s. However, the recession has also led to the cancellation of many field development projects. Given the rate of production decline from existing fields and the lead time required to develop new fields, this could lead to a near-term supply constraint should demand growth resume (IEA, 2008).

6.6 Summary

This section has summarised the various estimates of the global URR for conventional oil that have appeared over the last 50 years. It has examined the USGS World Petroleum Assessment in some detail and evaluated how more recent analysis and experience has modified the USGS estimates. It has also examined the implications of the global URR estimates for the future supply of conventional oil and the date of peak supply. The key conclusions are as follows:

- Estimates of the global URR for conventional oil vary widely in their methods, assumptions and results and the comparison between them is greatly complicated by the differing coverage of petroleum liquids - including in particular the inclusion or exclusion of NGLs. There is no universally agreed definition of ‘conventional oil’ and the more pessimistic estimates of URR result in part from an excessively narrow definition. Further difficulties arise from the use of competing reserve definitions in the estimation of cumulative discoveries, the uncertainty over the remaining reserves of various OPEC countries, the use of differing time-frames for the definition of URR and the treatment of reserve growth. All these factors contribute to considerable variability between different estimates.
- Global URR estimates have been trending upwards for the last 50 years and this trend shows little sign of diminishing. While many cluster in the range 2000-3000 Gb, credible estimates now exceed 3500 Gb.
- The USGS World Petroleum Assessment 2000 was a significant departure from previous studies, both in terms of the depth of the analysis and the size of the resulting estimates (a URR of 3345 Gb, or 47% larger than the previous USGS estimate). This estimate rested in part on some remarkably crude assumptions about reserve growth that have been widely criticised. Nevertheless, subsequent analysis indicates that these assumptions have proved broadly correct. Also, while the rate of new discoveries appears to be lower than anticipated by the USGS, this is partly a consequence of restrictions on exploration in certain regions and the failure to adjust these estimates to allow for future reserve growth. Hence, the repeated assertions that the USGS study is ‘discredited’ or ‘over-optimistic’ are at best premature.
- The recent IEA World Energy Outlook gives a slightly larger figure of 3577 Gb for the global URR of conventional oil. But when ‘conventional oil produced by unconventional means’ is included, the IEA report appears to suggest a figure of 4276 Gb. This larger estimate rests in part on contentious assumptions about reserve growth, undiscovered resources and/or the future potential of enhanced oil recovery.
- An update of the USGS study has also been provided by Aguilera, *et al* (2009). By including the resources contained with previously unassessed provinces, together with generous assumptions about reserve growth, they arrive at a estimate of 4233 Gb for the global URR. Since this is comparable to the IEA estimate, but arrived at by entirely

different means, it suggests that an even larger estimate is plausible. However, several of the assumptions used by Aguilera, *et al* appear very questionable.

- Contemporary estimates of the global URR for conventional oil fall within the range **2000-4300 Gb,** while the corresponding estimates of the quantity of remaining resources fall within the range **870 to 3170 Gb**. Hence, there is a factor of four difference between the lowest and highest estimate of remaining resources. While both the lower and higher end of this range seem questionable, the degree of uncertainty in URR estimates appears unlikely to be significantly reduced within the foreseeable future. This leads to a corresponding uncertainty in the projections of future global oil supply. Precise forecasts of the date of peak production appear unwarranted when there so much uncertainty over this key parameter.
- The implications of this uncertainty for the date of peak production may be demonstrated with the help of a simple logistic model. This suggests that increasing the global URR by one billion barrels would delay the date of peak production by only 4.7 days. To put this in perspective, it implies that the discovery of new resources equivalent to the UK portion of the North Sea would delay the peak by only six months. This result is not substantially changed if a more sophisticated model is used, that allows for varying degrees of asymmetry in the production cycle. For a wide range of assumptions about the size of the global URR and the rate of change of production before and after the peak, the date of peak production is found to lie between 2009 and 2031. Delaying the peak beyond 2030 requires very optimistic assumptions about the size of the global URR and also implies a relatively steep post-peak decline rate.
- These calculations do not consider the possibility that the date of peak production could be delayed by demand reductions or much slower demand growth, perhaps as a consequence of economic recession. However, by delaying upstream investments, economic recession could also contribute to near-term supply-constraints.
- Other things being equal, larger estimates of the global URR for conventional oil lead to more optimistic forecasts for future global oil supply. However, even if the larger estimates are correct, it does not necessarily follow that the resource can be accessed at the required rate to maintain global production at a particular level. Multiple political, economic and technological factors may prevent this occurring. In particular, the larger estimates may reflect the resources contained within small fields in relatively inaccessible regions, or rely upon optimistic assumptions about future improvements in the global average recovery rate. If these resources can only be accessed relatively slowly at high cost, or if their net energy yield is relatively low, they may have little or no influence on the date of global peak production. However, larger resources may contribute to a slower decline in production following the peak.

7 Summary and conclusions

The key lessons from this review may be grouped under three headings

7.1 Methods and principles

- There are a variety of methods for estimating URR and many variations on the basic techniques. 'Geological' techniques are more appropriate for relatively explored regions while 'extrapolation' techniques are more appropriate where exploration is advanced. The confidence bounds on these estimates are commonly very large and the few studies that compare different techniques show they can lead to quite different results. Accuracy can be improved through analysing disaggregate regions, but this is resource intensive and generally requires access to proprietary data. All estimation techniques have identifiable limitations and it is important that estimates are accompanied by confidence intervals and full details about the methodology and assumptions made.
- The extrapolation techniques differ in degree rather than kind and share many of the same strengths and weaknesses. But a key practical difference is that field-size distribution and discovery process techniques require data on individual fields, while simple curve-fitting only requires aggregate data. All assume a skewed field size distribution and diminishing returns to exploration, with the large fields being found relatively early. But these assumptions will only hold if depletion outweighs the effect of technical change and if the region is geologically homogeneous and has had a relatively unrestricted exploration history. This is frequently not the case.
- Assumptions about the field size distribution and discovery process underlie most of the extrapolation techniques. It is generally acknowledged that the majority of oil resources are contained in a small number of large fields, with around 100 oil fields accounting for up to half of global oil production and up to 500 fields accounting for two thirds of cumulative discoveries. Most of these fields are relatively old, many are well past their peak of production and most of the rest will begin to decline within the next decade or so. The remaining reserves at these fields, their future production profile and the potential for reserve growth is therefore of critical importance for future global supply.
- The proportion of total resources contained within small, undiscovered fields continues to be disputed. While the observed lognormal size distribution of discovered fields is likely to be the result of sampling bias, there is insufficient evidence to conclude whether a 'linear' or 'parabolic fractal' better describes the population size distribution. While technical improvements and higher prices should make more small fields viable, many will remain uneconomic to develop and the exploitation of the rest will be subject to rapidly diminishing returns. As a result, the competing estimates of the resources contained in small fields should be of less significance to future supply than the potential for increased recovery from the giant fields.

7.2 Curve fitting techniques

- The popularity of curve-fitting techniques to estimate URR derives from their simplicity and the relative availability of the required data. But many applications of curve-fitting take insufficient account of the weaknesses of these techniques, including: the inadequate theoretical basis; the sensitivity of the estimates to the choice of functional form; the risk of overfitting multi cycle models; the inability to anticipate future cycles of production or discovery; and the neglect of economic political and other variables. In general, these weaknesses appear more likely to lead to underestimates of the URR and have probably contributed to excessively pessimistic forecasts of oil supply.

- Curve fitting to discovery data introduces additional complications such as the uncertainty in reserve estimates and the need to adjust estimates to allow for future reserve growth. The common failure to make such adjustments is likely to have further contributed to underestimates of resource size.
- Tests of curve fitting techniques using illustrative data from a number of regions has shown how different techniques, functional forms, length of time series and numbers of curves can lead to inconsistent results. But although the results raise concerns about the reliability of curve-fitting estimates, the degree of uncertainty may be expected to decline in the future as exploration matures. Also, accuracy may be improved by using the lowest possible level of spatial aggregation, distinguishing between onshore and offshore regions and adjusting for future reserve growth using functions derive from the technical literature.
- The literature on curve-fitting techniques has generally paid insufficient attention to the statistical issues involved, such as goodness of fit, missing variables and serial correlation of the error terms. Where data is available, some of the limitations of curve fitting may be overcome with the use of hybrid models that incorporate relevant economic and political variables. But despite their better fit to historical data, such models may not lead to substantially different estimates of the URR.
- These limitations do not mean that curve fitting should be abandoned, but do imply that its applicability is more limited than commonly assumed and that the confidence bounds on the results are wider than is commonly assumed. Where possible, resource assessments should employ multiple techniques and sources of data and acknowledge the uncertainty in the results obtained.

7.3 Global Estimates

- Estimates of the global URR for conventional oil vary widely in their methods, assumptions and results. Comparison is complicated by the differing definitions of 'conventional oil' and the more pessimistic estimates of the global URR result in part from an excessively narrow definition. Further difficulties arise from the use of competing reserve definitions and differing time-frames for the definition of URR, together with uncertainty over OPEC reserves and the inconsistent treatment of reserve growth. The information currently available does not allow strong constraints to be placed on the last two variables.
- Estimates of the global URR of conventional oil have been trending upwards for the last 50 years and this trend shows little sign of diminishing. Contemporary estimates fall within the range 2000-4300 Gb, while the corresponding estimates of the quantity of remaining resources fall within the range 870 to 3170 Gb. This wide range leads to a corresponding uncertainty in the projections of future global oil supply and the date of peak production.
- The USGS estimated a global URR of 3345 Gb in 2000 and in 2008 the IEA revised this upwards to 3577 Gb. Despite being much larger than previous estimates, the repeated assertions that the USGS estimates are 'discredited' or 'over-optimistic' appear at best

premature. Global reserve growth appears to be matching the USGS assumptions, the size of recent discoveries may have been underestimated, there are continuing restrictions on exploration in the most promising areas and a more recent study by Aguilera *et al's* comes to comparably optimistic conclusions. However, the IEA estimate relies upon a large contribution from EOR that they anticipate will take decades to be realised while some of Aguilera *et al's* assumptions appear questionable.

- In a simple logistic model, increasing the global URR by one billion barrels would delay the date of peak production by only 4.7 days. This result is not substantially changed if a more sophisticated model is used, that allows for varying degrees of asymmetry in the production cycle (Kaufmann and Shiers, 2008). For a range of assumptions about the size of the global URR and the rate of change of production before and after the peak, the date of peak production is found to lie between 2009 and 2031. Delaying the peak beyond 2030 requires optimistic assumptions about the global URR combined with a relatively steep post-peak decline rate and/or slower rates of demand growth than are conventionally assumed. Forecasts that predict no peak before 2030 should be evaluated on this basis.
- Even if the larger URR estimates are correct, it does not necessarily follow that the resource can or will be accessed at the rate required to maintain global production at a particular level. If these resources can only be accessed relatively slowly at high cost, supply constraints could inhibit demand growth. Furthermore, if producers lack the incentive to maximize production, demand growth could be constrained further – especially in the importing countries. Hence, the primary issue for the period to 2030 is the *rate* at which the resource can be accessed and produced.

References

Adamic, L. A. and B. A. Huberman. (2002). "Zipf, power laws and the Internet." *Glottometrics*, **3**, pp. 143-50.

Adelman, M. A. (1991). "Oil Fallacies." *Foreign Policy*, **82**, pp. 3-16.

Adelman, M. A. (1993). *The Economics of Petroleum Supply: Papers, 1962-1993*: MIT Press.

Aguilera, R. F., R. G. Eggert, G. Lagos, and J. E. Tilton. (2009). "Depletion in the future availability of petroleum resources." *Energy Journal*, **30**:1, pp. 141-74.

Ahlbrandt, T. S. (2002). "Future petroleum energy resources of the world." *International Geology Review*, **44**:12, pp. 1092-104.

Ahlbrandt, T. S. and T. R. Klett. (2005). "Comparison of Methods Used to Estimate Conventional Undiscovered Petroleum Resources: World Examples." *Natural Resources Research*, **14**:3, pp. 187-210.

Ahlbrandt, T. S. and P. McCabe. (2002). "Global petroleum resources: a view to the future." *Geotimes*.

Ahmed, T. H. (2006). *Reservoir Engineering Handbook*: Gulf Professional Publishing.

Andrews, S. and R. Udall. (2003). "Oil prophets: looking at world oil studies over time." *ASPO Conference*: Paris, France.

Arps, J. J. (1945). "Analysis of decline curves." *Trans AIME*, pp. 160-247.

Arps, J. J. (1956). "Estimation of primary oil reserves." *Trans AIME*, **207**, pp. 182-91.

Arps, J. J., M. Mortada, and A. E. Smith. (1971). "Relationship Between Proved Reserves and Exploratory Effort." *Journal of Petroleum Technology*, pp. 671–75

Arps, J. J. and T. J. Roberts. (1958). "Economics of drilling for Cretaceous oil on the east flank of Denver-Julesberg basin " *AAPG Bulletin*, **42**:11, pp. 2549-66.

Arrington, J. R. (1960). "Predicting the size of crude reserves is key to evaluating exploration programs." *Oil and Gas Journal*, **58**:9, pp. 130-34.

Attanasi, E. D. and R. R. Charpentier. (2002). "Comparison of two probability distributions used to model sizes of undiscovered oil and gas accumulations-does the tail wag the assessment?" *Journal of mathematical Geology*, **34**:6, pp. 767-77.

Attanasi, E. D. and L. J. Drew. (1984). "Lognormal field size distributions as a consequence of economic truncation." *Mathematical Geology*, **17**:4, pp. 335-51.

Attanasi, E. D., L. J. Drew, and D. H. Root. (1981). "Physical variables and the petroleum discovery process," in J. Ramsay ed. *The economics of exploration for energy resources*. Greenwich JAI Press.

Attanasi, E. D., R. F. Mast, and D. H. Root. (1999). "Oil, gas field growth projections: Wishful thinking or reality?" *Oil and Gas Journal*, **97**:14, pp. 3.

Attanasi, E. D. and D. H. Root. (1994). "The enigma of oil and gas field growth." *AAPG Bulletin*, **78**:3, pp. 321-32.

Attansi, E. D. and D. H. Root. (1994). "The enigma of oil and gas field growth." *AAPG Bulletin*, **78**:3, pp. 321-32.

Ausloos, M. and M. Dirickx. (2005). *The logistic map and the route to chaos: from the beginnings to modern applications.* Berlin; London: Springer.

Barker, R. O., E. S. W. Jong, and J. G. Virues. (2004). "Understanding the enigma of reserves growth: the whys." *Canadian International Petroleum Conference*: Calgary, Alberta, Canada.

Barouch, E. and G. M. Kaufman. (1975). "Probablistic modelling of oil and gas discovery." *Energy: mathematics and models* SIAM Institute for Mathematics and Society, pp. 133-52.

Barouch, E. and G. M. Kaufman. (1978). "The interface between geostatistical modelling of oil and gas discovery and economics." *Mathematical Geology*, **10**:5, pp. 611-26.

Bartlett, A. A. (2000). "An analysis of US and world oil production patterns using Hubbert-style curves." *Mathematical Geology*, **32**:1, pp. 1-17.

Barton, C. C. and C. H. Scholz. (1995). "The fractal size and spatial distribution of hydrocarbon accumulation's," in C. C. Barton and P.R. La Pointe eds. *Fractals in Petroleum Geology and Earth Processes Springer.* Springer.

Bass, F. M. (1969). "A new product growth model for consumer durables." *Management Science*, **15**, pp. 215-27.

Bates, D. M. and D. G. Watts. (2007). *Nonlinear regression analysis and its applications*. New York: Chichester: Wiley.

Baxter, G. G., A. H. C. Cargill, P. E. Hart, G. M. Kaufmann, and F. Urquidi-Barrau. (1978). "Workshop on the Delphi method." *Journal of the International Association for Mathematical Geology*, **10**:5, pp. 581-87.

Bentley, R. W. (2002). "Global oil and gas depletion: An overview." *Energy Policy*, **30**:3, pp. 189-205.

Bentley, R. W. (2009). "A primer on oil peaking." Background note for the UKERC study on global oil depletion. UK Energy Research Centre: London.

Bentley, R. W., R. H. Booth, J. D. Burton, M. L. Coleman, B. W. Sellwood, and G. R. Whitfield. (2000). "Perspectives on future of oil." *Energy Exploration & Exploitation,* **18**:2-3, pp. 147-206.

Bentley, R. W., S. A. Mannan, and S. J. Wheeler. (2007). "Assessing the date of the global oil peak: the need to use 2P reserves." *Energy Policy*, **35**:12, pp. 6364-82.

Bentley, R. W., R. G. Miller, S. Wheeler, and G. Boyle. (2009). "Comparison of global oil supply forecasts." UKERC Review of Evidence on Global Oil Depletion: Technical Report 7. UK Energy Research Centre: London.

Bloomfield, P., M. Hudson, K. Kim, and G. S. Watsion. (1979). "The use of the lognormal distribution in estimating crude oil resources." Resource Estimation and Validation Project, Department of Statistics and Geology, Princeton University: Princeton, NJ.

Bowden, G. (1985). "The Social Construction of Validity in Estimates of US Crude Oil Reserves." *Social Studies of Science*, **15**:2, pp. 207-40.

BP. (2008). "Statistical Review of World Energy 2008 ". BP plc: London.

Brandt, A. R. (2007). "Testing Hubbert." *Energy Policy*, **35**:5, pp. 3074-88.

Brecha, R. J. (2008). "Emission scenarios in the face of fossil-fuel peaking." *Energy Policy*, **36**:9, pp. 3492-504.

Brecha, R. J., R. Berney, and B. Craver. (2007). "Revisiting Hafemeister's science and society tests." *American Journal of Physics*, **75**:10, pp. 916-30

Byrd, L. T., R. M. Kumar, A. F. Williams, and D. L. Moore. (1985). "US oil and gas finding and development costs 1973-1982, lower 48 onshore and offshore." *Journal of Petroleum Technology*, **31**, pp. 2040-48.

Caithamer, P. (2008). "Regression and time series analysis of the world oil peak of production: another look." *Journal of Mathematical Geosciences*, **40**:6, pp. 1874-.

Campbell, C. and J. Laherrère. (1995). "The world's supply of oil 1930-2050." Petroconsultants S.A.: Geneva.

Campbell, C. J. (1991). *The golden century of oil 1950:2050: the depletion of a resource*: Springer.

Campbell, C. J. (1996). "Status of world oil depletion at the end of 1995." *Energy Exploration & Exploitation*, **14**:1, pp. 63-81.

Campbell, C. J. (1997). *The Coming Oil Crisis*: Multi-science publishing & Petroconsultants.

Campbell, C. J. (2002). "The assessment and importance of oil depletion." *Energy Exploration & Exploitation*, **20-1**:6-1, pp. 407-35.

Campbell, C. J. and S. Heapes. (2008). *An atlas of oil and gas depletion*. Huddersfield: Jeremy Mills Publishing.

Campbell, C. J. and J. H. Laherrere. (1998). "The End of Cheap Oil." *Scientific American.*, **278**:3, pp. 60.

Capen, E. (1976). "The difficulty of assessing uncertainty." *Journal of Petroleum Technology*, **28**, pp. 843-50.

Carlson, W. B. (2007a). "Analysis of World Oil Production Based on the Fitting of the Logistic Function and its Derivatives." *Energy Sources Part B: Economics, Planning and Policy*, **2**:4, pp. 421-28.

Carlson, W. B. (2007b). "Sensitivity of Predicted Oil Production to the Sigmoid Function." *Energy Sources Part B: Economics, Planning and Policy*, **2**:4, pp. 321-27.

Cavallo, A. J. (2004). "Hubbert's Petroleum Production Model: An Evaluation and Implications for World Oil Production Forecasts." *Natural Resources Research*, **13**:4, pp. 211-21.

Cavallo, A. J. (2005a). "Hubbert's model: Uses, meanings, and limits-1." *Oil and Gas Journal*, **103**:21, pp. 22-24.

Cavallo, A. J. (2005b). "Hubbert's model: Uses, meanings, and limits - 2." *Oil and Gas Journal*, **103**:22, pp. 20-25.

Chang, C.-P. and Z.-S. Lin. (1999). "Stochastic analysis of production decline data for production prediction and reserves estimation." *Journal of Petroleum Science and Engineering*, **23**:3-4, pp. 149-60.

Charpentier, R. L. (2005). "Guiding principles of USGS methodology for assessment of undiscovered conventional oil and gas resources." *Natural Resources Research*, **14**:3, pp. 175-86.

Charpentier, R. L., G. L. Dolton, and G. F. Ulmishek. (1995a). "Annotated bibliography of methodology for an assessment of undiscovered oil and gas resources." *Nonrenewable Resources*, **4**:2, pp. 154-86.

Charpentier, R. R. (2003). "The future of petroleum: optimism, pessimism, or something else?" USGS Open File Report 03-137. US Geological Survey.

Charpentier, R. R., G. L. Dolton, and G. F. Ulmishek. (1995b). "Annotated bibliography of methodology for assessment of undiscovered oil and gas resources." *Natural Resources Research*, **4**:2, pp. 154-86.

Chaudhry, A. U. (2003). "Application of Decline Curve Analysis Methods." *Gas Well Testing Handbook*. Burlington Gulf Professional Publishing, pp. 637-63.

Cleveland, C. J. (1991). "Physical and economic aspects of resource quality: the cost of oil supply in the lower 48 United States, 1936-1988." *Resources and Energy*, **13**:2, pp. 163-88.

Cleveland, C. J. (1992a). "Energy surplus and energy quality in the extraction of fossil fuels in the U.S." *Ecological Economics*, **6**:2, pp. 139-62.

Cleveland, C. J. (1992b). "Yield per effort for additions to crude-oil reserves in the lower 48 united-states, 1946-1989." *AAPG Bulletin*, **76**:6, pp. 948-58.

Cleveland, C. J. and R. K. Kaufmann. (1991). "Forecasting ultimate oil recovery and its rate of production: incorporating economic forces into the models of M. King Hubbert." *The Energy Journal*, **12**:2, pp. 17-46.

Cleveland, C. J. and R. K. Kaufmann. (1997). "Natural gas in the US: how far can technology stretch the resource base." *The Energy Journal*, **18**:2, pp. 89-108.

Coustau, H. (1979). "Logique de distribution des tailles des champts dans les bassins." *Petrole et Techniques*, **May**.

Cramer Barton, C. and P. R. La Pointe. (1995). *Fractals in Petroleum Geology and Earth Processes* Springer.

Cuddington, J. T. and D. L. Moss. (2001). "Technical change, depletion and the US petroleum industry." *American Economic Review*, **914**, pp. 1135–48.

Davies, A. W. (1981). "Discussion of the paper by Mr Meisner and Mr Demirmen." *Journal of the Royal Statistical Society*, **144**:1, pp. 24-25.

Davies, J. C. and T. Chang. (1989). "Estimating potential for small fields are mature petroleum provinces." *AAPG Bulletin*, **73**:8, pp. 967-76.

Deffeyes, K. S. (2003). "Hubbert‘s Peak: The Impending World Oil Shortage." Princeton University Press.

Deffeyes, K. S. (2005). *Beyond oil; the view from Hubbert's peak*. United States (USA): Hill and Wang, New York, NY, United States (USA).

Divi, R. S. (2004). "Probabilistic methods in petroleum resource assessment, with some examples using data from the Arabian region." *Journal of Petroleum Science and Engineering*, **42**:2-4, pp. 95-106.

Dow, W. G. (1972). "Application of oil correlation and source rock data to exploration in Williston basin." *AAPG Bulletin*, **56**, pp. 615.

Drew, L. J. (1997). *Undiscovered Petroleum and Mineral Resources: assessment and controversy* Springer.

Drew, L. J., E. D. Attanasi, and J. H. Schuenemeyer. (1988). "Observed oil and gas field size distributions: a consequence of the discovery process and the prices of oil and gas." *Mathematical Geology*, **20**:20, pp. 939-53.

Drew, L. J. and J. C. Griffiths. (1965). "Size, shape and arrangement of some oilfields in the USA." *1965 symposium, Computer applications in the mineral industries*: FF-1-31: Pennsylvania State University.

Drew, L. J. and J. H. Schuenemeyer. (1992). "A petroleum discovery rate forecast revisited-the problem of field growth." *Nonrenewable Resources*, **1**:1, pp. 51-60.

Drew, L. J. and J. H. Schuenemeyer. (1993). "The evolution and use of discovery process models at the US Geological Survey." *AAPG Bulletin*, **77**:3, pp. 467-78.

Drew, L. J., J. H. Schuenemeyer, and W. J. Bawiec. (1982a). "Estimation of the future rate of oil and gas discoveries in the Gulf of Mexico." Professional Paper 1252. US Geological Survey.

Drew, L. J., J. H. Schuenemeyer, and W. J. Bawiec. (1982b). "Estimation of the future rates of discovery of oil and gas in the Gulf of Mexico." Professional Paper 1252. US Geological Survey.

Drew, L. J., J. H. Schuenemeyer, and R. F. Mast. (1995). "Application of the modified Arps-Roberts discovery proces model to the 1995 U.S. National Oil and Gas Assessment " *Natural Resources Research*, **4**:3, pp. 242-52.

EIA. (2008). "International Energy Outlook 2008 " DOE/EIA-0484(2008). Energy Information Administration: Washington, DC, USA.

Energy information Administration. (1990). "US oil and gas reserves by year of field discovery." DOE/EIA-0534.

Engle, R. F. and C. W. J. Granger. (1997). "Cointegration and error correction: representation, estimation and testing." *Econometricaa*, **55**, pp. 251-76.

Forbes, K. F. and E. M. Zampelli. (2000). "Technology and the exploratory success rate in the U.S. offshore." *Energy Journal*, **21**:1.

Forman, D. and A. L. Hinde. (1985). "Improved statistical method for assessment of undiscovered petroleum resources." *AAPG Bulletin*, **69**:1, pp. 106-18.

Gately, D. (1992). "Imperfect price reversed ability of US gasoline demand: asymmetric responses to price increases and declines." *Energy Journal*, **13**:4, pp. 179-207.

Gautier, D. L. (2004). "Oil and natural gas resource assessment: geological methods," in C.J. Cleveland ed. *The Encyclopedia of Energy*. Cheltenham Edward Elgar.

Gautier, D. L., G. L. Dolton, K. I. Takahashi, and K. L. Varnes. (1995). "National assessment of United States oil and gas resources-results, methodology and supporting data." US Geological Survey Digital Data Series DDS-30.

Gautier, D. L. and T. R. Klett. (2005). "Reserve growth in chalk fields of the North Sea; Petroleum geology; north-west Europe and global perspectives; proceedings of the 6th Petroleum geology conference." *6th Petroleum geology conference, London, United Kingdom, Oct. 6-9, 2003*, **6**, pp. 169-75.

Gautier, D. L., T. R. Klett, and B. S. Pierce. (2005). "Global significance of reserve growth." 25. U. S. Geological Survey, Reston, VA, United States (USA): United States (USA).

Gess, G. and C. Bois. (1977). "Study of petroleum zones - a contribution to the appraisal of hydrocarbon resources," in R.F. Meyer ed. *The future of supply of nature-made petroleum and gas*. New York Pergamon Press, pp. 155-78.

Gompertz, B. (1825). "On the nature of the function expressive of the law of human mortality; and on a new mode of determining the value of life contingencies." In a letter to Francis Baily, Printed by W. Nicol: London.

Gowdy, J. and J. Roxana. (2007). "Technology and petroleum exhaustion: Evidence from two mega-oilfields." *Energy*, **32**:8, pp. 1448-54.

Grubb, M. J. (1995). "Asymmetrical price elasticities of energy demand," in T. Barker, P. Ekins and N. Johnstone eds. *Global warming and energy demand*. London and New York Routledge.

Harbaugh, J. W., J. C. Davies, and J. Wendebourg. (1995). *Computing risk for oil prospects*. Oxford: Pergamon Press.

Harris, d. P. (1977). "Conventional crude oil resources of the United States: recent estimates, method for estimation and policy considerations." *Materials and Society*, **1**, pp. 263-86.

Haun, J. D. (1981). "Future of Petroleum exploration in the United States." *AAPG Bulletin*, **65**:1720-1727.

Hendricks, T. A. (1965). "Resources of oil, gas and natural gas liquids in the United States and the world." US Geological Survey Circular 522.

Herbert, J. H. (1982). "A Review and Comparison of Some Commonly Used Methods of Estimating Petroleum Resource Availability." *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, **6**:4, pp. 293-320.

Hewett, D. F. (1929). "Cycles in metal production." Technical Publication 183. 31. The American Institute of Mining and Metalurgical Engineers.

Hirsch, R. L. (2008). "Mitigation of maximum world oil production: Shortage scenarios." *Energy Policy*, **36**:2, pp. 881-89.

Hirsch, R. L., R. Bezdek, and R. Wendling. (2005). "Peaking of world oil production: impacts, mitigation, & risk management." 91. US Department of Energy, National Energy Technologies Laboratory.

Höök, M., B. Söderbergh, K. Jakobsson, and K. Aleklett. (2009). "The evolution of giant oil field production behaviour." *Natural Resources Research*, **18**:1, pp. 39-56.

Houghton, J. C., G. L. Dolton, R. F. Mast, C. D. Masters, and D. H. Root. (1993). "US Geological Survey estimation procedure for accumulation size distributions by play." *AAPG Bulletin*, **77**:454-486.

Hubbert, M. K. (1956). "Nuclear energy and the fossil fuels." *Meeting of the Southern Dsitrict, Division of production, American Petroleum Institute*. Shell Development Company: San Antonio, Texas.

Hubbert, M. K. (1959). "Techniques of prediction with application to the petroleum industry." *44th Annual meeting of the American Association of Petroleum Geologists*: 43. Shell Development Company: Dallas, TX.

Hubbert, M. K. (1962). "Energy Resources: A Report to the Committee on Natural Resources of the National Academy of Sciences-National Research Council." 153p. United States.

Hubbert, M. K. (1966). "Reply to J.M. Ryan." *AAPG Bulletin*, **18**, pp. 284-86.

Hubbert, M. K. (1967). "Degree of Advancement of Petroleum Exploration in United States." *AAPG Bulletin*, **51**, pp. 2207-27.

Hubbert, M. K. (1974). "U.S. Energy Resources: A Review as of 1972 " Background paper prepared for the U.S. Senate Committee on Interior and Insular Affairs, 93rd Congress, 2nd Session, Serial 93-40 (92-75). Washington: U.S. Government Printing Office.

Hubbert, M. K. (1982). "Techniques of prediction as applied to the production of oil and gas; Oil and gas supply modeling; proceedings of a symposium." *Oil and gas supply modeling, Washington, DC, United States, June 18-20, 1980*:631, pp. 16-141.

IEA. (2008). "World Energy Outlook 2008." International Energy Agency, OECD: Paris.

Iledare, O. and A. G. Pulsipher. (1999). "Sources of change in productivity of petroleum exploration and development in the U.S. onshore Louisiana, 1977-1994." *Energy Economics* **21**:3, pp. 261-72.

Imam, A., R. A. Startzman, and M. A. Barrufet. (2004). "Multicyclic Hubbert model shows global conventional gas output peaking in 2019." *Oil & Gas Journal*, **102**:31.

IPCC. (2007). "Climate Change 2007: Synthesis Report." Fourth Assessment Report (AR4). Intergovernmental Panel on Climate Change.

Ivanhoe, L. F. and G. G. Leckie. (1993). "Global oil, gas fields, sizes tallied, analyzed." *Oil & Gas Journal*, **1993**:February 15, pp. 87-91.

Jones, R. W. (1975). "A quantitative geologic approach to prediction of petroleum resources," in J.D. Haun ed. *Methods of Estimating the Volume of Undiscovered Oil and Gas Resources* American Association of Petroleum Geologists, Studies in Geology No. 1.

Kaufman, G. M. (1963). *Statistical decision and related techniques in oil and gas exploration*. Englewood Cliffs, New Jersey: Prentice-Hall.

Kaufman, G. M. (1975a). "Models and methods for estimating undiscovered oil and gas: what they do when do not do." *First IIASA conference on Energy Resources*: 237-49. International Institute for Applied Systems Analysis: Laxenburg, Austria.

Kaufman, G. M. (1975b). "A probabilistic model of oil and gas discovery," in J.D. Haun ed. *Methods of Estimating the Volume of Undiscovered Oil and Gas Resources: AAPG Studies in Geology Series 1*, pp. 113-42.

Kaufman, G. M. (1993). "Statistical issues in the assessment of undiscovered oil and gas resources." *Energy Journal*, **14**:1, pp. 183-213.

Kaufman, G. M. (2005). "Where have we been? Where are we going?" *Natural Resources Research*, **14**:3, pp. 145-51.

Kaufmann, R. K. (1991). "Oil production in the lower 48 states : Reconciling curve fitting and econometric models." *Resources and Energy*, **13**:1, pp. 111-27.

Kaufmann, R. K. and C. J. Cleveland. (1991). "Policies to Increase us Oil Production: Likely to Fail, Damage the Economy, and Damage the Environment." *Annual Review of Energy and the Environment*, **16**, pp. 379-400.

Kaufmann, R. K. and C. J. Cleveland. (2001). "Oil production in the lower 48 states: Economic, geological, and institutional determinants." *Energy Journal*, **22**:1, pp. 27-49.

Kaufmann, R. K. and L. D. Shiers. (2008). "Alternatives to conventional crude oil: When, how quickly, and market driven." *Ecological Economics*, **67**:3, pp. 405-11.

Kemp, A. G. and A. S. Kasim. (2005). "Are decline rates really exponential? Evidence from the UK continental shelf." *The Energy Journal*, **26**:1, pp. 27-50.

Kennedy, P. (2003). *A Guide to Econometrics*. Oxford: Blackwell.

Klemme, H. D. (1984). "Field size distribution related to basin characteristics," in C.D. Masters ed. *Petroleum resource Assessment* International Union of Geological Sciences, pp. 95-121.

Klett, T. R. (2004). "Oil and natural gas resource assessment: classifications and terminology," in C.J. Cleveland ed. *Encyclopaedia of Energy* Elsevier, pp. 595-605.

Klett, T. R. (2005). "United States Geological Survey's reserve-growth models and their implementation." *Natural Resources Research*, **14**:3, pp. 249-64.

Klett, T. R. and D. L. Gautier. (2005). "Reserve growth in oil fields of the North Sea." *Petroleum Geoscience*, **11**:2, pp. 179-90.

Klett, T. R., D. L. Gautier, and T. S. Ahlbrandt. (2005a). "An evaluation of the U. S. Geological Survey world petroleum assessment 2000." *AAPG Bulletin*, **89**:8, pp. 1033-42.

Klett, T. R., D. L. Gautier, and T. S. Ahlbrandt. (2005b). "An evaluation of the US Geological Survey World Petroleum Assessment 2000." *AAPG Bulletin*, **89**:8, pp. 1033-42.

Klett, T. R., D. L. Gautier, and T. S. Ahlbrant. (2007). "An evaluation of the World Petroleum Assessment 2000: supporting data." USGS Open File a report 2007-1021. US Geological Survey.

Klett, T. R. and J. W. Schmoker. (2003). "Reserve growth of the world's giant oil fields; Giant oil and gas fields of the decade, 1990-1999." *AAPG Memoir*, **78**, pp. 107-22.

Klett, T. R. and M. K. Verma. (2004). "Reserve growth in oil and gas fields of Mexico (abst.): Am." *Assoc. Petroleum Geologists Intern. Conf. and Exhibition (Cancun, Mexico)*.

La Pointe, P. R. (1995). "Estimation of undiscovered hydrocarbon potential through fractal geometry," in C. C. Barton and P.R. La Pointe eds. *Fractals in Petroleum Geology and Earth Processes Springer.* Springer.

Laherrere, J. (2003). "Future of oil supplies." *Energy exploration & exploitation.*, **21**, pp. 227-68.

Laherrère, J. (1996). "Distributions de type “fractal parabolique” dans la Nature " *Comptes Rendus de l’Académie des Sciences*, **2a**:7.

Laherrère, J. (2000a). "Distribution of field sizes in a petroleum system; parabolic fractal, lognormal or stretched exponential?" *Marine and Petroleum Geology*, **17**:4, pp. 539-46.

Laherrère, J. (2001a). "Estimates of Oil Reserves." *Paper presented at the EMF/IEA/IEW meeting*.

Laherrère, J. (2001b). "Is the USGS 2000 Assessment Reliable?" published on the cyberconference of the WEC.

Laherrère, J. (2002a). "Forecasting future production from past discovery." *International Journal of Global Energy Issues*, **18**:2-4, pp. 218-38.

Laherrère, J. (2003). "Future of oil supplies." *Energy Exploration and Exploitation*, **21**:3, pp. 227-67.

Laherrère, J. (2004). "Oil and natural gas resource assessment: production growth cycle models," in C.J. Cleveland ed. *Encyclopaedia of Energy*. Amsterdam Elsevier.

Laherrère, J. H. (1997). "Production decline and peak reveal true reserve figures." *World Oil*, **218**:12, pp. 77.

Laherrère, J. H. (1999a). "Reserve Growth: Technological Progress, or Bad Reporting and Bad Arithmetic?" *Geopolitics of Energy*, **22**, pp. 7-16.

Laherrère, J. H. (1999b). "World oil supply - what goes up must come down, but when will it peak?" *Oil and Gas Journal*, **97**:5, pp. 57-64.

Laherrère, J. H. (2000b). "Learn strengths, weaknesses to understand Hubbert curve." *Oil and Gas Journal*, **98**:16, pp. 63-64.

Laherrère, J. H. (2002b). "Modelling future liquids production from extrapolation of the past and from ultimates." *Energy Exploration & Exploitation*, **20**:6, pp. 457-79.

Laherrère, J. H. (2005). "Forecasting production from discovery." *ASPO IV International Workshop on Oil and Gas Depletion*. Association for the Study of Peak Oil: Lisbon.

Lee, P. J. (2008). *Statistical Methods for Estimating Petroleum Resources*: Oxford University Press.

Lee, P. J. and P. C. C. Wang. (1983). "Probabilistic formulation of a method for the evaluation of petroleum resources." *Journal of the International Association of Mathematical Geology,* **15**:1, pp. 163-82.

Lee, P. J. and P. C. C. Wang. (1985). "Prediction of oil and gas pools sizes when the discovery record is available." *Mathematical Geology*, **17**:2, pp. 95-113.

Lee, P. J. and P. C. C. Wang. (1986). "Evaluation of petroleum resources from pool size distributions," in D.D. Rice ed. *Oil and gas and assessment: methods and applications* American Association of Petroleum Geologists, Studies in Geology, No. 21, pp. 33-4 to.

Li, K. and R. N. Horne. (2007). "Comparison and verification of production prediction models." *Journal of Petroleum Science and Engineering*, **55**, pp. 213-20.

Lotka, A. J. (1925). *Elements of Physical Biology*. Baltimore, MD, USA: Williams and Wilkins.

Lynch, M. C. (1998). "Crying Wolf: Warnings about oil supply ". MIT: Boston.

Lynch, M. C. (1999). "The wolf at the door or crying wolf: Fears about the next oil crisis." *Advances in the Economics of Energy and Resources*, **11**, pp. 117-42.

Lynch, M. C. (2002). "Forecasting oil supply: theory and practice." *The Quarterly Review of Economics and Finance*, **42**:2, pp. 373-89.

Magoon, L. B. and R. M. O. Sanchez. (1995). "Beyond the petroleum system." *AAPG Bulletin*, **79**:12, pp. 1731-36.

Managi, S. (2002). "Technological change, depletion and environmental policy in offshore oil and gas industry." University of Rhode Island.

Managi, S., J. J. Opaluch, D. Jin, and T. A. Grigalunas. (2004). "Technological change and depletion in offshore oil and gas " *Journal of Environmental Economics and Management*,, **47**, pp. 388-409.

Managi, S., J. J. Opaluchb, D. Jinc, and T. A. Grigalunas. (2005). "Technological change and petroleum exploration in the Gulf of Mexico " *Energy Policy*, **33**:5, pp. 619-32.

Mandlebrot, B. B. (1962). "Statistics of natural resources and the law of Pareto." International Business Machines Research Note NC-146, June 29, reprinted in C.C Barton and P.R. La Pointe (eds) Fractals in Petroleum Geology and Earth Processes, Plenum Press, New York and London

Mandlebrot, B. B. (1977). *Fractals: form, chance and dimension*. San Francisco: W.H. Freeman.

Mast, R. F. e. a. (1989). "Estimates of undiscovered conventional oil and gas resources in the United States." US Geological Survey and US Minerals Management Service.

McKelvey, V. E. (1972). "Mineral resource estimates and public policy." *American Scientist*, **60**, pp. 32-40.

Mcrossan, R. G. (1969). "An analysis of size frequency distribution of oil and gas reserves of Western Canada." *Canadian Journal of Earth Sciences*, **6**:201, pp. 201-11.

Meade, N. M. (1984). "The use of growth curves in forecasting market development - a review and appraisal." *Journal of Forecasting*, **3**, pp. 49-451.

Meisner, J. and F. Demirmen. (1981). "The creaming method: a procedure to forecast future oil and gas discoveries in mature exploration provinces." *Journal of the Royal Statistical Society A*, **144**:1, pp. 1-31.

Menard, H. W. and G. Sharman. (1975). "Scientific uses of random drilling models." *Science*, **190**, pp. 337.

Merriam, D. F., L. J. Drew, and J. H. Schuenemeyer. (2004). "Zipf's law: a viable geological paradigm?" *Natural Resources Research*, **13**:4.

Meyer, P. S. (1994). "Bi-logistic growth." *Technological Forecasting and Social Change*, **47**, pp. 89-102.

Meyer, P. S., J. W. Yung, and J. H. Ausbel. (1999). "A primer on logistic growth and substitution: the mathematics of the loglet lab software." *Technological Forecasting Social Change*, **61**, pp. 247-71.

Meynerd, H. W. and G. Shaman. (1975). "Scientific Uses of Random Drilling Models." *Science*, **190**, pp. 337–43.

Miller, R. G. (1992). "The global oil system: the relationship between all generation, loss, half life and the world crude oil resource." *AAPG Bulletin*, **76**:4, pp. 489-500.

Miller, R. G., S. Sorrell, and J. Speirs. (2009). "Decline rates and depletion rates." UKERC Review of evidence on Global Oil Depletion: Technical Report 4. UK Energy Research Centre: London.

Mills, R. M. (2008). *The myth of the oil crisis*. Westport, CT, USA: Praeger Publishers.

Minerals Management Service. (1996). "An assessment of the undiscovered hydrocarbon potential of the nation's outer continental shelf." Report MMS 96-0034.

Mohr, S. and G. Evans. (2008). "Peak Oil: Testing Hubbert's Curve via Theoretical Modeling." *Natural Resources Research*, **17**:1, pp. 1-11.

Mohr, S. H. and G. M. Evans. (2007). "Mathematical model forecasts year conventional oil will peak." *Oil and Gas Journal*, **105**:17, pp. 45-46.

Moore, C. L. (1962). "Method for evaluating US crude oil resources and projecting domestic crude oil availability." US Department of the Interior.

Morehouse, D. F. (1997). "The intricate puzzle of oil and gas "reserve growth": Natural Gas Monthly, July 1997." *Energy Information Administration*, **20**.

Motulsky, H. and A. Christopoulos. (2004a). *Fitting models to biological data using linear and nonlinear regression : a practical guide to curve fitting*. Oxford; New York: Oxford University Press.

Motulsky, H. and A. Christopoulos. (2004b). *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting*. New York: Oxford University Press.

Muller, K. and A. Sturm. (2000). "Standardised eco-efficiency indicators." Ellipson AG: Basel.

Myers, R. H., D. C. Montgomery, and G. G. Vining. (2002). *Generalised Linear Models: with Applications in Engineering and the Sciences*. New York: John Wiley & Sons.

Nehring, R. (2006a). "How Hubbert method fails to predict oil production in the Permian Basin." *Oil and Gas Journal*, **104**:15, pp. 30-35.

Nehring, R. (2006b). "Post-Hubbert challenge is to find new methods to predict production, EUR." *Oil and Gas Journal*, **104**:16, pp. 43-46.

Nehring, R. (2006c). "Post Hubbert challenge is to find new ways to predict production, EUR." *Oil and Gas Journal*, **104**:April 24, pp. 43-51.

Nehring, R. (2006d). "Two basins show Hubbert's method underestimates future oil production." *Oil and Gas Journal*, **104**:13, pp. 37-42.

Nehring, R. (2006e). "Two basins shows Hubbert's method underestimates future oil production." *Oil and Gas Journal*, **104**:April 3, pp. 37-44.

Nelder, J. A. (1971). "The fitting of a generalisation of the logistics curve." *Biometrics*, **17**, pp. 89-110.

Nickerson, D. M. and B. C. Madsen. (2004). "Nonlinear regression and ARIMA models for precipitation chemistry in East Central Florida from 1978 to 1997." *Environmental Pollution*, **135**:3, pp. 371-79.

NPC. (1995). "Research development and demonstration needs of the oil and gas industry." National Petroleum Council: Washington, DC.

NPC. (2007). "Topic paper #10: Geologic Endowment:." Working document of the NPC Global Oil and Gas Study. National Petroleum Council: Washington, DC.

Odell, P. R. (1973a). "The future of oil: a rejoinder." *The Geographical Journal*, **139**:3, pp. 436-54.

Odell, P. R. (1973b). "The future of oil: a rejoinder." *Geographical Journal*, **139**:3, pp. 436-54.

Odell, P. R. (1973c). "The Future of Oil: A Rejoinder." *The Geographical Journal*, **139**:3, pp. 436-54.

Odell, P. R. (2004). *Why carbon fuels will dominate the 21st century's global energy economy*. Brentwood, England: Multi-Science Pub. Co.

Odell, P. R. and K. E. Rosing. (1980a). *The future of oil : a simulation study of the inter-relationships of resources, reserves, and use, 1980-2080*. London; New York: Kogan Page ; Nichols Pub. Co.

Odell, P. R. and K. E. Rosing. (1980b). *The future of oil : a simulation study of the interrelationships of resources, reserves, and use, 1980-2080*. New York: Nichols Publishing Co.

Pareto, V. (1987). *Cours d'Economie Politique*. paris: Roque and Cie.

Patzek, T. W. (2008). "Exponential growth, energetic Hubbert cycles and the advancement of technology." *Archives of mining sciences of the Polish Academy of Sciences*, **22**:Accepted for publication May 3, 2008.

Perron, P. (1997). "Further evidence on breaking trend functions in macroeconomic variables." *Journal of Econometrics*, **80**, pp. 355-85.

Pesaran, M. H. and H. Samiei. (1995). "Forecasting ultimate resource recovery." *International Journal of Forecasting*, **11**:4, pp. 543-55.

Pike, R. (2006). "Have we underestimated the environmental challenge?" *Petroleum Review*, **June**, pp. 26-27.

Pindyck, R. S. and D. L. Rubinfeld. (1998). *Econometric models and economic forecasts*. Boston, MA: McGraw-Hill

Power, M. (1992). "Lognormality in the observed size distribution of oil and gas pools as a consequence of sampling bias." *Mathematical Geology*, **24**:8.

Power, M. and J. D. Fuller. (1992). "A comparison of models for forecasting the discovery of hydrocarbon deposits." *Journal of Forecasting*, **11**, pp. 183-93.

Pratt, W. E. (1942). *Oil in the earth*: Lawrence, University of Kansas Press.

Rabinowitz, D. (1991). "Using exploration history to estimate undiscovered resources." *Mathematical Geology*, **23**:2, pp. 257-74.

Ramanathan, R. (2002). *Introductory econometrics with applications*. Manson, Ohio, USA: Southwestern.

Reynolds, D. B. (2002). "Using non-time-series to determine supply elasticity: how far do prices change the Hubbert curve?" *OPEC Review*, **26**:2, pp. 147-67.

Robelius, F. (2007). "Giant Oil Fields and Their Importance for Peak Oil." *Licentiate thesis from Uppsala University*.

Rogner, H. H. (1997). "An assessment of world hydrocarbon resources." *Annual Review of Energy and the Environment*, **22**, pp. 217-62.

Root, D. H. and E. D. Attanasi. (1993). "Small fields in the national oil and gas assessment." *AAPG Bulletin*, **77**:3, pp. 485-90.

Root, D. H. and R. F. Mast. (1993). "Future growth in know oil and gas fields." *AAPG Bulletin*, **77**:3, pp. 479-84.

Rosing, K. E. and P. R. Odell. (1984). "Estimating the ultimate economically recoverable hydrocarbon reserves of the North Sea province: phase 1, the theory." *The Statistician*, **33**, pp. 75-90.

Ryan, J. M. (1965). "National Academy report on energy resources: discussion of limitations of logistic projections." *Bulletin of the Association of Petroleum Geologists*, **49**, pp. 1713-27.

Ryan, J. M. (1966). "Limitations of statistical methods for predicting petroleum and natural gas reserves and availability." *Journal of Petroleum Technology*, pp. 231–37.

Ryan, J. T. (1973). "An analysis of crude oil discovery rates in Alberta." *AAPG Bulletin*, **21**:219-235.

Salvador, A. (2005). "Energy: A Historical Perspective and 21st Century Forecast. ." AAPG Studies 54 American Association of Petroleum Geologists.

Schuenemeyer, J. H. and L. J. Drew. (1983). "A procedure to estimate the parent population of the size of oil and gas fields as revealed by a study of economic truncation." *Journal of mathematical Geology*, **15**:1, pp. 145–61.

Schuenemeyer, J. H. and L. J. Drew. (1994). "Description of a discovery process modelling procedure to forecast future oil and gas incorporating field growth." ARDS v 5.01. 98-111. US Geological Survey, Open File Report.

Schuenemeyer, J. H. and L. J. Drew. (2004). "Uncovering Influences on the Form of Oil and Gas Field Size Distributions." *Natural Resources Research*, **8**:1, pp. 37-47.

Sem, T. and D. Ellerman. (1999). "North Sea reserve appreciation, production and depletion." MIT-CEEPR 99-011. Centre for Energy Environmental Policy Research, Massachusetts Institute of Technology: Cambridge.

Simmons, M. R. (2002). "The world's giant oil fields:." Simmons and Company International: Houston, Texas, USA.

Slade, M. (1982). "Trends in natural resource commodity prices: an analysis of the time domain." *Journal of Environmental Economics and Management*,, **9**:2, pp. 132-37.

Smith, F. E. (1963). "Population dynamics in Daphnia magna and a new model for population growth." *Ecology*, **44**, pp. 651-63.

Smith, J. L. (1980). "A probabilistic model of oil discovery." *Review of Economics and Statistics*, **62**, pp. 587–94.

Smith, J. L. and J. L. Paddock. (1984). "Regional modelling of oil discovery and production." *Energy Economics*, **6**:1, pp. 5-13.

Smith, J. L. and G. L. Ward. (1981). "Maximum likelihood estimates of the size distribution of North Sea oil fields " *Journal Mathematical Geology*, **13**:5, pp. 399-413.

Sneddon, J. W., F. Sarg, and X. Ying. (2003). "Exploration play analysis from a sequence stratigraphic perspective." *Search and Discovery*:Article 40079.

Speirs, J. and S. Sorrell. (2009). "Data Sources and Issues." UKERC review on Global Oil Depletion: Technical Report 1. UK Energy Research Centre: London.

Stark, P. H. and K. Chew. (2005). "Global oil resources: issues and implications." *The Journal of Energy Development*, **30**:2, pp. 159-70.

Stitt, W. (1982). "Current problems in oil and gas modelling," in S.I. Gass ed. *Oil and gas supply modelling*. Maryland National Bureau of Standard Special Publication 631, pp. 142-70.

Strahan, D. (2007a). "In praise of the USGS." *Petroleum Review*, **61**:723, pp. 18-19.

Strahan, D. (2007b). "The Last Oil Shock; a survival guide to the imminent extinction of petroleum man." John Murray: London.

Thompson, E. (2008). "Reserve Definitions." Report for the UKERC review of Global Oil Depletion. UK Energy Research Centre: London.

Thompson, E., S. Sorrell, and J. Speirs. (2009a). "Definition and interpretation of reserve estimates." UKERC review on Global Oil Depletion: Technical Report 2. UK Energy Research Centre: London.

Thompson, E., S. Sorrell, and J. Speirs. (2009b). "The nature and importance of reserve growth." UKERC review of evidence on Global Oil Depletion: Technical Report 3 UK Energy Research Centre: London.

Towler, B. F. and S. Bansal. (1993). "Hyperbolic decline-curve analysis using linear regression." *Journal of Petroleum Science and Engineering*, **8**:4, pp. 257-68.

Tsoularis, A. and J. Wallace. (2002). "Analysis of logistic growth models." *Mathematical Biosciences*, **179**:1, pp. 21-55.

USGS. (1995). "National oil and gas assessment." United States Geological Survey Reston, VA, USA.

USGS. (2000a). "U.S. Geological Survey World Petroleum Assessment 2000; description and results." U. S. Geological Survey Reston, VA, USA.

USGS. (2000b). "USGS World Petroleum Assessment 2000: description and results by USGS World Energy Assessment Team." USGS Digital Data Series DDS-60 (four CD-ROM set). U.S. Geological Survey: Reston, VA, USA.

USGS. (2007). "Assessment of undiscovered oil and gas resources of the East Greenland Rift Basins Province." United States Geological Survey: Reston, VA, USA.

USGS. (2008). "Circum-Artic resource appraisal: estimates of undiscovered oil and gas north of the Artic Circle." Factor Sheet 2008-3049. United StatesGeological Survey.

Verma, M. K. (2005). "A new reserve growth model for United States oil and gas fields." *Natural Resources Research (New York, N.Y.)*, **14**:2, pp. 77-89.

Verma, M. K. and G. F. Ulmishek. (2003). "Reserve growth in oil fields of West Siberian Basin, Russia." *Natural Resources Research (International Association for Mathematical Geology)*, **12**:2, pp. 105-19.

Walls, M. A. (1992). "Modeling and forecasting the supply of oil and gas: A survey of existing approaches." *Resources and Energy*, **14**:3, pp. 287-309.

Walls, M. A. (1994). "Using a 'hybrid' approach to model oil and gas supply: a case study of the Gulf of Mexico outer continental shelf." *Land Economics*, **70**:1, pp. 1-19.

Watkins, G. C. (2002). "Characteristics of North Sea oil reserve appreciation." *The Quarterly Review of Economics and Finance*, **42**:2, pp. 335-72.

Weeks, L. G. (1952). "Factors of sedimentary basin development that control oil occurence." *AAPG Bulletin*, **36**, pp. 2017-124.

Weeks, L. G. (1975). "Potential petroleum resources - classification, estimation and status," in J.D. Haun ed. *Methods of Estimating the Volume of Undiscovered Oil and Gas Resources* American Association of Petroleum Geologists, Studies in Geology No. 1.

Wendebourg, J. and C. Lamiraux. (2002). "Estimating the ultimate recoverable reserves of the Paris Basin, France." *Oil and Gas Science and Technology*, **57**:6, pp. 621-29.

White, D. (1920). "The petroleum resources of the world." *Annals of the American Academy of Political and Social Science*, **89**, pp. 111-34.

White, D. A. and H. M. Gehman. (1979). "Methods of estimating oil and gas resources." *AAPG Bulletin*, **63**, pp. 2183-92.

Wiorkowski, J. J. (1981). "Estimating volumes of remaining fossil fuel resources: a critical review " *Journal of the American Statistical Association*, **76**:375, pp. 534-47.

Wood, J. H., G. R. Long, and D. F. Morehouse. (2003). "World conventional oil supply expected to peak in 21st century." *Offshore*, **63**:4, pp. 90, 92, 94, 150.

Woods, T. J. (1985). "Long-term trends in oil and gas discovery rates in lower 48 United States." *The American Association of Petroleum Geologists Bulletin*, **69**:9, pp. 1321-26.

Wooldridge, J. M. (2003). *Introductory Econometrics: a modern approach*. Manson, Ohio: Thompson: Southwestern.

WPC. (2007). "Petroleum Resources Management System." World Petroleum Council.

Zapp, A. D. (1961). "World petroleum resources." *Domestic and world resources of fossil fuels, radioactive minerals and geothermal energy:*.

Zapp, A. D. (1962). "Future petroleum producing capacity of the United States." 1142-H. 36. United States Geological Survey: Washington, D.C.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least-Effort*: Addison-Wesley.