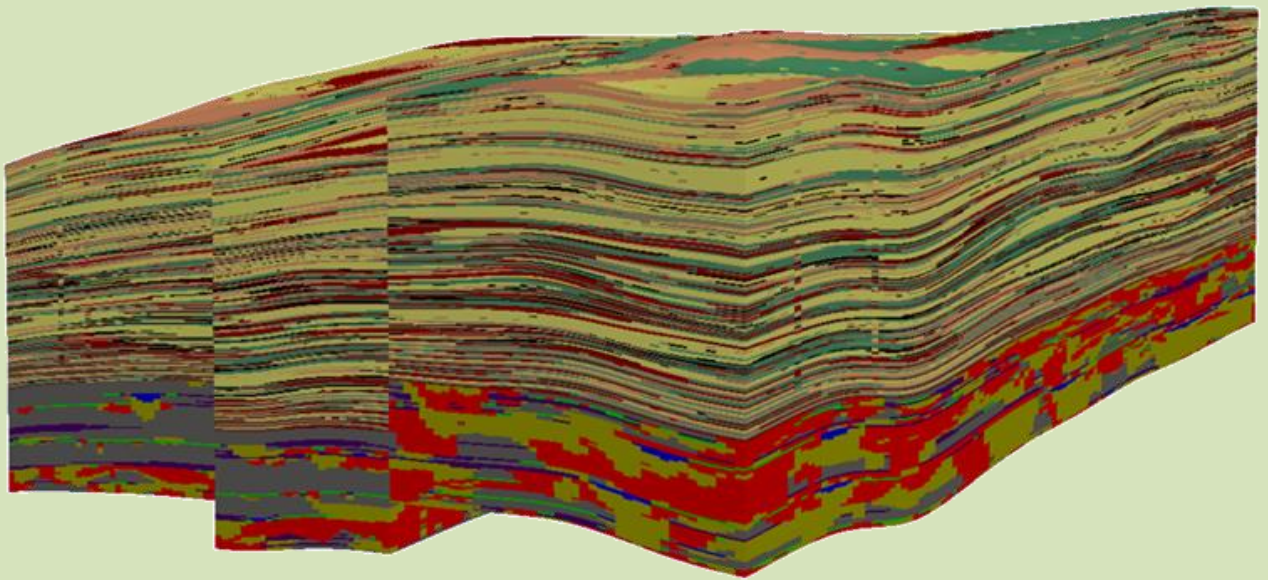


# Calibration and Uncertainty Analysis for Complex Environmental Models



**PEST: complete theory and what it means for modelling the real world**

**John Doherty**

# **Calibration and Uncertainty Analysis for Complex Environmental Models**

**PEST: complete theory and what it means for modelling the real world**

**John Doherty**

Watermark Numerical Computing,  
Brisbane, Australia

This edition first published 2015 by Watermark Numerical Computing

*Registered office* 336 Cliveden Avenue  
Corinda, 4075 Brisbane, Australia  
[www.pesthomepage.org](http://www.pesthomepage.org)

Copyright © 2015 by Watermark Numerical Computing

All rights reserved. This book or any portion thereof may not be reproduced or used without the express written permission of the publisher except for the use of brief quotations in a book review.

# A Message for Your Conscience

This book is very cheap. But it is not free.

If you are reading this book and have not paid for it, then I urge you to do so. This book is the product of many month's work. The story which it tells is the story of PEST, which is the outcome of a lifetime's work. PEST is free. This book is not - for one very practical reason. It is because the author must, like everyone else, earn a living.

This book can be purchased from the PEST web site at:

<http://www.pesthomepage.org>

# Preface

PEST stands for “parameter estimation”. When it was originally written in 1994, this is all that PEST did. However over the 20 years that have elapsed since then, the capabilities of the PEST suite of software have expanded enormously. The emphasis has shifted from inversion to inversion-constrained model parameter and predictive uncertainty analysis.

The purpose of this document is to present, in one place, the theory that underpins PEST and that underpins the plethora of utility software that supports and complements PEST. In doing this, it serves the same purpose for the next generation of PEST-like software. This includes PEST++ and pyEMU. As such, it is hoped that this book provides a valuable resource for those who wish to understand inversion and inversion-constrained uncertainty analysis as it pertains to environmental models. These type of models include (but are not limited to) petroleum and geothermal reservoir models, groundwater models and surface water hydraulic and hydrologic models.

As well as presenting theory, an equally important role of this book is to draw some important conclusions from the theory. These conclusions pertain to real world model usage. Some of them question modelling practices that are commonplace today. Others suggest roles for models in environmental management beyond those which they presently play.

Although a basic knowledge of matrix algebra, and of statistical and geostatistical concepts, is required in order to follow the discussion provided herein, mathematical prerequisites are not high. Most readers will have acquired the knowledge to understand the following theory at school or through undergraduate university courses. The early part of the book reminds readers of a few basic linear analysis and statistical concepts that they may have forgotten. Refer to the many good books on these topics for further details.

# Table of Contents

1. Introduction.....	1
1.1 PEST: The Past .....	1
1.2 PEST: The Future.....	3
1.3 The Present Document .....	3
1.4 A Philosophical Note .....	3
2. Vectors and Matrices .....	6
2.1 Introduction.....	6
2.2 Definitions.....	6
2.3 Addition and Scalar Multiplication of Matrices.....	7
2.4 Multiplication of Matrices .....	8
2.5 Partitioning of a Matrix .....	9
2.6 The Inverse of a Matrix.....	10
2.7 Determinant of a Matrix.....	11
2.8 The Null Space.....	12
2.9 The Generalized Inverse .....	12
2.10 Orthogonality and Projections.....	14
2.11 Eigenvectors, Eigenvalues and Positive-Definiteness .....	15
2.12 Singular Value Decomposition .....	15
2.13 More on Positive-Definite Matrices.....	20
3 Some Basic Probability Theory .....	22
3.1 A Random Variable.....	22
3.2 Some Commonly Used Density Functions .....	24
3.2.1 Gaussian or Normal Distribution .....	24
3.2.2 The $t$ or Student's $t$ Distribution.....	25
3.2.3 The Chi-Squared Distribution .....	25
3.2.4 The F or Fisher Distribution.....	25
3.3 Random Vectors.....	26
3.4 Dependence and Independence .....	27
3.5 Covariance and Correlation.....	30
3.6 The Multinormal Distribution .....	31
3.7 Bayes Equation.....	32
3.8 Conditional Covariance Matrix.....	33
3.9 Propagation of Covariance .....	33
3.10 Principal Components .....	34
3.11 Spatial Correlation .....	35
3.12 Modern Geostatistical Methods .....	38
4. Situation Statement .....	40
4.1 Model Parameters.....	40
4.2 History-matching.....	41
4.3 Sampling the Posterior .....	45
4.3.1 Rejection Sampling .....	45
4.3.2 Markov Chain Monte Carlo .....	46
4.3.3 Highly Parameterized Bayesian Methods .....	48

## Table of Contents

---

4.3.4 Other Bayesian Methods .....	49
4.4 Model Calibration .....	50
4.4.1 What is Calibration? .....	50
4.4.2 Regularization .....	53
4.5 Model Linearization .....	55
4.5.1 The Jacobian Matrix .....	55
4.5.2 Linear Model Analysis .....	56
5. Manual Regularization .....	58
5.1 Formulation of Equations .....	58
5.2 Well-posed Inverse Problem .....	61
5.2.1 Data without Noise .....	61
5.2.2 Data with Noise .....	62
5.2.3 Prior Information .....	64
5.2.4 Objective Function Contours .....	66
5.2.5 Probability Density Contours of Posterior Parameter Error .....	67
5.2.6 Residuals .....	68
5.3 Post-Calibration Analysis .....	69
5.3.1 Problem Well-Posedness .....	69
5.3.2 Analysis of Residuals .....	74
5.3.3 Influence Statistics .....	76
5.4 Nonlinear Models .....	79
5.4.1 The Jacobian Matrix .....	79
5.4.2 Nonlinear Parameter Estimation .....	80
5.5 Critique of Manual Regularization .....	87
5.5.1 Man-Made Structural Noise .....	88
5.5.2 Parameter and Predictive Error Variance at the Scale that Matters .....	91
5.5.3 What is being Estimated? .....	94
5.5.4 Cooley's Strategy .....	95
5.5.5 The Role of Prior Information .....	97
5.5.6 Information Criteria Statistics .....	98
6. Mathematical Regularization .....	100
6.1 General .....	100
6.1.1 Formulation of Equations .....	100
6.1.2 Measurement Noise .....	100
6.1.3 Parameter and Predictive Error Variance .....	101
6.1.4 Overview of Mathematical Regularization .....	104
6.2 Singular Value Decomposition .....	106
6.2.1 Estimation of Parameters .....	106
6.2.2 An Example .....	107
6.2.3 Relationship between Estimated and Real Parameters .....	109
6.2.4 Kahunen-Loève Transformation .....	111
6.2.5 Predictive Error Variance .....	114
6.2.6 Well-posedness and Flow of Information .....	118
6.2.7 Strengths and Weaknesses of Singular Value Decomposition .....	121
6.3 Tikhonov Regularization .....	123

*Table of Contents*

---

6.3.1 Concepts .....	123
6.3.2 Formulation of Equations .....	126
6.3.3 Optimization of Regularization Weight Factor .....	127
6.3.4 Relationship between Estimated and Real Parameters.....	129
6.4 Practical Implementation .....	131
6.4.1 Nonlinear Model Behaviour.....	131
6.4.2 Solution of Equations .....	132
6.4.3 Multiple Regularization Weight Factors .....	134
6.4.4 Working with “Super Parameters” .....	135
6.4.5 Optimizing the Level of Fit .....	136
7. Linear Uncertainty and Error Analysis .....	139
7.1 General .....	139
7.2 Post-Calibration Subspace-Based Analyses.....	140
7.2.1 Parameter Identifiability.....	140
7.2.2 Other Subspace-Based Statistics .....	142
7.3 Post-Calibration Uncertainty Analysis.....	147
7.3.1 Posterior Parameter Uncertainty .....	147
7.4 Predictions.....	150
7.4.1 Predictive Uncertainty and Error Variance .....	150
7.5 Minimizing the Cost of Parameter Reduction.....	156
7.5.1 Background .....	156
7.5.2 Estimating Simplification Residuals .....	157
8. Nonlinear Uncertainty and Error Analysis .....	159
8.1 Some General Comments.....	159
8.1.1 Model Parameterization .....	159
8.1.2 Looking Ahead.....	162
8.2 Calibration-Constrained Monte-Carlo .....	163
8.2.1 Linear-Assisted Monte Carlo .....	163
8.2.2 Null Space Monte Carlo .....	165
8.3 Objective Function Thresholds .....	166
8.3.1 Some Theory .....	166
8.3.2 Repercussions for Objective Functions.....	168
8.4 Constrained Predictive Maximization/Minimization.....	170
8.4.1 Concepts .....	170
8.4.2 Implementation.....	172
8.4.3 Some Practicalities .....	173
8.4.4 Handling Large Numbers of Parameters .....	176
8.5 Model-Based Hypothesis-Testing.....	179
8.5.1 Calibration as Hypothesis-Testing .....	179
8.5.2 The Decision-Making Context .....	182
8.5.3 Configuring a Model for Hypothesis-Testing .....	183
9. Model Defects.....	187
9.1 Introduction.....	187
9.2 An Example.....	187
9.3 Mathematical Formulation of the Problem .....	192



*Table of Contents*

---

9.3.1 A Defective Model .....	192
9.3.2 Parameters and Residuals .....	194
9.3.3 Unavoidable “Structural Noise” .....	197
9.4 Defence against Structural Noise .....	198
9.4.1 Theory .....	198
9.4.2 Practice .....	199
9.5 Defect-Inducing Model Simplification: a Pictorial Representation .....	202
9.6 Defect-Induced Predictive Error .....	208
9.6.1 Linear Analysis .....	208
9.6.2 Nonlinear Analysis .....	213
10 Other Issues .....	217
10.1 Introduction .....	217
10.2 Objective Function Formulation .....	217
10.2.1 General .....	217
10.2.2 Some Issues .....	218
10.3 Derivatives .....	220
10.4 Proxy and Surrogate Models .....	222
10.5 Conclusions .....	222
11. References .....	224