A METHODOLOGY FOR HYPOTHESIS TESTING IN CONCEPTUAL CATCHMENT MODELLING

Gregory Paul Raper BSc (Hons)

Doctor of Philosophy

July 1998
I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a degree to any other University or Institution.

______________________________

Gregory Paul Raper
ACKNOWLEDGEMENTS

Like every PhD candidate, I owe my greatest vote of thanks to my supervisor. George Kuczera was always full of encouragement, even when I perhaps did deserved a hard word or two. George always expressed such a level of faith in me that I felt I would let him down if I didn’t achieve at the highest possible level, and I’m grateful for the motivation that this gave me.

I would like to thank Gary Willgoose for putting up with all my silly questions when he was trying to achieve his own goals, which were high and which also provided me with further inspiration.

I would like to thank John Ruprecht, Greg May and Bob Stokes of the Water Authority of W.A. for providing catchment streamflow and climate data. David Williamson, Jeff Turner and Colin Johnson, of CSIRO Water Resources, also provided catchment data and discussion about catchment processes. I also wish to thank Murugesu Sivapalan and Neil Viney for making the LASCAM model available for this work. The Land and Water Resources Research and Development Corporation funded a portion of this work under project 88/71, for which I am grateful. I would like to thank Marek Mroczkowski for his work on extending the methodology presented here to catchments undergoing landuse change. I acknowledge the great support of my colleagues in Agriculture Western Australia, especially Richard George, who provided me with support and inspired me with their dedication and breadth of knowledge.

Lastly, I thank my wife, Maria, and sons, Kyle and Cameron, for their seemingly endless love and patience.
# TABLE OF CONTENTS

1 Introduction.......................................................................................................................................1
  1.1 What is a Conceptual Catchment Model?......................................................................................2
  1.2 Thesis Objective ............................................................................................................................3
  1.3 Thesis Overview .............................................................................................................................4

2 Applications of Lumped Catchment Models ........................................................................6
  2.1 Evolution of Lumped Catchment Models.................................................................................6
  2.2 Prediction of Streamflow from Gauged and Ungauged Catchments ........................................7
  2.3 Prediction of Unobserved Catchment Responses .....................................................................7
  2.4 Prediction of Catchment Water Quality .....................................................................................10
  2.5 Summary ......................................................................................................................................12

3 Problem Definition .....................................................................................................................13
  3.1 Klemes’ Hierarchical Scheme for Testing of Hydrologic Models ......................................13
    3.1.1 Stationary Conditions ...........................................................................................................15
    3.1.2 Non-Stationary Conditions ..................................................................................................16
    3.1.3 Some Contemporary Applications of Klemes’ Hierarchical Scheme for Testing of Hydrologic Models ......................................................................................17
  3.2 Difficulties Commonly Encountered in Model Calibration .....................................................18
    3.2.1 The Assumption of Perfect Knowledge .............................................................................18
    3.2.2 Multiple Optima and Parameter Identifiability ..................................................................19
  3.3 Sources of Error in Catchment Model Predictions .....................................................................22
  3.4 Summary ......................................................................................................................................23

4 Proposed Methodology .............................................................................................................24
  4.1 Popper’s Model of The Scientific Method .................................................................................25
    4.1.1 Degrees of Falsifiability ......................................................................................................27
  4.2 A Model Filter .............................................................................................................................28
  4.3 The Principle of Parsimony .......................................................................................................30
  4.4 Statistical Techniques for Model Calibration and Hypothesis Testing ..................................31
  4.5 Increasing the Power of Data: Tracers and Other Hydrologic Data .......................................33
  4.6 Choosing an Approach to Hypothesis Testing .........................................................................35
  4.7 Summary ......................................................................................................................................37

5 Statistical Framework .................................................................................................................39
  5.1 Basic Principles ..........................................................................................................................40
  5.2 Error Models ..............................................................................................................................40
    5.2.1 Least Squares Error Model .................................................................................................40
    5.2.2 Error Model Diagnostics ....................................................................................................41
    5.2.3 A More General Error Model ............................................................................................42
5.3 Parameter Optimisation using Bayesian Inference ...................................................44
  5.3.1 Multiple Response Optimisation ........................................................................46
  5.3.2 Parameter Transformations ...............................................................................47
5.4 Response Surface Plots ..........................................................................................48
5.5 Prediction and Confidence Limits ..........................................................................48
5.6 Summary ..................................................................................................................49

6 Case Study: Development of the CATPRO Model ..................................................51
6.1 Model Rationale and Philosophy ...........................................................................51
6.2 Catchment Water Balance Model ........................................................................53
  6.2.1 Rainfall Interception .......................................................................................53
  6.2.2 A Store ............................................................................................................54
  6.2.3B Store ..............................................................................................................59
  6.2.4 Groundwater Store .........................................................................................60
6.3 Catchment Salinity Balance Model .......................................................................62
  6.3.1 Chloride as a Natural Tracer ............................................................................62
  6.3.2 Model Formulation .........................................................................................63
6.4 Summary ..................................................................................................................66

7 Case Study Catchment ...............................................................................................68
7.1 Rainfall and Saltfall ...............................................................................................71
7.2 Streamflow and Chloride Export ...........................................................................73
7.3 Evapotranspiration ...............................................................................................73
7.4 Groundwater Levels .............................................................................................75
7.5 Profile Chloride Storage ........................................................................................77
7.6 Stable Isotope Observations ................................................................................78
7.7 Other Hydrologic Data .........................................................................................78
7.8 Summary ..................................................................................................................79

8 Streamflow Calibration ..............................................................................................80
8.1 Base Case Streamflow Calibration .........................................................................80
8.2 Effect of Box-Cox Transformations on Predicted Responses and Parameter Identiifiability ...........................................................................................................90
  8.2.1 Box-Cox $\lambda = 0.5$ .......................................................................................90
  8.2.2 Box-Cox $\lambda = 0.25$ .....................................................................................96
8.3 Split Sample Test ....................................................................................................103
8.4 Prediction of Unobserved Catchment Responses ................................................108
  8.4.1 Groundwater Recharge ..................................................................................109
  8.4.2 Actual Evapotranspiration ............................................................................110
8.5 Effect of Relaxing Some Constraints on Model Parameters ................................111
  8.5.1 Groundwater Leakage Unknown ..................................................................111
  8.5.2 Throughfall Parameters Unknown ................................................................114
8.6 Recasting Some Model Hypotheses - Assessment of Streamflow Generation Mechanisms ........................................................................................................116
8.7 Summary ..................................................................................................................118

9 Multi-Response Calibration .........................................................................................120
9.1 Baseflow Conceptualisation ....................................................................................121
9.2 Model Calibration and Predictions .........................................................................122
9.3 Calibration Statistics and Diagnostics .......................................................................127
  9.3.1 Streamflow ........................................................................................................131
  9.3.2 Chloride Export ................................................................................................134
  9.3.3 Groundwater Storage ......................................................................................136
  9.3.4 Objective Function Response Surface ..............................................................139
9.4 Prediction of Unobserved Catchment Responses ....................................................141
9.5 Reassessment of a Model Hypothesis ......................................................................144
9.6 Summary ..................................................................................................................145

10 An Alternative Hydrosalinity Model .......................................................................149
10.1 Model Description ..................................................................................................150
  10.1.1 Water Balance Model .....................................................................................152
  10.1.2 Chloride Balance Model ................................................................................157
10.2 Multi-Response Calibration ....................................................................................159
  10.2.1 Calibration Strategy .......................................................................................159
  10.2.2 Model Predictions ..........................................................................................163
  10.2.3 Calibration Diagnostics ..................................................................................169
  10.2.4 Objective Function Response Surface ..........................................................177
  10.2.5 Prediction of Unobserved Catchment Responses ..........................................180
10.3 Degree of Model Falsifiability ...............................................................................183

11 Conclusions ..............................................................................................................186

12 References ................................................................................................................190
LIST OF FIGURES

Figure 6.1: Schematic of CATPRO conceptual catchment water balance model. ...........................................................................................................54

Figure 6.2: Conceptualisation of CATPRO A store showing perched water table. .................................................................................................55

Figure 7.1: Location map for the Salmon Catchment in the catchment of the Wellington Reservoir which is fed by the Collie River in the south-west of Western Australia, rainfall isoheys are also shown. Other experimental catchments in the Collie River Basin are also marked (after Williamson et al., 1987). .................69

Figure 7.2: Salmon Catchment, showing topographic contours, stream gauging station, pluviographs and piezometer locations (after Stokes, 1985). ........................................................................................................71

Figure 7.3: Monthly rainfall data for Salmon Catchment, south-west Western Australia, derived by Thiessen-weighting of daily rainfalls observed at three pluviometers. (Source: Water Authority of W.A.) .................................................................72

Figure 7.4: Time series of monthly Class A pan evaporation and Morton’s wet environment evapotranspiration for Salmon Catchment, April 1974 to March 1983. .................................................................75

Figure 7.5: Time series of piezometric heads for four representative piezometers at Salmon Catchment between April 1974 and March 1983, inclusive (data courtesy of CSIRO Division of Water Resources). .................................................................76

Figure 7.6: Time series of catchment average piezometric head for Salmon Catchment. .................................................................................................77

Figure 8.1: Monthly streamflow for Salmon Catchment predicted by CATPRO calibrated to streamflow data alone, Box-Cox $\lambda = 1.0$. .................................................................................................................................82

Figure 8.2: Time series of predicted contributions to streamflow for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 1.0$. .................................................................................................................................82

Figure 8.3: Time series of CATPRO conceptual stores predicted for Salmon Catchment following calibration to streamflow data using Box-Cox $\lambda = 1.0$. .................................................................................................................................84

Figure 8.4: Scatter plot of residuals against predicted streamflow for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 1.0$. .................................................................................................................................85
Figure 8.5:  Residual time series plot for streamflow predicted by CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 1.0$. ...........................................................................................86

Figure 8.6:  Residual autocorrelation plot for streamflow predicted by CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 1.0$. ...........................................................................................86

Figure 8.7:  Residual normal probability plot for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 1.0$. ...........................................................................................87

Figure 8.8:  Residual cumulative periodogram plot for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 1.0$. .........................................................................................................87

Figure 8.9:  CATPRO response surface plot, rank 1 and rank 2 principal components for calibration to Salmon streamflow data using Box-Cox $\lambda = 1.0$. ...................................................................................88

Figure 8.10:  CATPRO response surface plot, rank 2 and rank 3 principal components for calibration to Salmon streamflow data using Box-Cox $\lambda = 1.0$. ...................................................................................89

Figure 8.11:  CATPRO response surface plot, rank 6 and rank 7 principal components for calibration to Salmon streamflow data using Box-Cox $\lambda = 1.0$. ...................................................................................90

Figure 8.12:  Monthly streamflow for Salmon Catchment predicted by CATPRO calibrated to streamflow data alone, Box-Cox $\lambda = 0.5$. .........................................................................................................91

Figure 8.13:  Scatter plot of residuals against predicted streamflow for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 0.5$. .........................................................................................................91

Figure 8.14:  Residual normal probability plot for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 0.5$. .........................................................................................................92

Figure 8.15:  Residual cumulative periodogram plot for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 0.5$. .........................................................................................................93

Figure 8.16:  Time series of predicted contributions to streamflow for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 0.5$. .........................................................................................................94

Figure 8.17:  Time series of CATPRO conceptual stores predicted for Salmon Catchment following calibration to streamflow data using Box-Cox $\lambda = 0.5$. .........................................................................................................94

Figure 8.18:  CATPRO response surface plot, rank 4 and rank 5 principal components for calibration to Salmon streamflow data using Box-Cox $\lambda = 0.5$. .........................................................................................................96
Figure 8.19: CATPRO response surface plot, rank 1 and rank 2 principal components for calibration to Salmon streamflow data using Box-Cox $\lambda = 0.5$, with parameters $bf$, $bf\_crit$ and $gws0$ fixed.................................................................97

Figure 8.20: Monthly streamflow for Salmon Catchment predicted by CATPRO calibrated to streamflow data alone, Box-Cox $\lambda = 0.25$.................................................................97

Figure 8.21: Scatter plot of residuals against predicted streamflow for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 0.25$.................................................................98

Figure 8.22: Residual normal probability plot for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 0.25$.................................................................99

Figure 8.23: Residual cumulative periodogram plot for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 0.25$.................................................................99

Figure 8.24: Time series of predicted contributions to streamflow for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 0.25$.................................................................101

Figure 8.25: Time series of CATPRO conceptual stores predicted for Salmon Catchment following calibration to streamflow data using Box-Cox $\lambda = 0.25$.................................................................101

Figure 8.26: CATPRO response surface plot, rank 2 and rank 3 principal components for calibration to Salmon streamflow data using Box-Cox $\lambda = 0.25$.................................................................103

Figure 8.27: CATPRO response surface plot, rank 1 and rank 2 principal components for calibration to Salmon streamflow data using Box-Cox $\lambda = 0.25$, with parameter $bf\_crit$ fixed.......................104

Figure 8.28: Monthly streamflow for Salmon Catchment predicted by CATPRO calibrated to the first 5 years of streamflow data as a split sample test, Box-Cox $\lambda = 0.5$. .........................................................105

Figure 8.29: Time series of predicted contributions to streamflow for Salmon Catchment following calibration of CATPRO to the first 5 years of streamflow data as a split sample test, Box-Cox $\lambda = 0.5$. .........................................................105

Figure 8.30: Time series of CATPRO conceptual stores predicted for Salmon Catchment following calibrated to the first 5 years of streamflow data as a split sample test, Box-Cox $\lambda = 0.5$. .................106

Figure 8.31: Monthly streamflow for Salmon Catchment predicted by CATPRO calibrated to the first 5 years of streamflow data as a split sample test, Box-Cox $\lambda = 0.5$ and groundwater store inactive.........................................................108
Figure 8.32: Time series of predicted contributions to streamflow for Salmon Catchment following calibration of CATPRO to the first 5 years of streamflow data as a split sample test, Box-Cox $\lambda = 0.5$ and groundwater store inactive. ...........................................108

Figure 8.33: Time series of unobserved groundwater recharge and 90% confidence limits predicted for Salmon Catchment by CATPRO calibrated to streamflow data, Box-Cox $\lambda = 0.5$ ..................110

Figure 8.34: Time series of unobserved catchment averaged evapotranspiration and 90% confidence limits predicted for Salmon Catchment by CATPRO calibrated to streamflow data, Box-Cox $\lambda = 0.5$ ........................................................................111

Figure 8.35: Time series of predicted streamflow and 90% prediction limits produced for Salmon Catchment by CATPRO calibrated to streamflow data without prior knowledge of groundwater leakage, Box-Cox $\lambda = 0.5$ ..................................................................................113

Figure 8.36: Time series of unobserved catchment averaged recharge and 90% confidence limits predicted for Salmon Catchment by CATPRO calibrated to streamflow data without prior knowledge of groundwater leakage, Box-Cox $\lambda = 0.5$ ......................113

Figure 8.37: Time series of unobserved catchment averaged evapotranspiration and 90% confidence limits predicted for Salmon Catchment by CATPRO calibrated to streamflow data without prior knowledge of groundwater leakage, Box-Cox $\lambda = 0.5$ ..................................................................................114

Figure 8.38: Time series of predicted streamflow and 90% prediction limits produced for Salmon Catchment by CATPRO calibrated to streamflow data without prior knowledge of throughfall, Box-Cox $\lambda = 0.5$ ..................................................................................115

Figure 8.39: Time series of unobserved catchment averaged recharge and 90% confidence limits predicted for Salmon Catchment by CATPRO calibrated to streamflow data without prior knowledge of throughfall, Box-Cox $\lambda = 0.5$ .................................................116

Figure 8.40: Time series of unobserved catchment averaged evapotranspiration and 90% confidence limits predicted for Salmon Catchment by CATPRO calibrated to streamflow data without prior knowledge of throughfall, Box-Cox $\lambda = 0.5$ ..................................................................................116

Figure 9.1: Groundwater storage - discharge relationship showing the effects of model parameters $bf_{max}$ and $bf_{crit}$ .................................................121

Figure 9.2: Monthly streamflow for Salmon Catchment predicted by CATPRO calibrated to streamflow, stream chloride export and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5$, and 1.0 respectively. .................................................................123
Figure 9.3: Monthly chloride exported in streamflow for Salmon Catchment predicted by CATPRO calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5, \text{ and } 1.0$ respectively. ................................................123

Figure 9.4: Calculated catchment averaged piezometric heads for the deep aquifer at Salmon Catchment and groundwater storage predicted by CATPRO calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5, \text{ and } 1.0$ respectively. ...........................................................124

Figure 9.5: Time series of predicted contributions to streamflow for Salmon Catchment made by CATPRO calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5, \text{ and } 1.0$ respectively. .........................................................................................126

Figure 9.6: Time series of predicted contributions to stream chloride export for Salmon Catchment made by CATPRO calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5, \text{ and } 1.0$ respectively. .........................................................................................126

Figure 9.7: Time series of predicted chloride concentrations of contributions to streamflow for Salmon Catchment made by CATPRO calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5, \text{ and } 1.0$ respectively. .........................................................................................127

Figure 9.8: Time series of predicted mass of chloride in model stores for Salmon Catchment made by CATPRO calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5, \text{ and } 1.0$ respectively. .........................................................................................127

Figure 9.9: Scatter plot of residuals against predicted streamflow for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5, \text{ and } 1.0$ respectively. .................................................................132

Figure 9.10: Normal probability plot of streamflow residuals for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5, \text{ and } 1.0$ respectively. .................................................................132

Figure 9.11: Cumulative periodogram of streamflow residuals for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5, \text{ and } 1.0$ respectively. .................................................................133
Figure 9.12: Autocorrelation plot for streamflow residuals for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5,$ and $1.0$ respectively.................................................................133

Figure 9.13: Scatter plot of residuals against predicted streamflow chloride export for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5,$ and $1.0$ respectively.........................................................................................134

Figure 9.14: Normal probability plot of streamflow chloride export residuals for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5,$ and $1.0$ respectively. .........................................................................................135

Figure 9.15: Cumulative periodogram of streamflow chloride export residuals for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5,$ and $1.0$ respectively. .........................................................................................135

Figure 9.16: Autocorrelation plot for streamflow chloride export residuals for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5,$ and $1.0$ respectively. .........................................................................................136

Figure 9.17: Scatter plot of residuals against predicted changes in groundwater storage for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5,$ and $1.0$ respectively. .........................................................................................137

Figure 9.18: Normal probability plot of change in groundwater storage residuals for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5,$ and $1.0$ respectively. .........................................................................................137

Figure 9.19: Cumulative periodogram of change in groundwater storage residuals for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5,$ and $1.0$ respectively. .........................................................................................138

Figure 9.20: Autocorrelation plot for change in groundwater storage residuals for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5,$ and $1.0$ respectively. .........................................................................................138
Figure 9.21: CATPRO response surface plot, rank 1 and rank 2 principal components for calibration to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5$, 0.5, and 1.0 respectively. ................................................................. 139

Figure 9.22: CATPRO response surface plot, rank 2 and rank 3 principal components for calibration to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5$, 0.5, and 1.0 respectively. ......................................................................................... 140

Figure 9.23: CATPRO response surface plot, rank 12 and rank 13 principal components for calibration to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5$, 0.5, and 1.0 respectively. ......................................................................................... 141

Figure 9.24: Time series of unobserved groundwater recharge and 90% confidence limits predicted for Salmon Catchment following calibration to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5$, 0.5, and 1.0 respectively. ......................................................................................... 142

Figure 9.25: Time series of unobserved catchment averaged evapotranspiration and 90% confidence limits predicted for Salmon Catchment following calibration to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5$, 0.5, and 1.0 respectively. ................................................................. 143

Figure 10.1: Schematic of LASCAM water balance model showing conceptual stores and fluxes (after Viney and Sivapalan, 1995). See the text for explanations of the model fluxes................. 150

Figure 10.2: Monthly streamflow for Salmon Catchment predicted by LASCAM calibrated to streamflow, stream chloride export and changes in piezometric head, Box-Cox $\lambda = 0.25$, 0.25, and 1.0 respectively................................................................. 165

Figure 10.3: Monthly stream chloride export for Salmon Catchment predicted by LASCAM calibrated to streamflow, stream chloride export and changes in piezometric head, Box-Cox $\lambda = 0.25$, 0.25, and 1.0 respectively................................................................. 166

Figure 10.4: Calculated catchment averaged piezometric heads for the deep aquifer at Salmon Catchment and groundwater storage predicted by LASCAM calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25$, 0.25, and 1.0 respectively. ................................................................. 166
Figure 10.5: Time series of predicted contributions to streamflow for Salmon Catchment made by LASCAM calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25, \text{ and } 1.0$ respectively. .................................................................167

Figure 10.6: Time series of predicted model storages for Salmon Catchment made by LASCAM calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25, \text{ and } 1.0$ respectively. ...............................................................................167

Figure 10.7: Time series of predicted contributions to stream chloride export for Salmon Catchment made by LASCAM calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25, \text{ and } 1.0$ respectively. ...............................................................................168

Figure 10.8: Time series of predicted chloride concentrations of contributions to streamflow for Salmon Catchment made by LASCAM calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25, \text{ and } 1.0$ respectively .................................................................168

Figure 10.9: Time series of predicted mass of chloride in model stores for Salmon Catchment made by LASCAM calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5, \text{ and } 1.0$ respectively. ...............................................................................169

Figure 10.10: Scatter plot of residuals against predicted streamflow for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25, \text{ and } 1.0$ respectively. .................................................................170

Figure 10.11: Normal probability plot of streamflow residuals for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25, \text{ and } 1.0$ respectively. .................................................................171

Figure 10.12: Cumulative periodogram of streamflow residuals for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25, \text{ and } 1.0$ respectively. .................................................................171

Figure 10.13: Autocorrelation plot for streamflow residuals for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25, \text{ and } 1.0$ respectively. .................................................................172
Figure 10.14: Scatter plot of residuals against predicted chloride export for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25,$ and $1.0$ respectively. ............................................173

Figure 10.15: Normal probability plot of chloride export residuals for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25,$ and $1.0$ respectively. ............................................173

Figure 10.16: Cumulative periodogram of chloride export residuals for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25,$ and $1.0$ respectively. ............................................174

Figure 10.17: Autocorrelation plot for chloride export residuals for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25,$ and $1.0$ respectively. ............................................174

Figure 10.18: Scatter plot of residuals against predicted changes in groundwater storage for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25,$ and $1.0$ respectively. ............................................175

Figure 10.19: Normal probability plot of changes in groundwater storage residuals for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25,$ and $1.0$ respectively. ............................................176

Figure 10.20: Cumulative periodogram of changes in groundwater storage residuals for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25,$ and $1.0$ respectively. ............................................176

Figure 10.21: Autocorrelation plot for changes in groundwater storage residuals for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25,$ and $1.0$ respectively. ............................................177

Figure 10.22: LASCAM response surface plot, rank 3 and rank 4 principal components for calibration to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25,$ and $1.0$ respectively. ............................................178
Figure 10.23: LASCAM response surface plot, rank 4 and rank 5 components for calibration to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25,$ and 1.0 respectively. .................................................................179

Figure 10.24: LASCAM response surface plot, rank 7 and rank 8 principal components for calibration to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25,$ and 1.0 respectively. .................................................................180

Figure 10.25: Time series of unobserved groundwater recharge and 90% confidence limits predicted for Salmon Catchment following calibration of LASCAM to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25,$ and 1.0 respectively. ..............................................................................182

Figure 10.26: Time series of unobserved catchment averaged evapotranspiration and 90% confidence limits predicted for Salmon Catchment following calibration of LASCAM to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25,$ and 1.0 respectively. ..............................................................................182

Figure 10.27: Time series of unobserved catchment averaged evapotranspiration predicted by CATPRO and by LASCAM for Salmon Catchment following calibration to streamflow, stream water salinity and changes in piezometric head. .........................183
LIST OF TABLES

Table 5.1: Parameter transformations available in NLFIT Bayesian nonlinear regression suite and their effects, where $\beta$ and $\beta_T$ are the parameter and its transformed value respectively. .................................47

Table 6.1: CATPRO model parameter definitions and units........................................67

Table 8.1: Posterior parameter distribution for calibration of the CATPRO model to Salmon streamflow data for a 10 year record using Box-Cox $\lambda = 1.0$. .................................................................83

Table 8.2: Posterior correlation matrix of parameters for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 1.0$. .........................................................................................84

Table 8.3: Posterior parameter distribution for calibration of the CATPRO model to Salmon streamflow data for a 10 year record using Box-Cox $\lambda = 0.5$. .................................................................95

Table 8.4: Posterior parameter distribution for calibration of the CATPRO model to Salmon streamflow data for a 10 year record using Box-Cox $\lambda = 0.25$. .................................................................102

Table 8.5: Posterior parameter distribution for calibration of the CATPRO model to Salmon streamflow data for the first 5 years of record as a split sample test, Box-Cox $\lambda = 0.5$.................................107

Table 9.1: Calibration statistics for CATPRO calibrated to streamflow, stream water salinity and changes in piezometric head at Salmon Catchment, Box-Cox $\lambda = 0.5$, 0.5, and 1.0 respectively. .........................................................................................124

Table 9.2: Posterior parameter distribution for CATPRO hydrosalinity model calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5,$ and 1.0 respectively. .................................................................129

Table 9.3: Posterior correlation matrix of parameters for CATPRO calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5,$ and 1.0 respectively. .........................................................................................130

Table 10.1: LASCAM model parameters and definitions (after Sivapalan et al., 1996a, 1996b and Viney and Sivapalan, 1995). ........................................................................................................151
Table 10.2: Posterior parameter distribution for LASCAM hydrosalinity model calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25$, 0.25, and 1.0 respectively. Five parameters not subject to calibration due to large standard deviations relative to expected values are shown as fixed. ..........................................162

Table 10.3: LASCAM parameters fixed due to large standard deviations relative to expected values. See text for explanation.........................163

Table 10.4: Calibration statistics for LASCAM calibrated to streamflow, stream water salinity and changes in piezometric head at Salmon Catchment, Box-Cox $\lambda = 0.25$, 0.25, and 1.0 respectively. ........................................................................................................163

Table 10.5: Posterior correlation matrix of parameters for LASCAM model calibrated to streamflow, stream water salinity and changes in piezometric head at Salmon Catchment, Box-Cox $\lambda = 0.25$, 0.25, and 1.0 respectively. Correlation coefficients $> 0.9$ are highlighted.................................................................164
ABSTRACT

During the latter part of the 1980s there was some debate in the literature that hydrology was not practised as a science but as an engineering technology. Hydrologic modelling was seen as devoid of a scientific basis, conceptual catchment modelling in particular, because the practice of calibrating a model to the response that was the subject of investigation was seen as “fudging” a catchment water balance. Presented here is a methodology, based on Popper’s (1959) theory of the nature of the scientific method, to more rigorously test the hypotheses about the catchment water balance that constitute a conceptual catchment model. The methodology has three main elements. Firstly, the calibration of a conceptual catchment model to multiple catchment responses to constrain the model behaviour. Secondly, the application of established statistical techniques to critically assess model predictions and errors; and, thirdly the principle of parsimony, the search for the simplest model that fully explains the data. The methodology is demonstrated by way of a case study in which a conceptual catchment model is proposed and rigorously tested. This is followed by a comparison with another model developed from the same conceptual base, but without reference to the proposed methodology.
The bulk of the research reported in this thesis was carried out between March 1990 and July 1992. The application of the statistical techniques described in Chapter 5 and the prediction of unobserved catchment responses, such as catchment averaged groundwater recharge, are described in Raper and Kuczera (1991). The use of multi-response time series data to provide greater power to falsify model hypotheses is reported in Raper and Kuczera (1993a, 1993b).

The methodology laid down in this thesis has been further strengthened by introducing the more demanding tests of model hypotheses that can be exerted by calibration to multi-response time series data that cover a period of catchment response to changes in landuse. Examples of the power of multi-response data and non-stationary hydrologic regime are given in the subsequent work of Kuczera et al. (1993) and more recently by Mroczkowski et al. (1997).
1 INTRODUCTION

During the latter part of the 1980's there appeared in the hydrology and water resources literature a number of review articles in which critical assessments were made of the way hydrology was, and by and large still is, practised (see for example Dooge, 1986; Beven, 1987, 1989; Klemes, 1986a). The standing of hydrology, as a science, was questioned. The main point of discontent was the widespread use of hydrologic models either devoid of physical meaning or based upon untested hypotheses about the processes that determine catchment dynamics.

In a philosophical and highly critical review of the state of hydrology as a science Beven (1987) deplored the apparent lack of procedures for testing the performance of hydrologic models in general. Specifically, he stated the need for a methodology of distinguishing between the effects of observation error, initial conditions, parameter values and model structure on the performance of hydrologic models in reproducing observed catchment responses. He argued that the practice of model calibration to data, and initial and boundary conditions, that are not error-free precludes any such analysis. He states:

There has been no adequate hypothesis testing in hydrology outside of the laboratory; and certainly not in the study of catchment responses. Consequently, we have inherited and continue to develop a plethora of incompatible and inadequate models of catchment response without the tools to choose between them in any rigorous way. They all work (to some extent); equally, they are all falsified.

Beven's (1987) observations about the lack of a methodology for the critical assessment of the effectiveness of hydrologic models in predicting catchment responses echoed what Dooge (1986) had stated earlier. Dooge (1986) accused the developers of hydrologic models of being seduced, as Pygmalion had, by the beauty of their own creations, as a consequence, he asserted, they have failed to objectively test the predictive ability of their models.

Klemes (1986a, 1988) has been equally scathing in his criticism of the way in which hydrology is practised. He argued that hydrology, by and large, is practised as a
technology rather than as a science and that hydrologists fudge a water balance rather than study the relationships between its components. He was particularly damning of the widespread, and apparently unquestioning, application of “black-box” models to hydrologic prediction. He argued that the inappropriate use of statistical techniques in hydrologic modelling is a poor substitute for a lack of hydrologic insight and quoted Box (1976) who expressed similar concerns about the over use of statistical techniques.

Hydrologists and engineers have used a range of models for varying purposes and no class of model was spared these criticisms as the same fundamental problems existed in all (Beven, 1987). The aim of this thesis is to redress this poor state of the art of hydrologic modelling, in particular, the state of the art of lumped or conceptual catchment modelling.

1.1 WHAT IS A CONCEPTUAL CATCHMENT MODEL?

The term conceptual model is often applied to that class of model in which the processes governing hydrologic responses are described by equations that mimic the bulked behaviour of a hydrologic system. The model equations do not necessarily describe the fundamental physics believed to be controlling catchment processes (Sorooshian, 1985, 1991; Kachroo and Natale, 1992). By necessity, the parameters of such a model are determined by calibration to the system response that is the object of prediction. In practice, most so-called distributed, physically based or deterministic models are also calibrated in the same manner. For the purpose of this thesis, the terms conceptual catchment model or lumped conceptual model will be used to describe the class of hydrologic models which describe the bulk or catchment averaged responses of a catchment to rainfall in a non-distributed way. The aim is to reproduce the main causal relationships that dominate catchment response to rainfall rather than to fully describe the micro-scale physical processes responsible. The use of the term conceptual model should not be confused with the three dimensional, mental picture developed as the first step of establishing a distributed, deterministic model for a particular hydrologic system or catchment sometimes used in the groundwater literature (Anderson and Woessner, 1992).
A conceptual catchment model is, at best, an incomplete conceptualisation of the dominant processes governing the spatially averaged catchment response to rainfall. At worst, it is a transfer function, devoid of any physical meaning, that reproduces an observed streamflow hydrograph to some level of acceptability. Beven (1987) argues that often “our standards of acceptability are not very high”.

Many early conceptual catchment models were devoid of any physically meaningful structure; consisting of conceptual stores used only to delay the rise in a predicted stream hydrograph to match the observed streamflow hydrograph. Model equations were solely empirical in nature, often with the conservation of mass being the only physical principle explicitly applied. From this position they have progressed to conceptualisations of what are considered the dominant processes governing streamflow generation at the hillslope or catchment scale. A catchment model of this type represents an attempt to lump or average the net effect of micro-scale processes that are known to vary spatially and sometimes temporally. In this process, model equations have sometimes tended toward simplifications of the equations describing micro-scale processes, derived from fundamental physical principles. Beven (1987) argues strongly against such application of micro-scale theory to the problem of catchment scale modelling. Rather, a conceptual model should be formulated in as parsimonious a fashion as possible while reproducing the essential dynamics of the observed catchment responses at the scale at which they are observed.

1.2 **Thesis Objective**

Conceptual catchment models have routinely been used with little regard for the manner in which they reproduce observed catchment responses, usually streamflow. It will be argued here that the scientific method, as proposed by Popper (1959), can be applied to the development of a conceptual catchment model. Popper’s model of the scientific method requires that hypotheses be formulated in terms that allow them to be tested against observation and demands that testing is actively pursued. Popper also stipulates that scientific hypotheses can be falsified, but never proven, and only rigorous testing will admit a hypothesis to the body of what is considered knowledge.
An effective methodology for model calibration and rigorous testing of model hypotheses will be presented. The methodology will rely directly on Popper’s model of the scientific method. The methodology will be demonstrated by way of a case study in which a conceptual catchment model is developed and calibrated to multiple catchment response data. The post calibration diagnostic statistics will be used to actively challenge model hypotheses and reject those that are not supported by the data.

1.3 **Thesis Overview**

This thesis will explore some of the reasons for the apparent lack of scientific rigour in hydrologic modelling. In the next chapter, the current state of the art of conceptual catchment modelling will be discussed. This will begin with a review of Klemes’ (1986b) scheme for the testing of hydrologic models in general; we start at this point because the scheme is hierarchically based on the proposed application of the model. The discussion of Klemes’ (1986b) scheme will serve two purposes. Firstly it will be used later as a benchmark against which the methodology proposed in Chapter 4 will be assessed. Secondly, it will serve as an introduction to a review of the applications to which conceptual catchment models are commonly put. The scheme is based on model application, so it will allow the review that follows to be brief. Discussion can then move onto the more important consideration of the practical problems encountered in conceptual catchment modelling and the extent to which they are causes or consequences of the current state of the science of hydrology. This will be followed by a discussion of the sources of error inherent in the use of conceptual catchment models.

Next, in Chapter 4, Popper’s (1959) theory of the scientific method will be discussed to outline a philosophical framework within which a methodology for testing model hypotheses will be developed. The essential elements of the systematic methodology for hypothesis testing in conceptual catchment model development will then be presented. The proposed methodology is based upon the calibration of a conceptual catchment model to multiple time series data and the application of established statistical techniques.
Chapter 5 introduces the statistical framework for model calibration and the powerful post calibration diagnostic tools that are an essential component of the proposed methodology. The basic principles are discussed and the NLFIT suit of software introduced.

In Chapters 6 and 7 a case study will be outlined. The catchment is Salmon Catchment in the south-west of Western Australia for which long time series of high quality data are available. The development of CATPRO, a conceptual catchment model, will then be traced through Chapters 8 and 9. CATPRO includes both catchment water and salt balances to extract information about the catchment water balance from streamflow, groundwater and stream water salinity data. The model was exposed to tests of increasing power to identify failed model hypotheses.

In Chapter 10, an alternative independently developed conceptual water and salinity balance model are examined and subjected to the same tests as CATPRO. The results of this examination are used to further explore some of the issues raised in Chapters 8 and 9, especially the relationship between model complexity and falsifiability.

Chapter 11 restates the main elements of the proposed methodology for hypothesis testing in conceptual catchment modelling. In addition, the lessons learnt through the application of this methodology to the development and assessment of the CATPRO model are reinforced. Furthermore, suggestions for the further development of the methodology are presented.
2 APPLICATIONS OF LUMPED CATCHMENT MODELS

Lumped catchment models were first developed to deal with, what are now considered routine, engineering type problems. Over time, however, they have been put to an increasing variety of uses.

In all areas of hydrologic prediction, the objective of a modelling exercise is paramount in determining the resources and, indeed, the choice of model applied to a given problem (Grayson and Chiew, 1994). It follows that the more demanding the objective of a hydrologic modelling exercise the more demanding the standard of the model used. Klemes (1986b) used a hierarchical scheme, based on the demands of a modelling exercise to determine the minimum degree of testing required of a hydrologic model. The demands placed on hydrologic models are critical in determining the degree of testing required of them. A full appreciation of the objectives lumped catchment models are commonly expected to meets is therefore essential to an appreciation for the need to apply a scientifically rigorous methodology to testing the hypotheses on which they are based.

Here we will briefly trace the evolution of lumped catchment models and explore the breadth and depth of the range of uses to which they are put. The great array of physical process which lumped catchment models are used to mimic will also become apparent. This will serve two related purposes: it will illuminate the need for a rigorous methodology for their testing; and, it will define the current state of the art of lumped catchment modelling.

2.1 EVOLUTION OF LUMPED CATCHMENT MODELS

Todini (1988) and O'Connell (1991) provide summaries of the parallel development of lumped catchment models and black-box statistical models from the 1940s to the 1980s. Lumped catchment models were first used to help solve engineering problems such as prediction streamflows for reservoir spillway design and drainage system design (O'Connell, 1991). They were developed as potential alternatives to
the instantaneous unit hydrograph approach to predicting the streamflow response to rainfall. The instantaneous unit hydrograph was widely used but was based on the invalid assumption that streamflow is linearly dependent on rainfall volume (Todini, 1988; O’Connell, 1991).

2.2 **Prediction of Streamflow from Gauged and Ungauged Catchments**

Lumped catchment models are often the most efficient means of predicting streamflow for flood mitigation or flood prediction purposes. For example, O’Connell (1991) describes how Lambert (1972) used simple lumped sub-catchment models linked to a streamflow routing algorithm to develop a real-time flood forecasting scheme for the River Dee, U.K. The scheme also used radar-derived estimates of rainfall and real-time flow gauging data to provide model inputs and parameters to continually update streamflow predictions using a recursive optimisation scheme.

The RORB model (Laurenson and Mein, 1985) is widely used for flood studies in Australia. The model uses simple lumped sub-catchment modules linked together to route streamflow through large catchments. Macoun and Lyall (1991), for example, used RORB to estimate peak flood discharges and durations on the Bogan River, N.S.W., for the design of flood mitigation works in the town of Nyngan. Streamflow data from three gauging stations in the catchment were used to manually calibrate the model.

2.3 **Prediction of Unobserved Catchment Responses**

Knapp *et al.* (1975) used a 40 parameter lumped catchment model coupled to a 1,232 element, finite difference groundwater flow code to predict streamflow, groundwater levels and groundwater recharge and discharge for a 1,300 square mile (3,370 km²) catchment near Wichita, Kansas over a 25 year period. The surface drainage component of the model was manually calibrated to daily streamflow data to minimise a logarithmic least squares objective function. The groundwater flow component was calibrated to streamflow data for periods when surface runoff and shallow sub-surface flows were considered negligible, to piezometric heads observed
at 16 bores in the catchment and to a 'basin-wide survey of hydraulic heads' carried out by the U.S. Geological Survey in 1971. The model matched the streamflow record well and made acceptable predictions of hydraulic head data.

Mather (1981) applied the monthly catchment water balance model developed by Thornthwaite and Mather (1955) to 28 catchments in New Jersey. He found that for two catchments where the model did not reproduce the observed monthly hydrographs well, the incorporation of a correction for recharge to a deep, regional aquifer greatly improved streamflow predictions. Under-prediction of streamflows was found to be more pronounced with increasing flow and this was attributed to leakage from the catchment caused by induced recharge as a result of groundwater pumping for the nearby city of Dover. Mather (1981) made a percentage correction that produced an average loss equivalent to the average annual recorded groundwater abstraction rate, however, he claims that the quantity of water pumped from the aquifer, and therefore the induced recharge, could have been inferred from the difference between the predicted and observed streamflows.

Bergström and Sandberg (1983) used a conceptual rainfall-runoff model to predict groundwater recharge to three Swedish aquifers. The conceptual model included snow melt, evapotranspiration and soil moisture accounting algorithms. A different conceptualisation of aquifer dynamics was appended to the conceptual model to suit the aquifer geology in each case. One of the case study aquifers was confined and recharge was modelled using a conceptual soil moisture store. Discharge from the confined aquifer was assumed to take place laterally, below the confining layer, and from its up gradient boundary whenever a storage threshold parameter was exceeded. Of the two unconfined aquifers, one exhibited a variable storage - discharge relationship that was modelled by the application of a different effective porosity and recession coefficient for each of three aquifer layers. Between two and seven groundwater parameters where employed depending on the aquifer type modelled. Parameters for the snow melt and soil moisture stores were chosen "by experience from previous work on conceptual runoff modelling", in all but one of the case studies in which snow melt parameters were calibrated. This implies that for the unconfined cases predicted recharge to the groundwater store from the soil moisture
store obtained from the use of the selected soil moisture parameters was assumed correct and not subject to calibration. The discharge parameters were optimised to produce the highest coefficient of correlation between conceptual groundwater storage and piezometric observations for each observation point. Effective porosity and reference levels were then determined by linear regression. This process was repeated interactively using sensitivity analyses and subjective trial and error calibration until an acceptable match between model predictions and observed piezometric heads was obtained. Bergström and Sandberg (1983) report no calibration to, or comparison with streamflow hydrographs. The authors had intentionally isolated the determination of groundwater response parameters from the estimation of other model parameters.

Alley (1984) as part of a critical comparison of five lumped catchment models operating on a monthly time-step, sought to determine the relationship between model state variables for groundwater storage and measurements of water levels in nearby wells. One of the models, the “abcd’ model developed by Thomas (1981), was found to predict groundwater storage values that matched the seasonal trend in water levels observed in wells near the catchment. The correlation coefficients between predicted groundwater storage and observed piezometric heads ranged from 0.55 to 0.90, only slightly less than the correlation between the heads observed at the different wells (0.77 to 0.93). However, the autocorrelation between predicted values of groundwater storage was considerably higher than the autocorrelation between observed heads. The purpose of Alley’s (1984) work was not to predict recharge rates but to use the prediction of groundwater storage as a check on model consistency with all available data on catchment responses. Although he had demonstrated that a conceptual model calibrated to streamflow could make qualitatively acceptable predictions of groundwater recharge, he suggested caution in attaching physical significance to the state variables of such models as indicators of actual catchment storage.

Chiew and McMahon (1990a, 1990b) demonstrated that a simple lumped-parameter catchment model could be used to predict regional recharge by calibrating the HYDROLOG catchment model to streamflow data for a 1,415 km$^2$ portion of the
Campaspe River Basin. However, they failed to fully explore the effects of parameter uncertainty on the accuracy of the recharge estimates they made. Their recharge estimates were produced after manually calibrating HYDROLOG to three years of streamflow data and testing simulated streamflows for the following three years against observed values. They then used the nonlinear least squares optimisation algorithm NLFIT (Kuczera, 1989) to calculate optimum parameter values, coefficients of variation and a parameter correlation matrix, and proceeded to a sensitivity analysis using sensitivity plots. The result of this procedure was that recharge predictions were made with one parameter set, determined by manual calibration, and the sensitivity analysis was then performed with another. Any discrepancy between the two parameter sets was not reported. Furthermore, the capability of the optimisation algorithm to produce confidence limits on the predicted recharge was not exploited and recharge rates were presented without any indication of the sampling errors associated with their generation. Instead, sensitivity plots were presented showing that streamflow and recharge predictions were sensitive to the same parameters. Sensitivity plots, however, only show the effect of varying individual parameters in isolation, they provide only limited information on prediction uncertainty because they make no allowance for the sources of uncertainty or the effect of parameter correlation, a criticism of conventional sensitivity analysis echoed by Beck (1987).

2.4 PREDICTION OF CATCHMENT WATER QUALITY
Hopkins (1984) describes a physically based, catchment scale model developed to predict streamflow, groundwater movement and chloride cycling for catchments of the northern jarrah forest of Western Australia. The model includes algorithms for interception, infiltration, evapotranspiration, perched and deep aquifer flows, stream-zone swamp processes and channel routing. Chloride cycling is modelled assuming complete mixing in each of the model stores and that atmospheric chloride is the only chloride input to the model. Chloride mobilisation from the unsaturated zone to the deep groundwater is determined by specifying a rate constant. The model is shown to match observed streamflow and chloride export extremely well for some catchments. However, its large field data requirements and number of parameters
makes calibration and application difficult and it has seen little operational use
(Ruprecht, pers. comm.).

Ruprecht and Sivapalan (1991) present a simplified streamflow and chloride model
for Darling Range catchments which has the same basic structure as that described by
Hopkins (1984). They present predicted chloride export for a catchment in the south-
west of Western Australia which matches the observed cumulative values very well
but which exhibit marked under-prediction for early flow events and over-prediction
of chloride export during the spring recession. This model was further developed by
Sivapalan et al. (1996a, 1996b, 1996c) and will be discussed in detail in Chapter 10.

Stone and Seip (1989) and Lundquist et al. (1990) describe applications of the
Birkenes catchment water balance and stream acidification model. The model takes
its name from the small (0.4 km$^2$) catchment in southern Norway for which it was
first developed in the late seventies by Lundquist (1976, 1977; Christophersen and
Neal, 1990), Christophersen and Wright (1981) and Christophersen et al. (1982). To
simulate the catchment water balance, the Birkenes model uses two linear stores in
series to produce quickflow and baseflow responses to rainfall (Beck et al., 1987 and
Christophersen and Neal, 1990). The contribution to streamflow from each store is
determined by a rate constant and a threshold parameter (Stone et al., 1990). A snow
store is included when required. Complete mixing of chemical species is assumed to
occur in the baseflow store. Piston flow is assumed to operate between the quickflow
and baseflow stores, so that an addition of water to the baseflow store from the
quickflow store will result in an equal volume of water being discharged to the
stream, which changes the chemical composition of the predicted streamflow. The
model has been used to predict the streamflow loads of both exchangeable and
conservative species, including $H^+$, $SO_4^{2-}$, $Cl^-$, $HCO_3^-$, $Na^+$, $Ca^{2+}$, $Mg^{2+}$ and $Al^{3+}$.
Rate constants and equilibrium concentrations are required model inputs for products
of chemical weathering.

Stone et al. (1990) describe the application of the Birkenes model to the Allt a'
Mharcaidh catchment in north-east Scotland in the Cairngorm Mountains. Their aim
was to test the Birkenes model in a physical environment different to that in which it
was developed. The Allt a' Mharcaidh catchment exhibits streamflow and stream water quality characteristics quite distinct from those observed at the Birkenes catchment due to its larger size (10 km$^2$ compared to 0.4 km$^2$) and greater range in elevation. Furthermore, Allt a' Mharcaidh is drained by a perennial stream where the Birkenes catchment is ephemeral. The model tended to underpredict peak flows for the Birkenes catchment and this was more pronounced for the Allt a' Mharcaidh catchment. Predicted stream chemical loads for the Allt a' Mharcaidh catchment matched the observed loads to a similar level as those predicted for the Birkenes catchment, long-term trends for most species were well reproduced but short term fluctuations were generally not well predicted.

2.5 SUMMARY

The uses to which relatively simple, lumped catchment models have been put has been briefly outlined. Originally, models of this type were used to fill in missing streamflow records or other relatively undemanding tasks. This is no longer the case and conceptual catchment models are now applied to problems as demanding as predicting the hydrologic impacts of landuse change.

Clearly, the expectations of conceptual catchment models demands a more rigorous method for testing model structure than has generally been used to date. In the following chapter some of the practical problems associated with calibrating and applying conceptual catchment models will be discussed. A methodology to rigorously challenge model hypotheses will then be proposed.
3 PROBLEM DEFINITION

In the Introduction to this thesis, the lack of a rigorous, systematic methodology for testing hydrologic models in general was discussed. In the previous section, some of the issues to which conceptual catchment models are applied were described so that the need for a methodology for their rigorous testing could be appreciated. In this section, the reasons for the lack of an appropriate methodology for this rigorous testing of conceptual catchment models will be summarised.

Klemes (1986b) outlined a hierarchical scheme for the testing of hydrologic simulation models. The scheme serves only to test the ability of a model to reproduce observations of a single catchment response and, as Klemes points out, it provides only the most basic level of model testing. However, it was the first enunciation of any systematic methodology for operational testing of hydrologic simulation models and it provides a benchmark against which other methodologies may be assessed.

Following a discussion of Klemes’ (1986b) hierarchical scheme, some of the more common difficulties encountered in the calibration and application of conceptual catchment models will be discussed. This will in turn be followed by a discussion of the sources and nature of the errors inherent in predictions of catchment responses made with conceptual catchment models. An appreciation of the difficulties of using conceptual catchment models in practice and of the nature of model errors is critical to defining the nature of any systematic methodology for hypothesis testing in conceptual catchment models.

3.1 KLEMES’ HIERARCHICAL SCHEME FOR TESTING OF HYDROLOGIC MODELS

Klemes (1986b) proposed a hierarchical approach to define the minimum requirements for the testing of hydrologic simulation models before their application to specific classes of problems. Before considering the difficulties encountered in the use of conceptual catchment models in depth, it is necessary to understand Klemes’ (1986b) scheme of model testing.
Firstly, Klemes (1986b) clearly differentiates between hydrologic simulation models and hydrologic forecasting models. Klemes defines a hydrologic simulation model as a mathematical model aimed at predicting some hydrologic variable $Y$, for a period $T$, from concurrent records of other variables $X, Z, \ldots$. In this context, the model is only useful if it can reliably predict a hydrologic variable $Y$ that would otherwise be unavailable. To satisfy this criteria a model does not have to be physically sound or defensible, it must merely be capable of producing estimates of the desired variable that fall within some acceptable limits. Generally the predictions, so produced would be for some purpose outside the realm of hydrology as a science, perhaps to answer a question of water resource allocation.

A hydrologic forecasting model is defined as a model aimed at producing predictions of some hydrologic variable $Y$ in an interval $\Delta T$, from available records of the same variable $Y$ and/or other variables $X, Z, \ldots$, in an immediately preceding period $T$. Under this definition, a hydrologic simulation model can operate in forecasting mode whenever it is used to predict future values of $Y$ from forecasted values of the independent variables $X, Z, \ldots$. It should be noted that for this application, the true future values of the predicted variable $Y$ will later be available and may be compared with model predictions.

This distinction between forecasting and simulation is important to the development of Klemes’ scheme, which relates to hydrologic simulation, but not to the central argument of this thesis. This thesis is fundamentally concerned with the structure of models rather than the requirements of the classes of applications for which they are used.

Before discussing the precise nature of Klemes’ hierarchical scheme for model testing, it is worth formally defining two important terms. These are verification and validation. Refsgaard and Knudsen (1996) define verification as the testing of the model algorithm against analytical or other numerical solutions. This is similar, but more demanding than, checking the model code for “bugs” because it makes the assumption that the analytical solution is correct. In fact, it says nothing about the status of the analytical solution, which is, in reality, only another mathematical
description of the same problem. This definition also implies that verification is only applicable to physically based or deterministic models. They define validation as the act of demonstrating that a site-specific model, including the appropriate parameter values, is capable of making accurate predictions for periods outside the calibration period. Klemes (1986b) does not formally define the term validation before he uses it. His purpose, however, is served by a definition similar to that of Refsgaard and Knudsen (1996), differing only in that in some instances validation will require the capability of making accurate predictions for a catchment other than that for which the model is calibrated. Because Klemes (1986b) makes no distinction about the form of the hydrologic model, no reference to the parameter values is made, though the implication is that they should not change.

Klemes (1986b) ranks the classes of applications to which hydrologic simulation models can be put in increasing complexity and therefore increasing demands on model performance. The resulting classes of applications and the minimum standards for operational testing are as follows:

3.1.1 Stationary Conditions

The term stationary is used to describe physical conditions that do not change appreciably with time.

Application to the Calibration Catchment

Examples of this class of application include filling in or extending a streamflow record. The minimum required test is the standard split sample test in which half the record is used for calibration and the other half used for validation, the second portion should then be used for calibration and the first for validation. The model passes the test only if the two results are similar and the errors for both validation periods fall within acceptable limits.

Application to a Catchment Other than the Calibration Catchment

The simulation of a streamflow record for an ungauged catchment is a typical example of this class of application; the proxy-catchment test is the minimum
requirement. If a streamflow record is required for a catchment $C$, two catchments $A$ and $B$ from the same region, for which streamflow records are available should be chosen. The model is calibrated on the streamflow record for catchment $A$ and validated against the record for catchment $B$ and *vice versa*. The model passes the test only if the validation results are both acceptable and similar.

### 3.1.2 Non-Stationary Conditions

Non-stationary conditions include marked changes in climate or landuse.

*Application to the Calibration Catchment*

Here a typical example would be the prediction of a streamflow record for a catchment undergoing a marked change in climate. The required test is the differential split sample test. For this example, two periods from the historical climate record should be selected such that the climate variable of interest exhibits different extremes in the two periods. If simulation of streamflows under a regime of increased rainfall is required the model should be calibrated on a period of below average rainfall and validated on a period of above average rainfall and *vice versa* if predictions are required under conditions of a reduced rainfall regime. If periods of appropriate rainfall distribution are not available the model should be tested on data for a substitute catchment.

A substitute catchment will always be required when a simulation is required for a change in landuse. In this case, a substitute catchment undergoing the same landuse change as the catchment in question should be selected. The model is then calibrated on the record prior to the landuse change and validated on the portion of the record following the landuse change. When substitute catchments are required for testing, two substitute catchments should always be used, and the test considered satisfactory only if the two validation results are acceptable.

*Application to a Catchment Other than the Calibration Catchment*

A proxy-catchment differential split sample test is necessary for a model required to simulate the effects of a change in landuse or climate on an ungauged catchment. As
in the standard proxy-catchment test described above, to test for applicability to an ungauged catchment $C$, two catchments $A$ and $B$ from the same region, for which streamflow records are available should be chosen. The model is calibrated on the streamflow record for say a dry period at catchment $A$ and validated against the record for a dry period at catchment $B$ and vice versa. The model passes the test only if the validation results are both acceptable and similar.

3.1.3 Some Contemporary Applications of Klemes’ Hierarchical Scheme for Testing of Hydrologic Models

Although optimisation techniques have improved substantially since Klemes (1986b) proposed his hierarchical scheme it is still used to assess and compare models and the algorithms used to calibrate them. For example, Gan and Biftu (1996) used split sample tests to assess three commonly used automatic optimisation algorithms; the shuffled complex evolution algorithm (Duan et al., 1992), the simplex method (Nelder and Mead, 1965) and the multiple start simplex algorithm. The testing procedure involved using the three optimisation algorithms to calibrate four conceptual catchment models to data sets from eight different catchments. They concluded that better calibration results did not necessarily result in better performance in the validation step. Similar exercises were performed by Resgaard and Knudsen (1996) and Michaud and Sorooshian (1994) who compared the abilities of conceptual catchment models to reproduce observed streamflow hydrographs with those of simple distributed models and complex “physically-based” distributed models.

Klemes’ (1986b) hierarchical scheme of model testing has been used in conceptual catchment modelling testing and validation exercises beyond its intent and design. It continues to be used in such exercises simply because a rigorous scheme for the more demanding testing required for the validation of the structure of conceptual catchment models has not been stated or demonstrated.
3.2 **DIFFICULTIES COMMONLY ENCOUNTERED IN MODEL CALIBRATION**

In practice, little or no information will be available concerning catchment responses other than a streamflow hydrograph. Hydrologic fluxes such as hillslope runoff and recharge, and control volumes, such as groundwater and soil water storages typically remain unmeasured and can not contribute to the calibration of a catchment model. In general, a conceptual model is calibrated to a streamflow hydrograph and judgements on the reliability of its predictions are based solely on the match to the calibration time series and a validation time series used for a split-sample test.

3.2.1 **The Assumption of Perfect Knowledge**

The most dangerous of all the problems associated with the use of a conceptual catchment model is the assumption that the model components are truly representative of the components of the physical system for which responses to rainfall are predicted. This assumption is often associated with models where the stores or variables are assigned names suggestive of physically definable entities, such as groundwater storage, recharge or baseflow. Other parameters sometimes assumed to be above the calibration process are those associated with ensuring a long-term water balance, such as constants of proportionality applied to Class A pan evaporation as a measure of potential evapotranspiration (Lundquist et al., 1990).

The assumption is often implied by statements to the effect that some model parameters should not be adjusted by calibration to prevent the model application from becoming a curve fitting exercise. Ironically, such statements are often made in attempts to be more rigorous in the application and testing of a model (see for example Lundquist et al., 1990).

Whenever a parameter of a conceptual catchment model is assigned a fixed value and not subject to calibration along with the other model parameters the implicit assumption is that the value of that parameter is known with absolute accuracy and certainty, the assumption of perfect knowledge. This has two important consequences. Firstly, it violates the principle of parsimony, which will be discussed at length later, and on a more immediate and practical level it will distort the
covariance matrix of the posterior parameter distribution so that it will not reflect the true extent of parameter error on model predictions (Beck, 1987).

### 3.2.2 Multiple Optima and Parameter Identifiability

Johnston and Pilgrim (1976) set out to calibrate a conceptual model (Boughton, 1965, 1966) to 80% of all gauged rural catchments in Australia with an area of less than 10 square miles (25.9 km\(^2\)). Their aim was to correlate parameter values to catchment characteristics and to test the derived relationships by applying them to the prediction of streamflow hydrographs for the remaining gauged catchments. However, they encountered great difficulty in determining optimum parameter values even for a group of five of the catchments. They document, in great detail, the problems of inadequate model structure, multiple optima and high parameter correlation noted by many others. The fact that they spent almost two years in an effort to achieve adequate calibrations is often quoted as an example of the poor state of the art of conceptual catchment modelling.

Ibbitt and O’Donnell (1974) presented one of the earliest reviews of the range of practical problems associated with using statistical optimisation techniques to calibrate conceptual catchment models. They list the existence of multiple local optima as the most important problem associated with calibrating conceptual models, other problems are associated with irregularities in the shape of the response surface caused by high degrees of parameter interaction and the effects of threshold parameters.

Alley (1984) systematically addressed some of the fundamental practical and philosophical questions central to the critical application of conceptual catchment models. He posed the following questions, in relation to five such models operating on a monthly time-step:

1. Are the optimised parameter values reasonable and consistent between models?
2. Do any of the models result in consistently lower calibration and prediction errors?
3. How accurately can the parameter values be determined and what are their covariances?

4. What is the relationship between model state variables for groundwater storage and measurements of water levels in nearby wells?

Alley found that all models predicted monthly and annual streamflows well when calibrated to ten-year records and then used to predict streamflows for the remaining forty years of the fifty years of record available. However, he identified gross inconsistencies in the prediction of evapotranspiration and soil water storage between the models due to the different conceptualisations used to partition monthly precipitation between increases in soil water storage, groundwater recharge and actual $\text{Et}$. In addition, high correlation was found to exist between initial storage values and maximum storage parameters when initial conditions were included in the model calibration as opposed to being set to zero with the use of a one-year warm-up period. One model, the “abcd” model, was found to predict groundwater storage values that matched the seasonal trend in water levels observed in wells near one of the experimental catchments. The other models where found to be lacking in this regard. Although Alley (1984) selected several models that some would consider outdated (Thornthwaite and Mather, 1955; Palmer, 1965) for critical assessment, and although many of the results were at least qualitatively predictable, his contribution is valuable because of the concise, explicit manner in which he addressed these issues.

Kleissen et al. (1990) and Gan and Burges (1990a, 1990b) provide concise, up to date summaries of the problems involved in the calibration of a conceptual catchment model to a streamflow hydrograph. They echo what Johnston and Pilgrim (1973) and Alley (1984) had earlier been stated; the most important problems faced are the existence of multiple optima and the presence of high degrees of correlation between some fitted model parameters. As a result there may be many sets of parameter values that produce almost identical streamflow predictions although the relative contributions of the fluxes that make up the streamflow may vary greatly. In addition, a high degree of correlation between model parameters usually leads to poor identifiability of some model parameters (Kleissen, et al., 1990). Poor identifiability implies great uncertainty in the calibrated parameter value and may result in
increased uncertainty in the prediction of fluxes other than the streamflow time series used in calibration.

Gan and Burges (1990b) demonstrate that, conceptual models often provide unreliable estimates of extreme hydrologic events, especially floods; the maximum capacities of the conceptual storages are often unrelated to the physical storages of a real world catchment; and model parameters are often climate dependent. Blackie and Eeles (1985) also discussed the last of these problems. They expressed the problem in terms of system, and therefore, parameter stability. An example of the climate dependence of model parameter values is evident in results presented by Hooper et al. (1988) who obtained markedly different values for some of the parameters of a conceptual catchment model when it was calibrated to streamflow time series for three different years.

Bates and Davies (1988) undertook a study to determine the effect of the choice of baseflow separation technique on the calibration and predictions made by the RORB runoff routing model (Laurenson and Mein, 1983). They found that calibrated model parameters were highly dependent upon baseflow separation technique for catchments where the model response was nonlinear. Having already established that “baseflow separation is an arbitrary procedure” they then concluded that extending the study to different surface runoff models and baseflow separation procedures should substantiate their findings. The authors could not recommend a baseflow separation procedure to overcome the problem; in fact they did not even raise the possibility that such a procedure may exist. This is not meant to be a criticism of the work of Bates and Davies (1988), they had identified a problem and recommended research to ascertain the extent to which that problem was encountered in standard engineering practice. The critical point is that the problem was identified, its source was obvious but no practical remedy could be envisaged. The problem here is that the baseflow parameters were determined in an arbitrary manner and then used in a conceptual model which was subsequently calibrated to individual flood hydrographs. If the model had been calibrated to a continuous streamflow record the extra data may have been found to contain information that would allow a more complete conceptual model to predict the baseflow component.
3.3 SOURCES OF ERROR IN CATCHMENT MODEL PREDICTIONS

Predictions of catchment fluxes made by a conceptual catchment model will be subject to two types of error, random error and systematic error. Random error originates from measurement error in both input and response data used in model calibration. A second source of random error is the uncertainty in the values of calibrated model parameters. Systematic error may result from the use of an incomplete and/or incorrect conceptualisation of the catchment water balance. Systematic errors will also be present in the measurement of hydrologic data (Essery, 1992).

Random errors lead to model predictions that are imprecise. That is, the statistical uncertainty in the estimate may be large. Systematic errors, on the other hand, will lead to biased model predictions that may be precise, but wrong. That is, the statistical uncertainty in the predicted values may be small but the estimates will be far from the true values, which in all probability, will not be known. The commonly encountered problem of multiple optimum sets of parameter values described above is a manifestation of the existence of systematic error and an inability to deal with it.

Furthermore, once calibrated a catchment model may reproduce the catchment response to which it is calibrated, usually a streamflow hydrograph, with a high degree of precision. However, a large degree of uncertainty may exist in some model parameters. In such a case, a high degree of uncertainty may be associated with any prediction of unobserved fluxes or with any extrapolation of the time series used in calibration beyond the range of flows in the calibration record (Sorooshian et al., 1983). Random errors propagating through the model can be quantified from information in the posterior distribution of model parameters. However, systematic error arising from model inconsistencies cannot be quantified in this way. Systematic error can not be identified or measured directly unless the true catchment response is known.

3.4 SUMMARY

The preceding discussion concerning the general lack of a systematic approach to the testing of conceptual catchment model structure and the practical difficulties
encountered in their application has provided an outline of what is required of a methodology to rigorously test model hypotheses. Klemes’ (1986b) hierarchical scheme for testing of hydrologic models has been discussed because it provides a benchmark against which any proposed methodology may be assessed. Indeed, Klemes’ scheme has continually found applications beyond its original intent and has served the science of hydrology well despite Klemes’ assertion that it provides only the minimum level of model testing.

Systematic errors in conceptual catchment model predictions provide the heaviest demands on any method of model testing. The presence of systematic error in hydrologic models is the most challenging issue to be raised during the debate on the status of hydrology as a science that took place during the 1980s (Beven, 1987, 1989; Klemes, 1986a, 1986b, 1988; Dooge, 1986).

In the following section, a systematic methodology to more rigorously test conceptual catchment model hypotheses will be presented. By necessity, the methodology will outline how both random and systematic errors can be dealt with. Popper’s (1959) theory of the falsifiability of scientific hypotheses will be used as a basis.
4 PROPOSED METHODOLOGY

Both Dooge (1986) and Beven (1987) refer to Popper's (1959) *The Logic of Scientific Discovery* as providing a basis for the critical assessment of hydrologic models. Popper (1959) proposed the falsification of scientific hypotheses as a basis for the advancement of scientific knowledge. The development of a methodology for hypothesis testing in conceptual catchment model development must logically start with a summary of Popper’s (1959) argument. This will be followed by a discussion of how Popper’s ideas on the constructs of the scientific method in general can be related specifically to the problem of falsifying the hypotheses of a conceptual catchment model. This will be followed by a brief discussion of the principle of parsimony, which is a critical issue because the simplicity of conceptual catchment models is often considered an advantage over deterministic models. More importantly, the notion of simplicity is linked to the degree of falsifiability of a scientific hypothesis.

The practical tools for the implementation of the proposed methodology are the statistical techniques used to calibrate conceptual catchment models and the observations of catchment responses to which they are calibrated and against which they are judged. The statistical tools commonly used to calibrate conceptual catchment models and assess their behaviour will be discussed and their demonstrated benefits listed. The power of different types of hydrologic data to falsify model hypotheses will then be reviewed. Finally, some practical considerations about how the available statistical techniques and empirical data can be combined to test model hypotheses will be considered before the case study presented in the following section is introduced.

In *The Logic of Scientific Discovery* (1959), Popper uses the terms “theory” and “hypothesis” interchangeably and makes no distinction between them, this convention will also apply in the following discussion.
4.1 Popper’s Model of the Scientific Method

Popper’s (1959) theory of the nature of the scientific method is based on the principle of the uniformity of nature. Natural laws are considered invariant with respect to time and space and there are no exceptions. These basic postulates are fundamental to the existence of what is called the scientific method and are beyond the scope of this thesis. Indeed, Popper does not expend great effort in discussing these issues in *The Logic of Scientific Discovery* (1959), preferring to concentrate on the issue of how hypotheses are admitted into the ranks of what is considered scientific knowledge.

Popper (1959) opens his discussion not by putting forward his own views on the nature of the scientific method but with a rebuttal of the principle of induction as a basis for the scientific method. The principle of induction is based on the development of universal statements from collections of singular statements, or the development of hypotheses from the results of observations and experiments.

Popper (1959) argues that although induction may be involved in the process of formulating scientific hypotheses it cannot form the basis on which hypotheses can be accepted or rejected. He distinguishes clearly between the act of conceiving a scientific theory, or hypothesis, and the practice of the scientific method in testing the theory. He uses the familiar example of deriving the statement “All swans are white” from many observations of white swans. Clearly, many observations of swans that are white may lead one to believe that all swans are white but they do not rule out the possibility of swans that are not white. Observation of a single swan that is, say, black would disprove the theory that “All swans are white”. Popper (1959) therefore, argues that scientific theories are differentiated from other forms of statements by the fact that they cannot be proved but only disproved, falsified.

To further his argument Popper (1959) postulates that natural laws, or theories, can be expressed in terms of the negation of a strictly existential statement. He uses the example of the law of conservation of energy, which he equates to a statement to the effect that “there is no perpetual motion machine.” This is a fundamental point,
because it follows that natural laws are falsifiable because they rule out the existence of certain things or the occurrence of certain events.

Popper (1959) goes to great lengths to define the terms in which a theory should be formulated to differentiate between theories and other classes of statements. Fundamental to Popper's argument is the empirical basis of the practice of science and his formal definition of what constitutes a theory depends on this (Popper, 1959; pp 86):

A *theory is empirical if it divides the class of all possible basic statements unambiguously into the following two non-empty subclasses, first the class of all basic statements with which it is inconsistent (prohibits): the potential falsifiers of the theory, second, basic statements which it does not contradict, furthermore the theory makes assertions only about its potential falsifiers, nothing about the permitted statements, in particular it does not say they are true.*

Popper (1959) defines a basic statement as a statement of singular fact that may be used to falsify a theory. A basic statement may therefore be considered a description of an event that will falsify a theory, if it does in fact occur.

Having defined what constitutes a scientific theory, Popper (1959) concerns himself with determining how theories may be falsified and how the degree of testability of theories may be compared. Part of this discussion involves the principle of parsimony and how simplicity and falsifiability are related. Popper's views on parsimony are summarised as follows

\[
\text{testability} = \text{high prior improbability} = \text{paucity of parameters} = \text{simplicity} \quad [4.1]
\]

The concept of high prior improbability is used to relate the idea that the form of the statements used to express a theory are important and the fact that theories are testable because they rule out certain occurrences. The more possible occurrences a theory rules out, the higher its level of prior improbability. So, paradoxically perhaps, a “good” theory, from the point of view of testability, is one that is highly improbable, one that rules out many possible occurrences.
A further issue raised by Popper (1959) in his discussion of testability and simplicity is the concept of auxiliary, or *ad hoc*, hypotheses, terms for which he provides no formal definition. Although he doesn’t formally define what he means by an auxiliary hypothesis, Popper (1959) asserts strongly, that auxiliary hypotheses should only be accepted if they increase the degree of falsifiability of a theory, that is, auxiliary hypotheses can strengthen a theory if after their addition, the theory rules out more. Conversely, auxiliary theories should be avoided if they act to defend a theory from falsification by contradictory observations.

### 4.1.1 Degrees of Falsifiability

The need to compare the degrees of falsifiability of scientific hypotheses arises from the fact that they cannot be proven. Popper (1959) goes to considerable length to provide a framework to facilitate the comparison of the degrees of falsifiability or testability of hypotheses. He outlines three methods for the comparison of the degree of falsifiability of different hypotheses, two of which are directly transferable to the testing of conceptual catchment models. These are the comparison of degrees of testability by the method of subclass relations, and by reference to the dimension of the hypotheses.

Popper (1959) proposes the application of simple set theory as one means of comparing the degrees of falsifiability of hypotheses. The test is applied to the potential falsifiers of the hypotheses; the sets of circumstances or occurrences that are disallowed by the hypotheses. Consider two hypotheses $X$ and $Y$; hypothesis $X$ is falsifiable to a higher degree than hypothesis $Y$ if, and only if, the class of potential falsifiers of $X$ includes the class of the potential falsifiers of $Y$ as a proper subclass.

In simple terms this means that hypothesis $X$ is more highly testable than hypothesis $Y$, if it disallows all that is disallowed by $Y$, and more. That is, a hypothesis that applies greater constraints on the possible outcomes of experiments tells us more about the physical world.
Extending this concept to conceptual catchment models; model $U$ is falsifiable to a higher degree than model $V$ if, and only if, the set of falsifiable predictions of $U$ includes the set of falsifiable predictions of $V$ as a proper subset.

The second method for comparing hypotheses proposed by Popper (1959) requires reference to the dimension of their possible falsifiers. Consider a falsifying basic statement as the conjunction of initial conditions with the negation of a prediction based on the hypothesis in question. The initial conditions will need to be unambiguously defined and the dimension of the basic statement depends on the level to which the initial conditions are defined. The fewer parameters to be quantified to describe the initial conditions the more general the basic statement that constitutes the potential falsifier of the hypothesis. Therefore, potential falsifiers of low dimension, that is, with few parameters required to specify initial conditions, impart greater generality and hence greater falsifiability on hypotheses.

Popper (1959) relates this concept to the principle of parsimony and the use of auxiliary hypotheses. Essentially a hypothesis is simple if its potential falsifiers are of low dimension and therefore exhibit a high degree of generality. Conversely, a hypothesis is considered complex if it is held and protected from falsification by auxiliary hypotheses, which result from potential falsifiers of high dimension.

Popper’s (1959) concept of simplicity and auxiliary hypotheses and how they relate to the testability of theories can be directly translated into the development of conceptual catchment models. This will be done in Section 4.2, which describes a model filter for conceptual catchment models based on Popper’s (1959) theory and Section 4.3 which discusses the principle of parsimony in more depth.

### 4.2 A Model Filter

In a review of the development of rainfall-runoff modelling, Todini (1988) divided present models into four groups: stochastic, lumped integral, distributed integral and distributed differential. The first group includes what are commonly referred to as “black-box” models, the next two groups encompass lumped or conceptual models, and the final group includes models formulated in terms of the differential equations
assumed to govern the flow of water and, in some cases, solutes on the micro-scale. Todini (1988) also classifies models according to the method of parameter estimation used. He argued that parameter optimisation by statistical techniques is equivalent to using a stochastic model and that a “great deal” of physical meaning inherent in model structure and/or parameter values is lost in the process. This may be true if a statistical parameter optimisation scheme is used without critical assessment of the derived parameter values and model output. However, it is proposed here that the critical use of statistical tools is essential not only to effective parameter optimisation but also to the testing and development of model structure which has physical meaning and which is also consistent with field observations. That is, statistical tools are necessary in devising a filter that separates model hypotheses that are falsified by observation from those which are not.

Beven (1987) and Todini (1988) have both been critical of the way that catchment modelling has been carried out in the past. The tools necessary to redress the situation have been available for some time though their full potential is far from being realised. Outlined here is a methodology to allow more rigorous testing of conceptual model structure and the evaluation of both random and systematic errors associated with model predictions.

Beven (1987) is correct in so far as no conceptual model can ever be “right”. By their very nature conceptual catchment models are at best a gross simplification of what are assumed to be the dominant processes governing catchment dynamics without regard for the effects of plot or micro-scale geologic and biological heterogeneity.

Beven (1987) did not criticise the state of hydrology without offering some guidance as to how he thought the problems which he so graphically illustrated may be solved. Beven contends that:

*Hydrology in the future will require a macroscale theory that deals explicitly with the problems posed by spatial integration of heterogeneous nonlinear interacting processes (including the effects of preferential flow paths) to provide a rigorous basis for both “lumped” and “physically-based” predictions. Such a theory will be inherently stochastic and will deal with the*
value of observations and qualitative knowledge in reducing predictive uncertainty; the interactions between parameterizations and uncertainty; and the changes in hydrological response to be expected as spatial scale increases. Such a theoretical framework should initiate new lines of thought, and innovative methods of measurement, analysis and hypothesis testing to be developed.

The purpose of a model filter is to identify, in some sense, the best model or models; clearly required to fulfil Beven’s (1987) vision. In the crudest terms, a scientifically acceptable model filter involves iteration of the following two steps (Medawar, 1984). First, the modeller conceptualises a model structure or, more accurately, offers testable hypotheses. This is a creative act based on the modeller’s world view which is much affected by current paradigms and by his experience during the second step. Second, the modeller attempts to falsify the hypothesis by comparing model predictions against independently observed data. This is the basis of the scientific method as espoused by Popper (1959).

One may argue that catchment hydrologists have not been derelict by ignoring the scientific method. Yet the multiplicity of ill-posed models suggests there are shortcomings in the model filter outlined above. Falsification of testable hypotheses can only eliminate hypotheses inconsistent with the evidence. Because noise or error corrupts hydrologic data we must use statistical concepts for rejecting hypotheses. We may find that data with low power to discriminate between hypotheses leave many hypotheses unrejected.

### 4.3 The Principle of Parsimony

The model filter needs to be strengthened by explicit application of the principle of parsimony which Box and Jenkins (1976) argue should lead to the development of the simplest possible model consistent with the evidence. This principle is deeply embedded in the practical application of the scientific method as described above. Consider the simple elegance of the Copernican model of the solar system as opposed to the ungainly complexity of the Ptolemaic system based on a stationary Earth and epicycles, an auxiliary hypothesis to defend the theory against falsification.
In the hydrologic context this principle has been somewhat neglected, possibly because reductionism coupled with a deep belief in the correctness of scaling up "small-scale physically-based" models (Beven, 1987) has been the dominant paradigm. Recently, Jakeman and Hornberger (1993), drawing on earlier work as well as their own, have forcibly advocated the principle of parsimony. They argued that if only streamflow response data are available, simple models with four or five parameters based on a quick- and slow-flow conceptualisation provide an adequate fit to the data. Furthermore, they argued that addition of greater model complexity and associated parameters requiring calibration leads to no significant improvement in fit yet introduces poorly identified parameters. In essence, Jakeman and Hornberger (1993) articulate the principle of parsimony in terms of a well-posed model. This principle must be embodied in the model filter.

4.4 STATISTICAL TECHNIQUES FOR MODEL CALIBRATION AND HYPOTHESIS TESTING

Hydroclimatic data are typically corrupted by error requiring the adoption of statistical techniques to test model hypotheses. It is therefore, a logical extension of the statistical techniques used for model calibration that the calibration statistics be used to gain insight into model structure and to test model hypotheses. Mein and Brown (1978) presented one of the first examples of how calibration statistics may be used to gain insight into model structure. They used parameter variances and the parameter covariance matrix to make inferences about the relationship between fitted model parameters and measurable catchment characteristics. Another example is provided by Gupta and Sorooshian (1983) who demonstrated the use of response surface plots to aid in the reparameterisation of a rainfall-runoff model calibrated to a streamflow hydrograph. Reparameterisation led to the identification of a unique set of optimum model parameters that was not previously identifiable.

Beck (1987) reviewed statistical techniques for parameter optimisation and prediction uncertainty estimation in water quality and ecosystem modelling. He provides examples of how control theory and recursive parameter estimation algorithms can be used in model identification. Testing for time variance in model
parameters identified flawed model hypotheses. Sloan et al. (1994), who used the extended Kalman filter algorithm to optimise the parameters of a simple stream nitrate concentration model, describe a recent example of this technique. The model used rainfall, streamflow, temperature and rainfall nitrate concentration as inputs to predict the stream nitrate concentration under forested or cleared conditions. An inconsistency in the function used to represent the effects of deforestation and subsequent regeneration on the catchment nitrate balance was manifest through an unstable estimate of one of the model parameters. After the biomass change function had been manually altered to produce stable parameter estimates the model was capable of matching the seasonal trends in stream nitrate concentration following catchment deforestation. However, it failed to match short-term stream nitrate concentration variability.

Wheater et al. (1986) experimented with what they called the formal use of prior subjective interpretation to improve the performance of a conventional ordinary least squares calibration algorithm for optimising the water balance parameters of the Birkenes model (Christophersen and Wright, 1981). They subjectively assigned different objective functions to individual parameters or pairs of parameters by relating different portions of the streamflow record to specific model hypotheses and their related parameters, for example the baseflow threshold parameter was assigned an objective function based on minimum flow periods. Each parameter was then optimised in turn to its specific objective function in an iterative manner.

They tested the methodology by generating two synthetic streamflow time series with the Birkenes model to which they then calibrated the model. Further tests were performed by introducing random errors in both the input and output time series data and by summing two synthetic streamflow time series in unequal proportion. The method accurately located the parameter set used to generate the synthetic data in seven out of eight cases for which there was no data error. For the case where data errors were introduced the method performed less convincingly. Overall the method improved the identifiability of several of the more difficult parameters but offered no improvement over conventional techniques for dealing with correlated parameters.
The GLUE (Generalised Likelihood Uncertainty Estimation) method is a statistical technique developed explicitly to deal with the problems of multiple local optima parameter sets (Beven and Binley, 1993). It is a Bayesian Monte Carlo based technique which permits the calculation of uncertainty bounds on model output to deal with sets of model parameters and their associated interactions rather than determining optimum values for each parameter individually and then dealing with parameter correlations separately. The GLUE method was derived for use with spatially distributed models (e.g. Freer et al., 1996) but the principles and the technique are equally applicable to conceptual catchment models.

4.5 Increasing the Power of Data: Tracers and Other Hydrologic Data

The principle of calibration to multi-response data relates to Popper’s (1959) ideas about auxiliary hypotheses. If the introduction of an auxiliary hypothesis increases the degree of falsifiability of a theory then it has strengthened the theory. If the introduction of a catchment solute balance to a conceptual catchment model allows it to be more easily falsified then it has strengthened the model; the model describes more of the catchment behaviour.

If the scope of the available data increases then one can reasonably expect the data to exert more power to discriminate between hypotheses. Suppose, in addition to streamflow data, spatially integrated or distributed data such as soil moisture, piezometric levels, or chemical stream loads, are observed. Obviously then model hypotheses pertaining to the prediction of these data can be tested. Furthermore, because stream chemical loads are dependent on flow paths through a catchment, testable hypotheses about how flow paths and chemical loads are related may offer greater insight into catchment processes then either could independently.

Kuczera (1983b) demonstrated how throughfall and soil moisture data can be pooled with streamflow data to reduce the uncertainty in predictions made by a catchment model calibrated with the aid of Bayesian nonlinear regression. He also demonstrated (Kuczera, 1990) how response surface plots can be used to assess
model nonlinearity and how it is related both to model structure and the level of information available for model calibration. Soil moisture data was also used to supplement or check the validity of models calibrated to streamflow by Johnston and Pilgrim (1976) and Langford et al. (1978).

To date only limited use of isotopic and solute tracer data in the development of catchment models has been reported. Brown (1986) demonstrated how information on the dissolved solids load of the Housatonic River, Connecticut, could be used to infer values for two of the three parameters of a model calibrated to streamflow data.

The Birkenes model was developed as part of a programme to understand the acidification of surface waters in southern Norway (Christophersen et al., 1982) (see also Section 2.4). In the first of a two paper set de Grosbois et al. (1988) demonstrated how concurrent calibration of the Birkenes catchment model to streamflow and synthetic $^{18}$O data could reduce the uncertainty in fitted model parameters as compared to calibrating to streamflow data alone. An automatic calibration scheme was used which incorporated both ordinary least squares and weighted least squares objective functions. When the methodology was applied to field data from the Birkenes catchment parameter uncertainty was reduced for three of the five calibrated parameters (Hooper et al., 1988). The model produced predictions of streamflow that matched the observed data very well. However, the predicted $^{18}$O concentration of the streamflow failed to match the observed variability over almost half of the calibration record. Over the remainder of the record the predicted variability far outweighed that observed. The authors concluded that the two-store Birkenes model was over-parameterised, and that only one conceptual store could be justified given the data available. The model violated the principle of parsimony.

Stone and Seip (1989) and Lundquist et al. (1990) present updated conceptualisations for the Birkenes model based on the research of Hooper et al. (1988). The revised model is capable of matching the observed streamflow very well and improvements in predicted streamflow solute concentrations have been noted. However, it is considered that flow paths within the catchment are still not well simulated.
Lundquist et al. (1990) attempted to further develop the use of isotopic tracers in the application and development of the Birkenes model by forcing the chemical composition of water stored in different conceptual model stores to match the chemical composition of the different soil horizons of the Birkenes Catchment. They manually calibrated a modified version of the Birkenes water balance model to a one-year record of daily streamflow data. They then tested two hypotheses concerning the degree of mixing of precipitation with soil and groundwater by comparing the predicted concentrations of $^{18}$O in the stream flow with the observed concentrations. The assumption that precipitation is well mixed with both soil water and groundwater provided the best match to observed streamflow $^{18}$O concentrations. However, the predicted concentrations of Al, $\text{H}^+$, and Ca did not match the observed concentrations well and Lundquist et al. (1990) recommended further field observations to determine whether the chemical characteristics of the different soil horizons are stable during the course of individual storm events.

In recognition of the power of multi-response time series to provide more rigorous tests of model structure, Yapo et al., (1998) described a new technique, the MOCOM-UA algorithm, to increase the efficiency of the calibration of conceptual catchment models using more than one objective function. The technique is based on the Shuffled complex evolution method (Duan et al., 1992, 1994; Yapo et al., 1996). They demonstrated the method with a case study in which the Sacramento water balance model was calibrated to daily streamflow data for a single year of record but in which two objective functions were formulated from the data.

4.6 CHOOSING AN APPROACH TO HYPOTHESIS TESTING

There are several formal approaches for testing model hypotheses when $m$ time series of response data $D = \{D_1, D_2, \ldots, D_m\}$ are available. In the simplest case the catchment model can be described by the regression model

$$ q_t = f(x_t, \beta) + \epsilon_t \quad t = 1, \ldots, n \quad [4.2] $$
where $\mathbf{q}_t$ is an $m$-vector of observed responses for time step $t$; $f()$ is the catchment model used to predict $\mathbf{q}_t$ given $\mathbf{x}_t$, the vector of measured catchment inputs, and parameter vector $\mathbf{\beta}$; and $\mathbf{e}_t$ is an $m$-vector of random errors. Consider three Bayesian approaches for testing whether the model $f()$ is consistent with the data $\mathbf{D}$:

i) Fit the model to each response time series yielding $m$ posterior probability density functions $\xi(\mathbf{\beta}_i \mid \mathbf{D}_i)$, $i = 1, \ldots, m$. Then test the hypothesis that $\mathbf{\beta}_1 = \mathbf{\beta}_2 = \ldots = \mathbf{\beta}_m$.

ii) Fit the model to the $i$th response time series yielding the posterior probability density function $\xi(\mathbf{\beta}_i \mid \mathbf{D}_i)$. Then compute the prediction limits for the other responses. Check whether the observed responses are consistent with the prediction limits.

iii) Jointly fit the model to the $m$ response time series yielding the posterior probability density function $\xi(\mathbf{\beta} \mid \mathbf{D}_1, \ldots, \mathbf{D}_m)$ with most probable parameter vector $\hat{\mathbf{\beta}}$. Estimate the residual time series

$$\mathbf{e}_t = \mathbf{q}_t - f(\mathbf{x}_t, \hat{\mathbf{\beta}}), \quad t = 1, \ldots, n$$

and examine the residuals looking for non-random behaviour. If the residual time series appear consistent with white noise then one can conclude there is insufficient evidence to reject the catchment model hypotheses $f()$.

The first approach, first introduced by Kuczera (1983b), is based on the notion of the compatibility of the response data $\{\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_m\}$ with respect to the model. Kuczera (1983b) showed that pooling incompatible data to calibrate a lumped catchment model may introduce bias into the posterior parameter vector $\hat{\mathbf{\beta}}$. More importantly, he points out that any incompatibility between the posterior parameter distributions derived from fitting to different time series data indicates a flaw in the model structure. Unfortunately, compatibility testing is dependant on the assumption that the probability density function $\xi(\mathbf{\beta}_i \mid \mathbf{D}_i)$ can be approximated by a multivariate normal distribution. It should also be noted that Kuczera’s aim was to improve the precision of poorly identified model parameters as a step to finding regional relationships between model parameters and catchment characteristics, not to test the model hypotheses.
The second approach also relies on the assumption that the probability density function can be approximated by a multivariate normal distribution. Therefore, of the three approaches, only the third is sufficiently robust in practice. Given the evidence regarding multiple optima and nonlinear response surfaces (Ibbitt and O’Donnell, 1974; Alley, 1984; Gan and Burges, 1990a, 1990b) the assumption of a multivariate normal distribution for the probability density function is unlikely to be satisfactory. Only the third approach requires no such assumption.

4.7 SUMMARY

Conceptual catchment models have been valuable tools in the practice of engineering hydrology for many decades and new applications are currently under study. However, many such models are unreliable when used to make hydrologic predictions beyond the range of the data used in calibration. In addition, many are lacking in physical meaning both in terms of model structure and in the parameter values required to achieve an acceptable prediction of a streamflow hydrograph.

Some authors, chiefly Beven (1987) and Klemes (1988), are highly critical of the lack of physical meaning in conceptual catchment models and of the lack of a scientifically sound basis for the testing of the hypotheses upon which such models are based.

A relatively small body of research has demonstrated that conceptual catchment models can have a physical basis and that this physical basis can be tested and falsified. Statistical optimisation techniques have been widely used to calibrate conceptual catchment models and the resulting statistics have been demonstrated to yield important information about model structure and individual model hypotheses. To date, however, no systematic methodology has been presented that explicitly shows how all these techniques should be used in a systematic way in developing a conceptual catchment model from first principles. Nor has any methodology to quantify the benefits of rigorous testing of conceptual model structure and consistency with observed catchment responses been proposed.
Proposed here is a methodology to more rigorously test conceptual model structure and consistency with field observations, in line with the scientific method as proposed by Popper (1959). The objective is to filter out as many models as possible using two guiding principles: 1) Strive for parsimonious models which are demonstrably well posed; and, 2) Use data sets with the maximum power for discriminating between model hypotheses. It will be demonstrated that the proposed methodology can be used to quantify and eliminate random error from model predictions. Furthermore, though systematic error cannot be quantified, it will be argued that application of the methodology will lead to the reduction of systematic error and increase the confidence that may be placed in model predictions. The methodology is independent of the specific algorithm used to find the posterior parameter distribution, provided that the global optimum is found. Post calibration diagnostic tests are more critical to the testing of model hypotheses than the method used to locate the mode of the posterior parameter distribution.

In the following sections, the filter methodology will be demonstrated by tracing the development of a conceptual catchment model. The model structure is described and then actively scrutinised by calibration to streamflow, stream salinity and piezometric head time series data and the application of established statistical techniques, used to their full, so far unrealised, potential.
5 STATISTICAL FRAMEWORK

In the previous Chapter a systematic methodology to rigorously test the hypotheses that constitute a conceptual catchment model was outlined. The methodology requires the use of statistical tools to deal with errors in the data used to assess conceptual catchment models and to deal with the practical problems associated with identifying an optimal parameter set. Before embarking on a case study to demonstrate how these statistical tools can be used to implement the proposed methodology the statistical framework for dealing with these issues needs to be discussed.

NLFIT is a suite of Bayesian, nonlinear regression programs (Kuczera, 1989, 1994) that includes statistical tools to perform all the practical tasks associated with the implementation of the proposed methodology. The case study model, presented in the next chapter, will need to be calibrated to observations of catchment responses. There will be a need to concurrently calibrate the model to multiple response time series and NLFIT allows for this.

This Chapter will describe the principles of the calibration procedure and the assumptions made in its implementation. Diagnostic procedures used to test the validity of those assumptions will also be presented. Valuable information about model structure and the relative value of different sources of data may be inferred from the results of these diagnostics and this will also be discussed. In particular, the nature of the objective function response surface with respect to the parameters will need to be examined to assess model linearity and this will be discussed. A description of how the calibration statistics may be used to produce a quantitative measure of the uncertainty inherent in model predictions of future or unobserved responses follows.

5.1 BASIC PRINCIPLES

The regression model offers the simplest description of causal and stochastic interactions affecting catchment response. The regression model is
\[ q_t = f(x_t, \beta) + \epsilon_t \quad t = 1, \ldots, n \]  

where \( q_t \) is the observed catchment response for time-step \( t \), \( f() \) is the catchment model used to predict \( q_t \) given \( x_t \), the vector of measured catchment inputs, and parameter vector \( \beta \), and \( \epsilon_t \) is an additive random error. The structure and validity of the catchment model \( f() \) and the appropriate values of its parameter vector \( \beta \) are the primary objects of interest but their characteristics cannot be ascertained without consideration of \( \epsilon_t \) and this will be explored first.

### 5.2 Error Models

#### 5.2.1 Least Squares Error Model

The simplest error model applicable to the regression model is the least squares error model. It is assumed that:

1) the catchment model \( f() \) predicts \( q_t \) without systematic bias and the expected value of \( \epsilon_t \) is zero;
2) the variance of \( \epsilon_t \) is constant, \( \sigma^2 \);
3) the errors are statistically independent of each other; and,
4) the errors are normally distributed with mean 0 and variance \( \sigma^2 \):

\[
\epsilon_t \sim N(0, \sigma^2) \quad [5.2]
\]

The assumption that \( f() \) predicts \( q_t \) without systematic bias is the most important in this context. If the model does not correctly predict the value of the observed response on average then it is not a valid model of the system. Diagnostic tools to assist in the identification of systematic error are required and where systematic errors are identified remedial procedures to eliminate or circumvent them are also necessary.

#### 5.2.2 Error Model Diagnostics

The adequacy of the least squares error model must be assessed against assumptions 1 to 4 above and this is most simply done with diagnostic residual plots (Draper and
Smith, 1981). Any suspected violation of assumptions can then be investigated more fully with specialised diagnostic tests. A residual $\hat{\varepsilon}_t$ is an estimate of the random error $\varepsilon_t$ and the standardised residual $z_t$ is defined as

$$z_t = \frac{\hat{\varepsilon}_t}{\sigma}$$  \hspace{1cm} [5.3]

If the least squares error model is adequate the standardised residuals should be distributed as

$$z_t \sim \text{N}(0,1)$$  \hspace{1cm} [5.4]

and 95% of the $z_t$ values should lie between -2 and 2. Absolute values of $z_t$ outside this range suggest outliers or unusually large deviations that should be investigated.

When the $z_t$ are plotted as a function of the predicted values they should lie in a roughly parallel band about $z_t = 0$ if they indeed have constant error variance. Violations of the least squares assumptions are indicated by any non-random pattern in the plot such as snaking about the line $z_t = 0$ or an increasing spread of $z_t$ with increasing values of the predicted response.

Trends or periodicities in the residuals can be detected when the $z_t$ are plotted as a time series. Plots with long runs of either positive or negative values of $z_t$ require further investigation and NLFIT uses the non-parametric runs test of Mendenhall and Schaeffer (1973) which reports a Z-statistic which is normally distributed with zero mean and variance 1. If the absolute value of the Z-statistic exceeds 2 then there is a less than 5% chance that the residuals come from a statistically independent error process.

Normal probability plots also aid in the assessment of the adequacy of the assumption of normality of the residuals, which should lie on a straight line on such a plot. For small samples, however, sampling variability may cause deviations from a straight line that are not associated with non-normal behaviour, NLFIT applies the Kolmogorov-Smirnov test (Bickel and Doksum, 1977) to assist in the assessment.
The cumulative periodogram (Box and Jenkins, 1976) is very sensitive to periodicities in the residuals provided that their variance is constant. As with the normal probability plot, well behaved residuals will lie on a straight line but this may not be the case for very small samples and the Kolmogorov-Smirnov test is again applied to test significance of any deviations from a linear plot.

Time dependence of the residuals are assessed with autocorrelation and partial autocorrelation plots for which the lagged residuals should lie within the 95% confidence limits for independent data if they are statistically independent of each other.

### 5.2.3 A More General Error Model

Where the error model diagnostics indicate that the least squares error assumptions (1 to 4 above) are violated, parameter uncertainty measures will be distorted and a more general error model is required (Kuczera, 1983a, b). NLFIT allows for the application of an error model in which rather than being additive as shown in [5.1] the random error is embedded in an autoregressive-moving average process with p autoregressive and q moving average parameters ARMA(p,q) (Box and Jenkins, 1976):

$$\eta_t = \phi_1 \eta_{t-1} + \ldots + \phi_p \eta_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$  \[5.5\]

where $\phi_1, \ldots, \phi_p$ are autoregressive parameters, $\theta_1, \ldots, \theta_q$ are moving average parameters and $\eta_t$ is a transformed error defined by

$$\eta_t = Q_t - E(Q_t)$$  \[5.6\]

which is related to the observed response $q_t$ by the Box-Cox transformation (Box and Cox, 1964)

$$Q_t = \frac{(q_t + K)^\lambda - 1}{\lambda} \quad \text{if } \lambda \neq 0$$  \[5.7\]

$$Q_t = \log_e (q_t + K) \quad \text{otherwise}$$
where \( \lambda \) and \( K \) are transformation parameters, and to the predicted response also by the Box-Cox transformation

\[
E(Q_t) = \frac{(f(x_t, \beta) + K)^\lambda - 1}{\lambda} \quad \text{if } \lambda \neq 0 \tag{5.8}
\]

\[
E(Q_t) = \log_\lambda (f(x_t, \beta) + K) \quad \text{otherwise}
\]

The predicted response \( f(x_t, \beta) \) is no longer interpreted as the expected response but as the 50th percentile response at time-step \( t \).

It is assumed that the variance of the error \( \varepsilon_t \) is constant, therefore it follows from [5.5] that the variance of the transformed error \( \text{Var}(\eta_t) \) must also be constant. The Box-Cox transformation has the effect of transforming the observed data so that the \( \text{Var}(\varepsilon_t) \) and \( \text{Var}(\eta_t) \) are constant. It can be shown that

\[
\text{Var}(q_t) = [f(x_t, \beta) + K]^{2(\lambda - 1)} \text{Var}(\eta_t) \tag{5.9}
\]

which demonstrates that for the Box-Cox transformation the response variance \( \text{Var}(q_t) \) is proportional to some power of the predicted response \( f(x_t, \beta) \). When \( \lambda = 1 \), the variance of the observed response \( q_t \) is also constant, but for \( \lambda < 1 \), the variance of \( q_t \) increases as \( f(x_t, \beta) \) increases with the rate dependant on \( \lambda \).

Even though the Box-Cox transformation may stabilise the variance of the transformed error \( \eta_t \), it may still exhibit a time dependence that the predictive model \( f(x_t, \beta) \) fails to account for. The ARMA model is introduced to filter out this time dependence so that the random error \( \varepsilon_t \) behaves like a least squares error. Note that the general error model defined by [5.5] to [5.8] reduces to the least squares error model for an ARMA(0,0) process and Box-Cox \( \lambda = 1 \).

### 5.3 Parameter Optimisation Using Bayesian Inference

Having selected a predictive model, \( f(\cdot) \), and observed data \( D \), and considered an appropriate error model the next task is to determine the parameter vector \( \beta \) that, in some sense, produces the best fit to the data \( D \). Bayes Theorem provides a
mechanism of estimating the posterior distribution of $\beta$ given $D$ and any prior information about $\beta$.

$$\xi(\beta|D) \propto L(\beta|D) \xi(\beta)$$

[5.10]

where $\xi(\beta|D)$ is the posterior probability density function (pdf) of $\beta$, $L(\beta|D)$ is the likelihood function and $\xi(\beta)$ is the prior pdf of $\beta$. The prior pdf encapsulates all that is known or inferred about $\beta$ before the new data $D$ is considered, it may include experimentally determined values of physical catchment properties, the results of previous model calibrations or simply “best guesses” at model parameters.

The likelihood function modifies the prior pdf of $\beta$ given the prior knowledge about $\beta$ and the data $D$. To understand the nature of the likelihood function, consider the regression model (Equation [5.1]) and the error model outlined in Equations [5.5] to [5.8]. Form the augmented parameter vector

$$\gamma = \begin{pmatrix} \beta \\ \phi \\ \theta \end{pmatrix}$$

[5.11]

where $\phi$ is the vector of autoregressive parameters and $\theta$ is the vector of moving average parameters used in the generalised error model (Equation [5.5]).

If the prior pdf of $\gamma$ can be approximated by a multivariate normal distribution with mean $\gamma_0$ and covariance matrix $\Sigma_p$, then it can be expressed as

$$\xi(\gamma|\gamma_p, \Sigma_p) \propto \exp\left\{-\frac{1}{2} \left[(\gamma - \gamma_p) \Sigma_p^{-1} (\gamma - \gamma_p)\right]\right\}$$

[5.12]

Updating this prior pdf given the observed time series data $D$ yields the posterior pdf of $\gamma$. 

44
\[ \xi(\gamma|D, \gamma_0, \Sigma_p, \lambda, K) \propto \exp\left\{-\frac{1}{2} \psi(\gamma)\right\} \]  

where

\[ \psi(\gamma) = (n - p) \log \varepsilon^T (\gamma - \gamma_p) \Sigma_p^{-1} (\gamma - \gamma_p) \]  

where \(\varepsilon^T\) is the transpose of the error vector, \((\varepsilon_{p+1}, \ldots, \varepsilon_n)\) and \(p\) is the number of autoregressive parameters (Equations [5.5]). The posterior pdf of \(\gamma\) is conditioned on the Box-Cox transformation and ARMA(p,q) model specified \textit{a priori}.

To locate the vector of maximum posterior probability density for \(\beta\) requires the minimisation of \(\psi(\gamma)\) which becomes the optimisation objective function

\[ \min_{\gamma} \psi(\gamma) \]  

If there is no prior information on \(\gamma\), then the inverse of the inverse of the prior covariance matrix \(\Sigma_p^{-1}\) is zero. Assuming an ARMA(1,0) autoregressive model the optimisation problem reduces to

\[ \min_{\beta} \sum_{t=1}^{n} \varepsilon_t^2 = \min_{\beta} \sum_{t=2}^{n} \left(\left[q_t + K \right]^T \left[\left[f(x_t, \beta) + K \right] - \phi_1 \left[q_{t-1} + K \right] - \phi_2 \left[f(x_{t-1}, \beta) + K \right]\right]^T \right) \]  

Note that when no Box-Cox transformation is applied (\(\lambda = 1\) and no time dependence is present in the errors (\(\phi_1 = 0\)) [5.16] simplifies to

\[ \min_{\beta} \sum_{t=1}^{n} \varepsilon_t^2 = \min_{\beta} \sum_{t=1}^{n} \left[q_t - f(x_t, \beta)\right]^2 \]  

which is the least squares minimisation problem.

NLFIT solves this minimisation problem iteratively, using the gradient based, Gauss-Marquardt algorithm (Bard, 1974; Press \textit{et al.}, 1986). NLFIT allows the search direction to be varied between the steepest decent and Gauss-Newton directions for
each search iteration. A scaling parameter $\lambda_M$ controls the search direction. When $\lambda_M$ is zero, the search direction is given by the Gauss-Newton direction which assumes that the response surface contours are elliptical. As the value of $\lambda_M$ increases, the search direction approaches the direction of steepest decent and the step size decreases. Additionally, the user may manually vary the step size as a fraction of the full Gauss-Marquardt step.

### 5.3.1 Multiple Response Optimisation

NLFIT facilitates concurrent calibration to multi-response time series by calculating a disturbance covariance matrix $\Omega$. It is a condition of the regression model (Equation [5.1]) that the random error $\varepsilon$ for each response is cross correlated with the random errors of the other responses. The $(i, j)^{th}$ element of the covariance matrix is given by

$$\Omega_{ij} = \frac{1}{n} \sum_{t=1}^{n} \varepsilon_{it} \varepsilon_{jt}$$

where the $\varepsilon_{it}, i = 1, ..., m; t = 1, ..., n$ are the disturbances of the $m$ concurrent response time series. The inverse of $\Omega$ is called the disturbance weight matrix and is used to weight the contribution of each time series to the posterior distribution of the model parameters, $\beta$:

$$\xi(\beta|D_1, ..., D_m, \beta_p, \Sigma_p, \Omega) \propto \exp \left\{ -\frac{1}{2} \psi(\beta) \right\}$$

where $D_1, ..., D_m$ are the $m$ multi-response time series; $\beta_p$ is the $\beta$ given $D_1, ..., D_m, \Sigma_p$ is its covariance; and, $\psi(\beta)$ is the objective function to be minimised:

$$\psi(\beta) = \sum_{t=1}^{n} \varepsilon_{t}^{'} \Omega^{-1} \varepsilon_{t} + (\beta - \beta_p) \Sigma_p^{-1} (\beta - \beta_p)$$

The disturbances can be obtained either by firstly calibrating the model to each time series independently or by assuming some arbitrary values. In either case NLFIT will provide final values at the end of each calibration which should be used iteratively.
until the assumed disturbances are the same as the calculated values. Failure to specify appropriate weights can result in a poor match between predicted and observed values for some of the $D_m$.

### 5.3.2 Parameter Transformations

If the assumption that the disturbance vector distribution is approximately linear in the vicinity of the mean posterior parameter vector is violated, fitting to transformed parameter values rather than to their actual values may extend the region over which the linearity assumption holds. Furthermore, parameter transformations may be used to impose constraints on parameters to prevent infeasible values from being derived during unconstrained optimisation. Three useful parameter transformations are available in NLFIT and these are shown in Table 5.1.

<table>
<thead>
<tr>
<th>Transformation</th>
<th>Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logarithmic</td>
<td>$\beta_T = \log_e \beta$ $\beta_T$ can range from $-\infty$ to $\infty$ during unconstrained optimisation, with transformation $\beta$ can range from 0 to $\infty$</td>
</tr>
<tr>
<td>Binomial</td>
<td>$\beta_T = \log_e \left(\frac{\beta}{1-\beta}\right)$ $\beta_T$ can range from $-\infty$ to $\infty$ during unconstrained optimisation, with transformation $\beta$ can range from 0 to 1</td>
</tr>
<tr>
<td>Square Root</td>
<td>$\beta_T = \sqrt{\beta}$ The square root transformation sometimes helps to make the response surface contours more elliptical.</td>
</tr>
</tbody>
</table>

### 5.4 RESPONSE SURFACE PLOTS

Response surface plots provide a two dimensional view of the posterior distribution of model parameters. If the posterior distribution is multivariate normal, then the posterior density is constant over an ellipsoidal surface in the parameter space. The intersection of the ellipsoidal surface and the plane containing the feasible values of two parameters is the boundary of an ellipse. The interior of the ellipse is called the conditional probability region. If the parameter pair is multivariate normal then
contours of the objective function will be elliptical (Sorooshian and Gupta, 1983; Kuczera, 1990). Deviations from the multivariate normal distribution will be manifest as non-elliptical contours of the objective function value. Furthermore, the presence of multiple local optima may be visible in the response surface.

When significant correlation exists between parameters, the conditional probability region may become elongated in directions other than parallel to the plane defined by any two parameter axes. In such a case, objective function contours drawn on the plane parallel to parameter axes will include parameter values that are far from the optimum. In this case a response surface plotted on a plane parallel to the axes of two principal components will provide more information of the nature of the objective function in the vicinity of the optimum parameter values. The rank 1 principal component corresponds to the longest axis of the ellipsoid and the rank 2 principal component to the second longest axis of the ellipsoid and so on (Kuczera, 1990).

5.5 Prediction and Confidence Limits

Prediction and confidence limits provide information on the uncertainty of model predictions of individual catchment responses. Prediction and confidence limits are generated using a Monte Carlo approach in which parameter values are randomly sampled from their posterior distributions assuming that they are indeed normally distributed. This was preferred over making a linear approximation to the normal probability density function due to the extreme nonlinear behaviour commonly displayed by conceptual catchment models.

All confidence limits presented here were determined using \( n = 300 \) randomly selected samples using the following Monte Carlo algorithm:

1. Randomly sample \( n \) times a parameter vector from the approximate posterior distribution

\[
\beta_i \leftarrow N\left(\hat{\beta}, \Sigma_\beta\right), \quad i = 1, \ldots, n
\]  

[5.21]
2. For each sampled parameter vector $\beta_i$ generate $n$ responses \( \{ q_i, i = 1, \ldots, n \} \).

\[
q_i \leftarrow f(x, \beta_i), \quad i = 1, \ldots, n
\]  

(5.22)

3. Rank \( \{ q_i, i = 1, \ldots, n \} \) to obtain selected percentiles of the Bayesian distribution.

Note that in equation [5.1] the regression model includes a random error $\varepsilon_i$. Where the random error is not included in the calculation of the predicted response, as is the case here, the ranked responses only reflect the effect of parameter uncertainty. If the random error $\varepsilon_i$ is included in the ranked responses prediction limits are generated which fully describe the uncertainty in predicted responses. Here confidence limits will be used to give a measure of the uncertainty on the prediction of unobserved catchment responses, clearly no estimate of $\varepsilon_i$ is possible if the true value of the predicted response is not known.

5.6 Summary

In the previous chapter a methodology to rigorously test conceptual model structure and consistency with field observations was proposed. The methodology is based on the notion of the scientific method as proposed by Popper (1959).

A lumped catchment model can be couched in terms of a simple regression model. The careful selection of an appropriate error model is then critical to assessing the ability of the model to extract all the information about the model parameters from calibration data. The Bayesian nonlinear regression suite, NLFIT, provides both the means to locate the global optimum posterior parameter set and the post-calibration diagnostics required to test the validity of assumptions made about the nature of model errors.

In the following chapters these techniques will be used to vigorously challenge the assumptions made in the development of a conceptual catchment model. The rationale and structure of a conceptual model for first order catchments in the south-west of Western Australia will first be described. This will be followed by a description of the Salmon Catchment, which was chosen as a case study.
6 CASE STUDY: DEVELOPMENT OF THE CATPRO MODEL

Popper’s (1959) theory of the scientific method requires that a hypothesis must be formulated in terms that allow or invite falsification. Complexity, in the form of auxiliary hypotheses, which serve to defend it from falsification, is to be avoided. This implies that the narrower the range of possibilities that a hypothesis allows, the more easily it is to falsify, the higher is its degree of testability. To be accepted into the realm of what is considered scientific knowledge the hypothesis must then be exposed to and withstand active and rigorous testing. Translating these requirements into a predictive hydrologic model requires that the simplest model that is consistent with observation is to be preferred. Furthermore, the more a model predicts about the physical system it is intended to simulate the more easily it is falsified; a catchment model calibrated to more than one catchment response is expected to exhibit a higher degree of testability than a model calibrated to a streamflow hydrograph alone.

Described here is a conceptual catchment model developed in line with Popper’s (1959) theory of the scientific method. Hypotheses about the generation of streamflow by different physical mechanisms are formulated from physical evidence and clearly stated as required by Popper (1959). Parsimonious functions describing hypotheses about streamflow generation at the catchment or hillslope scale are then employed to construct the model. In later sections the hypotheses formulated here will be actively scrutinised against empirical observations using the most powerful statistical tools available.

6.1 MODEL RATIONALE AND PHILOSOPHY

It has been established that the existence of an ephemeral, perched water table at shallow depth plays an important role in the hydrology of temperate and semiarid Australian catchments (see for example O’Loughlin, 1981, 1992; Smith and Hebbert, 1983; Moore et al., 1986, George and Conacher, 1993). Subsurface saturation may be caused by the presence of a highly impermeable layer; an exponential decrease in
hydraulic conductivity with depth (Beven, 1984); or in anisotropic soil profiles, by
the convergence of lateral unsaturated flow (Zaslavsky and Rogowski, 1969). The
perched water table may intersect the soil surface and produce saturation overland
flow (Dunne and Black, 1970). It may be the primary source of streamflow and the
factors that affect its spatial and temporal extent will also directly affect recharge to
any deep, regional aquifer, especially if vertical flow is dominated by flow through
macropores (Johnston, 1987a).

Smith and Hebbert (1983) demonstrated the highly dynamic nature of the response of
a perched water table to individual storm events. This dynamic behaviour can be
modelled with the use of finite difference or finite element techniques (e.g. Freeze,
1971; Beven, 1977; Smith and Hebbert, 1983; Hurley and Pantelis, 1985; Dawes and
Hatton, 1993) or, more recently with boundary integral methods (Volker and Read,
1990). Sloan and Moore (1984) demonstrated how relatively simple, storage-
discharge models can be used to predict the lateral discharge from a hillslope and to
predict the approximate shape of the water table under idealised conditions. Knapp
et al. (1975) used a logarithmic recession curve method to simulate a dynamic
subsurface lateral flow component in their catchment model but did not link the
process to a variable area of surface saturation. In contrast Federer and Lash (1978)
developed a conceptual catchment model in which the saturated fraction of the
catchment is an exponential function of the soil water storage but in which perched
groundwater flow (interflow) is a constant proportion of the unsaturated storage
below the root zone. In the HYDROLOG model, as used by Chiew and McMahon
(1990a, b), interflow is assumed proportional to the product of the relative soil
moisture storage and the daily infiltration; no allowance is made for the possibility of
a perched water table intersecting the soil surface.

Ruprecht and Sivapalan (1991) and Sivapalan et al. (1996a) describe a conceptual
catchment model developed for the south-west of Western Australia in which a
perched groundwater system in the vicinity of the stream channel contributes to
streamflow generation. This model will be discussed further in Chapter 10.
Given the importance of shallow perched aquifers to temperate and semiarid Australian catchment hydrology it was considered essential that the model explicitly account for the dynamic nature of the area of subsurface saturation and its inter-relationship with other catchment processes, especially recharge. A parsimonious model structure was sought to balance the inherent conflict between the need to adequately describe hydrologic processes and the need to calibrate to limited data.

6.2 CATCHMENT WATER BALANCE MODEL

The CATPRO model is a one-dimensional conceptualisation of what are considered to be the dominant hydrologic processes operating in temperate and semiarid Australian catchments. It has a quasi two-dimensionality in that an attempt is made to model the dynamic aspects of subsurface saturation. The model consists of three conceptual stores and the equations that govern the flow of water between them. A schematic representation is presented in Figure 6.1. The required inputs to the model are daily rainfall and potential evapotranspiration. A multiplier \( et\_mult \), the value of which is determined as part of the calibration process, is applied to evapotranspiration to ensure a long-term water balance. This is necessary because the values of meteorological variables used in the calculation of potential evapotranspiration measured at ground level may not be representative of those existing within a tall forest canopy (Jarvis, 1980). A monthly streamflow record is the minimum calibration data requirement. Data on other catchment responses may also be used; these will be discussed later. The model runs on a daily time-step but an algorithm to increase the number of time-steps performed on days of very high rainfall is included. A list of model parameters and their definitions is given in Table 6.1.

6.2.1 Rainfall Interception

Rainfall must pass through the leaf canopy before arriving at the ground surface and a proportion is trapped and lost to evaporation. Daily throughfall \( tf \) (mm) is assumed to be linearly dependent on the daily rainfall \( P \) (mm),
\[ tf = tfm \times P + tfb \quad \text{if } tfm \times P + tfb > 0 \]
\[ tf = 0 \quad \text{otherwise} \]  \[ \text{[6.1]} \]

where \( tfm \) and \( tfb \) are parameters to be determined by calibration.

**Figure 6.1:** Schematic of CATPRO conceptual catchment water balance model.

### 6.2.2 A Store

The A horizon is conceptualised as an inclined rectangular cuboid of dimensionless length \( L \) and depth \( D \) with \( D \ll L \), as shown in Figure 6.2; note that the vertical scale in the figure is exaggerated for clarity. The maximum water storage of the cuboid per unit width \( pch_{\text{max}} \) (mm) is

\[ pch_{\text{max}} = L \times D \times n \]  \[ \text{[6.2]} \]
where \( n \text{ (mm}^3\text{.mm}^{-3}) \) is the porosity of the soil material. To further simplify the problem it is assumed that the phreatic surface is flat and moves at a constant inclination to the horizontal, the degree of inclination being some unknown function of the soil properties and hillslope angle.

**Figure 6.2:** Conceptualisation of CATPRO A store showing perched water table.

If the storage \( \text{perch} \) (mm) is less than zero the store is unsaturated. If \( \text{perch} > 0 \) then a perched water table exists and extends some dimensionless distance \( W \) upslope and to a height \( X \) from the lower, downslope corner of the store. We define \( \text{wl}gth = \frac{W}{L} \) to be the dimensionless wetted length of the bottom surface of the cuboid \((0 \leq \text{wl}gth \leq 1)\); \( \text{wl} = \frac{W_0}{L} \) to be the value of \( \text{wl}gth \) at which the phreatic surface first intersects the soil surface; and; \( h = \frac{X}{D} \) to be the dimensionless height of the phreatic surface above the bottom of the cuboid at its downslope end. The storage at which the phreatic surface first intersects the soil surface \( pch\_crit \) (mm) is then

\[
pch\_crit = \frac{D \times W_0 \times n}{2} \tag{6.3}
\]

However, by definition, when the phreatic surface first intersects the soil surface
\[ W_0 = L \times w_l \]  

[6.4]

By substituting equations [6.2] and [6.4] into [6.3] we obtain \( pch_{\text{crit}} \) as a function of model parameters \( pch_{\text{max}} \) and \( w_l \),

\[ pch_{\text{crit}} = \frac{pch_{\text{max}} \times w_l}{2} \]  

[6.5]

If \( perch > pch_{\text{crit}} \) then an area of surface saturation will extend some relative distance \( sf \) along the upper surface of the store (\( 0 \leq sf \leq 1 \)). For any positive storage, values for \( sf, wlgth \) and \( h \) are then determined by similar triangles as follows:

For \( perch < pch_{\text{crit}} \):

\[ sf_r = 0 \]  

[6.6]

\[ wlgth_r = \left( \frac{2 \, w_l \times perch_{r-1}}{pch_{\text{max}}} \right)^{\frac{1}{2}} \]  

[6.7]

\[ h_r = \frac{wlgth_r}{w_l} \]  

[6.8]

For \( pch_{\text{crit}} \leq perch \leq pch_{\text{max}} - pch_{\text{crit}} \):

\[ sf_r = \frac{pch_{r-1}}{pch_{\text{max}}} - \frac{w_l}{2} \]  

[6.9]

\[ wlgth_r = sf_r + w_l \]  

[6.10]

\[ h_r = 1 \]  

[6.11]

For \( perch > pch_{\text{max}} - pch_{\text{crit}} \):

\[ sf_r = 1 - \left( \frac{2 \, w_l (pch_{\text{max}} - perch_r)}{pch_{\text{max}}} \right)^{\frac{1}{2}} \]  

[6.12]
Equations [6.5] to [6.14] provide a conceptual description of the relationship between the dynamic behaviour of the area of subsurface saturation and the volume of water in the A store. In addition, if a deeper regional aquifer is contributing to streamflow, then there may be an area of surface saturation adjacent to the stream channel associated with groundwater discharge \( sf_{min} \). The fraction of the hillslope producing saturation overland flow \( sf_{face} \) then becomes

\[
sf_{face} = sf + sf_{min}
\]  

and the daily saturation overland flow \( roff \) (mm) is calculated from

\[
roff = tf \times sf_{face}
\]  

It is assumed that the flow lines through the perched aquifer are parallel to the top and bottom of the cuboid and that the hydraulic gradient is equal to the hillslope gradient. In this case the kinematic wave model first proposed by Henderson and Wooding (1964) and latter extended by Beven (1981) and Sloan and Moore (1984) describes the relationship between storage and discharge.

\[
Q = KH \sin \theta
\]  

where \( Q \) (mm.d\(^{-1}\)) is the saturated subsurface discharge per unit hillslope width; \( K \) (mm.d\(^{-1}\)) is the saturated hydraulic conductivity; \( H \) (mm) is the saturated thickness at the downslope end of the hillslope; and, \( \theta \) is the angle of the hillslope to the horizontal. In terms of the conceptual storage model presented here

\[
iflo = T_{pch} \times h
\]  

where \( iflo \) (mm.d\(^{-1}\)) is the subsurface lateral discharge, \( T_{pch} \) (mm.d\(^{-1}\)) is the product of the saturated hydraulic conductivity and the hillslope gradient and \( h \) is the
nondimensional depth of saturation at the downslope end of the hillslope as defined above.

Vertical flow of water from the A store is assumed to be dominated by flow through macropores that may be, for example, old root channels, geological discontinuities or the result of soil heterogeneity. It is assumed that the macropores are uniformly distributed and act as pipes; that is, they transmit water at a significant rate only when saturated and only those in hydraulic contact with the perched aquifer are active at any time. It is also assumed that the depth of water perched in the shallow aquifer is small in comparison to the vertical distance to the deeper aquifer, in which case the storage of the perched aquifer has no effect on the potential gradient between the two aquifers. The volume of water leaving the A store via macropores $pref$ (mm.d$^{-1}$) is therefore calculated from

$$pref = \text{wlgth} \times k_\text{prf}$$  \[6.19\]

where $k_\text{prf}$ (mm.d$^{-1}$) is the product of the mean vertical conductance of the macropores and the potential gradient between the perched and deep aquifers.

If a perched water table exists evapotranspiration ($etA$) is extracted from it at the potential rate until either the demand is meet or the store becomes unsaturated. If the demand is not met from saturated storage then evapotranspiration proceeds from the unsaturated portion of the store at a rate determined by linear interpolation between the potential rate and zero at an unsaturated storage ($perch < 0$) determined by parameter $etAmin$, which is analogous to the permanent wilting point.

$$et_\text{unsat} = etstp \left(1 + \frac{perch}{etAmin}\right)$$  \[6.20\]

where $et_\text{unsat}$ (mm.d$^{-1}$) is the actual evapotranspiration from the unsaturated portion of the store, and $etstp$ (mm.d$^{-1}$) is the unsatisfied evaporative demand for the time-step. The variable $etstp$ is used to keep account of the unfulfilled portion of the evaporative demand for each time-step after evapotranspiration is subtracted from each conceptual store.
Evapotranspiration is subtracted from storage immediately and any unsatisfied potential evapotranspiration is carried forward to be met from the B store and groundwater store.

The storage in the A store at the end of the time-step $t$ is then

$$perch_t = perch_{t-1} + tf - roff - iflo - pref$$  \[6.21\]

If the storage at the end of the time-step exceeds the maximum allowable then the excess is assumed to contribute to saturation overland flow and $perch$ is set to $pch_{max}$ for the start of the next time-step.

### 6.2.3 B Store

Water may move radially from the macropores into the B store under a gradient in soil water potential. The rate of radial flow will be at a maximum at the commencement of flow and decrease as adjacent wetting fronts intersect then collapse to zero when the wetting fronts completely overlap. This behaviour is approximated by

$$percl_t = wlgth_t \times K_b \left(1 - \frac{B_t}{B_{max}}\right)$$  \[6.22\]

where $percl$ (mm.d$^{-1}$) is the radial flow into the B store from macropores; $K_b$ (mm.d$^{-1}$) is the horizontal hydraulic conductivity of the B store; $B$ (mm) is the water in storage at time $t$; and, $B_{max}$ (mm) is the maximum storage of the B store. $percl$ is limited to the volume of water passing through the macropores per unit time and is subtracted from the preferential recharge to the deep aquifer calculated by [6.19]. The storage is then calculated from

$$B_t = B_{t-1} + percl$$  \[6.23\]

The B store operates as a linear store from which water drains under the influence of gravity alone and evapotranspiration may be removed from it in direct proportion to its storage.
After recharge via *percl* is added, a check is performed to ensure there is sufficient storage before evapotranspiration is removed. Drainage *deep_seep* from the store then contributes to recharge to the deep aquifer at a rate fixed by model parameter *deepa*.

For $B > et\_B$

$$B_i = B_i - et\_B \quad [6.25]$$

$$B_i = B_i - deepa \quad [6.26]$$

$$deep\_seep = deepa \quad [6.27]$$

For $B \leq et\_B$

$$et\_B = B_i \quad [6.28]$$

$$B_i = 0 \quad [6.29]$$

$$deep\_seep = 0 \quad [6.30]$$

### 6.2.4 Groundwater Store

The groundwater storage *gws* is updated at the beginning of each time-step by adding the recharge that occurs via macropores and from deep drainage;

$$gws_i = gws_{i-1} + pref + deep\_seep \quad [6.31]$$

The deep aquifer acts as a linear store discharging whenever a flow threshold is exceeded, however because the native vegetation is capable of extending roots to great depth (Kimber, 1974), any unsatisfied evaporative demand is first fulfilled:
For \( gws > etstp \)

\[
et_\text{gws} = etstp
\]  

[6.32]

\[
gws_i = gws_i - etstp
\]  

[6.33]

\[
gflo = bf \left( gws - bf_{\text{crit}} \right)
\]  

[6.34]

\[
gws_i = gws_i - gflo
\]  

[6.35]

For \( gws \leq etstp \)

\[
et_\text{gws} = gws
\]  

[6.36]

\[
gws_i = 0
\]  

[6.37]

\[
gflo = 0
\]  

[6.38]

where \( et_{\text{gws}} \) is the evapotranspiration from the groundwater store, \( gflo \) is the groundwater discharge, including any loss to a regional aquifer, \( bf_{\text{crit}} \) is the threshold storage below which no groundwater discharge takes place and \( bf \) is a proportionality constant. The proportion of the groundwater discharge that may contribute to streamflow is then calculated from

\[
bflo = gflo \left( 1 - (1 - 2 \text{ leak}) \right)
\]  

[6.39]

where the parameter \( \text{leak} \) determines the proportion of groundwater discharge that leaks to or from a regional aquifer. For \( 0 \leq \text{leak} < 0.5 \) groundwater leaks from the store to a regional aquifer; for \( \text{leak} = 0.5 \) there is no leakage; and for \( 0.5 < \text{leak} \leq 1 \) leakage from outside the groundwater store contributes to streamflow.

Whenever there is baseflow an area of surface saturation \( sfmin \) exists adjacent to the stream channel regardless of whether \( perch \) is positive or not. Streamflow is then calculated from

\[
sflo = bflo + roff + iflo
\]  

[6.40]
and net groundwater recharge \( rchge \) is then calculated from

\[
rchge = \text{pref} + \text{deep_seep} - \text{et_gws}
\]  

Net recharge for any time-step may be negative, indicating evapotranspiration from the store in excess of the recharge reaching it from the \( A \) and \( B \) stores.

The change in groundwater storage between consecutive observations \( \Delta gws \) is related to the change in observed catchment averaged piezometric heads \( \Delta h \) by

\[
\Delta h = \frac{\Delta gws \times 10^{-3}}{gwsFit}
\]

where \( gwsFit \) is a proportionality constant that is equivalent to the specific yield of the deep aquifer and the factor of \( 10^{-3} \) is applied because predicted change in groundwater storage is in mm and piezometric heads are usually measured in metres relative to some datum. Using the change in storage rather than the actual storage avoids the necessity for a second nuisance parameter to allow for any offset between predicted storages and absolute values of the piezometric heads.

6.3 CATCHMENT SALINITY BALANCE MODEL

6.3.1 Chloride as a Natural Tracer

Chloride is a natural, conservative tracer of water. It is highly soluble though largely inert (Faust and Aly, 1981) and is only absorbed by plant roots in very small quantities. In Australia, chloride of oceanic origin, arriving in rainfall and as dry fallout is the dominant source of catchment chloride (Blackburn and McLeod, 1983; Keywood et al., 1997). In the south-west of Western Australia chlorine has accumulated in the deeply weathered soil profiles (Johnston et al., 1983) and is present in natural groundwaters at concentrations ranging from 40 to 63,000 mg.L\(^{-1}\) (Mazor & George, 1992).

Vertical profiles of chloride distribution are influence by the downward flux of water and water extraction by the deep rooted native vegetation (Allison, 1988; Johnston et al., 1983). Stream salinity is therefore, a reflection of the path and travel time of water through a catchment. Its temporal variation is a function of the relative
contributions of the different sources of streamflow operating in a catchment. Analysis of this variation can reveal much about catchment hydrology and the mechanisms of streamflow generation (see for example Turner et al., 1987; Johnston, 1987a, 1987b; Stokes and Loh, 1982). Chloride balance data therefore, can be a valuable source of additional information on mechanisms of catchment response to rainfall and hence can lead to increased confidence in the water balance predicted by a catchment model.

6.3.2 Model Formulation

The salinity balance model includes chloride balances for the saturated and unsaturated portions of both the perched aquifer (A) and deep groundwater stores. Both portions of the A store and the saturated portion of the deep groundwater store are treated as fully mixed with respect to chloride. The regolith profile above the deep water table however, is divided into layers of arbitrary thickness and a chloride balance maintained for each.

The chloride balance model parameters, their definitions and appropriate units are shown in Table 6.1.

**A Store**

Before chloride input from rainfall $cl_{in}$ (kg.m$^{-2}$) is added to the A store the quantity exported directly to streamflow by surface runoff is determined from

$$\text{cl}_\text{off} = \text{cl}_{in} \times \frac{\text{roff}}{\text{tf}}$$  \hspace{1cm} [6.43]

the remainder, $cl_{add}$ is added to the chloride storage of the A store. Note that chloride storage in the forest canopy is not modelled, implying that the chloride storage of foliage is either zero or in equilibrium with the incoming chloride. Chloride arriving with throughfall is only concentrated in the unsaturated portion of the A store, $cl_a$ when the store is wholly unsaturated (i.e. $\text{perch} \leq 0$)
\[ cl\_a_i = cl\_a_{i-1} + cl\_add \]  

When saturation does occur chloride is mobilised in proportion to the ratio of the saturated water storage to the maximum allowable

\[ cl\_pch = cl\_a \times \frac{perch}{pch\_max} \]  

As any throughfall that infiltrates into the A store when perching exists goes directly into saturated storage any chloride carried with it also goes directly into the saturated portion of the store. Chloride losses from the store are determined in direct proportion to the relative magnitudes of the fluxes from it and the storage at the end of the time-step

\[ cl\_iflo_i = cl\_pch_{i-1} \times \frac{iflo_i}{perch_i + iflo_i + pref_i} \]  
\[ cl\_pref_i = cl\_pch_{i-1} \times \frac{pref_i}{perch_i + iflo_i + pref_i} \]  
\[ cl\_pch_i = cl\_pch_{i-1} \times \frac{perch_i}{perch_i + iflo_i + pref_i} \]

where \( cl\_iflo \) is the mass of chloride in the subsurface lateral discharge and \( cl\_pref \) is the mass of chloride contributed to the deep groundwater store by preferential flow.

**Groundwater Store**

The chloride storage of the profile immediately above the deep aquifer is initially uniform and is determined by the proportion of chloride immobilised by a falling water table \( cl\_frac \). Chloride stored above the saturated zone is mobilised in proportion to the increase in groundwater storage and the salt storage of the layer into which the water table rises

\[ cl\_gws_i = cl\_gws_{i-1} + \Delta gws_i \times \frac{cl(layr)_{i-1}}{zlayr - gws_{i-1}} \]
where \( cl_{gws_t} \) is the chloride storage of the groundwater store at time \( t \); \( \Delta gws_t \) (mm) is the change in groundwater storage between time \( t-1 \) and \( t \); \( cl(\text{layr}) \) is the chloride storage of the layer immediately above the water table; and \( z\text{layr} \) is the height of each layer in the deep aquifer store. Any chloride arriving in recharge mixes with groundwater before any export is allowed. It should be noted that the layers are numbered consecutively from the bottom of the store and \( z\text{layr} \) is defined in terms of water storage (mm\(^3\).mm\(^{-2}\)), not in terms of absolute height above some datum.

In the case of a reduction in groundwater storage, the model parameter \( cl\_frac \) determines the mass of chloride that is left behind in the unsaturated layer or layers above

\[
cl\_avlb = cl\_frac \times cl\_gws_{t-1} \times \frac{\Delta gws_t}{gws_{t-1}}
\]

\( cl\_avlb \) is then distributed among the layers left unsaturated by the falling water table in proportion to the depth of water lost from each layer.

Chloride exported from the groundwater store is then partitioned between the remaining storage and groundwater discharge

\[
cl\_gflo_t = cl\_gws_{t-1} \times \frac{gflo_t}{gws_t + gflo_t}
\]

\[
cl\_gws_t = cl\_gws_{t-1} \times \frac{gws_t}{gws_t + gflo_t}
\]

The chloride lost to streamflow \( cl\_ex \) is then

\[
cl\_ex = cl\_bflo + cl\_roff + cl\_iflo
\]

where \( cl\_bflo \) is determined from

\[
cl\_bflo = cl\_gflo \times \frac{bflo}{gflo}
\]
6.4 Summary

The CATPRO hydro-salinity balance model described here is an attempt to construct a conceptual catchment model consistent with observed and inferred mechanisms of streamflow generation and water and salt pathways through natural catchments in south-western Australia. Model hypotheses about the nature of the relationships between the catchment water and salt balances are explicitly stated so that they may be critically assess in the light of field observations. In the following sections, the model hypotheses will be subjected to tests of increasing power. These tests will include calibration to multiple response time series and statistical tests on model behaviour and checking against prior knowledge on physical processes and model parameters.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>et_mult</td>
<td>Et multiplier to force long-term water balance</td>
<td>-</td>
</tr>
<tr>
<td>tfm</td>
<td>gradient of linear relationship between rainfall and throughfall</td>
<td>-</td>
</tr>
<tr>
<td>tfb</td>
<td>intercept of linear relationship between rainfall and throughfall</td>
<td>mm</td>
</tr>
<tr>
<td>pch_max</td>
<td>maximum saturated storage of perched aquifer</td>
<td>mm</td>
</tr>
<tr>
<td>wl</td>
<td>wetted length of A store at which water table first intersects soil surface</td>
<td>-</td>
</tr>
<tr>
<td>T_pch</td>
<td>product of lateral conductance of macropores and hillslope gradient</td>
<td>mm.d^{-1}</td>
</tr>
<tr>
<td>sfmin</td>
<td>proportional area of surface saturation due to baseflow</td>
<td>-</td>
</tr>
<tr>
<td>etAmin</td>
<td>absolute value of storage in A horizon at which Et ceases</td>
<td>mm</td>
</tr>
<tr>
<td>k_prf</td>
<td>preferential recharge parameter</td>
<td>mm.d^{-1}</td>
</tr>
<tr>
<td>bfm</td>
<td>gradient of linear relationship between bf &amp; gws</td>
<td>mm^{-1}</td>
</tr>
<tr>
<td>bf_max</td>
<td>maximum proportion of groundwater storage contributing to baseflow</td>
<td>-</td>
</tr>
<tr>
<td>bf_crit</td>
<td>baseflow threshold, groundwater storage below which no baseflow occurs</td>
<td>mm</td>
</tr>
<tr>
<td>pch0</td>
<td>initial storage of perched aquifer</td>
<td>mm</td>
</tr>
<tr>
<td>gws0</td>
<td>initial groundwater storage</td>
<td>mm</td>
</tr>
<tr>
<td>cl_a0</td>
<td>initial chloride storage per unit depth of A horizon</td>
<td>kg.m^{-2}.mm^{-1}</td>
</tr>
<tr>
<td>cl_gws0</td>
<td>initial chloride storage per unit depth of groundwater</td>
<td>kg.m^{-2}.mm^{-1}</td>
</tr>
<tr>
<td>cl_frac</td>
<td>fraction of chloride left in profile by falling water table</td>
<td></td>
</tr>
<tr>
<td>leak</td>
<td>leakage parameter for baseflow:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$0 \leq \text{leak} &lt; 0.5$, leakage to a regional aquifer</td>
<td></td>
</tr>
<tr>
<td></td>
<td>\text{leak} = 0.5, no leakage,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$0.5 &lt; \text{leak} \leq 1$, regional aquifer contributes to streamflow</td>
<td></td>
</tr>
<tr>
<td>gwsfit</td>
<td>specific yield or storage coefficient of deep aquifer</td>
<td>-</td>
</tr>
</tbody>
</table>

**Table 6.1**: CATPRO model parameter definitions and units.
7 CASE STUDY CATCHMENT

Salmon Catchment was chosen as a case study because long time series of high quality data on rainfall, salt input, streamflow, stream water quality and groundwater piezometric levels are available. This data will be used to actively challenge model hypotheses about the catchment water and salt balances and hence to illustrate the model filter methodology described in Chapter 4.

Salmon (33°25' S, 115°59' E) is typical of first order catchments of the western portion of the Darling Range, south-west Western Australia (Figure 7.1). The catchment is 81.6 ha in area and lies within the watershed of the Wellington Reservoir. The elevation ranges from 190 m AHD (Australian Height Datum) at the catchment outlet to 302 m AHD over a distance of approximately 1,500 m (Stokes, 1985). It has a Mediterranean climate, with cool, wet winters and warm to hot, dry summers. The average annual rainfall is 1,220 mm, 80% of which falls between May and October (Williamson, et al., 1987). Average annual class A pan evaporation is 1,600 mm (Stokes and Loh, 1982). The average maximum temperatures for January and July at Collie (16 km to the west north-east) are 31.2°C and 15.7°C; the average daily minima are 14°C and 5°C, respectively (Stokes, 1985).

The granitic gneiss basement is overlain by a lateritic regolith of up to 30 m in depth; basement topography approximates the surface topography. An A horizon of sandy gravels to gravelly loams, ranging in depth from a few centimetres to several metres overlie a clayey B horizon or a discontinuous duricrust. The hydraulic conductivity of Darling Range topsoils typically range from 1 to 20 m.d⁻¹ (Sharma et al., 1980; Ruprecht and Schofield, 1993). A perched aquifer may develop in the A horizon, especially in convergent terrain and lower slope landscape positions.

Dye tracer experiments (Johnston et al., 1983) and an infiltration study (Ruprecht and Schofield, 1993) demonstrated that the duricrust exhibits macroporosity at two scales. Large infilled holes of about 1 m² occupy approximately 10% of the area (K_sat = 10 m.d⁻¹), and small macropores of ≤ 1.0 x 10⁻⁴ m² (K_sat = 2.4 m.d⁻¹). The bulk hydraulic conductivity of the duricrust is approximately 3 m.d⁻¹ (Ruprecht and
Schofield, 1993). Perching rarely occurs on the duricrust, except in areas in which it is more massive, if perching occurs it is usually due to clayey sub-soils beneath the duricrust.

Figure 7.1: Location map for the Salmon Catchment in the catchment of the Wellington Reservoir which is fed by the Collie River in the south-west of Western Australia, rainfall isohyets are also shown. Other experimental catchments in the Collie River Basin are also marked (after Williamson et al., 1987).

Beneath the duricrust lies a mottled zone of sandy clays and clays which gradually grades to a pallid zone of highly variable kaolinitic clays and silt loams. The pallid zone grades to a weathering zone, and finally to basement at depth (Bettenay et al., 1980). The deep, permanent aquifer occupies the weathering zone and the lower
portion of the pallid zone in the upper reaches of the catchment (Stokes, 1985). The hydraulic conductivity of these layers is extremely low, in the order of $3.0 \times 10^{-3} \text{ m.d}^{-1}$ (Peck et al., 1980). In the lower reaches of the catchment, the deep aquifer approaches the soil surface and piezometric heads may be temporarily artesian in some areas (Stokes, 1985).

The catchment is drained by an ephemeral stream originating as two small, parallel channels incised into an area of convergent topography approximately 1,100 m from the catchment outlet. The stream flows over exposed basement near the catchment outlet, implying that streamflow is the only avenue via which groundwater discharge leaves the catchment (Bettenay et al., 1980). There are also other areas of outcrop and shallow soils especially on the eastern side of the stream channel downstream of S9. There is also a swampy area in the middle reaches of the catchment, through which the stream passes (S12 in Figure 7.2).

An open sclerophyllous forest consisting of an upper storey of 20 to 35 m, dominated by jarrah (*Eucalyptus marginata*) but also containing other eucalypts, covers the catchment. An under storey and groundcover lie below (Williamson et al., 1987). The jarrah tree has an extensive horizontal root system which occupies the surficial sandy gravel soil, sinker roots extend vertically from it and some occupy channels through the indurated duricrust and may extend to the deep permanent water table (Kimber, 1974; Carbon et al., 1980).

Hydrologic data have been collected at Salmon catchment since April 1974 (Williamson et al., 1987; Stokes, 1985). All data are reported on the basis of a water year running from April to March because the landscape is normally at its driest at this time of the year due to the long seasonal drought from November to April which is characteristic of the Mediterranean climate.
Figure 7.2: Salmon Catchment, showing topographic contours, stream gauging station, pluviographs and piezometer locations (after Stokes, 1985).

7.1 RAINFALL AND SALTFALL

Five storage rain gauges and three pluviometers were installed in or near Salmon catchment and are located so as to measure rainfall and saltfall both in clearings and beneath the forest canopy. Catchment averaged rainfall was determined by Thiessen-
weighting the observed daily rainfalls. With the large number of rain gauges used for such a small catchment area, errors in rainfall input may safely be assumed to be of low significance compared to other observational errors. This would not be the case for larger catchments with a lower density of rain gauges.

The period from April 1974 to April 1984 (Figure 7.3) represents a range of rainfall extremes for the catchment; it contains three of the seven years since 1899 that the rainfall at Collie was below the ten-percentile value. In addition, the annual rainfall for 1974 was 1,479 mm compared to a long-term median value of 1,220 mm (Williamson et al., 1987).

![Figure 7.3: Monthly rainfall data for Salmon Catchment, south-west Western Australia, derived by Thiessen-weighting of daily rainfalls observed at three pluviometers. (Source: Water Authority of W.A.).](image)

Williamson et al. (1987) determined the annual interception ratio for each of five sites at Salmon catchment for the period 1974 to 1981. The mean interception ratio was calculated from these data, to be 0.14 with a standard deviation of 0.0029.

Spatially averaged annual chloride concentrations in rainfall for Salmon catchment ranging from 4.8 mg L\(^{-1}\) to 10.1 mg L\(^{-1}\) with a mean of 6.7 mg L\(^{-1}\) (Williamson et al., 1987). These values were used directly to calculate daily chloride inputs to the catchment for model calibration.
Williamson *et al.* (1987) demonstrated that the enhanced saltfall they observed under the forest canopy was mainly the effect of salt extracted from the soil by vegetation then washed from the leaves by rainfall. The alternative, oceanic aerosols being caught by the canopy, was considered unlikely because the catchment is 250 m above sea level and some 40 km from the coast.

### 7.2 Streamflow and Chloride Export

Continuous streamflow data was recorded using a sharp crested v-notch weir constructed on basement outcrop as indicated above. A float operated Leopold and Stevens graphical recorder was used and less than 2% of the record for the ten year period from April 1974 to March 1984 was lost. In 1983, Stokes (1985) installed a small portable weir to measure storm event streamflows at two points upstream of the permanent weir to study of streamflow generation processes.

Due to the extremes in rainfall that occurred over the period, streamflow extremes were also recorded. The average annual streamflow for the period was 124.7 mm. The annual streamflow for 1974 was 367.4 mm, of which 155.1 mm was recorded in July alone. In comparison, the annual streamflows for 1976, 1979 and 1982 were 20.4 mm, 17.1 mm and 66.6 mm, respectively.

Stream water quality was determined from streamflow samples collected with an automatic pumping sampler. Daily streamflow chloride export was determined by the measurement of the electrical conductivity of the water samples and converted to a chloride concentration by linear regression ($R^2 = 0.97$) with laboratory determined samples (Stokes, 1985). A continuous record of electrical conductivity has been collected with a toroidal cell since 1980.

### 7.3 Evapotranspiration

Climatological data for the calculation of potential evapotranspiration $E_T$ is not measured on site and the closest site at which measurements have been made over a significant proportion of the time since Salmon was instrumented is Bridgetown, some 66 km to the south. The climatological variables measured are daily maximum and minimum air temperature, 0900 and 1500 wet and dry bulb temperatures and
cloud cover. Bridgetown is only a low priority station and observations are only available in digital format to July 1986. Gentilli (1989) presents a climatic classification of the two sites which suggests that the climatic regimes and grades of forest at the two sites are comparable. These data were used to calculate daily wet environment $E_t$ values from the complementary method as described by Morton (1978). This was used as a surrogate for Penman's potential $E_t$ as proposed by Sukvanachaikul and Laurenson (1983). Despite the similarity in the climate regimes and forest characteristics to the Bridgetown site Salmon catchment is within a kilometre of Wellington Reservoir which must certainly have a significant effect on the local microclimate.

The collection of class A pan evaporation at the town of Collie ceased in 1975, a year after data collection commenced at Salmon catchment. However, historical data is available and mean daily pan evaporation for each month was also used as a measure of potential $E_t$ at the site. In addition to the problem of using data representing only mean historical trends in evaporation, there is the further concern about the applicability of ground level observations to the determination of evapotranspiration from a forest canopy tens of metres high (Jarvis, 1980). Sharma (1984), for example, presents evidence that winter evapotranspiration at the site may be as much as three times pan evaporation.

Figure 7.4 compares the class A pan and wet environment evapotranspiration. The wet environment evapotranspiration is systematically below the Class A pan evapotranspiration with the greatest departures occurring during summer months.
7.4 GROUNDWATER LEVELS

Water levels in seven piezometers drilled to basement and slotted over the bottom 2 to 3 m of the deep aquifer were observed at approximately monthly intervals. The locations (Figure 7.2) of the piezometers were surveyed so that the observed water levels could be reduced to the Australian Height Datum (AHD). Piezometric observations were available from 11th May 1972 to the 14th March 1984. Time series of observed piezometric levels for four of the piezometers are shown in Figure 7.5. All piezometers exhibited similar long-term trends in piezometric head although the magnitude of long-term changes and seasonal fluctuations in head varied from piezometer to piezometer.

Piezometric heads in piezometer 1251, just upslope of the head of one of the twin stream channels (Figure 7.2), were almost 3 m above the ground surface in 1974 and 2.4 m below it in 1980. This indicates a significant shift in deep groundwater discharge regime over the course of the period for which calibration data are available.

Figure 7.4: Time series of monthly Class A pan evaporation and Morton's wet environment evapotranspiration for Salmon Catchment, April 1974 to March 1983.
Head observations from the seven piezometers were averaged to provide a mean reduced level of the deep aquifer piezometric surface and linear interpolation was used to estimate a mean reduced level for the piezometric surface at the end of each month (Figure 7.6). Piezometer 1151 is located less than 120 m from the catchment outlet and therefore piezometric heads observed there were significantly lower (195 m AHD) than the other piezometers which are located in the upper reaches of the catchment (Figure 7.2). Consequently there is a shift in the mean piezometric heads shown in Figure 7.6 compared to the levels shown in Figure 7.5.

In 1983, a transect of six shallow piezometers (< 3 m) was installed over a 150 m section of the hillslope immediately above the headwaters of the main stream channels (Figure 7.2). Perched aquifer water levels were observed on 16 occasions between 29th June 1983 and 8th November 1983. Manual observations of water quality and temperature were also made at those times (Stokes, 1985). Perched water salinities ranged from 25 to 50 mg.L$^{-1}$, some observations of water quality were discarded due to contamination of water by salts leaching from bentonite used to seal the annulus of some piezometers (Stokes, 1985).

![Piezometric Head Graph](image)

**Figure 7.5:** Time series of piezometric heads for four representative piezometers at Salmon Catchment between April 1974 and March 1983, inclusive (data courtesy of CSIRO Division of Water Resources).
Figure 7.6: Time series of catchment average piezometric head for Salmon Catchment.

7.5 PROFILE CHLORIDE STORAGE

Soil cores were collected at five sites on Salmon catchment by drilling to bedrock using hollow-stemmed augers and wire-line recovery equipment (Johnston, 1987a, 1987b). All drilling operations were completed prior to instrumentation of the catchment. Chloride storage profiles were then determined by regression using the measured electrical conductivity of the supernatant obtained by washing 25 g soil samples with distilled water; allowance was made for the initial water content as determined by drying at 105°C for 18 hours.

Johnston (1987a) reports average chloride storages for both saturated (deep groundwater) and unsaturated zones for each sampling site. The mean chloride storage of the saturated zone determined from Johnston's (1987a) data was 0.156 kg.m\(^{-3}\) and the standard deviation was 0.0795.

Johnston (1987a) also reported unsaturated zone chloride storages. However, these values will be heavily influenced by the high chloride storages of the pallid zone. Because CATPRO does not model water or chloride fluxes through this part of the profile these values do not represent the appropriate chloride storage for the 'A' store and were not used as such.
7.6 Stable Isotope Observations

The deuterium and $^{18}$O composition of rainfall, deep and perched groundwaters and stream water were monitored over a 21-month period, from April 1984 to December 1985 (Turner et al. 1987). Analyses of the stable isotope data was combined with chloride data for four storm events during the period of observation to assess the contributions of event and pre-event water on streamflow generation. Streamflow isotopic composition matched that of the perched groundwater system for almost the entire winter of 1985.

Turner et al. (1987) concluded that pre-event water stored in the perched groundwater system contributed between 60% and 95% of the streamflow generated during individual storm events. These observations were made after the period of continuous streamflow, stream water salinity and piezometric data used for model calibration (Chapters 8, and 9) and were therefore used only in a qualitative manner to test model structure, particularly the predicted contributions of different mechanisms of streamflow generation.

7.7 Other Hydrologic Data

The catchment has also been instrumented with shallow piezometers and 6 m neutron moisture meter access tubes (Williamson et al., 1987; Sharma et al., 1987). Results from the neutron moisture meter observations have been published elsewhere. The degree of spatial variability was high. Furthermore, some of the neutron moisture meter calibration curves were unreliable because of the high iron contents of the gravely soils and the highly heterogeneous nature of the soil profiles. For this reason, the neutron moisture meter data were not considered as model calibration time series.

Stokes (1985) conducted a visual survey of stream zone source areas and seepage areas (Figure 7.2) during the winter of 1980. He later used manual measurements of seepage water quality and temperature along with short time series of observations from the shallow piezometer transect and data from the temporary weir (Figure 7.2) to infer streamflow generation mechanisms for the upper reaches of the catchment. Stokes (1985) concluded that at least 40% of the streamflow for the two storm events he studied in detail originated from the shallow perched groundwater system.
surrounding the headwaters of the stream channel. A lack of piezometers (deep and shallow) in the middle and lower reaches of the catchment precluded any estimation of the contributions of the different streamflow generation mechanisms for the whole catchment. Stokes’ (1985) conclusions were used in a qualitative manner to assess CATPRO model hypotheses about the contributions of streamflow generation mechanisms.

7.8 SUMMARY

Salmon is a small first order experimental catchment, typical of the Darling Range in south-western Australia. Time series of high quality, streamflow, stream water quality and piezometric head data are available for the catchment. The ten year streamflow record contains zero flows in every year and three years where rainfall was below the 10 percentile value plus one very high rainfall year, this range of flow and climatic regimes should provide calibration time series demonstrating the full range of catchment responses to rainfall.

These data sets will be used to demonstrate the methodology, proposed in Chapter 4, as the hypotheses of a case study model (Chapter 6) are rigorously tested and new model hypotheses proposed.
8 STREAMFLOW CALIBRATION

This Chapter describes the application of the methodology described in Chapter 6 to the CATPRO water balance model as described in Chapter 1. The objective is to demonstrate the proposed methodology of model hypothesis testing as opposed to identifying an optimum posterior parameter distribution for the prediction of future streamflow responses. To this end, the results of several optimisation runs will be presented and compared, the results of a traditional split sample test will also be examined.

8.1 BASE CASE STREAMFLOW CALIBRATION

The CATPRO water balance model was calibrated to monthly streamflow data for the period April 1974 to March 1984, inclusive. The streamflow record contains a significant run of zero flow months each summer. Furthermore, the period contains three years where the rainfall was below the 10 percentile value and one year, the first of the calibration period, in which the rainfall was well above average. This wide range of flow regimes would be expected to provide a high level of information to aid in parameter identifiability (Gupta and Sorooshian, 1985).

No Box-Cox transformation was applied to the streamflow time series \( \lambda = 1.0 \), consistent with the assumption that the residuals have constant variance. A plot of the observed and predicted monthly streamflow values is shown in Figure 8.1. The model predicted the magnitudes and timings of the peak flows well but several small peaks during the recessions are not well matched or were missed. The coefficient of determination was 0.97. The mean observed and predicted streamflow values were 9.98 and 9.76 mm.mth\(^{-1}\), respectively with observed and predicted standard deviations of 21.86 and 21.78, indicating that the predicted streamflow response had a distribution very closely resembling the observed streamflow time series.

For this calibration, all the parameters associated with the permanently unsaturated B store were redundant. No parameter values could be found which resulted in the B store contributing to the simulated catchment water balance. Continued attempts to
locate a posterior parameter distribution which resulted in the B store contributing to the water balance always resulted in one or more of the B store parameters becoming redundant and zero fluxes from the B store.

The posterior distribution of fitted model parameters is summarised in Table 8.1, it shows that three parameters, \(bf\), \(bf_{crit}\) and \(gws0\); the baseflow proportionality constant, the baseflow threshold parameter and the initial groundwater storage, were estimated with such high degrees of uncertainty as to be almost unidentifiable.

The initial conditions of conceptual model storages are often poorly defined as discussed in Section 3.2 and are generally considered nuisance parameters. Note, however, that in this instance, the initial condition for the A store, \(pch0\), was moderately well identified by the available data. Furthermore, baseflow was predicted to contribute only a small portion of the total predicted streamflow for the first year of simulation after which time no baseflow was predicted (Figure 8.2). The streamflow data for the calibration period therefore does not contain a sufficiently large range of observed values to fully exercise the baseflow algorithm in its current form given the ordinary least squares error model (see also Sorooshian et al., 1983).

Examination of the posterior correlation matrix (Table 8.2) shows that \(bf_{crit}\) is strongly correlated to the initial groundwater storage parameter, \(gws0\), this is also a consequence of the inability of the calibration data to adequately exercise the baseflow algorithm of the CATPRO model. The only other parameter pair to exhibit even a moderate degree of correlation was the evapotranspiration parameters, \(et_{mult}\) and \(et_{Amin}\).

Although the initial groundwater storage and baseflow parameters were not well identified in this calibration, the time series of predicted groundwater storages (Figure 8.3) exhibits a long-term trend similar to that observed in the piezometric head data for Salmon Catchment shown in Figure 7.5 and Figure 7.6.
Figure 8.1: Monthly streamflow for Salmon Catchment predicted by CATPRO calibrated to streamflow data alone, Box-Cox $\lambda = 1.0$.

Figure 8.2: Time series of predicted contributions to streamflow for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 1.0$. 
Table 8.1: Posterior parameter distribution for calibration of the CATPRO model to Salmon streamflow data for a 10 year record using Box-Cox $\lambda = 1.0$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Transform</th>
<th>Mean</th>
<th>Std dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>etmult</td>
<td>None</td>
<td>0.8408</td>
<td>0.08075</td>
</tr>
<tr>
<td>tfin</td>
<td>None</td>
<td>0.86</td>
<td>*</td>
</tr>
<tr>
<td>tfb</td>
<td>None</td>
<td>0.00</td>
<td>*</td>
</tr>
<tr>
<td>pch_max</td>
<td>Loge</td>
<td>5.145</td>
<td>0.08666</td>
</tr>
<tr>
<td>wl</td>
<td>Binomial</td>
<td>0.5603</td>
<td>0.1934</td>
</tr>
<tr>
<td>$T_{pch}$</td>
<td>Loge</td>
<td>-0.3357</td>
<td>0.1006</td>
</tr>
<tr>
<td>etAmin</td>
<td>Loge</td>
<td>5.832</td>
<td>0.1087</td>
</tr>
<tr>
<td>k_prf</td>
<td>Loge</td>
<td>1.437</td>
<td>0.04620</td>
</tr>
<tr>
<td>k_B</td>
<td>None</td>
<td>0.00</td>
<td>‡</td>
</tr>
<tr>
<td>Bmax</td>
<td>None</td>
<td>0.00</td>
<td>‡</td>
</tr>
<tr>
<td>deep</td>
<td>None</td>
<td>0.00</td>
<td>‡</td>
</tr>
<tr>
<td>bf</td>
<td>Binomial</td>
<td>-4.008</td>
<td>4.596</td>
</tr>
<tr>
<td>bfcrit</td>
<td>Loge</td>
<td>6.043</td>
<td>4.531</td>
</tr>
<tr>
<td>pch0</td>
<td>None</td>
<td>-129.8</td>
<td>20.40</td>
</tr>
<tr>
<td>Bwater0</td>
<td>None</td>
<td>0.00</td>
<td>‡</td>
</tr>
<tr>
<td>gws0</td>
<td>Loge</td>
<td>3.277</td>
<td>72.04</td>
</tr>
<tr>
<td>leak</td>
<td>Binomial</td>
<td>0.00</td>
<td>*</td>
</tr>
</tbody>
</table>

* denotes parameter value fixed to value determined from field data  
‡ denotes parameter value fixed due to unidentifiability

Figure 8.3: Time series of CATPRO conceptual stores predicted for Salmon Catchment following calibration to streamflow data using Box-Cox $\lambda = 1.0$. 

82
Diagnostic plots indicate that the least squares error assumption is violated. Figure 8.4, a plot of residuals verses predicted streamflow, reveals systematic error in the prediction of low flows which is probably due to the high number of zero flows. However, the spread of residuals beyond this region does not increase greatly suggesting that the constant variance assumption may be met in this region. A residual time series plot (Figure 8.5) indicates some short runs of consecutive negative residuals but autocorrelation of the residuals is not indicated by this plot or by the autocorrelation function (Figure 8.6). The normal probability plot of the residuals shown in Figure 8.7 is strongly non-linear indicating a non-normal distribution, this is confirmed by the Kolmogorov-Smirnov test and by the cumulative periodogram of residuals shown in Figure 8.8.

Table 8.2: Posterior correlation matrix of parameters for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 1.0$.

<table>
<thead>
<tr>
<th></th>
<th>etMult</th>
<th>pchMax</th>
<th>wl</th>
<th>T_pch</th>
<th>etAmin</th>
<th>k_prf</th>
<th>bf</th>
<th>bf_crit</th>
<th>pch0</th>
<th>gws0</th>
</tr>
</thead>
<tbody>
<tr>
<td>etMult</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pchMax</td>
<td>0.4965</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wl</td>
<td>-0.2537</td>
<td>-0.4729</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T_pch</td>
<td>0.2076</td>
<td>-0.1898</td>
<td>0.3651</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>etAmin</td>
<td>0.8207</td>
<td>-0.6445</td>
<td>0.1228</td>
<td>0.0660</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k_prf</td>
<td>-0.5397</td>
<td>-0.3445</td>
<td>-0.4177</td>
<td>-0.0554</td>
<td>0.6402</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bf</td>
<td>0.4925</td>
<td>0.4460</td>
<td>-0.1420</td>
<td>0.0935</td>
<td>-0.4715</td>
<td>-0.3560</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bf_crit</td>
<td>-0.0022</td>
<td>0.0031</td>
<td>-0.0002</td>
<td>-0.0011</td>
<td>0.0003</td>
<td>0.0004</td>
<td>0.0061</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pch0</td>
<td>0.3683</td>
<td>0.6423</td>
<td>-0.0606</td>
<td>-0.0379</td>
<td>-0.5091</td>
<td>-0.4059</td>
<td>0.3213</td>
<td>0.0040</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>gws0</td>
<td>0.0022</td>
<td>-0.0031</td>
<td>0.0002</td>
<td>0.0011</td>
<td>-0.0003</td>
<td>-0.0004</td>
<td>-0.0061</td>
<td>0.9998</td>
<td>-0.0040</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
**Figure 8.4:** Scatter plot of residuals against predicted streamflow for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 1.0$.

**Figure 8.5:** Residual time series plot for streamflow predicted by CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 1.0$. 
Figure 8.6: Residual autocorrelation plot for streamflow predicted by CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 1.0$.

Figure 8.7: Residual normal probability plot for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 1.0$. 
The unidentifiability of the baseflow parameters has resulted in gross model non-linearity as revealed in the response surface plot shown in Figure 8.9. Note that even though the plot is aligned with the principal components of the linearised conditional probability region (Kuczera, 1990), the response surface contours are extremely elongated with respect to the 90% probability region. This extreme non-linearity is evident throughout the parameter space as demonstrated by the rank 2, 3 and rank 6, 7 principal component response surface plots shown in Figure 8.10 and Figure 8.11, respectively.

Figure 8.8: Residual cumulative periodogram plot for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 1.0$. 
Figure 8.9: CATPRO response surface plot, rank 1 and rank 2 principal components for calibration to Salmon streamflow data using Box-Cox $\lambda = 1.0$.

Figure 8.10: CATPRO response surface plot, rank 2 and rank 3 principal components for calibration to Salmon streamflow data using Box-Cox $\lambda = 1.0$. 
8.2 EFFECT OF BOX-COX TRANSFORMATIONS ON PREDICTED RESPONSES AND PARAMETER IDENTIFIABILITY

Figure 8.4, Figure 8.7 and Figure 8.8 revealed that the variance of the residuals for the base case calibration was not stationary and that this may be stabilised by using a Box-Cox $\lambda < 1.0$. The use of the Box-Cox transformation increases the weight given to low flows in the objective function by assuming that the variance in the observed streamflow increases as the predicted streamflow increases. Two calibrations were performed with Box-Cox $\lambda = 0.5$ and $\lambda = 0.25$ to explore the effect of the transformation on the ability of the single response calibrations to discriminate between good and bad model hypotheses.

8.2.1 Box-Cox $\lambda = 0.5$

Figure 8.12 shows the streamflow time series produced following calibration of CATPRO to the Salmon data using Box-Cox $\lambda = 0.5$. Comparison of Figure 8.12 with Figure 8.1 reveals only marginal improvements in the match between the
smaller annual streamflow peaks and recessions and a marginally less satisfactory match to the observed streamflow for the first year of simulation. Figure 8.13, however, shows that the Box-Cox transformation had the desired effect on the variances of the residuals with a much more even spread of residuals about the zero line.

Figure 8.12: Monthly streamflow for Salmon Catchment predicted by CATPRO calibrated to streamflow data alone, Box-Cox $\lambda = 0.5$.

Figure 8.13: Scatter plot of residuals against predicted streamflow for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 0.5$. 
Figure 8.14: Residual normal probability plot for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 0.5$.

The normal probability plot shown in Figure 8.14 demonstrates a small improvement in the behaviour of the residuals with respect to the assumption of normality over the base case calibration (Figure 8.7). The cumulative periodogram shown in Figure 8.15 demonstrates a marked improvement over Figure 8.8 and the Kolmogorov-Smirnov test for this plot indicates that the residuals are independent and that their variance is constant.

A plot of the predicted mechanisms of streamflow generation (Figure 8.16) and the predicted conceptual model storages (Figure 8.17) show that the application of the Box-Cox transformation has resulted in physically unrealistic prediction of catchment responses other than the streamflow hydrograph used for calibration. The baseflow time series contributes an insignificant volume of water to the monthly streamflow volumes and the groundwater store (Figure 8.17) shows a long-term rising trend inconsistent with a forested catchment in dynamic equilibrium, and with observed piezometric heads for Salmon Catchment shown in Figure 7.5. The predicted dominance of surface runoff in the generation of streamflow for Salmon
Catchment is also inconsistent with field observations (Stokes, 1985) and stable isotope studies (Turner et al., 1987).

The posterior parameter distribution (Table 8.3) shows that the parameters associated with the groundwater store, $bf$, $bf_{crit}$ and $gws0$, are almost totally unobservable given the available streamflow record and the error model assumed. Most other parameters are well identified, the exceptions being those associated with the generation of interflow, $wl$ and $T_{pch}$, due to the small contribution this process made to the predicted streamflow.

**Figure 8.15**: Residual cumulative periodogram plot for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 0.5$.

The baseflow parameters were included in the calibration process for demonstration purposes, but in any practical application, these parameters would have to be excluded from the calibration. The most appropriate option would be to assign zero values to all these parameters reducing the model to a single conceptual store. The effect of fixing redundant parameters at non-zero values is equivalent to the assumption of perfect knowledge of their true values. This violates the principle of
parsimony in model structure as the baseflow conceptualisation becomes an auxiliary hypothesis (Popper 1959) introduced to defend the model against falsification.

Figure 8.16: Time series of predicted contributions to streamflow for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 0.5$.

Figure 8.17: Time series of CATPRO conceptual stores predicted for Salmon Catchment following calibration to streamflow data using Box-Cox $\lambda = 0.5$. 
Response surface plots for the rank 1, 2 and 3 principal components could not be generated for this calibration because of the almost total unobservability of the baseflow parameters when they were included in the calibration process. Figure 8.18 shows the response surface for the rank 4 and rank 5 principal components and reveals a reduced degree of non-linearity compared to the base case calibration. Figure 8.19 illustrates that when the baseflow parameters are fixed, model non-linearity is further reduced, in this instance they were fixed at the values shown in Table 8.3 although this should not be done in practice for the reasons described above.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Transform</th>
<th>Mean</th>
<th>Std dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>etmult</td>
<td>None</td>
<td>0.6283</td>
<td>0.03328</td>
</tr>
<tr>
<td>tfm</td>
<td>None</td>
<td>0.86</td>
<td>*</td>
</tr>
<tr>
<td>tfb</td>
<td>None</td>
<td>0.00</td>
<td>*</td>
</tr>
<tr>
<td>pch_max</td>
<td>Loge</td>
<td>5.379</td>
<td>0.09834</td>
</tr>
<tr>
<td>wl</td>
<td>Binomial</td>
<td>0.3537</td>
<td>0.2884</td>
</tr>
<tr>
<td>T_pch</td>
<td>Loge</td>
<td>-1.098</td>
<td>0.1321</td>
</tr>
<tr>
<td>etAmin</td>
<td>Loge</td>
<td>5.902</td>
<td>0.1491</td>
</tr>
<tr>
<td>k_prf</td>
<td>Loge</td>
<td>1.404</td>
<td>0.08745</td>
</tr>
<tr>
<td>k_B</td>
<td>None</td>
<td>0.00</td>
<td>‡</td>
</tr>
<tr>
<td>Bmax</td>
<td>None</td>
<td>0.00</td>
<td>‡</td>
</tr>
<tr>
<td>deep</td>
<td>None</td>
<td>0.00</td>
<td>‡</td>
</tr>
<tr>
<td>bf</td>
<td>Binomial</td>
<td>-19.06</td>
<td>168.7</td>
</tr>
<tr>
<td>bfcrit</td>
<td>Loge</td>
<td>9.741</td>
<td>1070</td>
</tr>
<tr>
<td>pch0</td>
<td>None</td>
<td>-99.31</td>
<td>23.91</td>
</tr>
<tr>
<td>Bwater0</td>
<td>None</td>
<td>0.00</td>
<td>‡</td>
</tr>
<tr>
<td>gws0</td>
<td>Loge</td>
<td>11.18</td>
<td>253.2</td>
</tr>
<tr>
<td>leak</td>
<td>Binomial</td>
<td>0.00</td>
<td>*</td>
</tr>
</tbody>
</table>

* denotes parameter value fixed to value determined from field data
‡ denotes parameter value fixed due to unidentifiability

Table 8.3: Posterior parameter distribution for calibration of the CATPRO model to Salmon streamflow data for a 10 year record using Box-Cox $\lambda = 0.5$. 
Figure 8.18: CATPRO response surface plot, rank 4 and rank 5 principal components for calibration to Salmon streamflow data using Box-Cox $\lambda = 0.5$.

8.2.2 Box-Cox $\lambda = 0.25$

Figure 8.20 shows the streamflow time series produced following calibration of CATPRO to the Salmon data using Box-Cox $\lambda = 0.25$. Comparison of Figure 8.20 with Figure 8.1 reveals only marginal improvements in the match between the smaller annual streamflow peaks and recessions and minimal changes to the matches with the observed streamflow for the first two years of simulation. Visual differences between the streamflows predicted in this calibration (Figure 8.20) and those predicted using $\lambda = 0.5$ (Figure 8.12) are minimal except for the second year of the simulation. Figure 8.21 shows that the Box-Cox transformation has resulted in a more even spread of residuals compared to the base case calibration (Figure 8.4). Compared to the calibration performed using $\lambda = 0.5$ the spread of residual has improved marginally.
Figure 8.19: CATPRO response surface plot, rank 1 and rank 2 principal components for calibration to Salmon streamflow data using Box-Cox $\lambda = 0.5$, with parameters $bf$, $bf_{crit}$ and $gws0$ fixed.

Figure 8.20: Monthly streamflow for Salmon Catchment predicted by CATPRO calibrated to streamflow data alone, Box-Cox $\lambda = 0.25$. 
Figure 8.21: Scatter plot of residuals against predicted streamflow for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 0.25$.

Together, Figure 8.22, a normal probability plot of the residuals, and Figure 8.23, a residual cumulative periodogram demonstrate a marked improvement over the base case calibration (Figure 8.7 and Figure 8.8) in the validity of the assumption that the residuals are normally distributed. This is particularly so for the cumulative periodogram.
Figure 8.22: Residual normal probability plot for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 0.25$.

Figure 8.23: Residual cumulative periodogram plot for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 0.25$. 
The improvement achieved in the behaviour of the residuals with the application of the Box-Cox transformation was offset by the change in the predicted contributions of the different mechanisms of streamflow generation. Figure 8.24 shows that surface runoff is predicted to be the dominant mechanism of streamflow generation, as was the case when a Box-Cox $\lambda = 0.5$ was applied. This result is in direct contradiction with field evidence presented by Stokes (1985) and Turner et al. (1987). Furthermore, baseflow is predicted to occur continuously throughout the simulation but at a rate that contributes an insignificant volume of water to the monthly streamflow.

Figure 8.25 shows that the predicted storage of the deep groundwater system maintains a pattern of seasonal variability but that a long-term trend of increasing storage is evident. This long-term increase in groundwater storage is not consistent with the observed trend in piezometric heads shown in Figure 7.5. Together with the low, continuous level of baseflow this implies that the groundwater store has become a sink for water not required to reproduce the observed streamflow data. This is confirmed by the posterior distribution of model parameters shown in Table 8.4. The table reveals that the identifiability of $bfcrit$ has deteriorated from the already low level demonstrated in the base case calibration (Box-Cox $\lambda = 1.0$; Table 8.1) to such a level as to totally outweigh the improved identifiability of $bf$. In practice this level of observability can only be dealt with by fixing the parameters and making the groundwater store redundant, this was not done for demonstration purposes. The level of observability of the initial condition for the A store, $pch0$, has also deteriorated from that for the calibration using $\lambda = 1.0$, though parameters $wl$ and $T_{pch}$ are identifiable to a higher degree than in the calibration using Box-Cox $\lambda = 0.5$.

The response surface plot for the rank 1, 2 principal components could not be produced because of the extreme level of unobservability of the baseflow parameters. Either parameter would assume a physically infeasible value when reduced by less than one standard deviation as shown in Table 8.4. This is further evidence that, in practice these parameters should be excluded from the calibration.
Figure 8.26 shows the response surface plot for the rank 2 and rank 3 principal components. Although the degree of non-linearity is high, it is arguably not as extreme as that demonstrated by the rank 2 and rank 3 principal components for the calibration without the Box-Cox transformation applied (Figure 8.10). Figure 8.27 demonstrates the effect of fixing model parameter $bf_{crit}$, which was the least well observed parameter in this calibration, the degree of non-linearity is markedly reduced.

![Figure 8.24](image1.png)

**Figure 8.24:** Time series of predicted contributions to streamflow for CATPRO calibrated to Salmon streamflow data using Box-Cox $\lambda = 0.25$.

![Figure 8.25](image2.png)

**Figure 8.25:** Time series of CATPRO conceptual stores predicted for Salmon Catchment following calibration to streamflow data using Box-Cox $\lambda = 0.25$. 

99
Table 8.4: Posterior parameter distribution for calibration of the CATPRO model to Salmon streamflow data for a 10 year record using Box-Cox $\lambda = 0.25$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Transform</th>
<th>Mean</th>
<th>Std dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>etmult</td>
<td>None</td>
<td>0.7009</td>
<td>0.04523</td>
</tr>
<tr>
<td>tfm</td>
<td>None</td>
<td>0.86</td>
<td>*</td>
</tr>
<tr>
<td>tfb</td>
<td>None</td>
<td>0.00</td>
<td>*</td>
</tr>
<tr>
<td>pch_max</td>
<td>Loge</td>
<td>6.023</td>
<td>0.1013</td>
</tr>
<tr>
<td>wl</td>
<td>Binomial</td>
<td>-0.8613</td>
<td>0.2459</td>
</tr>
<tr>
<td>T_pch</td>
<td>Loge</td>
<td>-2.588</td>
<td>0.2196</td>
</tr>
<tr>
<td>etAmin</td>
<td>Loge</td>
<td>5.382</td>
<td>0.06633</td>
</tr>
<tr>
<td>k_prf</td>
<td>Loge</td>
<td>1.600</td>
<td>0.1328</td>
</tr>
<tr>
<td>k_B</td>
<td>None</td>
<td>0.00</td>
<td>‡</td>
</tr>
<tr>
<td>Bmax</td>
<td>None</td>
<td>0.00</td>
<td>‡</td>
</tr>
<tr>
<td>deep</td>
<td>None</td>
<td>0.00</td>
<td>‡</td>
</tr>
<tr>
<td>bf</td>
<td>Binomial</td>
<td>-17.83</td>
<td>5.970</td>
</tr>
<tr>
<td>bfcrit</td>
<td>Loge</td>
<td>4.628</td>
<td>2294</td>
</tr>
<tr>
<td>pch0</td>
<td>None</td>
<td>-43.68</td>
<td>31.82</td>
</tr>
<tr>
<td>Bwater0</td>
<td>None</td>
<td>0.00</td>
<td>‡</td>
</tr>
<tr>
<td>gws0</td>
<td>Loge</td>
<td>5.132</td>
<td>1387</td>
</tr>
<tr>
<td>leak</td>
<td>Binomial</td>
<td>0.00</td>
<td>*</td>
</tr>
</tbody>
</table>

* denotes parameter value fixed to value determined from field data  
‡ denotes parameter value fixed due to unidentifiability
Figure 8.26: CATPRO response surface plot, rank 2 and rank 3 principal components for calibration to Salmon streamflow data using Box-Cox $\lambda = 0.25$.

8.3 SPLIT SAMPLE TEST

For the split sample test, a Box-Cox $\lambda$ of 0.5 was selected because this transformation resulted in the best compromise between matching the observed monthly streamflow volumes and the requirement for independent, constant variance residuals. It could be argued that a Box-Cox $\lambda$ of 1.0 should be used because the base case calibration produced predictions of streamflow generation mechanisms that were more consistent with the results of other studies not used in the calibration (Stokes, 1985; Turner et al., 1987). This is true, but to assess the value of the split sample test methodology, as it is generally used in practice, the choice of Box-Cox transformation was considered overriding. Information of the type gained from the detailed process studies carried out at Salmon Catchment is not commonly available for calibration of lumped catchment models, and was therefore not considered in this case.
Figure 8.27: CATPRO response surface plot, rank 1 and rank 2 principal components for calibration to Salmon streamflow data using Box-Cox $\lambda = 0.25$, with parameter $bf_{crit}$ fixed.

Figure 8.28 shows the results of a split sample test in which CATPRO was calibrated to the monthly streamflow data for the period from April 1974 to March 1979 and the derived parameter values used to predict the streamflows for the following five year period. The match between the predicted and observed streamflows are considerably worse than those shown in Figure 8.12 for the calibration to the full ten year streamflow record using Box-Cox $\lambda = 1.0$.

The posterior parameter distribution is shown in Table 8.5. The initial groundwater storage parameter, $gws0$, was unidentifiable and was not included in the calibration process. Three other parameters, $wl$, $T_{pch}$ and $bf$, were almost unidentifiable given the shorter calibration time series.

The time series of predicted contributions to streamflow and conceptual model storages are shown in Figure 8.29 and Figure 8.30, respectively. They are actually
more consistent with the results of the process studies of Stokes (1985) and Turner et al. (1987) than the results of the calibration to the full streamflow time series shown in Figure 8.16 and Figure 8.17 for which Box-Cox $\lambda = 0.5$.

The poor identifiability of $w_l$, $T_{pch}$ and $bf$ and the total unobservability of the initial groundwater storage, $gws0$, suggests a simpler model structure may suffice to replicate the observed streamflow times series. Figure 8.31 shows the predicted and observed streamflows obtained when CATPRO was reduced to a single store model by fixing the values of $k_{prf}$, $bf$, $bf_{crit}$ and $gws0$ to zero and calibrating it to the first five years of streamflow data. The results are significantly better than those shown in Figure 8.28 for the two store model with a baseflow contribution. The streamflow contributions of surface runoff and perched groundwater flow are shown in Figure 8.32. For a split sample test, applied without reference to the results of process studies (Stokes, 1985; Turner et al., 1987) a simpler model structure, for which only six model parameters are calibrated, performed better than the two store CATPRO model as Jakeman and Hornberger (1993) suggest as a general rule.

Figure 8.28: Monthly streamflow for Salmon Catchment predicted by CATPRO calibrated to the first 5 years of streamflow data as a split sample test, Box-Cox $\lambda = 0.5$. 

![Diagram showing streamflow data for two periods: Calibration and Prediction. The graph plots streamflow in millimeters against months from April 1974 to April 1984. The data points for both predicted and observed streamflows are marked with diamonds.](image-url)
**Figure 8.29:** Time series of predicted contributions to streamflow for Salmon Catchment following calibration of CATPRO to the first 5 years of streamflow data as a split sample test, Box-Cox $\lambda = 0.5$.

**Figure 8.30:** Time series of CATPRO conceptual stores predicted for Salmon Catchment following calibrated to the first 5 years of streamflow data as a split sample test, Box-Cox $\lambda = 0.5$. 
Table 8.5: Posterior parameter distribution for calibration of the CATPRO model to Salmon streamflow data for the first 5 years of record as a split sample test, Box-Cox $\lambda = 0.5$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Transform</th>
<th>Mean</th>
<th>Std dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>etmult</td>
<td>None</td>
<td>0.7956</td>
<td>0.09925</td>
</tr>
<tr>
<td>tfm</td>
<td>None</td>
<td>0.86</td>
<td>*</td>
</tr>
<tr>
<td>tfb</td>
<td>None</td>
<td>0.00</td>
<td>*</td>
</tr>
<tr>
<td>pch_max</td>
<td>Log e</td>
<td>4.937</td>
<td>0.1226</td>
</tr>
<tr>
<td>wl</td>
<td>Binomial</td>
<td>1.693</td>
<td>1.075</td>
</tr>
<tr>
<td>T_pch</td>
<td>Log e</td>
<td>-0.02439</td>
<td>0.1810</td>
</tr>
<tr>
<td>etAmin</td>
<td>Log e</td>
<td>5.907</td>
<td>0.1299</td>
</tr>
<tr>
<td>k_prf</td>
<td>Log e</td>
<td>1.404</td>
<td>0.07356</td>
</tr>
<tr>
<td>k_B</td>
<td>None</td>
<td>0.00</td>
<td>‡</td>
</tr>
<tr>
<td>Bmax</td>
<td>None</td>
<td>0.00</td>
<td>‡</td>
</tr>
<tr>
<td>deep</td>
<td>None</td>
<td>0.00</td>
<td>‡</td>
</tr>
<tr>
<td>bf</td>
<td>Log e</td>
<td>-0.9632</td>
<td>1.409</td>
</tr>
<tr>
<td>bfcrit</td>
<td>Log e</td>
<td>6.036</td>
<td>0.08089</td>
</tr>
<tr>
<td>pch0</td>
<td>None</td>
<td>-142.8</td>
<td>23.24</td>
</tr>
<tr>
<td>Bwater0</td>
<td>None</td>
<td>0.00</td>
<td>‡</td>
</tr>
<tr>
<td>gws0</td>
<td>Log e</td>
<td>2.614</td>
<td>‡</td>
</tr>
<tr>
<td>leak</td>
<td>Binomial</td>
<td>0.00</td>
<td>*</td>
</tr>
</tbody>
</table>

* denotes parameter value fixed to value determined from field data
‡ denotes parameter value fixed due to unidentifiability
8.4 Prediction of Unobserved Catchment Responses

When a lumped catchment model is used to predict an unobserved catchment response implicit assumptions are made about the internal structure of the model. In the absence of supporting data, any estimation of the errors associated with such predictions represent the effects of parameter uncertainty only. Therefore, prediction limits and confidence limits represent the approximate effect of parameter uncertainty on predicted responses. They do not provide any information about
systematic error, that is, we may not assume that the model conceptualisation is correct, only that it has captured the essential dynamics of the calibration data. Model validity, with regard to the estimation of unobserved responses, can only be addressed by testing model predictions against other kinds of information such as process based studies such as those reported by Stokes (1985) and Turner et al. (1987).

Here the effects of parameter uncertainty on the prediction of the unobserved responses of catchment averaged groundwater recharge and evapotranspiration are examined with the aid of confidence limits (Kuczera, 1988). As was the case for the split sample test, the posterior parameter distribution derived from the calibration using Box-Cox $\lambda= 0.5$ was selected for estimation of groundwater recharge and catchment averaged evapotranspiration.

### 8.4.1 Groundwater Recharge

Monthly rates of groundwater recharge predicted using the posterior distribution of parameter values obtained from calibrating to monthly streamflow, with 90% confidence limits, are shown in Figure 8.33. Confidence limits were calculated by performing 300 Monte Carlo simulations using parameter values chosen randomly from the posterior distribution of model parameters summarised in Table 8.3. The predicted recharge time series displays the same seasonal trend apparent in both rainfall and streamflow records and is qualitatively as would be expected. Note that net recharge is predicted by subtracting any evapotranspiration from the deep aquifer from the calculated accessions to the store, recharge values less than zero are therefore possible in months where evapotranspiration from the store exceeds the total accessions.
Figure 8.33: Time series of unobserved groundwater recharge and 90% confidence limits predicted for Salmon Catchment by CATPRO calibrated to streamflow data, Box-Cox $\lambda = 0.5$.

The greatest uncertainty in recharge predictions exist during times of net discharge, when streamflow, also is very low or zero. At these times recharge predictions will be dominated by model parameters affecting evapotranspiration from the groundwater, in particular the volume of water in the store. However, all the parameters associated with the groundwater store were poorly identified (Table 8.3) and therefore may be expected to wield a significant influence over the confidence limits on the predicted recharge.

8.4.2 Actual Evapotranspiration

Predicted monthly evapotranspiration and 90% confidence limits are shown in Figure 8.34. The parameters which may be assumed to have the greatest influence on predicted evapotranspiration are better defined in the model calibration than those affecting the groundwater store (Table 8.3). However, the 90% confidence limits on the predicted evapotranspiration are wide in relation to the 90% confidence limits on the predicted recharge (Figure 8.33). The expected value of parameter $etMult$ is 0.6283 with a standard deviation of 0.003328 and the expected value of $\text{Log}_e(etAmin)$ is 5.902 with a standard deviation of 0.1491. Even though no strong parameter correlations were noted for this calibration the cumulative effect of interactions between many parameters influences the confidence limits on the predicted evapotranspiration.
8.5 EFFECT OF RELAXING SOME CONSTRAINTS ON MODEL PARAMETERS

In the previous section, confidence limits were used to examine the approximate effects of parameter uncertainty on the prediction of unobserved catchment responses. In the case of an unobserved catchment response, no estimation of systematic error is possible. Here prediction limits and confidence limits are used to assess the value of prior knowledge on model parameters or catchment behaviour on the predictability of observed and unobserved catchment responses.

8.5.1 Groundwater Leakage Unknown

A conceptual groundwater store is included in almost every lumped or conceptual catchment model, and invariably all groundwater fluxes are assumed to contribute to streamflow. In reality, this is a rare occurrence. Even for a small, first order catchment, if a groundwater system exits below the stream channel, groundwater will leave the catchment without passing any gauging structure on the stream channel. In effect, the catchment has two unobserved mechanisms via which water is lost; evapotranspiration is usually identified as an unobserved sink for water and some measure of potential evaporation is used to predict evaporative losses. Groundwater flow or leakage from the catchment, on the other hand, is rarely acknowledged.
The Salmon Catchment data set offers the rare opportunity to quantitatively assess the effect of the normally unobservable consequences of failing to acknowledge groundwater losses from a catchment in the calibration of a conceptual catchment model. The v-notch weir at Salmon Catchment is constructed on basement outcrop so it may safely be assumed that groundwater leakage from the catchment is negligible. To assess the value of this prior knowledge, which is usually implicitly assumed, the posterior parameter distribution resulting from the streamflow calibration obtained with Box-Cox $\lambda = 0.5$ (Table 8.3) was perturbed with the NLFIT calibration module by freeing the leak parameter without permitting any of the other parameters to change. New standard deviations for each calibrated parameter were then calculated.

The predicted streamflow, recharge and evapotranspiration and 90% prediction or confidence limits produced by CATPRO without prior knowledge of the true value of the groundwater leakage parameter are shown in Figure 8.35, Figure 8.36 and Figure 8.37, respectively. The prediction limits on the streamflow (Figure 8.35) form a tight envelope about both the predicted and observed values except for the annual peaks where the upper prediction limits exceed the observed values by as much as 50 mm. Several of the observed values for secondary peaks during the recessions fall just outside the upper prediction limits.

The confidence limits on both the predicted recharge and evapotranspiration are both almost indistinguishable from those determined with the groundwater leakage parameter fixed (Figure 8.33 and Figure 8.34). The streamflow hydrograph contains little information about the behaviour of the groundwater system and without prior knowledge of the behaviour of the groundwater system, in the form of observations of piezometric head, the parameters associated with the groundwater store were poorly identified. Given the already poor degree of observability of the groundwater storage parameters a lack of prior knowledge about groundwater leakage from the catchment does not contribute greatly to the uncertainty of other catchment responses.
**Figure 8.35:** Time series of predicted streamflow and 90% prediction limits produced for Salmon Catchment by CATPRO calibrated to streamflow data without prior knowledge of groundwater leakage, Box-Cox $\lambda = 0.5$.

**Figure 8.36:** Time series of unobserved catchment averaged recharge and 90% confidence limits predicted for Salmon Catchment by CATPRO calibrated to streamflow data without prior knowledge of groundwater leakage, Box-Cox $\lambda = 0.5$. 
**Figure 8.37:** Time series of unobserved catchment averaged evapotranspiration and 90% confidence limits predicted for Salmon Catchment by CATPRO calibrated to streamflow data without prior knowledge of groundwater leakage, Box-Cox $\lambda = 0.5$.

### 8.5.2 Throughfall Parameters Unknown

In all calibration runs described above the throughfall parameters $t_{fm}$ and $t_{fb}$ were fixed at the values determined from the data presented by Williamson et al. (1987). If such information were not available, $t_{fm}$ and $t_{fb}$ would have to be calibrated to streamflow data with the other model parameters. Figure 8.38 shows the effect on the accuracy of streamflow predictions of not using this prior information on the throughfall parameters, in this case only $t_{fb}$ was fixed, to zero, and $t_{fm}$ was assumed unknown. Comparison with Figure 8.35 reveals a large increase in the uncertainty of the predicted streamflow. The upper prediction limits on all streamflow peaks have increased markedly, by up to almost 150 mm in some instances.

The extreme increases in the uncertainty of predicted streamflow resulting from a lack of prior information on throughfall is replicated in the uncertainty in the predicted recharge and evapotranspiration as shown in Figure 8.39 and Figure 8.40. Comparison of the confidence limits on these responses with and without prior information on throughfall show that a lack of prior knowledge results in an eight fold increase in the range of the confidence limits on the predicted recharge. The effect of the confidence limits on the predicted evapotranspiration is even more severe.
Without prior knowledge on the throughfall parameters the throughfall function apparently becomes the greatest contributor to measurable, parameter-induced uncertainty in the predicted catchment responses. Inclusion of a throughfall function in the model without prior information is akin to not knowing the rainfall but having only an estimation of its upper bound. These results suggest that, if no prior information were available on throughfall, then it may not be appropriate to include a throughfall mechanism in the model if uncertainty in the predicted responses is to be minimised. Without prior knowledge, the inclusion of a throughfall mechanism in the model becomes an auxiliary hypothesis, contributing nothing to the prediction of catchment responses and, in effect, reducing the testability of other model hypotheses (Popper, 1959).

Figure 8.38: Time series of predicted streamflow and 90% prediction limits produced for Salmon Catchment by CATPRO calibrated to streamflow data without prior knowledge of throughfall, Box-Cox $\lambda = 0.5$. 
Figure 8.39: Time series of unobserved catchment averaged recharge and 90% confidence limits predicted for Salmon Catchment by CATPRO calibrated to streamflow data without prior knowledge of throughfall, Box-Cox $\lambda = 0.5$.

Figure 8.40: Time series of unobserved catchment averaged evapotranspiration and 90% confidence limits predicted for Salmon Catchment by CATPRO calibrated to streamflow data without prior knowledge of throughfall, Box-Cox $\lambda = 0.5$.

8.6 Recasting Some Model Hypotheses - Assessment of Streamflow Generation Mechanisms

In all the calibrations presented above surface runoff was predicted to be the dominant mechanism of streamflow generation. This is inconsistent with prior information gained from field observations, stream water salinity and temperature data (Stokes, 1985) and analysis of stream water isotopic composition (Turner et al., 1987). The model hypothesis that perched groundwater flow occurs uniformly
throughout the A horizon is brought into question by the inability of the CATPRO model to correctly predict perched groundwater flow as the dominant streamflow generation mechanism for Salmon Catchment.

Stokes (1985) observed water entering the stream channel through macropores or pipes in the stream bank. A perched groundwater function that reproduces the effects of macropore flow on the generation of streamflow from the perched aquifer would be more consistent with Stokes’ (1985) field observations. It may also allow the A store to be more responsive to rainfall input and produce greater perched aquifer contributions to predicted streamflow.

Germann (1990) proposed a nonlinear kinematic wave relationship between storage and discharge for a hillslope dominated by macropore flow, a simplified form is:

$$iflo = T_{pch} \times h^{a_{pch}}$$

where, $iflo$, $T_{pch}$, and $h$ are as defined for Equation [6.18] and $a_{pch}$ is a dimensionless exponent, Germann (1990) suggests $a_{pch} = 2.5$. Note that if the data available for calibrating the model has insufficient power to allow all the perched aquifer parameters to be identifiable, $a_{pch}$ can be fixed to 1.0 and the model simplifies to its previous form and no additional, unsupported complexity will be imparted to the model.

In all the calibrations performed here the permanently unsaturated B store always became redundant and failed to contribute to the predicted catchment water balance. Jakeman and Hornberger (1993) argue that when only a streamflow hydrograph is available for calibration, a simple model with only five or six parameters is sufficient to reproduce the observed data. In the calibrations presented here, ten model parameters were fitted, three of which were generally poorly defined; leaving seven parameters well identified. In line with Popper’s (1959) theory, complexity that does not contribute to the testability of a hypothesis should be avoided and the redundant B store was removed from the CATPRO model.
In this Chapter, the CATPRO water balance model has been considered in isolation. It was at this point that the catchment salinity balance model described in section 6.3.2 was incorporated into the CATPRO model to exert additional power to actively challenge model hypotheses about water flow paths.

8.7 SUMMARY

The three store, CATPRO conceptual catchment model was calibrated to a ten year time series of monthly streamflow data for Salmon Catchment. The calibration period included zero flow periods every year and a range of rainfall extremes. A range of Box-Cox transformations were applied to the observed streamflow responses to improve the behaviour of the model residuals. In each case, the model parameters associated with the groundwater store and the generation of baseflow were poorly identified. Furthermore, parameters associated with a permanently unsaturated B store were always totally unobservable and the store was redundant. In the language of Popper, (1959), the existence of the B store was an auxiliary hypothesis that contributed nothing to the model’s ability to predict the observed streamflow response and should be removed.

Prior information on the contribution of the perched groundwater system to streamflow gained from process studies using field observations, perched water chloride concentration and temperature (Stokes, 1985) and stable isotope composition (Turner et al., 1987) was also used to assess model performance.

The use of Box-Cox $\lambda = 0.5$ improved the model behaviour in terms of producing constant variance residuals but did not improve the model predictions of the contributions of the different mechanisms of streamflow generation.

Response surface plots generated parallel to the principal component axes of the parameter space (Kuczera, 1990) were used to assess model linearity and in each calibration gross deviations from linearity were observed.

When a split sample test was performed, by calibrating the CATPRO model to only the first five years of the streamflow data, the model failed the test and parameter identifiability was further reduced. The split sample test is only a test of the ability of
the model to statistically reproduce the calibration time series, it is not a test of model structure. These results confirm Klemes’ (1986b) argument that a split sample test is the minimum requirement for the least demanding application of a hydrologic model.

Prediction limits and confidence limits on predicted catchment responses were used to assess the value of prior information relating to two important model hypotheses. Firstly, conceptual catchment models usually include a groundwater store that contributes baseflow to the predicted streamflow, the assumption that all water discharging from the groundwater store enters the stream channel is usually implicit but rarely acknowledged. Zero groundwater leakage from the Salmon Catchment can safely be assumed because the weir is constructed on basement outcrop and this provides an opportunity to assess the zero leakage assumption. Perhaps fortunately, the baseflow parameters of the CATPRO model were the least well identified by the calibration to the Salmon streamflow hydrograph. A lack of prior knowledge of the leakage parameter did not greatly affect prediction uncertainty. This may often be the case for conceptual catchment models calibrated to streamflow data alone.

Prior knowledge on the value of the throughfall parameter $tf_{in}$, however proved to be critical to the degree of uncertainty in model predictions. In the absence of prior information on throughfall, inclusion of a throughfall function in a conceptual catchment model produces extremely wide prediction and confidence limits on model predictions. Without prior knowledge, a throughfall function should be omitted from a conceptual catchment model.

Finally some model hypotheses were critically assessed in light of poor parameter identifiability and prior information gained from streamflow generation mechanism studies. The relationship between perched aquifer storage and discharge to the stream was revised; the totally redundant B store was omitted from the model; and, a chloride balance model was hypothesised to exert greater power in testing other model hypotheses about the relative contributions of different mechanisms of streamflow generation.
9 MULTI-RESPONSE CALIBRATION

In the previous chapter, the CATPRO water balance model was calibrated to a ten-year streamflow record for Salmon Catchment. The calibration statistics, diagnostic tests, and prior information on mechanisms of streamflow generation operating at the catchment were used to critically test model hypotheses in accordance with Popper’s (1959) theory of the scientific method. The model was found to exhibit all the problems commonly associated with conceptual catchment models (see section 3.2). Some parameters were poorly defined and there was a high degree of correlation between some parameter pairs.

The model matched the observed streamflow record well, despite the unidentifiability of some parameters but failed a split sample test. Furthermore, during the split sample test the shorter streamflow record used for calibration further compromised parameter identifiability.

Most importantly, the hypothesis that perched groundwater flows uniformly through the depth of the perched aquifer was challenged. Field observations (Stokes, 1985) identified macropores as producing a significant proportion of the perched water discharge to the stream channel. Furthermore, the model failed to produce perched groundwater discharge in a responsive enough manner for it to dominate over surface runoff as a contributor to streamflow generation (Stokes, 1985; Turner et al., 1987). The hypothesis about how water flows through the perched aquifer was recast in the light of field evidence and model performance and a catchment salt balance model was proposed to extract information on streamflow generation mechanisms from the stream chloride export time series.

Here the CATPRO hydrosalinity model will be concurrently calibrated to streamflow, stream chloride export, and piezometric head data for the Salmon Catchment. Again, statistical tools will be used to actively interrogate the calibration statistics for evidence of model hypotheses that are not supported by the calibration data and field observations.
9.1 **Baseflow Conceptualisation**

Several baseflow parameterisations were tested because of the low level of identifiability of the baseflow parameters in the streamflow calibrations. The baseflow parameterisation used in the calibrations reported here was

\[ bf = \min \left( bfm \times (gws - bf_{\text{crit}}) , bf_{\text{max}} \right) \]  \[ \text{(9.1)} \]

\[ gflo = \max \left( 0 , bf \times (gws - bf_{\text{crit}}) \right) \]  \[ \text{(9.2)} \]

where \( bf_{\text{crit}} \) is the threshold below which no groundwater discharge occurs, \( bfm \) is the gradient of the linear relationship between the storage above the threshold and the proportion of the excess that discharges, and \( bf_{\text{max}} \) sets an upper limit to the proportion that discharges. The form of the relationship is shown in Figure 9.1.

![Groundwater storage - discharge relationship showing the effects of model parameters \( bf_{\text{max}} \) and \( bf_{\text{crit}} \).](image)

**Figure 9.1:** Groundwater storage - discharge relationship showing the effects of model parameters \( bf_{\text{max}} \) and \( bf_{\text{crit}} \).

None of the baseflow conceptualisations used resulted in significant improvements in the identifiability of baseflow parameters over the others, even when the model was concurrently calibrated to three catchment response time series. Therefore, discussion will be restricted to the CATPRO model containing the baseflow conceptualisation described by Equations [9.1] and [9.2].
9.2 Model Calibration and Predictions

The Box-Cox transformations applied to each observed response will affect the disturbance weight matrix via its affect on the residuals of each response time series when the model is calibrated to the multi-response data. So the Box-Cox $\lambda$ that produces the best behaved residuals for an individual calibration time series may not necessarily produced the best result overall and this was the case for the CATPRO model.

Several trials were carried out to select the most appropriate values of Box-Cox $\lambda$ for the multi-response calibration. The choice of Box-Cox $\lambda$ for the change in groundwater piezometric head time series was not critical but the choice of $\lambda$ for the streamflow and stream chloride time series had significant impact on the match between the predicted and observed values and the nature of the residuals. The best matches between predicted and observed streamflows and stream water salinities were achieved with Box-Cox $\lambda = 0.5$ for both the streamflow and salinity time series; a Box-Cox $\lambda = 1.0$ was used for the change in piezometric head time series.

The time series of predicted and observed streamflows and chloride exported in streamflows are shown in Figure 9.2 and Figure 9.3, respectively. The match to the observed streamflows has deteriorated only slightly compared to that achieved in the calibration to the streamflow time series alone (Figure 8.1) with the most notable difference being in the substantial under-prediction of the peak discharge for the first year of record. The matches to some of the smaller peaks have also deteriorated to a lesser degree.

The plot of the predicted and observed masses of chloride exported in the streamflow shows an over-prediction in the first and last years of the simulation, which are the years of maximum streamflow and of chloride export. In most other years, catchment chloride export is under-predicted; the secondary peaks in chloride export are also under-predicted or missed.
Figure 9.2: Monthly streamflow for Salmon Catchment predicted by CATPRO calibrated to streamflow, stream chloride export and changes in piezometric head, Box-Cox $\lambda = 0.5$, 0.5, and 1.0 respectively.

Figure 9.3: Monthly chloride exported in streamflow for Salmon Catchment predicted by CATPRO calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5$, 0.5, and 1.0 respectively.

The time series of calculated catchment average piezometric heads is shown in Figure 9.4 along with the predicted groundwater storages. The seasonal dynamics are similar, as are the long-term trends but the magnitudes of the relative changes are not well reproduced by the model. Table 9.1 confirms what is shown in Figure 9.2, Figure 9.3 and Figure 9.4.
Figure 9.4: Calculated catchment averaged piezometric heads for the deep aquifer at Salmon Catchment and groundwater storage predicted by CATPRO calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5$, 0.5, and 1.0 respectively.

Table 9.1: Calibration statistics for CATPRO calibrated to streamflow, stream water salinity and changes in piezometric head at Salmon Catchment, Box-Cox $\lambda = 0.5$, 0.5, and 1.0 respectively.

<table>
<thead>
<tr>
<th>Response</th>
<th>Mean Observed</th>
<th>Mean Predicted</th>
<th>Standard Deviation Observed</th>
<th>Standard Deviation Predicted</th>
<th>Coefficient of Determination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Streamflow (mm)</td>
<td>9.98</td>
<td>9.70</td>
<td>21.9</td>
<td>21.1</td>
<td>0.927</td>
</tr>
<tr>
<td>Salt Export (kg.m$^{-2}$)</td>
<td>8.29x10$^{-4}$</td>
<td>7.31x10$^{-4}$</td>
<td>1.28x10$^{-3}$</td>
<td>1.39x10$^{-3}$</td>
<td>0.869</td>
</tr>
<tr>
<td>Change in Groundwater Level (m)</td>
<td>-1.02x10$^{-2}$</td>
<td>5.28x10$^{-3}$</td>
<td>0.137</td>
<td>0.111</td>
<td>0.671</td>
</tr>
</tbody>
</table>

Figure 9.5 shows that predicted interflow from the perched water store is the dominant streamflow mechanism in the CATPRO model for Salmon Catchment, and that it also contributes the greatest mass of chloride to the predicted streamflow (Figure 9.6) even though the concentration of chloride (mg.L$^{-1}$) in the water discharged from the store is usually less than the concentration of the baseflow. Baseflow contributed less water and less chloride despite its greater chloride concentration.
Stokes (1985) measured salinities of between 200 and 500 mg.L$^{-1}$ in the deep aquifer and salinities of between 25 and 50 mg.L$^{-1}$ in the perched groundwater system during 1983. The salinities predicted by CATPRO (Figure 9.7) are therefore of roughly the right order of magnitude although the very high salinity predicted for the perched groundwater contribution to stream flow in the first year of the simulation is questionable even though no observations were made in that year. Groundwater heads at piezometer 1251 were above ground level in 1974 and did not fall below the ground surface until 1977 (Stokes, 1985) indicating that a change in groundwater discharge regime is likely to have occurred over that time. This would have contributed to the somewhat inconsistent behaviour of the chloride dynamics predicted by the model over that period.

The time series of the mass of chloride in the model stores shown in Figure 9.8 are consistent with field observations except for the chloride held in the A store. The long-term trend of chloride held in the deep groundwater store is roughly constant, with the bulk of the chloride stored in the unsaturated zone above the permanent water table. However, the chloride storage of the A store shows a long-term decreasing trend consistent with a shortfall in groundwater chloride discharge to the stream in the first year of simulation for which the A store compensated. This is consistent with the discrepancy between the predicted groundwater storage and piezometric heads for the beginning of the simulation shown in Figure 9.4.
**Figure 9.5:** Time series of predicted contributions to streamflow for Salmon Catchment made by CATPRO calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5,$ and $1.0$ respectively.

**Figure 9.6:** Time series of predicted contributions to stream chloride export for Salmon Catchment made by CATPRO calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5,$ and $1.0$ respectively.
Figure 9.7: Time series of predicted chloride concentrations of contributions to streamflow for Salmon Catchment made by CATPRO calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5, \text{ and } 1.0$ respectively.

Figure 9.8: Time series of predicted mass of chloride in model stores for Salmon Catchment made by CATPRO calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5, \text{ and } 1.0$ respectively.

9.3 Calibration Statistics and Diagnostics

Thirteen model parameters were fitted in the multi-response calibration. The posterior parameter distributions shown in Table 9.2 reveal that the level of
parameter uncertainty has improved considerably over that found following the streamflow-only calibrations (Table 8.1, Table 8.3 and Table 8.4).

Although only one of the baseflow parameters, \(bfm\), could be identified its value is well defined by the information combined in the three calibration time series. The other two base flow parameters were set to values that prevented any influence over the model predictions; \(bfmax\) was fixed at a value of 1.0 and \(bfcrit\) was fixed at zero. This is equivalent to the hypothesis that baseflow occurs whenever there is some groundwater storage and that the discharge is proportional to the storage. This is the simplest possible hypothesis possible for the contribution of the groundwater store to streamflow. Certainly the assumption that there is some groundwater contribution to streamflow is consistent with the observation that the v-notch weir at Salmon is constructed on basement outcrop.

The only fitted parameter which is poorly identified is \(cl\_frac\), the proportion of chloride left in the unsaturated portion of the groundwater store under falling water table conditions. Note that parameters \(wl\), and \(sfmin\) were also set to values that prevent them from having any influence over model predictions.

The posterior correlation matrix of fitted model parameters is shown in Table 9.3, two parameter pairs exhibit high degrees of correlation, they are marked in the table. The maximum storage of the A store, \(pch\_max\) is strongly correlated to \(k\_pref\) this may be because, together they control recharge to the groundwater store and hence have an influence on the predicted baseflow and its chloride contribution to streamflow.

Note that, even though parameter \(wl\) was not identifiable, the other parameters associated with the generation of perched groundwater flow, \(pch\_max\), \(T\_pch\) and \(a\_pch\), were well identified and showed only low degrees of correlation. This supports the inclusion of the added complexity introduced into the hypothesis of the nature of perched groundwater contributions to streamflow, at least with the availability of the additional calibration data sets.
Table 9.2: Posterior parameter distribution for CATPRO hydrosalinity model calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5$, 0.5, and 1.0 respectively.

<table>
<thead>
<tr>
<th>Name</th>
<th>Transform</th>
<th>Mean</th>
<th>Std dev</th>
<th>Untransformed</th>
</tr>
</thead>
<tbody>
<tr>
<td>etmult</td>
<td>None</td>
<td>0.7900</td>
<td>0.01181</td>
<td>0.7900</td>
</tr>
<tr>
<td>tfm</td>
<td>None</td>
<td>0.86</td>
<td>*</td>
<td>0.86</td>
</tr>
<tr>
<td>tfb</td>
<td>None</td>
<td>0.00</td>
<td>*</td>
<td>0.00</td>
</tr>
<tr>
<td>pch_max</td>
<td>Log</td>
<td>8.928</td>
<td>0.5631</td>
<td>7541</td>
</tr>
<tr>
<td>wl</td>
<td>None</td>
<td>1.00</td>
<td>‡</td>
<td>1.00</td>
</tr>
<tr>
<td>T_pch</td>
<td>Log</td>
<td>6.788</td>
<td>1.195</td>
<td>886.8</td>
</tr>
<tr>
<td>a_pch</td>
<td>None</td>
<td>3.046</td>
<td>0.3970</td>
<td>3.046</td>
</tr>
<tr>
<td>sfmin</td>
<td>None</td>
<td>0.0</td>
<td>‡</td>
<td>0.0</td>
</tr>
<tr>
<td>etAmin</td>
<td>Log</td>
<td>5.912</td>
<td>0.05462</td>
<td>369.4</td>
</tr>
<tr>
<td>k_prf</td>
<td>Log</td>
<td>3.435</td>
<td>0.2857</td>
<td>31.03</td>
</tr>
<tr>
<td>bfm</td>
<td>Binomial</td>
<td>-13.69</td>
<td>0.1609</td>
<td>$1.133 \times 10^{-6}$</td>
</tr>
<tr>
<td>bfmax</td>
<td>None</td>
<td>1.0</td>
<td>‡</td>
<td>1.0</td>
</tr>
<tr>
<td>bf_crit</td>
<td>None</td>
<td>0.0</td>
<td>‡</td>
<td>0.0</td>
</tr>
<tr>
<td>pch0</td>
<td>None</td>
<td>-61.67</td>
<td>27.87</td>
<td>-61.67</td>
</tr>
<tr>
<td>gws0</td>
<td>Log</td>
<td>3.846</td>
<td>1.137</td>
<td>46.81</td>
</tr>
<tr>
<td>cla0</td>
<td>Log</td>
<td>-9.346</td>
<td>0.10330</td>
<td>$8.734 \times 10^{-5}$</td>
</tr>
<tr>
<td>clgws0</td>
<td>Log</td>
<td>-6.300</td>
<td>2.57</td>
<td>$1.835 \times 10^{-3}$</td>
</tr>
<tr>
<td>cl_frac</td>
<td>Binomial</td>
<td>-3.117</td>
<td>2.929</td>
<td>$4.239 \times 10^{-2}$</td>
</tr>
<tr>
<td>leak</td>
<td>Binomial</td>
<td>0.0</td>
<td>*</td>
<td>0.5</td>
</tr>
<tr>
<td>gws_fit</td>
<td>None</td>
<td>0.5026</td>
<td>0.03798</td>
<td>0.5026</td>
</tr>
</tbody>
</table>

* denotes parameter value fixed to value determined from field data
‡ denotes parameter value fixed due to unidentifiability
Table 9.3: Posterior correlation matrix of parameters for CATPRO calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5$, 0.5, and 1.0 respectively.

<table>
<thead>
<tr>
<th></th>
<th>et_mult</th>
<th>pch_max</th>
<th>T_pch</th>
<th>a_pch</th>
<th>etAmin</th>
<th>k_pref</th>
<th>bfm</th>
<th>pch0</th>
<th>gws0</th>
<th>cla0</th>
<th>clgws0</th>
<th>cl_frac</th>
<th>gws_fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>et_mult</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pch_max</td>
<td>0.1181</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T_pch</td>
<td>0.1011</td>
<td>0.7334</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a_pch</td>
<td>0.0357</td>
<td>0.0471</td>
<td>0.6958</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>etAmin</td>
<td>-0.2024</td>
<td>-0.1265</td>
<td>-0.3305</td>
<td>-0.4346</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k_pref</td>
<td>0.1467</td>
<td>0.9434</td>
<td>0.7071</td>
<td>0.0039</td>
<td>-0.0700</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bfm</td>
<td>0.3457</td>
<td>0.0294</td>
<td>0.1378</td>
<td>0.1176</td>
<td>0.0000</td>
<td>0.1172</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pch0</td>
<td>0.0478</td>
<td>0.1083</td>
<td>-0.0060</td>
<td>-0.0432</td>
<td>-0.1173</td>
<td>0.0305</td>
<td>-0.0795</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gws0</td>
<td>0.6718</td>
<td>-0.0333</td>
<td>0.1782</td>
<td>0.2798</td>
<td>-0.2812</td>
<td>-0.0444</td>
<td>0.3185</td>
<td>-0.3386</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cla0</td>
<td>-0.0933</td>
<td>-0.7241</td>
<td>-0.5681</td>
<td>-0.0975</td>
<td>0.0808</td>
<td>-0.6412</td>
<td>-0.0435</td>
<td>-0.1056</td>
<td>-0.0862</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>clgws0</td>
<td>0.2034</td>
<td>0.6179</td>
<td>0.3982</td>
<td>-0.0625</td>
<td>-0.0436</td>
<td>0.6066</td>
<td>0.2000</td>
<td>0.0273</td>
<td>-0.1052</td>
<td>-0.2152</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cl_frac</td>
<td>-0.2387</td>
<td>-0.5780</td>
<td>-0.3583</td>
<td>0.0776</td>
<td>0.0497</td>
<td>-0.5703</td>
<td>-0.2603</td>
<td>-0.0318</td>
<td>0.0795</td>
<td>0.1411</td>
<td>-0.9837</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>gws_fit</td>
<td>0.1326</td>
<td>0.0013</td>
<td>0.0274</td>
<td>0.0440</td>
<td>-0.2551</td>
<td>0.0636</td>
<td>0.0094</td>
<td>0.0675</td>
<td>-0.0009</td>
<td>0.0579</td>
<td>0.0117</td>
<td>-0.0189</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
Diagnostic plots, and associated statistical tests, produced by NLFIT were used to assess the validity of the error models for each calibration time series. As stated previously, the Box-Cox $\lambda$ that produced the residuals that best satisfied the assumption of normality for an individual response time series did not necessarily produced the best compromise match to all three calibration time series.

The disturbance variances for the three response time series were 2.76 for streamflow, $1.65 \times 10^{-4}$ for chloride export and $6.97 \times 10^{-3}$ for the change in groundwater storage time series. This means that the weights assigned to the observed response times series in the multi-response objective function were approximately $10^4:400:1$ for chloride export: change in groundwater storage: streamflow. The weights may be viewed as an indication of the sensitivity of the model parameters to the information content of the response time series (see Section 5.3.1).

### 9.3.1 Streamflow

The scatter plot of the streamflow residuals against the predicted values shown in Figure 9.9 indicates that the revised CATPRO model, calibrated to the multi-response data, correctly matches the annual sequences of zero flows significantly better than the previous version (Figure 8.4). Furthermore, the spread of residuals is reasonably even about the zero line though some significant outliers are present.

Figure 9.10, a normal probability plot of the streamflow residuals, and Figure 9.11, a cumulative periodogram both show marked deviations from normality, but are a marginal improvement over those for the base case streamflow calibration (Figure 8.7 and Figure 8.8). There was no indication of autocorrelation of the streamflow residuals (Figure 9.12). Together, these plots indicate a marginal improvement in the validity of the assumption of normality for the streamflow residuals over that shown in the base case streamflow calibration. However, the violation of the assumption of normality for the residuals indicates a structurally imperfect model in which some hypotheses about catchment behaviour are falsified by the available data.
Figure 9.9: Scatter plot of residuals against predicted streamflow for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5,$ and $1.0$ respectively.

Figure 9.10: Normal probability plot of streamflow residuals for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5,$ and $1.0$ respectively.
Figure 9.11: Cumulative periodogram of streamflow residuals for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5,$ and $1.0$ respectively.

Figure 9.12: Autocorrelation plot for streamflow residuals for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5,$ and $1.0$ respectively.
9.3.2 Chloride Export

Figure 9.13, a scatter plot of the chloride export residual as a function of the predicted value shows a strong systematic bias in the residuals. Comparison with Figure 9.2 and Figure 9.3 clearly shows a tendency to over-predict chloride export for the higher streamflow and chloride export peaks and to underestimate chloride export for years of lower chloride export and lower streamflow. This deviation from normality for the residuals is also reflected in the normal probability plot (Figure 9.14) and the cumulative periodogram (Figure 9.15). The autocorrelation plot (Figure 9.16) shows that while the residuals are not autocorrelated, there is a slight indication that the first lag residual shows some tendency to be correlated though it is not statistically significant.

![Figure 9.13: Scatter plot of residuals against predicted streamflow chloride export for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5$, 0.5, and 1.0 respectively.](image-url)
Figure 9.14: Normal probability plot of streamflow chloride export residuals for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5, \text{ and } 1.0$ respectively.

Figure 9.15: Cumulative periodogram of streamflow chloride export residuals for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5, \text{ and } 1.0$ respectively.
Figure 9.16: Autocorrelation plot for streamflow chloride export residuals for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5,$ and $1.0$ respectively.

9.3.3 Groundwater Storage

The scatter plot of the residuals for the change in groundwater storage time series is shown in Figure 9.17, the residuals form an even band about the zero line without a discernible pattern but there are significant outliers. Although the match between the predicted groundwater storage and observed piezometric heads (Figure 9.4) was poor, the assumption of normality for the residuals appears to have been met. The normal probability plot of the residuals (Figure 9.18) and the cumulative periodogram (Figure 9.19) both indicate normality or only a marginal deviation from it for the residuals of the change in groundwater storage response. No autocorrelation of the residuals is suggested by the autocorrelation plot (Figure 9.20).
**Figure 9.17:** Scatter plot of residuals against predicted changes in groundwater storage for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5$, 0.5, and 1.0 respectively.

**Figure 9.18:** Normal probability plot of change in groundwater storage residuals for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5$, 0.5, and 1.0 respectively.
Figure 9.19: Cumulative periodogram of change in groundwater storage residuals for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5$, 0.5, and 1.0 respectively.

Figure 9.20: Autocorrelation plot for change in groundwater storage residuals for CATPRO calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5$, 0.5, and 1.0 respectively.
9.3.4 Objective Function Response Surface

Three response surfaces for the multi-response calibration of CATPRO are shown. All indicate model nonlinearity in the neighbourhood of the 90% linearised probability regions of the selected principal components.

Comparison of the rank 1 and rank 2 principal component response surface plot (Figure 9.21) with the equivalent plot for the calibration of the unmodified CATPRO model to streamflow data alone using Box-Cox $\lambda = 1.0$ (Figure 8.9) shows a significant improvement in model linearity. Comparison to Figure 8.19, the equivalent plot for Box-Cox $\lambda = 0.5$ shows a marginal improvement in model linearity.

![Figure 9.21: CATPRO response surface plot, rank 1 and rank 2 principal components for calibration to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5$, 0.5, and 1.0 respectively.](image)

Figure 9.22 shows the response surface in the vicinity of the 90% linearised probability region of the rank 2 and rank 3 principal components. Comparison with Figure 8.10 and Figure 8.26 for the unmodified CATPRO model calibrated to streamflow data alone indicates significant improvements in model linearity, though not as marked as that for the rank 1 and rank 2 principle components.
Figure 9.23 indicates that model linearity extends throughout the parameter space. Though the degree of nonlinearity is not as severe as indicated in response surface plots for the unmodified CATPRO model calibrated to streamflow data alone.

**Figure 9.22:** CATPRO response surface plot, rank 2 and rank 3 principal components for calibration to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5$, 0.5, and 1.0 respectively.
9.4  PREDICTION OF UNOBSERVED CATCHMENT RESPONSES

Monthly recharge rates and catchment-averaged evapotranspiration predicted by the modified CATPRO model calibrated to the multi-response data are shown in Figure 9.24 and Figure 9.25, respectively. Two significant changes are evident between the confidence limits on the predicted recharge shown here and those shown for the unmodified CATPRO model calibrated to streamflow alone (Figure 8.33 and Figure 8.34). Firstly, the confidence limits about the annual recharge peaks have been significantly reduced. Secondly, and more significantly, the very broad confidence limits about the negative recharge values predicted during times of low or zero streamflows have been greatly reduced.

The model modifications and multi-response calibration strategy have resulted in better identified parameters and therefore a reduction in parameter uncertainty propagating through the model and affecting the uncertainty on the prediction of other catchment responses. It is reasonable to argue that the introduction of the groundwater storage time series has had a marked positive influence on the
identifiability of parameters associated with the groundwater store and accessions to it. Furthermore, because streamflow at Salmon Catchment is ephemeral, the streamflow record contains little information to constrain the model over the summer period. The piezometric head data set contains information, though perhaps limited, about catchment dynamics over the periods of annual zero streamflows.

![Graph](image)

**Figure 9.24:** Time series of unobserved groundwater recharge and 90% confidence limits predicted for Salmon Catchment following calibration to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5, 0.5,$ and $1.0$ respectively.

The 90% confidence limits on the predicted evapotranspiration time series (Figure 9.25) show an improvement over those generated by the streamflow-only calibration shown in Figure 8.34. The unmodified CATPRO model, calibrated to streamflow alone, produced confidence limits that were widest over the summer months when streamflow is low or zero. The modified CATPRO model calibrated to multi-response time series produced 90% prediction limits that tightly bound the predicted values through most of the simulation period.

The modification of the CATPRO model and multi-response calibration strategy has resulted in reduced uncertainty in the prediction of the unobserved catchment responses, of recharge and evapotranspiration. The 90% confidence limits on recharge and evapotranspiration predictions shown in Figure 9.24 and Figure 9.25 are a measure of random error resulting from uncertainty in the model parameters.
Confidence limits provide no information about systematic error in predicted catchment responses arising from the use of an inappropriate model.

Figure 9.25: Time series of unobserved catchment averaged evapotranspiration and 90% confidence limits predicted for Salmon Catchment following calibration to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5$, 0.5, and 1.0 respectively.

Despite the calibration to the multi-response time series failing to determine a posterior parameter distribution consistent with all the calibration data, uncertainty in model predictions arising from random error due to parameter uncertainty has been reduced. This is demonstrated in the 90% confidence limits on the predicted recharge and evapotranspiration. It may be argued that the multi-response calibration has reduced systematic error in model predictions, because auxiliary hypotheses (Popper, 1959), in the form of a catchment chloride balance and a relationship between observed piezometric heads and predicted groundwater storages, have added to the testability of the model. The degree of nonlinearity in the model residuals following modification and calibration to the multi-response time series data was of a similar order to that observed following calibration to streamflow data alone. Therefore, despite the fact that the model has failed the test imposed by the multi-response calibration, it is argued that systematic error has been reduced. The demands of this test on model structure were far more demanding than that imposed by calibration to streamflow data alone, which the previous model had failed.
9.5 Reassessment of a Model Hypothesis

Here, the CATPRO model has been modified following rejection of a model hypothesis about the generation of perched groundwater contributions to streamflow and it has been calibrated to multi-response time series data. This has resulted in a reduction in model nonlinearity as demonstrated in response surface plots of the 90% linearised probability regions of the highest ranked principal components. However, some model hypotheses are still not consistent with the available data as indicated by the violation of the assumption of normality for both the streamflow and stream chloride export responses. The assumption of normality for the groundwater storage residuals could not be rejected on the available evidence, even though the match between predicted and observed values was poor.

From the point of view of the model, the streamflow and chloride export time series provide conflicting information about catchment dynamics. Consequently the multi-response calibration represents a compromise between satisfying the power of the two data sets over the nature of the posterior distributions of the model parameters. Such a compromise is necessary because one or more of the hypotheses about catchment water and chloride dynamics is incorrect or incomplete. The model has been falsified.

Clues to the nature of the failed model hypothesis that contributed to the violation of normality for the streamflow and chloride export residuals and the poor match between predicted and observed chloride export can be gained by analysing the calibration diagnostics and field observations together. Firstly, consider the observations of Stokes (1985) that groundwater heads at piezometer 1251, about 40 m upslope of the stream channel head (see Figure 7.2), were almost 3 m above ground level during the first year of the calibration period and remained artesian until 1977. By 1980 the piezometric surface at 1251 was 2.4 m below ground level. Add to this the extremely poor match between the initial predicted groundwater storage and initial catchment averaged piezometric heads despite the high degree of identifiability of parameter gws0.
In addition, CATPRO under-estimated the streamflow peak for the first year of the simulation but over-predicted the chloride export. The perched groundwater store contributed most of the chloride exported in that first year. Annual chloride export peaks were then under-predicted for the following six years, while the chloride storage of the A store showed a long-term decreasing trend. Chloride storage of both the saturated and unsaturated portions of the deep groundwater store exhibited annual oscillations but no significant long-term trends.

The accumulated evidence is that the model hypothesis of the relationship between groundwater storage and baseflow, and its associated contribution of chloride to stream salinity, and the course of baseflow to the stream channel is at fault. During the opening months of the simulation period, actual piezometric heads were at a level above which baseflow was apparently generated in subsequent years. However, no streamflow (predicted or observed) was generated until the second month of the simulation period. Baseflow was predicted to contribute little more than 0.6% of the streamflow for that month. The predicted contribution made by baseflow to the streamflow increased each month during the first years of simulation until the eighth month (November) in which it contributed all the predicted streamflow.

It appears that the predicted baseflow should have contributed a significantly greater proportion of the streamflow, and stream chloride export, during the first year of simulation. The perched water store compensated for the low level of predicted baseflow contributions, especially the chloride export, which caused the long-term trend in predicted chloride storage for the A store shown in Figure 9.3. A conceptualisation of baseflow generation that matches this hypothesis needs to be formulated and tested in the same way that other model hypotheses have been critically scrutinised here.

9.6 SUMMARY
The CATPRO model structure was revised following its failure to correctly predict the perched aquifer as the dominant source of streamflow at the Salmon Catchment. At the same time a catchment chloride balance model was hypothesised to extract information on catchment dynamics from the stream chloride export time series. The
revised model was then concurrently calibrated to streamflow, stream chloride export and groundwater piezometric head data.

Two of the three baseflow parameters, and two other water balance parameters, were excluded from the multi-response calibration process due to their total unidentifiability. Model parameters, $bf_{max}$, $bf_{crit}$, $wl$ and $sf_{min}$ were set to values that excluded their influence on model predictions. This ensured that they did not contribute to model complexity while not contributing to the ability of the model to reproduce the calibration time series. The model equations were, by design, formulated in such a way as to allow this to be done in accord with Popper’s (1959) argument concerning the nature of auxiliary hypotheses and the testability of scientific theories.

All model parameters, bar one, other than those fixed during calibration were identified with a low degree of uncertainty. Furthermore, the correlation matrix of fitted parameters showed only two parameter pairs that were highly correlated. This was a marked improvement over the calibration to streamflow data only.

The selection of Box-Cox transformations applied to the different response time series had a profound influence on the degree to which the predicted responses matched the data. It also profoundly affected the validity of the assumption of normality for the residuals. The Box-Cox transformations also had an effect on the weight assigned to each response in the multi-response objective function via their effect on the residual variances. The weights assigned to the observed response times series were approximately $10^4:400:1$ for chloride export: change in groundwater storage: streamflow. By considering Popper’s (1959) definition of complexity, as it refers to a hypothesis to be tested, the weights assigned to the responses used in the multi-response calibration may, conceptually, be interpreted as follows. The stream chloride export time series contributed significant information content to the objective function with the addition of only three extra parameters, one of which was poorly identified, one moderately, and one well identified by the multi-response calibration. It’s information content for testing model hypotheses was therefore high in relation to the addition degree of complexity required of the model. The
information content of the groundwater storage time series was low but the complexity cost to the model was also low because only one additional, well defined, parameter was required to extract useful information from the data. The streamflow time series was weighted lower than the groundwater storage response, even though its information content was higher, because ten of the thirteen fitted model parameters were required to match the response and exploit the information content of the streamflow data. In these terms, the water balance component of the model required to match the streamflow time series contributed the greatest degree of complexity to the overall model.

The multi-response calibration represents a compromise between the information on catchment water balance and chloride balance dynamics inherent in the three calibration time series and the model structure. The predicted streamflows did not match observations as well as the predicted values for the streamflow-only calibrations but parameter identifiability improved markedly and the nature of the residuals improved marginally over the earlier calibration. The model over-predicted the stream chloride export in the first year and then consistently under-predicted the chloride export peaks for most of the remainder of the calibration period. The assumption of normality of the residuals was violated for the chloride export time series. The predicted change in groundwater storages matched the observed piezometric heads only in general form, the coefficient of determination was 0.67. However, the assumption of normality of the residuals was not violated for the groundwater storage response.

The perched aquifer was correctly identified as the dominant source of streamflow by the revised CATPRO model calibrated to multi-response catchment data. Furthermore, the average relative chloride concentrations of the perched and deep groundwaters agreed moderately well with observations made by Stokes (1985). The main exception was an extreme spike in the predicted chloride concentration of the perched groundwater contribution to streamflow in the first year of simulation.

The inconsistent chloride concentrations predicted for the perched aquifer, other field observations and an analysis of the calibration diagnostics were used to deduce which
of the model hypotheses was most likely responsible for the poor calibration to the multi-response data. The conceptualisation of the relationship between groundwater storage and baseflow and the chloride load it contributes to stream salinity was determined to be the falsified model hypotheses.

Despite the failure of CATPRO to match the chloride export and piezometric head data well it is argued that the reparameterisation of the perched groundwater flow system and the compromise fit to the multi-response time series data has removed systematic as well as random error from the model predictions. As a hypothesis of the dominant processes controlling the catchment water balance CATPRO exhibits a higher degree of testability when the catchment chloride balance is included as an auxiliary hypothesis and multi-response time series data is used for model calibration. Although the model ultimately failed this test, it was argued that the demands of the concurrent calibration to multi-response catchment data is significantly more demanding of the model than calibration to a streamflow hydrograph alone. Furthermore, the level to which the model reproduced the main dynamics of the three responses implies that systematic error has been reduced. This cannot, of course, be proven without observations of catchment responses not used in the calibration process.
10 AN ALTERNATIVE HYDROSALINITY MODEL

In this chapter an alternative conceptual catchment hydrosalinity model for catchments in the south-west of Western Australia will be presented and exposed to the same critical hypothesis testing methodology applied to the CATPRO model. This will allow a comparison to be made between a model developed in line with the methodology proposed in Chapter 4 and a model developed from a similar conceptual basis but without reference to the proposed methodology. The purpose of the comparison is not so much to compare the models, and their ability to reproduce streamflow and stream salinity time series, but to demonstrate the value of the methodology proposed in Chapter 4. The aim is to demonstrate the value of actively developing a model that is testable according to Popper’s, (1959) theory of the nature of the scientific method.

The LASCAM model (Ruprecht and Sivapalan, 1991; Viney and Sivapalan, 1995; Sivapalan et al., 1996a, 1996b, 1996c) was developed concurrently to, but independently of the CATPRO model. The aim of developing LASCAM was to provide a model that could predict the effects of landuse change on the streamflow and stream water salinity of large catchments in south-west Western Australia. LASCAM consists of a small scale, conceptual catchment water balance model, an accompanying chloride balance model, a streamflow routing algorithm and a subsurface water storage connection mechanism based on a topographic index. Assemblages of small-scale conceptual catchment models are used to represent large-scale catchments (the LArge Scale CAitcument Model). The methods use to assemble a model for a large catchment are described by Sivapalan et al. (1996c) but will not be discussed here because the model components to be compared to CATPRO are the small-scale conceptual catchment model described by Ruprecht and Sivapalan (1991) and Sivapalan et al. (1996a) and the small scale chloride balance model described by Sivapalan et al. (1996b). Although the whole model structure is referred to as LASCAM by Sivapalan et al. (1996a, 1996b, 1996c), for the sake of simplicity, here LASCAM will be used to refer only to the small scale conceptual water and chloride balance model.
**10.1 MODEL DESCRIPTION**

LASCAM consists of three interconnected stores, which accommodate both water and chloride, plus functions describing the relationships between their storage values and the fluxes of water and chloride between them and to the stream channel. A schematic is presented in Figure 10.1 and the model parameters are summarised in Table 10.1. The structures of LASCAM and the original three-store CATPRO model (Figure 7.1) are almost identical. This is because both models are based on the same process based research and the same conceptualisation of the dominant processes operating at the hillslope or small catchment scale in the south-west of Western Australia.

![Figure 10.1: Schematic of LASCAM water balance model showing conceptual stores and fluxes (after Viney and Sivapalan, 1995). See the text for explanations of the model fluxes.](image)

Figure 10.1: Schematic of LASCAM water balance model showing conceptual stores and fluxes (after Viney and Sivapalan, 1995). See the text for explanations of the model fluxes.
Table 10.1: LASCAM model parameters and definitions (after Sivapalan et al., 1996a, 1996b and Viney and Sivapalan, 1995).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$et_{mult}$</td>
<td>Et multiplier to force long-term water balance</td>
</tr>
<tr>
<td>$\alpha_e$</td>
<td>throughfall (intercept) parameter</td>
</tr>
<tr>
<td>$\beta_e$</td>
<td>throughfall (gradient) parameter</td>
</tr>
<tr>
<td>$\alpha_c$</td>
<td>sat’d overland flow parameter</td>
</tr>
<tr>
<td>$\beta_c$</td>
<td>sat’d overland flow exponent</td>
</tr>
<tr>
<td>$\alpha_{in}$</td>
<td>infiltn. potential parameter</td>
</tr>
<tr>
<td>$\beta_{in}$</td>
<td>infiltn. potential exponent</td>
</tr>
<tr>
<td>$\alpha_a$</td>
<td>subsurface flow parameter</td>
</tr>
<tr>
<td>$\beta_a$</td>
<td>subsurface flow exponent</td>
</tr>
<tr>
<td>$f_s$</td>
<td>infiltration capacity</td>
</tr>
<tr>
<td>$\alpha_f$</td>
<td>dependant on Ksat B horizon</td>
</tr>
<tr>
<td>$\beta_f$</td>
<td>describes exponential decrease in Ksat with depth</td>
</tr>
<tr>
<td>$\gamma_a$</td>
<td>A store Et factor</td>
</tr>
<tr>
<td>$\delta_a$</td>
<td>A store Et exponent</td>
</tr>
<tr>
<td>$\gamma_b$</td>
<td>Et parameter for groundwater store</td>
</tr>
<tr>
<td>$\delta_b$</td>
<td>groundwater store Et exponent</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Et parameter for unsat’d store</td>
</tr>
<tr>
<td>$\delta$</td>
<td>A store Et exponent</td>
</tr>
<tr>
<td>$t_d$</td>
<td>F store recharge parameter</td>
</tr>
<tr>
<td>$\alpha_b$</td>
<td>discharge to A store parameter</td>
</tr>
<tr>
<td>$\beta_b$</td>
<td>discharge to A store parameter</td>
</tr>
<tr>
<td>$A0$</td>
<td>initial storage of perched aquifer</td>
</tr>
<tr>
<td>$B0$</td>
<td>initial groundwater storage</td>
</tr>
<tr>
<td>$F0$</td>
<td>initial unsat’d zone storage</td>
</tr>
<tr>
<td>$A_{max}$</td>
<td>max. storage of perched aquifer</td>
</tr>
<tr>
<td>$A_{min}$</td>
<td>min. active storage of perched aquifer</td>
</tr>
<tr>
<td>$B_{max}$</td>
<td>max. groundwater storage deficit</td>
</tr>
<tr>
<td>$F_{max}$</td>
<td>max. storage of unsat’d zone</td>
</tr>
<tr>
<td>$S_{a0}$</td>
<td>initial salt storage of perched aquifer</td>
</tr>
<tr>
<td>$S_{b0}$</td>
<td>initial salt storage of groundwater</td>
</tr>
<tr>
<td>$S_{f0}$</td>
<td>initial salt storage of unsat’d zone</td>
</tr>
<tr>
<td>$S_{p0}$</td>
<td>initial salt storage of soil profile</td>
</tr>
<tr>
<td>$B_{ref}$</td>
<td>threshold for salt leaching to groundwater</td>
</tr>
<tr>
<td>$\kappa_f$</td>
<td>groundwater salt leaching parameter</td>
</tr>
<tr>
<td>$\kappa_m$</td>
<td>groundwater salt leaching parameter</td>
</tr>
<tr>
<td>$\kappa_{ref}$</td>
<td>groundwater salt leaching parameter</td>
</tr>
<tr>
<td>$S_{p0}$</td>
<td>initial salt storage of soil profile</td>
</tr>
<tr>
<td>gwsfit</td>
<td>relates change in B to change in piezometric surface</td>
</tr>
</tbody>
</table>
10.1.1 Water Balance Model

The three conceptual stores shown in Figure 10.1 represent a near-stream perched groundwater system (A store) which includes all water stored in the permeable surface soils even where the deep, permanent aquifer water table enters this horizon. The F store represents the upslope, soil profile through which water percolates to the B store which represents the deep, permanent aquifer. The F store may generate perched lateral groundwater flow which contributes to the near-stream, A store. The volume of water in each of the stores is denoted by $A$, $F$ and $B$, respectively.

The required daily inputs are rainfall, saltfall and potential evaporation. A multiplier $et\_mult$, was introduced to scale the potential evaporation to allow a long-term water balance in the same way as was specified for CATPRO. The parameter $gwsfit$ (Table 10.1) was also added to the LASCAM model to allow calibration to the groundwater piezometric data. The LASCAM model has 35 parameters, two of which, those that determine throughfall, can be fixed to values determined from observations made at Salmon Catchment (Williamson et al., 1987). A further two are described as geometrical parameters and can be fixed to values recommended by Sivapalan et al. (1996b). Including the two parameters introduced above, a total of 33 parameters require calibration.

Interception

The same linear relationship between rainfall and interception used in CATPRO is implemented in LASCAM, the relationship is

$$p_g = \max(\alpha_g + \beta_g p, 0)$$

[10.1]

where $p_g$ is throughfall and $\alpha_g$ and $\beta_g$ are coefficients dependant on the type and density of the vegetation cover. The values determined by Williamson et al. (1987) were used in the calibration of both CATPRO and LASCAM.

Surface Runoff Generation

Throughfall is partitioned into infiltration and surface runoff; infiltration excess and saturation excess runoff mechanisms are included. Saturation excess runoff is
generated when the perched aquifer intersects the soil surface forming a variable contributing area given by

\[
\phi_e = 0 \quad \text{for} \quad A \leq A_{\text{min}} \tag{10.2}
\]

\[
\phi_e = \alpha \left( \frac{A - A_{\text{min}}}{A_{\text{max}} - A_{\text{min}}} \right)^{\beta} \quad \text{for} \quad A > A_{\text{min}}
\]

where \(A_{\text{max}}\) and \(A_{\text{min}}\) define the range of possible storage values of the store, and \(\alpha\) and \(\beta\) are parameters. Sivapalan \textit{et al.} (1996a) describe \(\beta\) as a geometric parameter that according to the analysis of Sivapalan and Wittorff (1993) should be set to 1.0, this leaves \(\alpha\), \(A_{\text{max}}\) and \(A_{\text{min}}\) to be determined by calibration. Saturation excess runoff is then

\[
q_{se} = \phi_e p_g \tag{10.3}
\]

Infiltration excess runoff, \(q_{ie}\), is dependant on the infiltration capacity of the surface soil \(f_s^*\),

\[
q_{ie} = \max(p_g - q_{se} - f_s^*, 0) \tag{10.4}
\]

and infiltration, \(p_c\), is equal to

\[
p_c = p_g - q_{se} - q_{ie} \tag{10.5}
\]

**Subsurface Runoff Generation**

Perched groundwater flow is assumed to be generated at the top of the perching layer by either infiltration excess or saturation excess. Saturation excess interflow is generated on that part of the perching layer above which saturation has developed:

\[
\phi_{ss} = 0 \quad \text{for} \quad A \leq A_{\text{min}} \tag{10.6}
\]
\[
\phi_{ss} = \alpha_{ss} \left( \frac{A - A_{\min}}{A_{\max} - A_{\min}} \right)^{\beta_{ss}} \quad \text{for} \quad A > A_{\min}
\]

where \( \alpha_{ss} \) and \( \beta_{ss} \) are parameters, the recommended value for the geometric parameter, \( \beta_{ss} \) is 1.0 (Sivapalan and Wittorff, 1993).

Saturation excess interflow is then

\[
q_{ss} = p_c \frac{\phi_{ss} - \phi_c}{1 - \phi_c}
\]  \[10.7\]

Infiltration excess interflow, \( q_{sie} \), is dependant on the infiltration capacity which is a function of the cumulative infiltration as represented by the state of the infiltration store

\[
f_{ss}^* = \alpha_f \left( 1 - \frac{B}{B_{\max}} \right) \left( \frac{F}{F_{\max}} \right)^{-\beta_f}
\]  \[10.8\]

where \( \alpha_f \) is a parameter dependant on the average hydraulic conductivity of the perching layer, and \( \beta_f \) describes the exponential decrease in hydraulic conductivity with depth. The infiltration into the perching layer and the infiltration excess interflow are then

\[
f_a = \min \left( p_c \frac{1 - \phi_{ss}}{1 - \phi_c}, f_{ss}^* \right)
\]  \[10.9\]

and

\[
q_{sie} = \max \left( p_c \frac{1 - \phi_{ss}}{1 - \phi_c} - f_{ss}^*, 0 \right)
\]  \[10.10\]
The total subsurface runoff given by

\[ q_{ss} = q_{ste} + q_{slr} = p_c - f_a \]  \[10.11\]

is added to the A store and the actual infiltration \( f_a \) added to the F store.

**Generation of Interflow**

The perched groundwater contribution to streamflow is dependent on the volume of water in the A store and is given by

\[ q_A = \begin{cases} 0 & \text{if } A \leq A_{\text{min}} \\ \beta_A \left( \frac{A - A_{\text{min}}}{A_{\text{max}} - A_{\text{min}}} \right)^{\alpha_A} & \text{if } A > A_{\text{min}} \end{cases} \]  \[10.12\]

**Groundwater Recharge**

The deep, permanent aquifer is recharged from the infiltration store by applying a detention time parameter \( t_d \)

\[ r_F = \frac{F}{t_d} \]  \[10.13\]

and from the saturated portion of the A store at a rate determined by the infiltration rate into the perching layer

\[ r_A = \phi_{ss} f_{ss}^* \]  \[10.14\]

**Groundwater Discharge**

Groundwater discharge is assumed to take place into the perched aquifer system immediately adjacent to the stream channel and not directly to streamflow. Discharge to the A store is then
\[ q_B = \beta_B \left( \exp \left( \alpha_B \left( \frac{B}{B_{\text{max}}} \right) \right) - 1 \right) \]  

where \( \alpha_b \) and \( \beta_b \) are parameters to be determined by calibration.

**Evapotranspiration**

Evapotranspiration is extracted from all three model stores according to nonlinear equations relating actual evapotranspiration to the potential rate, \( e_p \), and the volume of water in storage. For the A store, evapotranspiration is extracted from the saturated fraction of the store at the potential rate and a lower rate from the remainder

\[ e_A = \phi_A e_p + \gamma_A e_p \left( \frac{A}{A_{\text{max}}} \right)^{\delta_A} \]  

Evapotranspiration from the groundwater and infiltration stores are calculated using similar equations

\[ e_B = \gamma_B e_p \left( \frac{B}{B_{\text{max}}} \right)^{\delta_B} \]  

and

\[ e_F = \gamma_F e_p \left( \frac{F}{F_{\text{max}}} \right)^{\delta_F} \]  

where \( \gamma_A, \gamma_B, \gamma_F, \delta_A, \delta_B \) and \( \delta_F \) are parameters dependant on the soil properties and the type and density of the vegetation.

**Store Water Balances**

The perched aquifer, represented by the A store, receives water by subsurface runoff generated above the perching layer by both infiltration excess and saturation excess mechanisms, and from groundwater discharge. The store also discharges water to the stream channel, provides recharge to the deep, permanent groundwater system and provides water for evapotranspiration.
\[ A_{t+1} = A_t + q_{se} - q_A + q_B - e_A - r_A \]  \[10.19\]

The water balance equations for the groundwater and infiltration stores are

\[ B_{t+1} = B_t - q_B - e_B + r_A + r_F \]  \[10.20\]

and

\[ F_{t+1} = F_t + f_u - e_F - r_F \]  \[10.21\]

where the \( t \) and \( t+1 \) refer to the time-step.

The total streamflow and evapotranspiration are given by

\[ q_T = q_{se} + q_{ie} + q_A \]  \[10.22\]

and

\[ e_T = e_A + e_B + e_F \]  \[10.23\]

10.1.2 Chloride Balance Model

The structure of the catchment salt balance model is identical to the structure of the water balance model. There are three salt stores \( S_A, S_B \) and \( S_F \) in which chloride is assumed well mixed at concentrations \( c_A, c_B, \) and \( c_F, \) respectively. The chloride concentration of surface runoff is assumed constant and equal to the chloride concentration of the throughfall, \( c_g. \)

There is an additional salt store \( S_P \) which represents the chloride storage of the unsaturated profile above the water table. Groundwater recharge and increases in groundwater levels are assumed to mobilise chloride from this store as follows:

\[ \Delta S_{B1} = \max(\kappa_f S_F (r_F - r_{ref}), 0) \]  \[10.24\]

\[ \Delta S_{B2} = \max(\kappa_p S_F (B - B_{ref}), 0) \]  \[10.25\]
where $r_F$ is the recharge rate as defined above and $r_{\text{ref}}$ is a threshold recharge rate, above which chloride leaching is assumed to occur, similarly, $B_{\text{ref}}$ is a threshold groundwater storage, above which chloride is mobilised by increases in groundwater level and $\kappa_1$ and $\kappa_m$ are proportionality constants.

The total chloride mobilised from the store is then

$$\Delta S_B = \Delta S_{B_1} + \Delta S_{B_2}$$ \hspace{1cm} [10.26]

The chloride balances for the conceptual stores are therefore

$$S_{A_{t+1}} = S_{A_t} + c_A q_{ss} - c_A q_A + c_B q_B - c_A r_A$$ \hspace{1cm} [10.27]

$$S_{B_{t+1}} = S_{B_t} - c_B q_B + c_A r_A + c_F r_F + \Delta S_B$$ \hspace{1cm} [10.28]

$$S_{F_{t+1}} = S_{F_t} + c_f f_a - c_F r_F$$ \hspace{1cm} [10.29]

$$S_{P_{t+1}} = S_{P_t} - \Delta S_B$$ \hspace{1cm} [10.30]

where the concentrations of chloride in the stores are given by

$$c_g = \frac{c_p P}{P_g}$$ \hspace{1cm} [10.31]

$$c_A = \frac{S_A}{A}$$ \hspace{1cm} [10.32]

$$c_B = \frac{S_B}{B}$$ \hspace{1cm} [10.33]

$$c_F = \frac{S_F}{F}$$ \hspace{1cm} [10.34]

where $c_p$ is the known concentration of chloride in the rainfall.

The daily chloride load and concentration in the streamflow are then

$$S_q = c_g (q_{se} + q_{le}) + c_A q_A$$ \hspace{1cm} [10.35]
and

\[ c_q = \frac{S_q}{q_T} \]  

[10.36]

10.2 Multi-Response Calibration

10.2.1 Calibration Strategy

Sivapalan et al. (1996b) recommend calibrating the LASCAM water balance model to streamflow data alone, then fixing the water balance parameters and independently calibrating the chloride balance parameters to observed trends in catchment chloride export. This strategy was not adopted here because it is not consistent with the methodology proposed in Chapter 4. The strategy proposed by Sivapalan et al. (1996b) is designed to minimise the number of parameters calibrated at any one time to make the calibration problem more tractable. This strategy, however, fails to take advantage of the information content about streamflow generation processes inherent in the stream-water chloride time series.

The calibration strategy implemented here is the same as for the multi-response calibration of the CATPRO hydrosalinity model. LASCAM was calibrated to the three catchment response time series concurrently with all parameters included, except those controlling throughfall and the geometric parameters \( \beta_c \) and \( \beta_{ss} \). The total number of parameters requiring calibration was 33. As was the case with CATPRO, the choice of Box-Cox \( \lambda \) for the streamflow and chloride export responses were critical in determining the match to the observed data and on the behaviour of the residuals.

With such a large number of parameters to be determined by calibration, locating a global optimum parameter set proved very difficult. During manually controlled Gauss-Marquardt optimisation, there appeared to be considerable interaction between the choice of Box-Cox \( \lambda \) for the streamflow and chloride export responses, the parameter values, parameter observability and the match between predicted and
observed responses. The final choices of Box-Cox transformations were, $\lambda = 0.25$, 0.25 and 1.0 for streamflow, chloride export and piezometric head, respectively.

The disturbance variances for the three response time series were 0.770 for streamflow, $6.86 \times 10^{-4}$ for chloride export and $8.46 \times 10^{-3}$ for the change in groundwater storage time series. The weights assigned to the observed response times series in the multi-response objective function were therefore $10^3:90:1$ for chloride export: change in groundwater storage: streamflow. The equivalent weight for the CATPRO multi-response calibration were $10^4:400:1$, indicating that the CATPRO model parameters were more sensitive to the stream salinity and piezometric head data than the LASCAM parameters. In practical terms, this implies that these additional time series data contributed more to improved parameter observability for CATPRO than for LASCAM.

During the course of the calibration procedure 18 parameters required fixing because they were unobservable (Table 10.2). A further five parameters were fixed following the initial calibration due to low levels of observability that resulted in very large standard deviations in relation to their expected values, in each case variations in the fitted parameter value of less than one standard deviation would result in infeasible parameter values.

Table 10.2 shows the posterior distribution of fitted model parameters, the values of parameters not included in the optimisation or fixed during the calibration process are also included. Important points to note from the table include: all the A store parameters are unobservable, $A_{\text{min}}$ required fixing to zero, and the initial conditions for both water and chloride, $A_0$ and $S_A$, required fixing at values that are effectively zero. Furthermore, the surface soil infiltration capacity $f_s$ was unobservable, implying that only one of the two surface runoff mechanisms included in LASCAM are required to match the streamflow and chloride export dynamics. All the initial chloride storage parameters required fixing except $S_B$. The parameters associated with chloride leaching from the P store were also unidentifiable. Parameters fixed due to unacceptably large standard deviations are shown in Table 10.3.
The calibration goodness of fit statistics are shown in Table 10.4, the streamflow and chloride export time series are moderately well matched (see also Figure 10.2 and Figure 10.3) but the change in groundwater storage time series (Figure 10.4) is not well predicted. Comparison with the CATPRO multi-response calibration (Table 9.1) indicates that LASCAM matched the chloride export dynamics better than CATPRO but not the streamflow or groundwater storage dynamics.

The posterior correlation matrix of model parameters (Table 10.5) revealed that four parameter pairs exhibited extreme degrees of correlation. Five individual parameters were involved in the four pairs for which correlation coefficients were greater than 0.9, of these four control the initial conditions, maximum storage and discharge from the groundwater store; the fifth parameter is $F_{\text{max}}$, the maximum storage of the unsaturated soil profile.
Table 10.2: Posterior parameter distribution for LASCAM hydrosalinity model calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25$, $0.25$, and $1.0$ respectively. Five parameters not subject to calibration due to large standard deviations relative to expected values are shown as fixed.

<table>
<thead>
<tr>
<th>Name</th>
<th>Transform</th>
<th>Mean</th>
<th>Std dev</th>
<th>Untransformed</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_{t}$</td>
<td>None</td>
<td>0.8436</td>
<td>4.073x10^{-2}</td>
<td>0.8436</td>
</tr>
<tr>
<td>$a_{c}$</td>
<td>None</td>
<td>0</td>
<td>*</td>
<td>0</td>
</tr>
<tr>
<td>$\beta_{c}$</td>
<td>None</td>
<td>0.86</td>
<td>*</td>
<td>0.86</td>
</tr>
<tr>
<td>$a_{c}$</td>
<td>Log</td>
<td>0.4629</td>
<td>†</td>
<td>1.589</td>
</tr>
<tr>
<td>$\beta_{c}$</td>
<td>None</td>
<td>1</td>
<td>⊗</td>
<td>1</td>
</tr>
<tr>
<td>$a_{a}$</td>
<td>Log</td>
<td>0.5954</td>
<td>†</td>
<td>1.814</td>
</tr>
<tr>
<td>$\beta_{a}$</td>
<td>None</td>
<td>1</td>
<td>⊗</td>
<td>1</td>
</tr>
<tr>
<td>$\alpha_{a}$</td>
<td>Log</td>
<td>0.7730</td>
<td>2.506x10^{-2}</td>
<td>2.166</td>
</tr>
<tr>
<td>$\beta_{a}$</td>
<td>Log</td>
<td>5.482</td>
<td>†</td>
<td>240.4</td>
</tr>
<tr>
<td>$\gamma_{c}$</td>
<td>Log</td>
<td>4.803</td>
<td>†</td>
<td>121.8</td>
</tr>
<tr>
<td>$\delta_{c}$</td>
<td>Log</td>
<td>2.194</td>
<td>†</td>
<td>8.968</td>
</tr>
<tr>
<td>$\beta_{c}$</td>
<td>Log</td>
<td>0.9898</td>
<td>5.655x10^{-2}</td>
<td>2.691</td>
</tr>
<tr>
<td>$\gamma_{e}$</td>
<td>Log</td>
<td>6.768</td>
<td>†</td>
<td>869.4</td>
</tr>
<tr>
<td>$\delta_{e}$</td>
<td>Log</td>
<td>4.400</td>
<td>†</td>
<td>81.42</td>
</tr>
<tr>
<td>$\gamma_{e}$</td>
<td>Log</td>
<td>100.0</td>
<td>†</td>
<td>$\rightarrow\infty$</td>
</tr>
<tr>
<td>$\delta_{e}$</td>
<td>Log</td>
<td>-38020</td>
<td>†</td>
<td>$\rightarrow0$</td>
</tr>
<tr>
<td>$\gamma_{f}$</td>
<td>Binomial</td>
<td>-1.143</td>
<td>†</td>
<td>0.2417</td>
</tr>
<tr>
<td>$\delta_{f}$</td>
<td>Binomial</td>
<td>-0.1506</td>
<td>†</td>
<td>0.4624</td>
</tr>
<tr>
<td>$\gamma_{h}$</td>
<td>Log</td>
<td>4.369</td>
<td>†</td>
<td>78.99</td>
</tr>
<tr>
<td>$\delta_{h}$</td>
<td>Log</td>
<td>0.7918</td>
<td>†</td>
<td>2.207</td>
</tr>
<tr>
<td>$\beta_{h}$</td>
<td>Binomial</td>
<td>-6.926</td>
<td>0.9873</td>
<td>9.810x10^{4}</td>
</tr>
<tr>
<td>$A_{0}$</td>
<td>Log</td>
<td>-21910</td>
<td>†</td>
<td>$\rightarrow0$</td>
</tr>
<tr>
<td>$B_{0}$</td>
<td>Log</td>
<td>9.248</td>
<td>0.8730</td>
<td>1.038x10^{4}</td>
</tr>
<tr>
<td>$F_{0}$</td>
<td>Log</td>
<td>5.168</td>
<td>0.1182</td>
<td>175.6</td>
</tr>
<tr>
<td>$A_{\text{max}}$</td>
<td>Log</td>
<td>5.907</td>
<td>†</td>
<td>367.7</td>
</tr>
<tr>
<td>$A_{\text{min}}$</td>
<td>None</td>
<td>0</td>
<td>†</td>
<td>0</td>
</tr>
<tr>
<td>$B_{\text{max}}$</td>
<td>Log</td>
<td>9.419</td>
<td>0.7245</td>
<td>1.232x10^{4}</td>
</tr>
<tr>
<td>$F_{\text{max}}$</td>
<td>Log</td>
<td>6.004</td>
<td>0.2996</td>
<td>405.1</td>
</tr>
<tr>
<td>$S_{0}$</td>
<td>Log</td>
<td>-38.04</td>
<td>†</td>
<td>3.022x10^{17}</td>
</tr>
<tr>
<td>$S_{0}$</td>
<td>Log</td>
<td>6.037</td>
<td>1.320</td>
<td>418.6</td>
</tr>
<tr>
<td>$S_{0}$</td>
<td>Log</td>
<td>-37.00</td>
<td>†</td>
<td>8.541x10^{17}</td>
</tr>
<tr>
<td>$B_{\text{ref}}$</td>
<td>Log</td>
<td>8.930</td>
<td>†</td>
<td>7.557x10^{17}</td>
</tr>
<tr>
<td>$K_{1}$</td>
<td>Log</td>
<td>-0.6931</td>
<td>†</td>
<td>0.5000</td>
</tr>
<tr>
<td>$K_{0}$</td>
<td>Log</td>
<td>-42.56</td>
<td>†</td>
<td>3.286x10^{17}</td>
</tr>
<tr>
<td>$r_{\text{ref}}$</td>
<td>Log</td>
<td>6.215</td>
<td>†</td>
<td>500.0</td>
</tr>
<tr>
<td>$S_{\text{max}}$</td>
<td>Log</td>
<td>-11.17</td>
<td>†</td>
<td>1.414x10^{13}</td>
</tr>
<tr>
<td>$gwsfit$</td>
<td>None</td>
<td>1.694x10^{-3}</td>
<td>1.336x10^{-4}</td>
<td>1.694x10^{-3}</td>
</tr>
</tbody>
</table>

* denotes parameter fixed to value determined from field data
⊗ denotes parameter fixed to value recommended by Sivapalan et al. (1996a)
‡ denotes parameter value fixed due to unidentifiability
Table 10.3: LASCAM parameters fixed due to large standard deviations relative to expected values. See text for explanation.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Transform</th>
<th>Mean</th>
<th>Std dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_c$</td>
<td>Log</td>
<td>0.4629</td>
<td>5.943</td>
</tr>
<tr>
<td>$\beta_a$</td>
<td>Log</td>
<td>5.482</td>
<td>12.90</td>
</tr>
<tr>
<td>$\alpha_f$</td>
<td>Log</td>
<td>2.194</td>
<td>7.798</td>
</tr>
<tr>
<td>$\alpha_b$</td>
<td>Log</td>
<td>0.7918</td>
<td>1.055</td>
</tr>
<tr>
<td>$A_{max}$</td>
<td>Log</td>
<td>5.907</td>
<td>5.942</td>
</tr>
</tbody>
</table>

Table 10.4: Calibration statistics for LASCAM calibrated to streamflow, stream water salinity and changes in piezometric head at Salmon Catchment, Box-Cox $\lambda = 0.25, 0.25, \text{ and } 1.0$ respectively.

<table>
<thead>
<tr>
<th>Response</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Coefficient of Determination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed</td>
<td>Predicted</td>
<td>Observed</td>
</tr>
<tr>
<td>Streamflow (mm)</td>
<td>9.98</td>
<td>12.8</td>
<td>21.9</td>
</tr>
<tr>
<td>Salt Export (kg.m$^{-2}$)</td>
<td>8.29x10$^{-4}$</td>
<td>8.43x10$^{-4}$</td>
<td>1.28x10$^{-3}$</td>
</tr>
<tr>
<td>Change in Groundwater Level (m)</td>
<td>-1.02x10$^{-2}$</td>
<td>-1.71x10$^{-2}$</td>
<td>0.137</td>
</tr>
</tbody>
</table>

10.2.2 Model Predictions

The predicted streamflow, stream water chloride export and groundwater storage time series are shown in Figure 10.2, Figure 10.3 and Figure 10.4, respectively. LASCAM produced a visually better fit to the observed streamflow than that produced by CATPRO when calibrated to the multi-response data (Figure 9.2). The fit to the chloride export time series is significantly better than that produced by CATPRO (Figure 9.3), especially during the middle part of the calibration period. The fit to the piezometric data is very different to that achieved by CATPRO and appears to match the actual values to a higher degree, though the final three years data is not well matched. The calibration statistics (Table 10.4) do not support this visual assessment.
Table 10.5: Posterior correlation matrix of parameters for LASCAM model calibrated to streamflow, stream water salinity and changes in piezometric head at Salmon Catchment, Box-Cox $\lambda = 0.25, 0.25,$ and $1.0$ respectively. Correlation coefficients $> 0.9$ are highlighted.

<table>
<thead>
<tr>
<th></th>
<th>et_mult</th>
<th>$\alpha_a$</th>
<th>$\beta_f$</th>
<th>$\beta_b$</th>
<th>$B0$</th>
<th>$F0$</th>
<th>$B_{\text{max}}$</th>
<th>$F_{\text{max}}$</th>
<th>$S_{00}$</th>
<th>gwsfit</th>
</tr>
</thead>
<tbody>
<tr>
<td>et_mult</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_a$</td>
<td>-0.227</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_f$</td>
<td>-0.197</td>
<td>0.166</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_b$</td>
<td>-0.101</td>
<td>-0.190</td>
<td>0.088</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B0$</td>
<td>0.403</td>
<td>0.154</td>
<td>-0.209</td>
<td>-0.486</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F0$</td>
<td>-0.215</td>
<td>-0.052</td>
<td>-0.047</td>
<td>-0.022</td>
<td>-0.111</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B_{\text{max}}$</td>
<td>0.447</td>
<td>0.130</td>
<td>-0.230</td>
<td>-0.477</td>
<td>0.998</td>
<td>-0.106</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F_{\text{max}}$</td>
<td>0.187</td>
<td>0.180</td>
<td>-0.278</td>
<td>-0.486</td>
<td>0.959</td>
<td>-0.080</td>
<td>0.943</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_{00}$</td>
<td>0.304</td>
<td>0.202</td>
<td>-0.184</td>
<td>-0.930</td>
<td>0.762</td>
<td>-0.029</td>
<td>0.760</td>
<td>0.726</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>gwsfit</td>
<td>-0.503</td>
<td>0.071</td>
<td>0.204</td>
<td>-0.034</td>
<td>-0.232</td>
<td>0.105</td>
<td>-0.256</td>
<td>-0.140</td>
<td>-0.108</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 10.5 shows time series of the predicted mechanisms of streamflow generation, interflow is correctly identified as the dominant source of streamflow in all years (Stokes, 1985; Turner et al., 1987). LASCAM predicted that surface runoff made a contribution to streamflow each year, though the magnitude of its contribution to streamflow is greater than that indicated by the field data. CATPRO, however, predicted no surface runoff in any year of the calibration period (Figure 9.5). LASCAM predicted extremely low rates of baseflow with little seasonal variation but a long-term decreasing trend (Figure 10.5). This behaviour is not consistent with the observations of piezometric head (Stokes, 1985), especially those at piezometer 1251, which is within 50 m of the stream headwaters, where heads were artesian at the commencement of the calibration period. The time series of A and groundwater stores (Figure 10.6) are qualitatively consistent with field observations and process studies in the region, though the predicted groundwater storage does not demonstrate the general upward trend evident over the last three years of record (Figure 7.6 and Figure 10.4). The large range over which the $F$ store varies is consistent with neutron moisture meter and tensiometer observations of deep soil profiles in the Darling Range (Sharma, 1984; Ruprecht and Schofield, 1990a, 1990b, 1993). However, the
slightly increasing trend is not consistent with the long-term trends in A store and groundwater storage.

The predicted chloride contribution of the perched aquifer to streamflow chloride (Figure 10.7) is consistent with the field observations of Stokes (1985). However, the long-term trends in baseflow chloride flux (Figure 10.7) and concentration (Figure 10.8) are consistent with the long-term decreasing trend in predicted groundwater storage which fails to match the observed recovery of piezometric levels over the final three years of the simulation period (Figure 10.4 and Figure 10.6). Together with the predicted chloride storages (Figure 10.9) they demonstrate a higher degree of internal consistency than apparent in the CATPRO predictions of the salt fluxes and concentrations shown in Figure 9.6, Figure 9.7 and Figure 9.8, though there is still significant differences between the observed and predicted trends. Note that because baseflow contributes water and salt to the A store rather than directly to streamflow the baseflow chloride flux shown in Figure 10.7 contributes to the chloride storage of the A store as shown in Figure 10.9. The predicted chloride storages are consistent with field observations (Stokes, 1985; Johnston, 1983, 1987a), however, the low level of chloride storage in the supplementary chloride store $S_p0$ (Table 10.2) is not consistent with these data.

![Figure 10.2: Monthly streamflow for Salmon Catchment predicted by LASCAM calibrated to streamflow, stream chloride export and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25$, and 1.0 respectively.](image)
Figure 10.3: Monthly stream chloride export for Salmon Catchment predicted by LASCAM calibrated to streamflow, stream chloride export and changes in piezometric head, Box-Cox $\lambda = 0.25$, 0.25, and 1.0 respectively.

Figure 10.4: Calculated catchment averaged piezometric heads for the deep aquifer at Salmon Catchment and groundwater storage predicted by LASCAM calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25$, 0.25, and 1.0 respectively.
Figure 10.5: Time series of predicted contributions to streamflow for Salmon Catchment made by LASCAM calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25$, 0.25, and 1.0 respectively.

Figure 10.6: Time series of predicted model storages for Salmon Catchment made by LASCAM calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25$, 0.25, and 1.0 respectively.
Figure 10.7: Time series of predicted contributions to stream chloride export for Salmon Catchment made by LASCAM calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25$, $0.25$, and $1.0$ respectively.

Figure 10.8: Time series of predicted chloride concentrations of contributions to streamflow for Salmon Catchment made by LASCAM calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25$, $0.25$, and $1.0$ respectively.
Figure 10.9: Time series of predicted mass of chloride in model stores for Salmon Catchment made by LASCAM calibrated to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.5$, 0.5, and 1.0 respectively.

10.2.3 Calibration Diagnostics

Diagnostic plots, and associated statistical tests, produced by NLFIT were used to assess the validity of the error models for each calibration time series.

Streamflow

Figure 10.10 shows the LASCAM streamflow residuals plotted against the predicted values. The residuals form a broad band concentrated below the zero line, furthermore, there is a spread of both positive and negative residuals along the ordinate axis. The negative residuals are a result of LASCAM predicting small positive streamflows in every month of the calibration period when none was observed, due to the unobservability of $A_{min}$ which determines the A store volume below which streamflow is not generated. The positive residuals are a result of systematic under-prediction of small observed streamflows.
Figure 10.10: Scatter plot of residuals against predicted streamflow for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25$, 0.25, and 1.0 respectively.

Figure 10.11 shows that the assumption of normally distributed residuals is violated for the LASCAM streamflow predictions, however, the degree of nonlinearity shown here is a marked improvement on that demonstrated by the CATPRO streamflow residuals (Figure 9.10), and the Kolmogorov-Smirnov test confirms this. Similarly, the cumulative periodogram (Figure 10.12) also shows a significant deviation from linearity but an improvement relative to the CATPRO residuals. (Figure 9.11). The residual autocorrelation plot (Figure 10.13) shows only marginal autocorrelation for the lag 1 streamflow residuals.
Figure 10.11: Normal probability plot of streamflow residuals for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25$, 0.25, and 1.0 respectively.

Figure 10.12: Cumulative periodogram of streamflow residuals for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25$, 0.25, and 1.0 respectively.
Figure 10.13: Autocorrelation plot for streamflow residuals for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25, \text{ and } 1.0$ respectively.

Chloride Export

The plot of chloride export residuals for LASCAM in Figure 10.14 shows a small dense cluster of residuals about the origin but an otherwise uniform distribution of residuals about the zero line. The systematic pattern apparent in the CATPRO chloride export residuals (Figure 9.13) is not replicated. The cause of the cluster of residuals about the origin is the prediction of very small streamflows during the summer months due to the unobservability of $A_{\text{min}}$, as described above.

The normal probability plot shown in Figure 10.15 indicates that the assumption of normality for the residuals has been violated. The Kolmogorov-Smirnov test statistic indicates a small improvement over the CATPRO chloride export normal probability plot. The cumulative periodogram (Figure 10.16) indicates normality, which is a significant improvement over the equivalent CATPRO residual diagnostic. No autocorrelation of the chloride export residuals is implied (Figure 10.17).
Figure 10.14: Scatter plot of residuals against predicted chloride export for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25, \text{ and } 1.0$ respectively.

Figure 10.15: Normal probability plot of chloride export residuals for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25, \text{ and } 1.0$ respectively.
Figure 10.16: Cumulative periodogram of chloride export residuals for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25,$ and $1.0$ respectively.

Figure 10.17: Autocorrelation plot for chloride export residuals for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25,$ and $1.0$ respectively.
Groundwater Storage

Figure 10.18 shows the LASCAM groundwater storage residuals plotted against the predicted values, the pattern of residuals is very similar to that produced by CATPRO (Figure 9.17) and no significant irregularities are indicated. The normal probability plot (Figure 10.19) shows no deviation from normality, as did the normal probability plot for the CATPRO groundwater storage residuals (Figure 9.18). The cumulative periodogram (Figure 10.20), however, indicates a strong periodicity in the groundwater storage residuals, which is confirmed by the Kolmogorov-Smirnov test statistic. This result is significantly more severe than that shown in Figure 9.19 for the CATPRO groundwater storage residuals. Again, no autocorrelation of the residuals was implied (Figure 10.21).

![Figure 10.18: Scatter plot of residuals against predicted changes in groundwater storage for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25,$ and 1.0 respectively.](image-url)
Figure 10.19: Normal probability plot of changes in groundwater storage residuals for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25$, $0.25$, and $1.0$ respectively.

Figure 10.20: Cumulative periodogram of changes in groundwater storage residuals for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25$, $0.25$, and $1.0$ respectively.
Figure 10.21: Autocorrelation plot for changes in groundwater storage residuals for LASCAM calibrated to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25$, 0.25, and 1.0 respectively.

10.2.4 Objective Function Response Surface

Response surface plots could not be generated for the posterior distribution of model parameters without fixing five parameters for which deviations of much less than one standard deviation would result in an infeasible parameter value (Table 10.3). The posterior distribution of model parameters shown in Table 10.2 includes only those parameters, which were sufficiently observable so as not to include infeasible values within one standard deviation of the expected value.

The response surface plots shown in Figure 10.22, Figure 10.23 and Figure 10.24 were produced by fixing the unobservable parameters to the values shown in Table 10.2. The response surfaces orientated along the rank 3 and rank 4 (Figure 10.22), and along the rank 4 and rank 5 (Figure 10.23) principal component directions still encountered infeasible parameter space and show strong deviations from linearity.
Figure 10.22: LASCAM response surface plot, rank 3 and rank 4 principal components for calibration to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25$, 0.25, and 1.0 respectively.
Figure 10.23: LASCAM response surface plot, rank 4 and rank 5 components for calibration to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25,$ and $1.0$ respectively.

The response surface orientated along the rank 8 and rank 9 principal components (Figure 10.24), shows almost linear behaviour. This kind of behaviour was only observed in response surfaces plotted along the lower ranked principal component directions and is one of the most nearly linear response surfaces produced for the LASCAM calibration to the multiple response time series data. The degree of linearity for these lower ranked principal component directions was, however, closer to linear than those observed for the lower ranked principal component directions of the CATPRO multi response objective function.
Figure 10.24: LASCAM response surface plot, rank 7 and rank 8 principal components for calibration to Salmon Catchment streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25, \text{ and } 1.0$ respectively.

10.2.5 Prediction of Unobserved Catchment Responses

Confidence limits on the unobserved catchment responses, groundwater recharge and evapotranspiration were generated from the same posterior parameter distributions used to generate the response surface plots shown above (Table 10.2). Prediction limits and confidence limits are an indication of the effect of random error resulting from uncertainty in the model parameters, they cannot provide information about systematic error in predicted catchment responses arising from the use of an inappropriate model. In this case, the confidence limits represent a lower limit to the effect of random errors on predicted catchment responses due to uncertainty in the least observable parameters not being considered because they were excluded from the posterior distribution (Table 10.3). Furthermore, the confidence limits do not convey any information about the systematic error demonstrated to be present in the model predictions as indicated by the nature of the groundwater storage residuals in particular.
Note that fixing the five least observable parameters to their optimised values resulted in the posterior distributions of ten parameters being considered in the calculation of prediction limits on recharge and evapotranspiration. This is three less than the number of parameters fitted in the CATPRO multi-response calibration.

The predicted time series of groundwater recharge and evapotranspiration, along with their 90% confidence limits, are shown in Figure 10.25 and Figure 10.26, respectively. The most obvious difference between the recharge and evapotranspiration predictions presented here and those made by CATPRO following multi-response calibration (Figure 9.24 and Figure 9.25) is the order of magnitude of the predicted recharge fluxes. CATPRO predicts monthly recharge fluxes ranging from approximately \(-80\) to \(160\) mm.m\(^{-1}\)y, whereas LASCAM predicts recharge fluxes in the range \(-6\) to \(4\) mm.m\(^{-1}\).y.

A fundamental difference in model structure accounts for the differences in predicted recharge. LASCAM includes a permanently unsaturated store that accepts infiltration from the surface soils and provides recharge to the deep, permanent aquifer where CATPRO does not. The LASCAM F store demonstrates soil water storage dynamics that are consistent with observations of deep unsaturated soil profiles elsewhere in the region (Johnston et al., 1983; Sharma, 1984; Ruprecht and Schofield, 1990a, 1990b, 1993) and its parameters are moderately well identified in the multi-response calibration (Table 10.2). The F store is an effective sink for water not contributing to the quick-flow responses of surface runoff and interflow and allows access to that water at times of high evaporative demand outside the winter streamflow periods. The parameters of the CATPRO B store, which may have served a similar function, were not identifiable and the store was removed from the model conceptualisation. Therefore CATPRO has only one store to accommodate water not contributing to quick-flow processes where LASCAM has two.

Figure 10.25 shows that the magnitude of the monthly recharge predictions made by LASCAM are consistent with estimates made by analysis of chloride storage profiles at Salmon Catchment performed by Johnston (1987a, 1987b). Monthly recharge
totals predicted by LASCAM were negative over the summer every year of the calibration period as was also the case for CATPRO.

![Figure 10.25](image)

**Figure 10.25:** Time series of unobserved groundwater recharge and 90% confidence limits predicted for Salmon Catchment following calibration of LASCAM to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25, \text{ and } 1.0$ respectively.

![Figure 10.26](image)

**Figure 10.26:** Time series of unobserved catchment averaged evapotranspiration and 90% confidence limits predicted for Salmon Catchment following calibration of LASCAM to streamflow, stream water salinity and changes in piezometric head, Box-Cox $\lambda = 0.25, 0.25, \text{ and } 1.0$ respectively.

As was the case for the CATPRO model, the confidence limits on the predicted recharge time series are widest over the summer, when streamflow is low or zero and the model behaviour is less constrained, less easily falsified, by the calibration data.
Figure 10.26 also shows a slight widening of the confidence limits on the predicted evapotranspiration time series over the summer months, which coincide with the times of maximum evaporative demand. The evapotranspiration predictions produced by LASCAM are compared with those produced by CATPRO in Figure 10.27, both systematic and random difference exist between the two time series. Given that LASCAM has the potential to hold more water in unsaturated storage than CATPRO, it may be reasonable to expect LASCAM to predict greater evapotranspiration rates than CATPRO and this was the case. The plot also shows that LASCAM predicts lower minimum monthly evapotranspiration rates during winter than CATPRO because water not contributing to streamflow can be absorbed by the F store and accessed during summer to partially satisfy the higher evaporative demand.

![Figure 10.27: Time series of unobserved catchment averaged evapotranspiration predicted by CATPRO and by LASCAM for Salmon Catchment following calibration to streamflow, stream water salinity and changes in piezometric head.](image)

### 10.3 Degree of Model Falsifiability

The critical application of established statistical techniques for examining the nature of the errors in model predictions, along with concurrent, multi-response calibration are proposed as two of the essential elements of a more rigorous methodology for hypothesis testing in conceptual catchment model development. The third element is
the principle of parsimony (Popper, 1959; Jakeman and Hornberger, 1993). Application of the first two elements of the methodology to the LASCAM model clearly indicated violation of the third. The LASCAM model structure is not parsimonious.

LASCAM has 35 model parameters, of which four could be fixed to values that either took advantage of prior information on catchment processes \((\alpha_g, \beta_g)\), or which reduced the degree of complexity of the model by eliminating auxiliary hypotheses \((\beta_c, \beta_{ss})\). This leaves 33 model parameters requiring calibration plus two addition parameters required to scale the potential evapotranspiration time series and to allow calibration to observed changes in catchment averaged piezometric head. During calibration the number of parameters that were statistically observable, given the multi-response time series data, was 15. Examination of the posterior distributions of the fitted parameters indicated that this should be reduced to ten. It is argued that, the 23 model parameters that were not statistically observable constitute auxiliary hypotheses about catchment water and chloride balance dynamics that contribute to model complexity and which act to defend it from falsification.

LASCAM produces predictions of some catchment responses that matched observations and the outcomes of process orientated research well. Calibration diagnostics revealed systematic error in LASCAM predictions of streamflow and stream water chloride of a lower order than those produced by CATPRO, however, LASCAM was less reliable than CATPRO in predicting changes in groundwater storage. LASCAM’s reduction in systematic error, over CATPRO, however was achieved by employing many more parameters, a majority of which where not statistically observable.

Furthermore, the calibration strategy suggested by Sivapalan et al. (1996b) in which the water balance and chloride balance parameters are fitted independently amounts to the two linked models becoming what may be called mutually auxiliary hypotheses. That is, one adds complexity to the other but does not increase its degree of falsifiability because interactions between the two sets of parameters are excluded by design.
Here, the proposed methodology to increase the rigor in conceptual catchment model development has been demonstrated on a complex, conceptual catchment model. Although critical examination of some model predictions indicate that LASCAM’s structure may be more representative of catchment streamflow generation processes and soil water dynamics, it exhibits a lower degree of testability than a simpler model (CATPRO). The proposed methodology has exposed the violation of the principle of parsimony as defined by Popper (1959).

In the remaining chapter, conclusions about what the application of the proposed methodology to hypothesis testing of the CATPRO and LASCAM models has demonstrated will be presented. Information presented here and in Chapter 9 on the degree of testability of the two models will be central to that discussion.
11 CONCLUSIONS

Presented here is a methodology to test the structure of conceptual catchment models in a more rigorous way than has been the norm. The development of this methodology is in response to a number of critical reviews of the way hydrology is practised as a science (Beven, 1987, 1989; Klemes, 1986a; Dooge, 1986). The methodology implements Popper’s (1959) theory of the nature of the scientific method. Popper argues that by their very nature scientific hypotheses cannot be proven to be true, they can only be falsified. Popper also stipulates that the falsification of hypotheses should be actively and aggressively pursued.

The proposed methodology has three main elements; 1) the principle of parsimony, the simplest model that fully describes the available data should be postulated; 2) the application of established statistical techniques to assess model predictions and errors; and, 3) the use of multi-response time series data for model calibration.

Popper (1959) equates parsimony, or a lack of complexity, to the concepts of auxiliary hypotheses and falsifiability. Auxiliary hypotheses add to the complexity of a theory if they reduce its degree of falsifiability. Alternatively, auxiliary hypotheses can strengthen a theory if after their addition, the theory rules out more. These points are directly applicable to conceptual catchment models; auxiliary hypotheses that add parameters requiring calibration but which do not improve the ability of the model to match the calibration data should be avoided. Furthermore, auxiliary hypotheses that introduce parameters that are not identifiable from the calibration data and which are set to arbitrary values reduce the degree of testability of the hypotheses that make up the catchment model.

The methodology has been demonstrated by way of a case study in which, CATPRO, a conceptual catchment model is proposed and its constituent hypotheses rigorously challenged with established statistical techniques and against field observations. Some model hypotheses were falsified and alternative hypotheses formulated. Finally, CATPRO was compared with LASCAM, another catchment model.
independently developed from the same conceptual basis but without reference to the hypotheses testing methodology proposed here.

In Chapter 8, the CATPRO model was carefully scrutinised following calibration to streamflow data alone. Post calibration diagnostics, which are a fundamental element of the proposed methodology, showed that CATPRO suffered from many of the problems commonly associated with conceptual catchment models. Perhaps more informative about the nature of conceptual catchment models was the examination of some assumptions commonly made in such models. The assumption that no groundwater leaves the catchment other than via baseflow was examined and found, in this case, to be of limited consequence when compared to the assumptions made about rainfall interception. If interception data are not available, the inclusion of an interception function in a conceptual catchment model leads to extreme uncertainty in model predictions. A lack of interception data was likened to not knowing the rainfall but having only an estimation of its upper limit.

The full power of the proposed methodology was demonstrated in Chapters 9 and 10. In Chapter 9 the CATPRO model was calibrated to streamflow, stream chloride export and piezometric head time series data. Post calibration diagnostics demonstrated that although the model could not match the streamflow and chloride export data particularly well the compromise forced by concurrent calibration resulted in a much higher degree of parameter identifiability than calibration to streamflow data alone. However, the same diagnostics also demonstrated that systematic error still existed in model predictions of streamflow, chloride export and piezometric heads. Examination of the predicted chloride concentrations of the different contributions to streamflow proved inconsistent with field observations. Plots of the chloride storage of the conceptual stores revealed an unacceptable trend in the behaviour of one of the model stores. This information was analysed along with the post calibration diagnostics to identify that a revised hypothesis of the relationship between groundwater storage and baseflow and the accompanying chloride load was required for the model to reproduce the observed catchment responses.
The observed reduction in parameter uncertainty brought about with multi-response calibration resulted in the narrowing of confidence limits on model predictions of the unobserved catchment responses of recharge and evapotranspiration over the streamflow-only calibration. Although systematic error cannot be quantified without knowledge of the true value of the predicted catchment response, it is argued that systematic error is reduced following multi-response calibration because of the more demanding nature of the test on model structure.

The LASCAM model was introduced and compared to CATPRO in Chapter 10. LASCAM produced visually better matches to the streamflow and chloride export time series than CATPRO but parameter identifiability was low. This is because, in Popper's (1959) language, LASCAM is burdened with auxiliary hypotheses which do little to improve predictability but which make the model less falsifiable. The application of statistical diagnostics revealed high levels of parameter correlation and systematic error in the predictions of streamflow and chloride export that were a consequence of this.

Ultimately, both models were falsified. Both produced predictions of catchment responses that included clearly identifiable systematic error: clearly identifiable though, only with the exhaustive application of the post calibration diagnostics which form a crucial element of the hypothesis testing methodology proposed in Chapter 4. LASCAM was shown to be overparameterised, the level of complexity in the model is not capable of being supported even with three catchment response time series for calibration. Parts of the model structure, however, were demonstrated to be superior to those included in CATPRO. The inclusion of a permanently unsaturated store was responsible for recharge predictions more consistent with other data than those made by CATPRO. The cost was a much higher degree of uncertainty in model predictions because of high parameter uncertainty.

Finally, the groundwater discharge conceptualisation used in LASCASM appears to match the data better than that used in CATPRO. The results of the post calibration diagnostics performed on the CATPRO model following multi-response calibration and the relative success of LASCAM in matching the chloride export time series
support this. The conceptualisation, used in LASCAM, whereby groundwater discharge contributes water to the saturated zone around the stream channel (A store) rather than directly to baseflow may be a more appropriate conceptualisation to that used in CATPRO; a hypothesis that withstands more rigorous testing.

A methodology for the rigorous testing of conceptual catchment models has been proposed and demonstrated. A case study has been presented in which a parsimonious catchment model was proposed and revised using the results of post calibration diagnostics. These post calibration diagnostics are based on established statistical techniques and multi-response calibration. The model was compared to a second which suffered from overparameterisation even though elements of its structure appears to be a better reflection of the processes governing streamflow generation in the study catchment.

The proposed methodology may be strengthened by one further addition. The addition of a test for a catchment undergoing significant landuse or climate change as proposed by Klemes’ (1986b). Subsequent to this work, the methodology has been extended to catchments undergoing extensive landuse change (Mroczkowski et al., 1997). Furthermore, the hypothesis that groundwater contributes to the variable source area adjacent to the stream channel rather than directly to baseflow was tested.
12 REFERENCES


189


