RECURSIVE PARAMETER ESTIMATION
OF CONCEPTUAL
WATERSHED RESPONSE MODELS

by
Harihar Rajaram
and
Konstantine P. Georgakakos

IIHR Report No. 319
Iowa Institute of Hydraulic Research
The University of Iowa
Iowa City, Iowa 52242
October 1987
RECURSIVE PARAMETER ESTIMATION
OF CONCEPTUAL WATERSHED RESPONSE MODELS

by

Harihar Rajaram
and
Konstantine P. Georgakakos

Sponsored by

United States Environmental Protection Agency
(Contract No. 812329-01-0)
and
Hydrologic Research Laboratory, National Weather Service, NOAA
(Cooperative Agreement No. NA86AA-H-HY126)

IIHR Report No. 319

Iowa Institute of Hydraulic Research
The University of Iowa
Iowa City, Iowa 52242

October 1987
EXECUTIVE SUMMARY

A nonlinear filtering approach to recursive parameter estimation of conceptual watershed response models has been developed. It allows for the incorporation of time-varying uncertainties in the parameter estimation framework in a manner that is in strong consistence with the dynamics of the watershed response. This is achieved by means of explicit parameterizations of input and parameter uncertainties in the noise covariance matrices. An algorithm has been developed for the adaptive estimation of model states and parameters, and of the noise statistical parameters. It uses a Gaussian minimum variance unbiased estimator for the former and a stochastic approximation estimator based on enforcement of compatibility between one-step predicted and observed residual-error covariances for the latter.

The above approach was applied to the parameter estimation of the Hydrology submodel of the Enhanced Trickle Down (ETD) model for the Woods Lake watershed. The results indicated considerable success achieved by the algorithm during periods without snowcover, and poor performance of the algorithm during periods with snowcover, owing to significant structural errors in the representation of snow related dynamics in the model. Post-calibration analyses indicated a high degree of multiparameter interaction and insensitivity of poorly inferred parameters.

It is concluded that the new approach offers great promise in the context of recursive state and parameter estimation of conceptual watershed response models, and provides a convenient framework for the analysis of the influence of input uncertainties on the state/parameter estimation. In terms of the application to the ETD model the following are significant findings:

1) The representation of snowmelt dynamics in the ETD model is inadequate for the reproduction of the actual character of the snowmelt process and needs to be improved upon.

2) The representation of unsaturated flow based on quasi-Darcian hypothesis does not appear to be satisfactory in all situations. A more physical approach to the formulation of unsaturated flow is likely to improve the predictive ability of the ETD model significantly.
ACKNOWLEDGEMENTS

This report essentially constitutes the Master of Science degree thesis of Harihar Rajaram. The research work was sponsored by the United States Environmental Protection Agency, Corvallis, Oregon, under Contract No. 812329-01-0 and by the Hydrologic Research Laboratory, National Weather Service, NOAA, under the cooperative agreement No. NA86AA-H-HY126. The research described in the report has not been subjected to EPA's required Peer and Policy Review. Therefore, it does not necessarily reflect the view of that Agency, and no official endorsement should be inferred.

The authors are grateful to Professors J.L. Schnoor and S. Dasgupta for their suggestions and comments. Discussions with Dr. N.P. Nikolaidis and Dr. S. Lee have been useful.

The excellent work of Twila Meder and Karen Nall in the typing of this report is also gratefully acknowledged.
ADDENDUM TO IIHR REPORT NO. 319

ERRATA

1. Page viii, Figure Caption No. 21:
   Substitute "Evaporation..." with "Temperature (°C)...".

2. Page 8, third line from the top:
   Substitute "...algorithm to parameter..." with "...algorithm for parameter...".

3. Page 10, second line, second paragraph:
   Substitute "...to a parameter..." with "...and a parameter...".

4. Page 59, eleventh line from the top:
   Substitute "...r-matrix..." with "...Q-matrix...".

5. Page 58, fourth line from the top:
   Substitute "...P uu and due..." with "...P uu due..."

6. Page 59, one to last line:
   Substitute "...mean discrete-time white noise,..." with "...mean and heteroscedastic,..."

7. Page 74, tenth item on list, fourth column:
   Substitute "X" with "Yes".

8. Page 89, second line, second paragraph:
   Substitute "...mean white noise sequence..." with "...mean sequence...".

9. Page 103, Figure 9b, Caption:
   Substitute "b. KPERC4" with "b. FRAX".

10. Page 103, Figure 9c, Caption:
    Substitute "c. FRAX" with "c. KPERC4".

11. Page 120, Figure 21, Caption:
    Substitute "Evaporation..." with "Temperature (°C)...".

12. Page 137, fifth reference item:
    Substitute "...1980" with "...1986".

13. Page 141, second paragraph, 3rd line:
    Substitute "...white noise..." with "...zero mean...".
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXECUTIVE SUMMARY</td>
<td>i</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF SIGNIFICANT SYMBOLS</td>
<td>ix</td>
</tr>
<tr>
<td>I.  INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>A.  Conceptual Models of Watershed Response</td>
<td>1</td>
</tr>
<tr>
<td>B.  Review of Literature in Parameter Estimation for Conceptual Models of Watershed Response</td>
<td>3</td>
</tr>
<tr>
<td>1.  Statement of the Parameter Estimation Problem</td>
<td>3</td>
</tr>
<tr>
<td>2.  Applications of Optimization Algorithms</td>
<td>4</td>
</tr>
<tr>
<td>to the Parameter Estimation Problem</td>
<td>4</td>
</tr>
<tr>
<td>3.  Applications of Recursive Parameter Estimation Algorithms to Simple Conceptual Rainfall-Runoff Models</td>
<td>6</td>
</tr>
<tr>
<td>C.  Motivations for and Objectives of Present Study</td>
<td>8</td>
</tr>
<tr>
<td>II.  RECURSIVE IDENTIFICATION: AN OVERVIEW</td>
<td>10</td>
</tr>
<tr>
<td>A.  Introduction</td>
<td>10</td>
</tr>
<tr>
<td>B.  Conceptual Bases for System Identification</td>
<td>11</td>
</tr>
<tr>
<td>1.  Error Representations</td>
<td>11</td>
</tr>
<tr>
<td>2.  Estimation Criteria and Cost Functions</td>
<td>13</td>
</tr>
<tr>
<td>3.  Convergence and Asymptotic Properties</td>
<td>15</td>
</tr>
<tr>
<td>4.  Data Requirements for System Identification</td>
<td>17</td>
</tr>
<tr>
<td>C.  Approaches to Recursive Identification (RI)</td>
<td>17</td>
</tr>
<tr>
<td>1.  Modifications of Offline Methods</td>
<td>17</td>
</tr>
<tr>
<td>2.  Nonlinear Filtering Methods</td>
<td>23</td>
</tr>
<tr>
<td>3.  Stochastic Approximations Methods</td>
<td>31</td>
</tr>
<tr>
<td>4.  Model Reference Approaches and Pseudolinear Regressions</td>
<td>35</td>
</tr>
<tr>
<td>D.  Unified Approach to Recursive Identification</td>
<td>36</td>
</tr>
<tr>
<td>E.  Convergence and Asymptotic Properties of Recursive Identification Algorithms</td>
<td>48</td>
</tr>
<tr>
<td>F.  Choice of a Recursive Identification Algorithm for Application to Conceptual Models of Watershed Response</td>
<td>49</td>
</tr>
<tr>
<td>III.  AN APPROACH TO RECURSIVE STATE AND PARAMETER ESTIMATION</td>
<td>51</td>
</tr>
<tr>
<td>OF CONCEPTUAL WATERSHED RESPONSE MODELS</td>
<td>51</td>
</tr>
<tr>
<td>A.  Introduction</td>
<td>51</td>
</tr>
<tr>
<td>B.  Adaptive Filtering Based on Explicit Parameterizations of Input and Parameter Uncertainty</td>
<td>55</td>
</tr>
<tr>
<td>1.  Formulation</td>
<td>55</td>
</tr>
</tbody>
</table>
2. Adaptive Estimation of State Estimator Parameters..................59
3. Summary..............................................................................66

IV. A CASE STUDY: ESTIMATION OF PARAMETERS OF THE HYDROLOGY
SUBMODEL OF THE ENHANCED TRICKLE DOWN (ETD) MODEL..............69

A. Introduction: The Enhanced Trickle-Down (ETD) Model.............69
B. Description of the Hydrology Submodel.................................70
   1. Snow Compartment......................................................70
   2. Soil Compartment......................................................77
   3. Unsaturated Zone Compartment.....................................77
   4. Lake Compartment.....................................................78
   5. Groundwater Compartment..........................................78
C. Data Requirements for the ETD Model....................................78
D. Modifications of the Form of the Equations of the ETD Model
   to Facilitate Applications of Nonlinear Filtering Methods.........79
E. Sources of Uncertainty in the Data and Their Incorporation
   in the Parameter Estimation Framework................................80
F. Description of the Study Area: Woods Lake
   Watershed in Adirondack Park, New York.............................85
G. A Stepwise Procedure for the Estimation of
   Parameters of the Hydrology Submodel...............................86
H. Description of the Computer Program
   Used in this Study.........................................................90

V. RESULTS AND DISCUSSION...................................................91

A. Estimation of Parameters for Periods without
   Snowcover..................................................................91
B. Evaluation of the Parameter Estimation for
   Periods with no Snowcover.............................................101
C. Estimation of Parameters of the Snow Model........................115
D. Evaluation of the Parameter Estimation of
   the Snow Model.........................................................121
E. Verification of the Model................................................128

VI. CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH......130

REFERENCES.................................................................134

APPENDIX A. DERIVATION OF THE STATE ERROR COVARIANCE
DIFFERENTIAL EQUATION..................................................141

APPENDIX B. DERIVATION OF THE UPDATE EQUATION FOR THE
DERIVATIVE MATRIX $\frac{\partial P}{\partial \alpha}$ ...............................................145

APPENDIX C. SUMMARY OF THE EQUATIONS INVOLVED IN THE RECURSIVE
ALGORITHM FOR STATE ESTIMATION AND ESTIMATION OF
PARAMETERS OF THE STATE ESTIMATOR.................................149
APPENDIX D. SMOOTHING FUNCTIONS USED TO REPLACE THE MODAL
CHARACTER OF THE EQUATIONS OF THE ETD MODEL.................152
A. A smoothing Function to Replace Threshold-Type
Behavior at Saturation..............................................152
B. A Smoothing Function for the Overland Flow
Representation.........................................................152
C. A Smoothing Function to Retard Percolation
when Lower Compartments Approach Saturation.............154
D. A Smoothing Function to Eliminate Evaporation
from Soil and Lake Compartments in Presence of
Snowcover...............................................................154

APPENDIX E. DERIVATIVES OF THE ETD MODEL EQUATIONS WITH RESPECT
TO STATE VARIABLES, FREE PARAMETERS, FIXED
PARAMETERS AND INPUT VARIABLES..............................159

APPENDIX F. BRIEF DESCRIPTION OF THE SUBROUTINES INVOLVED IN
THE ADAPTIVE ESTIMATION ALGORITHM IN APPENDIX C.............167
LIST OF TABLES

Table                                                                 | Page     
---                                                                 | ---       
I. Hydrologic Model Equations ..............................................72  
II. Listing of Input, Output Variables and Parameters of the Hydrologic Model ..............................................73  
III. Modified Form of the Equations for the Flows in the Hydrology Submodel of the ETD Model .................................81  
IV. Initial and Final Values of Model Parameters and Standard Deviations .........................................................94  
V. Results of the Estimation of $\alpha_u$ and $\alpha_p$ .......................................................................................95  
VI. Parameter Correlation Matrix at end of Parameter Estimation Run .................................................................106  
VII. Comparison of Least Squares Performance Indices Before and After Parameter Estimation - Periods Without Snowcover 111  
VIII. Least Squares Performance Indices for Simulation with Initial and Final Estimates of Snow Model Parameters ........126
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Schematic representation of the Hydrology submodel of the ETD model</td>
<td>71</td>
</tr>
<tr>
<td>2.</td>
<td>Precipitation data for the calibration period during periods without snowcover</td>
<td>92</td>
</tr>
<tr>
<td>3.</td>
<td>Evaporation data for the calibration period during periods without snowcover</td>
<td>93</td>
</tr>
<tr>
<td>4.</td>
<td>Traces of the parameters $\alpha_p$ and $\alpha_u$ during Run 1</td>
<td>96</td>
</tr>
<tr>
<td>5.</td>
<td>Traces of the parameters $\alpha_p$ and $\alpha_u$ during Run 2</td>
<td>97</td>
</tr>
<tr>
<td>6.</td>
<td>Traces of the parameters $\alpha_p$ and $\alpha_u$ during Run 3</td>
<td>98</td>
</tr>
<tr>
<td>7.</td>
<td>Normalized residuals during the final pass over the data in Run 2</td>
<td>100</td>
</tr>
<tr>
<td>8.</td>
<td>Evolution of the parameters of the soil compartment during the parameter estimation run</td>
<td>102</td>
</tr>
<tr>
<td>9.</td>
<td>Evolution of the parameters of the unsaturated zone compartment during the parameter estimation run</td>
<td>103</td>
</tr>
<tr>
<td>10.</td>
<td>Evolution of the parameters of the lake and groundwater compartments during the parameter estimation run</td>
<td>104</td>
</tr>
<tr>
<td>11.</td>
<td>Evolution of the parameters KLAT3 and KLAT4 during the experimental run</td>
<td>105</td>
</tr>
<tr>
<td>12.</td>
<td>Deterministic simulations during periods without snowcover, with initial values of parameters. Observations are signified by the thick solid line.</td>
<td>109</td>
</tr>
<tr>
<td>13.</td>
<td>Deterministic simulations during periods without snowcover, with estimated values of parameters. Observations are signified by the thick solid line.</td>
<td>110</td>
</tr>
<tr>
<td>14.</td>
<td>Sensitivity analysis of the parameters of the soil compartment</td>
<td>112</td>
</tr>
<tr>
<td>15.</td>
<td>Sensitivity analysis of the parameters of the unsaturated zone compartment</td>
<td>113</td>
</tr>
</tbody>
</table>
16. Sensitivity analysis of the parameters of the lake and groundwater compartments.................................................114
17. Interaction between the parameters KLAT3 and UZSSAT.........................116
18. Interaction between the parameters KLAT3 and KLAT4...............................117
19. Interaction between the parameters D1 and ALF2.....................................118
20. Precipitation data for the calibration period during periods with snowcover.................................................119
21. Evaporation data for the calibration period during periods with snowcover..................................................120
22. Traces of the parameters $\alpha$ and $\alpha$ during the simultaneous estimation run for periods with snowcover........122
23. Traces of the snow model parameters during the simultaneous estimation run..................................................123
24. Normalized residuals during the final pass over the data in the simultaneous estimation run..................................124
25. Deterministic simulations during periods with snowcover with initial values of parameters. Observations are signified by the thick solid line..................................................125
26. Deterministic simulations during periods with snowcover with final values of parameters. Observations are signified by the thick solid line..................................................125
27. Sensitivity analysis of the snow model parameters..................................127
28. Deterministic simulations during the verification period with estimated parameter values........................................129
29. Smoothing function for threshold type behaviour at saturation.................................................................153
30. Smoothing function for overflow representation..................................................155
31. Smoothing function for the retardation of percolation........................................156
32. Smoothing function for the elimination of evaporation from lake and soil in presence of snowcover..........................158
### LIST OF SIGNIFICANT SYMBOLS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_u$</td>
<td>input-uncertainty coefficient</td>
</tr>
<tr>
<td>$\alpha_p$</td>
<td>parameter-uncertainty coefficient</td>
</tr>
<tr>
<td>$\alpha_t$</td>
<td>time-varying weights</td>
</tr>
<tr>
<td>$\beta(t)$</td>
<td>Wiener process at time $t$</td>
</tr>
<tr>
<td>$E{\cdot}$</td>
<td>expectation operator</td>
</tr>
<tr>
<td>$\varepsilon(t, \theta)$</td>
<td>prediction error as a function of time $t$ and parameters $\theta$</td>
</tr>
<tr>
<td>$\varepsilon(t)$</td>
<td>error vector at time $t$</td>
</tr>
<tr>
<td>$\xi(t)$</td>
<td>sequence of instrumental variables</td>
</tr>
<tr>
<td>$g_M(\cdot)$</td>
<td>vector function of system response</td>
</tr>
<tr>
<td>$\xi(\theta)$</td>
<td>prior probability distribution of the parameter vector $\theta$</td>
</tr>
<tr>
<td>$\xi(\theta</td>
<td>z_t)$</td>
</tr>
<tr>
<td>$H(k)$</td>
<td>measurement equation coefficient matrix at time $t_k$</td>
</tr>
<tr>
<td>$J$</td>
<td>error norm</td>
</tr>
<tr>
<td>$L(\theta, \hat{\theta})$</td>
<td>loss criterion as a function of true $\theta$ and estimated $\hat{\theta}$</td>
</tr>
<tr>
<td>$L$</td>
<td>loss function or criterion</td>
</tr>
<tr>
<td>$\hat{\theta}(\cdot)$</td>
<td>parameter vector estimate</td>
</tr>
<tr>
<td>$\theta$</td>
<td>parameter vector</td>
</tr>
<tr>
<td>$\theta_o$</td>
<td>true value of parameter vector</td>
</tr>
<tr>
<td>$v(t)$</td>
<td>scalar equation error</td>
</tr>
<tr>
<td>$\tilde{v}(\cdot)$</td>
<td>normalized residual sequence</td>
</tr>
<tr>
<td>$\phi(t)$</td>
<td>memory vector of sufficient parameter information in $z_t$</td>
</tr>
<tr>
<td>$\phi(k+1,k)$</td>
<td>transition matrix for times $t_k$, $t_{k+1}$</td>
</tr>
<tr>
<td>$F(t)$</td>
<td>coefficient matrix of a continuous system at time $t$</td>
</tr>
<tr>
<td>$\Omega_\theta$</td>
<td>parameter space of physically realistic parameter-value ranges</td>
</tr>
<tr>
<td>$P(\cdot)$</td>
<td>state-estimate error covariance matrix</td>
</tr>
</tbody>
</table>
Q(t) covariance parameter matrix of model error at time t
R(t_k) covariance matrix of observation error at time t_k
t time variable
u(t) input vector at time t
v(t) scalar observation error process
V_N(θ) sum of squared errors criterion function for N observations
W(t) positive definite weighting matrix at time t
x(t) state vector at time t
\hat{x}() estimate of system state x()
y(t) scalar observation at time t
z(k) vector of measurements at time t_k
z^T vector of available data up to time t
z_T vector of input-output data over a period T

NOTE: Description of additional terms is provided in Table II.
I. INTRODUCTION

A. Conceptual Models of Watershed Response

A watershed is the area tributary to a given point on a surface water body such as a stream or lake and is separated from adjacent watersheds by a divide or ridge. All the surface water originating in the watershed is discharged either to the lowest point in the divide through which the main stream of the watershed passes or as groundwater export to adjacent watersheds.

Watershed systems are subjected to various input loadings such as precipitation, chemical constituents dissolved in precipitation and dry deposition of chemical substances on the watershed area, water transfer demands such as evapotranspiration and water demands from the animal life it sustains. In response, some of the watershed characteristics such as moisture contents in various soil horizons, elevation of the groundwater table and water levels in the surface water bodies and chemistry of the soil, porewater, groundwater and surface waters exhibit changes and the watershed generates sequences of outputs such as outflows in the main stream and concentrations of chemical constituents in the streamflow.

The study of watershed responses to various types of input loadings is crucial to various problems in water resources engineering and allied disciplines. Some of these problems are flood forecasting, generation of future sequences of streamflows for long term management of water resources systems, prediction of impact of acid deposition and precipitation on watershed chemistry, etc. Owing to the large spatial extents of typical watersheds, spatial heterogeneity in the watershed characteristics, spatial and temporal variability in input loadings and insufficient knowledge of some of the physical and chemical processes in the watershed, it is practically impossible to perform an exact analysis of watershed response. An expediency that has commonly been employed is to construct models of watershed response based on idealistic assumptions like the use of lumped representations of the spatially distributed processes in the watershed and mathematical equations that are thought to approximate the actual mathematical character of the dynamics of the watershed processes.
The steps involved in the development and use of such models can be summarized as follows:

1) Identification of the dominant processes involved in the watershed response of interest.

2) Development of mathematical equations describing each of the processes involved, based on physically realistic hypotheses.

3) Characterization of the parameters involved in these mathematical equations as being inferrable from readily available data such as topographic maps or as requiring estimation from sequences of input-response (output) data.

4) Estimation of parameters requiring estimation, from past records of input-response data.

Until the 1960's most of the conceptual watershed models used were black-box in character and had no obvious correspondence to the physical system. The 1960's however, saw the emergence of models of the physical-conceptual type, where the components of the models corresponded to units of the physical system. With the advent of powerful digital computers, there was a proliferation in the development and use of these models to study the hydrologic response of watersheds. Some of these models are the Stanford Watershed Model, SWM (Linsley et al., 1970), the soil-moisture accounting scheme of the National Weather Service River Forecasting System, NWSRFS (Brazil and Hundlow, 1981) and the Precipitation-Runoff modelling system PRMS (Leavesley et al., 1983) of the United States Geological survey. More recently, models of this type have been developed and used to predict the hydro-chemical response of watershed systems to acidic precipitation and dry deposition of acidic matter on watersheds. Some of these are the Integrated Lake-Watershed acidification study model (ILWAS, Chen, et al., 1980), the Trickle-Down model (Lin and Schnoor, 1985), the MAGIC model (Cosby, et. al., 1985) and the Enhanced Trickle Down model (Nikolaidis, et al., 1986). The problem of estimation of parameters for these conceptual models involves the use of input-response data in an attempt to arrive at parameter values such that the modelled watershed response closely approximates measured responses for a variety of conditions. In the next section the parameter estimation problem is formalized in mathematical terms and a brief review of literature addressing the problem of parameter estimation for watershed response models is presented.

1. Statement of the Parameter Estimation Problem  Let \( u(t) \) denote the sequence of the vector of inputs to the watershed system, \( x(t) \) that of the state variables describing the watershed characteristics of interest, \( q(t) \) that of the vector of measured output variables of interest and \( \theta \) the vector of parameters in the model equations requiring estimation. Let \( g_M(\cdot) \) denote the vector valued function that relates the model predictions of the output variables to the state variables, inputs and parameters. Then

\[
q(t) = g_M(x(t), u(t), \theta) + \varepsilon(t) \tag{I.1}
\]

where \( \varepsilon(t) \) is a time sequence of error or residual vectors that in general represents the errors involved in the measurement of inputs, the model structure (i.e., the functional form of the predictor \( g_M \)), the lumping of spatially variable processes and the values of the parameters in \( \theta \).

The parameter estimation problem is solved by finding a parameter vector \( \theta^* \) such that

\[
\theta^* = \text{Min}_{\theta \in \Omega_{\theta}} \{ [\varepsilon(t)/x(0), q(t), u(t), g_M(\cdot)], t \in T_c \} \tag{I.2}
\]

i.e. some loss-criterion function \( \ell(\cdot) \) of \( \theta \) in minimized in some parameter space \( \Omega_{\theta} \) (specified by knowledge of physically realistic values or prior knowledge) over some time period \( T_c \) (usually referred to as the calibration period), conditional on the form of the predictor \( g_M \), estimates of initial values of state variables, \( x(0) \), and the sequences of estimates of inputs and measurements of inputs and outputs.

There are many possible approaches to the parameter estimation problem, based on the choices of the loss-criterion \( \ell(\cdot) \) and the scheme for its minimization. Some of the choices for \( \ell(\cdot) \) reported in literature include a squared sum of the residual vectors, a squared sum of weighted residual vectors and likelihood functions, leading respectively to least squares (LS), weighted least squares (WLS) and maximum likelihood (ML) estimates. Approaches to the
minimization problem are based on commonly used approaches in optimization literature, and include trial and error (simulation type) approaches and use of optimization algorithms such as the Simplex algorithm, the Davidon-Fletcher-Powell (DFP) algorithm, etc. Both the above approaches have been used in an offline or batch processing mode. In addition, there have been instances of the use of recursive or on-line estimation algorithms to parameter estimation of simpler conceptual rainfall-runoff models. Some of these will be reviewed in the following.

2. Applications of Optimization Algorithms to the Parameter Estimation Problem

The earliest application of an optimization algorithm to the parameter estimation problem, reported in literature, is that of Dawdy and O'Donnell (1965), who used the simplex algorithm for the estimation of parameters of a model developed by O'Donnell. Since then, a variety of optimization algorithms have been applied to the problem.

Chapman (1970) compared three nonlinear optimization algorithms and concluded that a Simplex-type approach performed better than a steepest descent algorithm. Ibbitt and O'Donnell (1971) in a comprehensive analysis of the problems associated with the use of optimization algorithms traced the difficulties involved to the highly irregular response surfaces generated. Dawdy and Litchy (1968), Chapman (1970) and Clarke (1973) pointed out the inappropriateness of the use of a simple least squares (SLS) criterion function (applicable only if the error sequence $\varepsilon(t)$ is uncorrelated and homoscedastic) in the context of Watershed response models and suggested the use of transformations of $\varepsilon(t)$ to ensure homoscedasticity of the transformed sequence before application of the SLS criterion function. The U.S. Army Corps of Engineers flood hydrograph package HEC-1 (1973) used a weighted least squares (WLS) criterion function.

Ibbitt (1970), and Johnston and Pilgrim (1976) reported the failure of optimization algorithms to pick up true values of parameters even in simulation studies (where $q(t)$ is generated by the model $g_X(\cdot)$ using $\hat{\theta}_0$, a particular value of the parameter vector, and a parameter estimation exercise is performed in some parameter space $\Omega_{\theta}$, containing $\hat{\theta}_0$, starting with an arbitrary initial condition for $\theta$), working respectively with the Stanford Watershed model and the Boughton model. They traced the difficulties involved to the structural identifiability of the models.
Aitken (1973) analyzed the character of systematic errors in rainfall runoff (RROF) models and examined the causes for correlations in the error sequence. Ibbitt and O'Donnell (1974) suggested guidelines for the design of conceptual RROF model structures that permitted efficient applications of optimization algorithms to parameter estimation problems. Johnston and Pilgrim (1976) using the DFP algorithm to estimate parameters of the Boughton model, reported discontinuities in the response surface and its derivatives on account of threshold-type elements in the model structure. They suggested that this was the limiting factor in exploiting the power of gradient-based algorithms.

Sorooshian and Dracup (1980) used a maximum likelihood approach to derive a general criterion function for autocorrelated and heteroscedastic error cases. They reported that the response surfaces generated were far superior to those generated using an SLS criterion function and improved the performance of optimization algorithms. Troutman (1982) and Kuczera (1982) in independent analyses of the influence of input data errors on the parameter estimation problems, suggested that it was a source of bias in the parameter estimates. Kuczera (1982) also noted that the bias was likely to become more severe as systematic errors increased. Restrepo-Posada and Bras (1982) developed a maximum-likelihood estimation procedure, allowing for the incorporation of prior information about parameters. They used a nonlinear constrained optimization algorithm based on the DFP algorithm and Zoutendijk's projection algorithm with selective capabilities that made it possible to exclude certain parameters at certain times, to improve convergence. They concluded that high degrees of multiparameter interaction among the parameters of the upper zone of the NWSRFS model marred the performance of the algorithm.

Kuczera (1982, 1983a,b) examined the parameter estimation problem from a Bayesian viewpoint and addressed an approach directed towards combining different kinds of hydrologic data to improve parameter inference. He used a nonlinear regression approach with transformed errors. He examined the parameter uncertainties based on the posterior probability density function of the parameter estimates.

Sorooshian and Arfl (1982) and Sorooshian et al. (1983a,b) examined the issue of parameter sensitivity and observability and its influence on the
response surface. They suggested the use of the Heteroscedastic errors Maximum Likelihood Estimator (HMLE) criterion function (Sorooshian & Dracup, 1980) to expedite such analysis, since it led to better response surfaces. The latter work also examined the issue of parameter observability and uniqueness in relation to model structures. Gupta and Sorooshian (1983) examined the possibilities of reparameterization of models to improve identifiability.

Troutman (1985 a,b) reported improvement of parameter inference through use of a maximum likelihood criterion function. He examined the use of criterion functions based on storm peak magnitudes and total storm runoff values and reported that some parameters of the Precipitation-Runoff Modeling System model (developed by the U.S. Geological Survey) were better inferred based on storm peaks.

Sorooshian and Gupta (1985 a,b) examined the application of derivative-based algorithms to the parameter estimation problem exploiting the state-space structure of RROF models to obtain derivatives of model equations, in spite of threshold structures. They reported that while derivative algorithms were more efficient in that improved convergence rates resulted, they did not overcome the problems that were inherently due to poor structural identifiability. They suggested that the Hessian information available through use of derivative-based algorithms can be used as a tool for post-calibration analyses. They also developed a systematic procedure for the study of local structural identifiability of RROF models, based on a perturbation approach. They demonstrated the value of such analyses in achieving meaningful reparameterizations of equations to improve structural identifiability.

3. Applications of Recursive Parameter Estimation Algorithms to Simple Conceptual Rainfall-Runoff Models Recursive or online algorithms for parameter estimation, in contrast to offline or batch algorithms, involve updating of the parameter estimate vector while serially working through the data. Some of the advantages that recursive algorithms offer over offline algorithms are that only a small portion of data needs to be stored and processed at any time, reducing storage costs, and unlike offline algorithms they (because of their essentially iterative nature for nonlinear problems) do not require too many passes over a record of data. In addition, they provide a very strong conceptual framework for the analysis of the adequacy of model structure.
Sharp variations in parameters assumed time-invariant during the later stages of the estimation procedure (when the parameters have more or less converged) would be expected in the vicinity of periods where the model structure may be inadequate. This is of particular significance in the case of highly nonstationary systems such as hydrologic systems.

The literature is devoid of real-world applications of recursive parameter estimation algorithms to conceptual hydrologic models. A very recent real-world application of recursive identification algorithms to parameter estimation of the National Weather Service River Forecast System soil-moisture accounting model, has been reported by Georgakakos and Brazil (1987). It utilized the discrete form of the modified Sacramento soil-moisture accounting scheme and the Extended Kalman Filter. The filter parameters were obtained based on trial and error techniques and no physical interpretation of the model error covariance in terms of parameter errors was attempted.

In contrast to the existing state of affairs for conceptual hydrologic models, several applications of recursive identification algorithms to simpler "black-box" models have been reported in the literature. Some of these are reviewed in the following.

Hino (1970, 1973), in what were perhaps the earliest applications of Kalman filtering in the hydrosiences, used it to estimate unit hydrograph ordinates. The unit hydrograph theory is essentially a linear theory and hence a linear estimator was developed, requiring no approximations. He reported sensitivity of estimates to choice of noise covariance matrices and emphasized the subjectivity involved in the estimation procedure.

Katayama (1976) and Wood & Cooper (1983) used recursive estimation algorithms for ARMA and ARMAX models respectively using a maximum likelihood approach. They used concepts from Akaike Information theory for model order estimation. Wood and Szollosy-Nagy (1978) developed an adaptive estimation algorithm to assess short term structural and parameter changes in hydrologic prediction models and applied it to an ARMA model. Their algorithm included adaptive estimation of noise covariance matrices.

Duong et al. (1975) illustrated the application of modern control concepts to hydrologic systems. They performed simulation studies to investigate the effect of input uncertainties. They used a Gauss-Markov model of param-
eter evolution. They reported filter divergence in absence of adaptive filtering to estimate noise covariance matrices.

Moore and Weiss (1976) used the Extended Kalman filter (EKF) algorithm to parameter estimation of a simple nonlinear reservoir model incorporating an independently estimated channel delay element and an empirical relationship between measured and effective rainfall. They used an exponential age-weighting scheme to improve adaptive performance of the filter. Whitehead (1979) used an instrumental-variable-approximate maximum likelihood (Young and Jakesman, 1979) algorithm for recursive estimation of a time-varying ARMA hydrologic model.

Cooper (1982) presented a simple algorithm for the estimation of parameters of nonlinear hydrologic models based on a recursive gradient-based updating scheme. He examined the convergence properties of the estimates generated as related to the asymptotic properties of the derivatives of the loss criterion used with respect to the parameters.

Bras and Restrepo-Posada (1980) discussed the applications of online parameter estimation schemes to conceptual hydrologic models. They investigated the influence of model-structural errors on the convergence of parameter estimates and suggested the use of heuristic modifications to mathematically based techniques to overcome the problems arising on account of structural errors.

C. Motivations for and Objectives of Present Study.

It is clear from the review of the literature that there is a need for the development of physically based, statistically consistent, recursive state and parameter estimation algorithms suitable for use with large scale conceptual hydrologic models. The new algorithms should remove the need for extensive trial and error procedures for the identification of the estimator parameters while at the same time account for the highly nonstationary nature of the rainfall-runoff process. Recursive or adaptive algorithms provide an elegant framework for the analysis of nonstationarities in the system dynamics and in the uncertainties associated with the input variables, and the incorporation of these in the parameter estimation framework.
The main objective of the present study is to develop a comprehensive and systematic approach to the applications of recursive parameter estimation algorithms to large scale conceptual models of watershed response, allowing for physically consistent time varying representations of uncertainties in the model dynamics and their incorporation in the parameter estimation framework.

In Chapter II, an overview of recursive identification theory is presented and the various approaches to recursive identification are discussed. The choice of a nonlinear filtering-type algorithm as a candidate for application to conceptual watershed models is justified.

In Chapter III, the Extended Kalman Filter algorithm is briefly reviewed and a physically-based parametric approach to adaptive estimation of noise covariance matrices is presented. A technical approach to the step-wise procedure for the applications of these concepts to real-time watershed systems is discussed.

In Chapter IV, the Enhanced Trickle Down (ETD) model is briefly described, followed by a description of a stepwise procedure to the parameter estimation of the hydrology submodel of the ETD model.

In Chapter V, the results of the application of the procedures developed in Chapter IV to the Woods Lake Catchment in Adirondack Park, New York, are presented. Model validation and verification are performed and the results are analyzed in relation to the nuances of the physical system and processes involved.

Finally, Chapter VI contains the conclusions of this study and recommendations for further research.
II. RECURSIVE IDENTIFICATION: AN OVERVIEW

A. Introduction.

A mathematical model involves two primary features — a model structure and numerical values of parameters. The need for system identification arises when a mathematical model of a system cannot be specified completely and the model needs to be "identified" based on input-output data. The problem of system identification most generally involves the identification of model structure and numerical values of parameters. However, most often, a reasonable structure is hypothesized and assumed fixed while only numerical values of parameters are identified. Subsequent to a system identification exercise, the validity of the identified model is tested and if serious deficiencies are suspected, either in the structure or the parameter estimates, the system identification process needs to be repeated.

The process of identifying model parameters from input-output data can be viewed as the synthesis of a mapping between the "data space" to a parameter space, which translates the "information" contained in the data to obtain parameter estimates. This mapping can be described in most general terms as

\[ \hat{z}_T = \hat{\theta}(z_T, \ell); \hat{\theta} \in \Omega_\theta \]  

II.1

where \( \hat{z}_T \) represents the vector of available input-output data over a time period \( T \), \( \ell \) represents a loss criterion or more generally the structure of the identification algorithm used, \( \hat{\theta}(\cdot) \) is the estimate of the parameter vector obtained through the mapping and \( \Omega_\theta \) is the parameter space specifying physically realistic parameter value ranges. It is precisely the form of this mapping that characterizes a particular identification scheme. In case of offline or batch processing algorithms, the form of the mapping holds without modification, and in case of iterative (off-line) methods, it holds during each iteration. In case of recursive or on-line identification algorithms that process the information in the data record serially, additional restrictions are imposed on the form of the mapping. These can be described by

\[ \hat{z}_T + \hat{\phi}(t) + \hat{\theta}(t), \hat{\theta} \in \Omega_\theta, 0 < t < T \]  

(II.2)
where $\mathbf{z}_{t}$ represents the available data up to time $t$, $\mathbf{\phi}(t)$ is a memory vector of some fixed "small" dimension that summarizes all the information contained in $\mathbf{z}_{t}$ and $\hat{\mathbf{\theta}}(t)$ is the estimate of the parameter vector $\mathbf{\theta}$ at time $t$. $\mathbf{\phi}(t)$ and $\hat{\mathbf{\theta}}(t)$ are usually updated at each $t$. When all the data contained in the record are processed ($t = T$) the final parameter estimates are obtained.

There are various conceptual bases that have motivated choices of particular cases of the mappings in (II.1) and (II.2). Some of these will be briefly reviewed in the next section, which is based on the discussions in Schweppie (1973, Ch. 14), Eykhoff (1974), Young (1980) and Ljung and Söderström (1981).

**B. Conceptual Bases for System Identification**

Most system identification algorithms involve two characteristic features: a) definition of some error function representation of the discrepancy between the model and the physical system, and b) the formulation of an estimation criterion or cost function (usually scalar) in terms of some norm of the error function. Estimates of parameters are obtained by the minimization of this cost function. These are reviewed respectively in Sections II.2.1 and II.2.2. In Section II.2.3, convergence and asymptotic properties that help evaluate the success of a system identification exercise are discussed, and in II.2.4 issues related to the input-output data are presented.

1. **Error Representations** There are three widely used representations of errors in system identification. These are the equation error, output error and prediction error representations. Approaches to the system identification problem have been developed based on each of these representations.

**Equation Error (EE) Approach:** This approach, also known as direct, open-loop or explicit approach, involves the implementation of mathematical relations that can be solved directly to obtain parameter estimates. This approach derives from an analogy with the classical problem of regression in statistics. While the solution to the mathematical equations, by itself, may require iteration, the parameter estimation is not iterative per se.
Output Error (OE) Approach: This approach, also known as indirect, response error, implicit or model reference approach, involves the use of a model adjustment technique wherein the model parameters are adjusted iteratively until some measure of the discrepancy between the model and the system is minimized.

Prediction Error (PE) Approach: This approach, which can be cast in an OE or EE framework involves the interpretation of the model predictions (or output errors generated) or equation errors generated as "best predictions" conditional on some estimate of the parameters. PE methods cast in the OE framework are more common. In general, PE approaches are more complex for noisy systems, since complete knowledge or concurrent estimation of the noise model parameters is required.

Each of these approaches apply to both deterministic and stochastic situations. The deterministic EE is an algebraic function of the unknown parameters and its minimization is straightforward. The characteristics of this approach has been discussed by Young (1965, 1970) among others. In most situations, it leads to a general nonlinear regression problem. In case of applications to dynamic systems, generation of time derivatives of input and response processes is usually required for the implementation of identification algorithms based on this approach. This could lead to the obvious problems that arise from the differentiation of noisy signals. A common device employed to solve this problem is the use of state variable filters. EE approaches usually prove unsatisfactory in case of high noise levels, in stochastic situations, and an asymptotic bias in the estimates it generates results from the fact that models are structural, rather than simple regressions, and all variables are "measured" in the presence of noise (Young, 1968). One of the popular approaches to overcoming this problem is the use of Instrumental Variables (Young, 1976, Young and Jakeman, 1979), which is discussed further in Section II.3.

The deterministic OE approach, also popularly referred to as model reference approach, was motivated by problems in adaptive control, where control system parameters are adjusted recursively to produce desired outputs. A review of early developments in this area can be found in Landau (1972).
Later developments were reviewed by Narendra (1976) and Anderson (1977). Most of the recent research in this area is based on the concept of adaptive observers (e.g. Kriesselmeir, 1976, Anderson, 1977). In the stochastic situation, the OE approach offers certain conceptual advantages over EE approaches, in that there is no stochasticity associated with the model, per se. This has strong implications on the sensitivity of the algorithm to stochastic disturbances. Some approaches like the nonlinear filtering and stochastic approximations approaches are extensions of the general OE approach.

2. Estimation Criteria and Cost Functions The choice of an estimation criterion or cost function, depends to a large extent on the nature of the problem. In case of deterministic models, criteria of the least squares type are often used. In parameter estimation for stochastic models, most of the approaches used find their basis in statistical estimation theory. In this section, Least squares, Bayesian, Maximum likelihood and Maximum A-posteriori approaches to estimation are briefly described.

Least Squares Estimation: One of the cost functions most commonly used in parameter estimation for deterministic models is an integral weighted error norm of the form

\[ J = \int_{t_0}^{t_f} \{ \epsilon(t)^T W(t) \epsilon(t) \} dt \]  \hspace{1cm} \text{(II.3)}

where \( \epsilon(t) \) is a (vector) error function, \( W(t) \) is a time-varying positive definite weighting matrix and \( (t_0-t_f) \) is the time period over which data are available. Algorithms that use such a cost function are usually referred to as Weighted Least Squares (WLS) algorithms. In case the weighting matrix is always the identity matrix, the simple least squares (SLS) algorithms result. Least squares criteria have also been used extensively in stochastic EE frameworks. In these situations, SLS criteria apply only if the error sequence \( \epsilon(t) \) is homoscedastic and uncorrelated. The use of a time-varying weighting matrix accounts in part for heteroscedasticity, while a modified approach called the generalized least squares (GLS) approach is used in case of correlated errors. Strejc (1981) presents an extensive survey of the use of Least Squares and related regression methods in system identification.
Bayesian Estimation: This is based on the Bayesian approach in statistical estimation theory. In this approach, the parameter vector \( \theta \) is regarded as a random vector, with subjective beliefs of the modeller about the value range in which \( \theta \) lies being condensed into a probability distribution \( \xi(\theta) \) on the parameter space \( \Omega_{\theta} \), known as the prior distribution. Then a loss criterion \( L(\theta, \hat{\theta}) \) is defined, which measures the error incurred in some sense, when the true value of the parameters is \( \theta \) and the estimate is \( \hat{\theta} \). If observations \( z_t \) are now available, the posterior distribution \( \xi(\theta/z_t) \) of the parameters conditional on the observations can be derived by application of Bayes theorem, and the expected value of the loss criterion can be expressed as

\[
E[L(\theta, \hat{\theta})/z_t] = \int_{\Omega_{\theta}} L(\theta, \hat{\theta}) \cdot \xi(\theta/z_t) d\theta
\]  

(II.4)

The Bayes estimator of \( \theta, \hat{\theta}_{\text{b}} \) is that value in \( \Omega_{\theta} \) for which the expected value in Eq. (II.4) is minimized.

The Bayesian approach involves a fair amount of subjectivity in the specification of a prior distribution and the choice of a loss function. There is a long standing controversy among statisticians about whether the treatment of parameters as random variables is reasonable. Commonly used loss functions include a squared error loss function, when the Bayes estimator is the mean of the posterior distribution, and an absolute error loss function when the Bayes estimator is the median of the posterior distribution. Most recursive identification algorithms can be interpreted as being, at least in spirit, of the Bayesian type, since they involve sequential estimation. For this reason, the Bayesian concept is very important in this context.

Maximum Likelihood (ML) Estimation: Owing to the controversies over the interpretation of unknown parameter values as random variables and the difficulties in specifying prior distributions, an alternative approach that was introduced in statistical literature (by R.A. Fisher in 1912) is the Maximum likelihood approach. It has a strong intuitive appeal, generates estimators with excellent assymptotic properties and is perhaps the most widely used approach in statistics (DeGroot, 1986). This approach is based on the joint probability density function of the observations \( z_t \), regarded as a function \( f(z_t/\theta) \) of the parameter vector \( \theta \), \( (\theta \in \Omega_{\theta}) \) for which the likelihood function is maximized, i.e.
\[ f(z_t/\theta) = \max_{\theta \in \Theta} f(z_t/\theta) \] (II.5)

The ML approach overcomes the need to choose a loss function, as well, but the existence of an ML estimator is not always guaranteed. It has been widely used in system identification and is cited (Schweppe, 1973, Maybeck, 1982) as a very powerful approach.

**Maximum A-Posteriori (MAP) Estimation:** This approach, while being essentially Bayesian, is treated separately because of the close resemblance it bears to the ML approach. It provides a framework that is particularly convenient in recursive identification. The MAP estimate is that value of \( \theta \in \Theta \), for which the posterior distribution \( \xi(\theta/z_t) \) takes its maximum value. The posterior distribution is defined by Bayes theorem as

\[ \xi(\theta/z_t) = \frac{\xi(\theta) \cdot f(z_t/\theta)}{\int_{\Theta} \xi(\theta) \cdot f(z_t/\theta) d\theta} \] (II.6)

where \( \xi(\theta) \) is the prior distribution.

Since any \( \theta \) - dependence in the denominator is integrated out, \( \theta \) enters only the numerator, which differs from the likelihood function only because of the prior distribution term. In fact, this approach allows for the inclusion of prior information in an ML framework for the cases when significant information is indeed available. It should be noted, based on the discussions in Section II.2.1, that the Bayesian, ML and MAP approaches are essentially prediction error approaches, while the Least Squares approaches are not.

3. **Convergence and Asymptotic Properties** In order to be able to use any particular algorithm for parameter estimation with confidence, it is important to know the behavior of the estimates it generates as the data available increases indefinitely. It is these characteristics that are referred to as asymptotic properties. In case of models with time invariant parameters, the convergence of parameter traces to certain values signifies the culmination of the parameter estimation exercise. It is desirable that these values are close to the "true" parameter values and are inferred with a high
degree of precision. These issues are dealt with in the analysis of asymptotic properties. Some of the important asymptotic properties related to parameter estimation problems are now described.

Asymptotic Consistency: An estimator \( \hat{\theta}(t) \) is said to be an asymptotically consistent estimator of the parameter \( \theta \), if it converges in probability to the true (but unknown) value \( \theta_0 \) of the parameter as \( t \to \infty \).

Asymptotic Unbiasedness: An estimator \( \hat{\theta}(t) \) is said to be asymptotically unbiased if, as \( t \to \infty \), the expected value of \( \hat{\theta}(t) \), with respect to the joint probability density of the parameters and the data, \( h(\theta, z_t) \), tends to the true value \( \theta_0 \); i.e.,

\[
\lim_{t \to \infty} E_\theta(\hat{\theta}(z_t)) = \lim_{t \to \infty} \int \hat{\theta}(z_t) h(\theta, z_t) d\theta = \theta_0
\]

(II.7)

Asymptotic Efficiency: An estimator \( \hat{\theta}(\cdot) \) is said to be efficient, if it is unbiased, has finite error covariance and there exists no other unbiased estimator whose error covariance is "smaller". It is said to be asymptotically efficient if \( \hat{\theta}(t) \) becomes efficient as \( t \to \infty \).

In statistical estimation theory, an inequality called the Cramer-Rao inequality specifies the theoretical lower bound (the Cramer-Rao bound) of the variance that can be obtained in a linear estimation problem. Most of these concepts about asymptotic properties, arose in what is referred to as Fisher Information theory in statistical literature. A detailed treatment can be found in Rao (1973).

In regard to the estimators generated by the estimation criteria, specified in Section II.2.2, the following remarks are in order. Bayesian estimators are consistent estimators and if based on squared error loss criteria, unbiased. ML estimators (MLE's) are asymptotically consistent. If there exists an efficient estimator for a particular problem, it can be shown (Rao, 1973) that the MLE will be efficient. MLE's are also invariant (meaning that if \( \hat{\theta}_{ML} \) is the MLE of \( \theta \), \( g(\hat{\theta}_{ML}) \) is the MLE of \( g(\theta) \)) and if there exists a sufficient statistic (loosely, one that contains all the information in the data about \( \theta \)), the MLE is sufficient. The asymptotic distribution of the MLE is normal with the mean tending to the true value (if one exists) and covariance matrix tending to the Cramer-Rao bound.
In conclusion of this section, it should be noted that in most identification problems, there may be no "true" value of the parameter vector, $\theta$, for which the model gives the time description of the system, due to structural inadequacies in the model. Hence, the meaning of the asymptotic properties must be interpreted very carefully. Also, the analysis of asymptotic properties and convergence does not provide any hint at how large $t$ must actually be for the asymptotic properties to become applicable. Hence, analysis of asymptotic properties must always be complemented by simulation studies.

4. **Data Requirements for System Identification** One important feature of all identification methodologies is that "the final results of an identification exercise are only as good as the data is". Firstly, all measurements must be as accurate as possible and the characteristics of all sensor errors must also be known. Secondly, the inputs should be such that response modes corresponding to all parameters are excited adequately — such inputs are called persistently exciting inputs. In situations where inputs can be chosen by the user, they should be chosen to satisfy this criterion, which can be related to the identifiability criteria of the system. Similarly, identifiability analyses of models can also provide information on the optimum design of instrumentation systems for measurements. A detailed treatment of these concepts can be found in Mehra (1976, 1981).

C. **Approaches to Recursive Identification (RI)**

Most of the recursive identification algorithms that have been developed can be put in four major categories (after Ljung and Soderstrom, 1981). These are:

1) modifications of offline methods  
2) nonlinear filtering methods  
3) stochastic approximation methods  
4) model reference approaches and pseudolinear regressions.

These will be discussed in detail in this section.

1. **Modifications of Offline Methods** The motivation for this approach comes from the fact that there exist many time-tested satisfactory off-line
algorithms. While some of them can be cast in a recursive form that is exact, others cannot be cast in a recursive form unless some approximation is used.

Least Squares Methods:

The Recursive Least Squares (RLS) methods derive by direct analogy from the offline least squares and regression methods. Consider the linear model

\[ y(t) = \theta^T \phi(t) + \nu(t) \]  \hspace{1cm} (II.8)

where \( y() \) is a scalar representing the current observation, \( \theta \) is a \((n \times 1)\) parameter vector, \( \phi() \) is also a \((n \times 1)\) memory vector of past inputs and outputs and \( \nu() \) is a scalar equation error. A criterion function based on the sum of squared errors is

\[ V_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} \alpha_t (y(t) - \theta^T \phi(t))^2 \]  \hspace{1cm} (II.9)

where \( \alpha_t \) is a sequence of time-varying weights. Analytic minimization of Eq. (II.9), by setting its derivative with respect to \( \theta \) equal to zero, gives the least squares estimate

\[ \hat{\theta}(N) = \left[ \sum_{t=1}^{N} \alpha_t \phi(t) \phi^T(t) \right]^{-1} \sum_{t=1}^{N} \alpha_t \phi(t)y(t) \]  \hspace{1cm} (II.10)

assuming that the inverse exists.

Strejc (1981) and Ljung and Soderstrom (1981) present detailed derivations of the recursive form of the estimator in Eq. (II.10), based on partitioning of the matrix \([\phi(t)\phi^T(t)]\) and the vector \([\phi(t)y(t)]\). Defining

\[ R(t) = \sum_{k=1}^{t} \alpha_k \phi(k)\phi^T(k) \]  \hspace{1cm} (II.11)

and

\[ P(t) = R^{-1}(t) \]  \hspace{1cm} (II.12)

a convenient recursive form that results is

18
\[ \hat{\theta}(t) = \hat{\theta}(t-1) + L(t) \left[ y(t) - \hat{\theta}^T(t-1)\hat{\phi}(t) \right] \] (II.13.a)

\[ L(t) = \frac{P(t-1)\phi(t)}{1/\alpha_t + \phi^T(t)P(t-1)\phi(t)} = P(t-1)\phi(t) \] (II.13.b)

\[ P(t) = P(t-1) - \frac{P(t-1)\phi(t)\phi^T(t)P(t-1)}{1/\alpha_t + \phi^T(t)P(t-1)\phi(t)} \] (II.13.c)

with initial conditions on \( \hat{\theta}(0) \) and \( P(0) \) specified by the modeller. It should be noted that \( R(t) \) is not strictly invertible until \( n \) observations are available and an exact implementation of the algorithm should start at \( t_0 = n \), with

\[ P(t_0) = \left[ \sum_{k=1}^{t_0} \alpha_t \phi(k)\phi^T(k) \right]^{-1} \] (II.14)

and

\[ \hat{\theta}(t_0) = P(t_0) \cdot \left[ \sum_{k=1}^{t_0} \alpha_t \phi(k)y(k) \right] \] (II.15)

However, for any reasonable choices of initial conditions (e.g., \( \hat{\theta}(0) = 0 \) and \( P(0) = \delta I_n \), where \( \delta \) is a small positive number and \( I_n \) is the \( nn \) identity matrix), the online estimate approaches the offline estimate for large \( t \).

In case there exists a true value \( \theta_o \) for \( \theta \), the asymptotic properties of \( \hat{\theta}(t) \) can be investigated by writing

\[ y(t) = \theta_o^T \hat{\phi}(t) + v(t) \] (II.16)

as the true system, with \( v(t) \) denoting a scalar observation error process.

Substituting in (II.10) and simplifying yields

\[ \hat{\theta}(N) = \theta_o + \frac{1}{N} \sum_{t=1}^{N} \alpha_t \phi(t)\phi^T(t) \] \( -1 \) \[ \cdot \frac{1}{N} \sum_{t=1}^{N} \alpha_t \phi(t)v(t) \] (II.17)

From (II.17), it is obvious that \( \text{E}(\hat{\theta}(N)) = \theta_o \) as \( N \to \infty \), if and only if \( v(t) \) and \( \phi(t) \) are uncorrelated. This is true if \( v(t) \) is a zero mean white noise
sequence, for instance. It is also apparent that if \( \hat{\phi}(t) \) and \( v(t) \) are correlated, an asymptotic bias is introduced (characteristic of most stochastic EE representations). This problem has been overcome in the Instrumental Variable (IV) approaches (Mayne, 1967, Young, 1968, 1970, 1976,, Young and Jakeman, 1979) wherein the following estimator is used in place of (II.10).

\[
\hat{\theta}(N) = \sum_{t=1}^{N} \alpha_t \bar{z}(t)\hat{\phi}(t)^T [\sum_{t=1}^{N} \alpha_t \bar{z}(t)y(t)]^{-1} \sum_{t=1}^{N} \alpha_t \bar{z}(t)y(t)
\]  

(II.18)

where \( \bar{z}(t) \) is a sequence of instrumental variables that is chosen to be correlated enough with \( \hat{\phi}(t) \) so that the expectation of the inverse in (II.18) exists asymptotically and at the same time is uncorrelated with \( v(t) \), so that asymptotic bias is eliminated. A common choice has been (Young, 1965, 1968, Mayne, 1967, and Wang and Polak, 1967) to replace past outputs \( y(t) \) contained in \( \hat{\phi}(t) \) by \( y_M(t) \) the corresponding outputs from a deterministic system subject to the same inputs, i.e. \( \bar{z}(t) \) are such that

\[
y_M(t) = \hat{\theta}(t)^T \bar{z}(t)
\]  

(II.19)

Recursive versions of the estimator in (II.18) can be derived similar to (II.13).

An approach that develops estimates free of bias in case of colored (non white noise) error sequences \( v(t) \) is the Generalized Least Squares (GLS) method, based on the use of "whitening" prefilters. An offline estimator of this type was developed by Clarke (1967), and online versions by Hastings-James and Sage (1969). Strejc (1981) remarks that the GLS approach may not eliminate bias for high noise strengths (low signal to noise ratios).

Recursive Prediction Error Methods (RPEM)

These methods are based on extensions of the least squares methods to more general model structures and performance criteria. Consider a general scalar prediction model

\[
y(t/\theta) = g_M(\theta, t, \bar{z}_{t-1})
\]  

(II.19)
where the left hand side represents a prediction conditional on a particular value of \( \theta \), and the right hand side is the predictor function, in general a time varying function of the parameters and past data (inputs and perhaps outputs). Define the prediction error (PE)

\[
\varepsilon(t, \theta) = y(t) - \hat{y}(t/\theta)
\]

(II.20)

\( \varepsilon(t, \theta) \) may not in general be linear in \( \theta \), whence any PE criterion will also be nonlinear in \( \theta \) and may require iterative minimization based on numerical schemes. From a computational point of view, this is undesirable in recursive identification algorithms. Hence certain approximations may need to be introduced. Let \( V_t(\theta) \) represent a general PE criterion based on \( \varepsilon(t, \theta) \). An estimate \( \hat{\theta}(t) \) of \( \theta \) is sought such that \( V_t(\hat{\theta}) \) is minimized. Let \( \hat{\theta}(t-1) \) denote the previous estimate of \( \theta \). Expanding \( V_t(\theta) \) about \( \hat{\theta}(t-1) \) in a Taylor Series yields

\[
V_t(\theta) = V_t(\hat{\theta}(t-1)) + V'_t(\hat{\theta}(t-1))[\theta - \hat{\theta}(t-1)] +
\]

\[
\frac{1}{2} \cdot [\theta - \hat{\theta}(t-1)]^T V''_t(\hat{\theta}(t-1))[\theta - \hat{\theta}(t-1)] + \text{higher order terms}
\]

(II.21)

where

\[
V'_t(\hat{\theta}(t-1)) = \frac{\partial V_t(\theta)}{\partial \theta} \bigg|_{\theta = \hat{\theta}(t-1)} \quad \text{and} \quad V''_t(\hat{\theta}(t-1)) = \frac{\partial^2 V_t(\theta)}{\partial \theta^2} \bigg|_{\theta = \hat{\theta}(t-1)}
\]

If the higher order terms in (II.21) are neglected, a quadratic approximation to (II.21) is obtained, analytic minimization of which yields

\[
\hat{\theta}(t) = \hat{\theta}(t-1) - [V''_t(\hat{\theta}(t-1))]^{-1} V'_t(\hat{\theta}(t-1)) + \text{or} \ (|\hat{\theta}(t)-\hat{\theta}(t-1)|)
\]

(II.22)

where the last term on the right hand side denotes terms of orders of the difference between \( \hat{\theta}(t) \) and \( \hat{\theta}(t-1) \) and its powers. It should be noted that if \( \theta \) is a nxl vector, \( V'_t(\theta) \) is a nxl vector and \( V''_t(\theta) \) is a nxn matrix.

Define

21
\[
\dot{\psi}^T(t, \theta) = \frac{\dot{y}(t|\theta)}{d\theta} = -\frac{d\epsilon(t, \theta)}{d\theta}
\]  

(II.23)

If the sum of squared prediction errors is used as the criterion function, i.e.

\[
V_N(\theta) = \sum_{t=1}^{N} \epsilon^2(t, \theta)
\]  

(II.24)

then

\[
V^T_t(\theta) = -2 \sum_{k=1}^{t} \psi(k, \theta)\epsilon(k, \theta) = V^T_{t-1}(\theta) - 2\psi(t, \theta)\epsilon(t, \theta)
\]  

(II.25)

Differentiating (II.25) with respect to \( \theta \), gives

\[
V^T_t(\theta) = V^T_{t-1}(\theta) + \psi(t, \theta)\dot{\psi}^T(t, \theta) + \epsilon''(t, \theta)\epsilon(t, \theta)
\]  

(II.26)

To simplify these expressions, certain approximations need to be introduced.

1) Neglect the terms \( \{\dot{\theta}(t) - \dot{\theta}(t-1)\} \) in (II.22).

2) Assume that \( \dot{\theta}(t-1) \) is close enough to the optimum of \( V_{t-1}(\theta) \), so that \( V^T_{t-1}(\theta(t-1)) = 0 \).

3) Assume that \( E[\epsilon''(t, \theta)\epsilon(t, \theta)] = 0 \). (Close to the optimum, \( \epsilon(t, \theta) \) is "almost-white" noise and \( \epsilon''(t, \theta) \) is actually the negative second derivative of the prediction function, which depends upon past information.)

Defining \( R(t) = V''(\theta) \), using assumption (iii) and substituting in (II.26) with \( \theta \) replaced by \( \dot{\theta}(t-1) \),

\[
R(t) = R(t-1) + \psi(t, \dot{\theta}(t-1))\dot{\psi}^T(t, \dot{\theta}(t-1))
\]  

(II.27a)

Using assumption (ii) in (II.25),

\[
V^T_{t}(\dot{\theta}(t-1)) = -\psi(t, \dot{\theta}(t-1))\epsilon(t, \dot{\theta}(t-1))
\]  

(II.27b)
and from (II.22), using assumption (1) and (II.27b),

\[ \hat{\theta}(t) = \hat{\theta}(t-1) + R^{-1}(t)\Psi(t, \hat{\theta}(t-1))\epsilon(t, \hat{\theta}(t-1)) \]  

(II.27c)

Equations (II.27a, b and c) summarize the Recursive Prediction error algorithm that attempts to minimize the criterion function (II.24) sequentially. It should be noted though that (II.27) may not be strictly recursive, since \( \Psi() \) and \( \epsilon() \) in general depend on all past data. Approximations that allow for the estimation of these from fixed length memory vectors need to be introduced.

Based on a similar approach, other criterion functions could be used and RPEM derived. A commonly used criterion function is based on Maximum Likelihood estimation and was studied by Astrom and Kallstrom (1973) and Mehra and Tyler (1973) for a state space model.

The similarities between (II.27) and (II.13) are striking. Note for instance that \( \epsilon(t, \hat{\theta}(t-1)) \) replaces \([y(t) - \hat{\theta}^\top(t-1)\hat{\phi}(t)]\), \( L(t) \) is replaced by \( R^{-1}(t)\Psi(t, \hat{\theta}(t-1)) \), and, in the least squares, case \( \phi() \) is identically equal to \( \hat{\phi}() \) so that \( L(t) \) in (II.13) is a special case of \( R^{-1}(t)\psi(t, \hat{\theta}(t-1)) \).

2. Nonlinear Filtering Methods The nonlinear filtering approach is essentially a Bayesian approach, where the parameters are viewed as random variables and they are estimated indirectly from observations on variables that are correlated with them. More explicitly, the posterior distribution of the parameters conditional on the observations is used as the basis for estimation. Essential to this framework is the need to determine how the density in question evolves with time.

Within the framework of stochastic differential equations, the equations for the evolution of densities are the celebrated Kolmogorov equations (Jazwinski, 1970) that apply to Markov processes. While they apply to densities of very general form, the computations involved are rarely tractable for non Gaussian densities. In case of Gaussian densities describing initial conditions and all driving processes and linear differential equations, it can be proved (Jazwinski, 1970) that the evolved density always remains Gaussian, and, since Gaussian densities are completely specified by their first two moments, it suffices to consider the evolution of those moments.
The application of these general concepts to the problem of minimum variance state estimation of a linear stochastic dynamic system was first suggested by Kalman (1960) and Kalman and Bucy (1961), and the recursive filter they derived is commonly referred to as the Kalman filter. Extensions to nonlinear dynamic systems can be achieved by applying the algorithm to the linearized nonlinear systems (Jazwinski, 1970), leading to the Extended Kalman Filter (EKF) and a whole family of related algorithms.

Central to the development of the nonlinear filtering approach is the following linear state-space model structure (described for the discrete-time situation)

\[
\hat{x}(k+1) = \Phi(k+1,k)\hat{x}(k) + G(k)\omega_d(k) \tag{II.28.a}
\]

\[
\bar{z}(k) = H(k)\hat{x}(k) + \nu(k) \tag{II.28.b}
\]

where (II.28.a) is the vector stochastic difference equation that describes the linear system of interest, and (II.28.b) models the discrete-time measurements available from the system. \(\hat{x}(\cdot)\) is the state vector, \(\bar{z}(\cdot)\) the observations vector, \(\Phi(\cdot)\) is the transition matrix of the system, \(H(\cdot)\) is a coefficient matrix, \(G(\cdot)\) is a coefficient matrix and \(\omega_d(\cdot)\) is a sequence of model errors.

If the initial condition of \(x\), \(\hat{x}(o)\) has the following first and second moments:

\[
\hat{x}(o) = \mathbb{E}\{x(o)\} \tag{II.29}
\]

\[
P(o) = \mathbb{E}\{[x(o) - \hat{x}(o)][x(o) - \hat{x}(o)]^T\} \tag{II.30}
\]

then, straightforward use of the fact that (II.28.a) is linear in \(x(\cdot)\) leads to the following equations for the evolution of the first and second moments in time:

\[
\hat{x}(j+\Delta j) = \Phi(j+\Delta j,j)\hat{x}(j) + G(j)\mathbb{E}\{\omega_d(j)\} \tag{II.31}
\]

\[
P(j+\Delta j) = \Phi(j+\Delta j,j)P(j)\Phi^T(j+\Delta j,j) +
\]

24
\[ \Phi(j+\Delta j, j)E\{[x(j+\Delta j) - x(j)]w_d^T(j)\}G^T(j) + \]
\[ G(j)E\{w_d(j)[x(j) - x(j)]^T\} \Phi^T(j+\Delta j, j) + G(j)E\{w_d(j)w_d^T(j)\} G^T(j) \]

(II.32)

The nonstationary filtering problem that was solved by Kalman and Bucy (1961) involves the determination of the minimum variance estimate of \( x(j) \) conditional on the observation \( z(j) \). In the case of a Gaussian density for \( x(o) \), it is completely specified by its first two moments. If we now introduce the assumptions that \( w_d() \) and \( v() \) are mutually uncorrelated white-noise Gaussian zero-mean sequences with covariance matrices defined by

\[ E\{w_d(j)w_d^T(k)\} = Q_d(j)\delta_{jk} \forall k,j \]

(II.33)

\[ E\{v(j)v^T(k)\} = R(j)\delta_{jk} \forall k,j \]

(II.34)

then the recursive minimum variance estimator of \( x(k) \) has the following structure:

**Between Observations:** (PREDICTION)

\[ \hat{x}(k+1/k) = \Phi(k+1,k)\hat{x}(k/k) \]

(II.35.a)

\[ P(k+1/k) = \Phi(k+1,k)P(k/k)\Phi^T(k+1,k) + G(k)Q_d(k)G^T(k) \]

(II.36.b)

**At an Observation:** (UPDATE)

\[ \hat{x}(k+1/k+1) = \hat{x}(k+1/k) + K(k+1)[z(k+1) - H(k+1)\hat{x}(k+1/k)] \]

(II.35.c)

\[ K(k+1) = P(k+1/k)H^T(k+1)[H(k+1)P(k+1/k)H^T(k+1) + R(k+1)]^{-1} \]

(II.35.d)

\[ P(k+1/k+1) = [I - K(k+1)H(k+1)]P(k+1/k) \]

(II.35.e)

where \( \hat{x} \) and \( P \) denote best estimates of \( x \) and associated estimation covariance matrix, and "(i/j)" denotes the value of a quantity at time i, conditional on
the observations up to and including time j. In the continuous-time case with
discrete measurements, the update equations are unchanged and the prediction
equations are changed due to the fact that the white noise process in con-
tinuous time is nondifferentiable, so that a stochastic differential equation
driven by a white noise process is not integrable in the Reimann sense. This
problem can be overcome by the use of the Itô stochastic calculus (Jazwinski,
1970). For the stochastic differential equation

$$dx(t) = F(t)x(t)dt + G(t)dB(t)$$  \hspace{1cm} (II.36)

the prediction equations are

$$\frac{dx(t)}{dt} = F(t)x(t)$$  \hspace{1cm} (II.37)

$$\frac{dP(t)}{dt} = F(t)P(t) + P(t)F^T(t) + G(t)Q(t)G^T(t)$$  \hspace{1cm} (II.38)

with (II.37) replacing (II.35.a) and (II.38) replacing (II.35.b). Eq. (II.36)
is the linear stochastic differential equation of interest, with \( B(t) \) being a
Wiener process (whose formal derivative is often replaced by white-noise \( W(t) \)
in engineering notation), and \( F(t) \) the matrix of coefficients in the linear
system of differential equations. \( Q(t) \) is the covariance-parameter matrix of
the white-noise process. A formal treatment of these concepts can be found in

The parameters to be estimated in the model of (II.28) or (II.36) are
coefficients in the \( \Phi, G, \) and \( H \) or equivalently \( F, G, \) and \( H \) matrices. It should
be noted that in the equations (II.35.b) and (II.38), \( Q_d() \) and \( Q() \) respectively,
are assumed known. Also, \( R() \) in (II.35.d) is assumed known. Very often,
these matrices that represent the levels of uncertainty associated with the
model and the observations are not known a-priori and need to be estimated, as
well. This problem is known as Adaptive Filtering. Later in this work, the
issue of filter-parameter identification is examined. For the purpose of the
treatment to follow, however, the filter parameters will be assumed known.

Consider the use of the state estimator (II.35) for a nonlinear system,
such as

26
\[
\frac{dx(t)}{dt} = f(x(t), u(t), \theta) + G(t)w(t) \tag{II.39}
\]

where \( f() \) is a general vector-valued nonlinear function, \( x() \) is the state vector, \( u() \) a vector of known inputs, \( \theta \), a vector of parameters in the function \( f() \), and \( G() \) and \( w() \) are as defined before. State estimation is accomplished by linearizing (II.39) in some way and applying (II.35) to the linearized system. Expand \( f() \) about \( \hat{x}(t) (t) \) (= \( \hat{x}(t) \)), in a Taylor series as

\[
\frac{dx(t)}{dt} = f(\hat{x}(t), u(t), \theta) + G(t)w(t) + F(t) \cdot (x(t) - \hat{x}(t))
\]

+ higher order terms \tag{II.40}

where

\[
F(t) = \left. \frac{\partial f(x, u, \theta)}{\partial x} \right|_{\hat{x}(t)}
\tag{II.41}
\]

If all the terms of order higher than 1, are neglected in (II.40), the resulting equation for the evolution of the state mean between observations is (to a first order linear approximation):

\[
\frac{\hat{x}(t)}{dt} = f(\hat{x}(t), u(t), \theta) \tag{II.42}
\]

Note that it involves the replacement of the expected value of a nonlinear function of \( x() \) with the nonlinear function of its expected value, which does not strictly hold and introduces a bias in the propagation. Some other forms of linearization (e.g. statistical linearization, Gelb, 1974) overcome this.

The recursive state estimator that uses (II.42), (II.38) and (II.35.c-e) is the EKF algorithm. Several variants of it are described in Jazwinski (1970). Consider now the Equation (II.36). If it is desired to estimate some parameters in the matrices \( F() \) and \( H() \), one way to proceed is to consider the augmented state

\[
\xi(t) = [x(t) \mid \theta(t)]^T \tag{II.43}
\]

27
with \( \hat{\theta}(t) \) being the vector of parameters to be estimated. While Equation (II.36) is linear in \( x(t) \), it is nonlinear in \( \xi(t) \). Thus recursive estimation of \( \xi(t) \) (which would provide estimates of \( \hat{\theta}(t) \)) can be achieved by the use of an algorithm such as the EKF that applies to nonlinear systems. Perhaps the earliest instance of this approach to parameter estimation of linear systems is that of Kopp & Orford (1963). The resulting combined state-parameter estimator is presented below.

Consider a continuous-time system, such as in (II.36), with discrete time measurements \( z(t_k) \) related to the states and parameters by

\[
\begin{align*}
    z(t_k) &= h(x(t_k), \theta(t_k)) + v(t_k) \\
    \hat{x}(t/t_k) &= E\{x(t)/z(t_k)\}, \quad t \in [t_k, t_{k+1}] \\
    \hat{\theta}(t/t_k) &= E\{\theta(t)/z(t_k)\}, \quad t \in [t_k, t_{k+1}] \\
    \hat{x}^- &= \hat{x}(t_{k+1}/t_k), \quad \hat{\theta}^- = \hat{\theta}(t_{k+1}/t_k) \\
    \hat{x}^+ &= \hat{x}(t_{k+1}/t_{k+1}), \quad \hat{\theta}^+ = \hat{\theta}(t_{k+1}/t_{k+1}) \\
    P_{xx}(t/t_k) &= E\{[x(t) - \hat{x}(t/t_k)] [x(t) - \hat{x}(t/t_k)]^T\} \\
    P_{x\theta}(t/t_k) &= E\{[x(t) - \hat{x}(t/t_k)] [\theta - \hat{\theta}(t/t_k)]^T\} \\
    P_{\theta\theta}(t/t_k) &= E\{[\theta - \hat{\theta}(t/t_k)] [\theta - \hat{\theta}(t/t_k)]^T\}
\end{align*}
\]

and similarly \( P_{xx}^-, P_{xx}^+, P_{x\theta}^-, P_{x\theta}^+, P_{\theta\theta}^- \) and \( P_{\theta\theta}^+ \).

Let

\[
\begin{align*}
    F_x(t) &= \frac{\partial f(\cdot)}{\partial x(\cdot)} \bigg|_{x(t/t_k)} \\
    F_\theta(t) &= \frac{\partial f(\cdot)}{\partial \theta(\cdot)} \bigg|_{x(t/t_k)}
\end{align*}
\]
\[ H_x(t_k) = \frac{\partial h(t)}{\partial x(t)} \bigg|_{\hat{x}(t_k/t_{k-1}), \hat{\theta}(t_k/t_{k-1})} \]  

(II.53)

\[ H_\theta(t_k) = \frac{\partial h(t)}{\partial \theta} \bigg|_{\hat{x}(t_k/t_{k-1}), \hat{\theta}(t_k/t_{k-1})} \]  

(II.54)

The application of the EKF to the parameter-state estimation problem results in the following algorithm:

**Between Observations:**

\[ \frac{dx(t/t_k)}{dt} = f(\hat{x}(t/t_k), u(t), \hat{\theta}(t_k/t_k)) \]  

(II.55.a)

\[ \frac{d\theta(t/t_k)}{dt} = 0 \]  

(II.55.b)

\[ \frac{dP_{xx}(t/t_k)}{dt} = F_x(t)P_{xx}(t/t_k) + P_{x\theta}(t/t_k)F_{\theta}(t) + F_{\theta}(t)P_{\theta\theta}(t/t_k) + G(t)Q(t)G^T(t) \]  

(II.55.c)

\[ \frac{dP_{x\theta}(t/t_k)}{dt} = F_x(t)P_{x\theta}(t/t_k) + F_{\theta}(t)P_{\theta\theta}(t/t_k) \]  

(II.55.d)

\[ \frac{dP_{\theta\theta}(t/t_k)}{dt} = 0 \]  

(II.55.e)

**At an Observation:**

\[ S(t_{k+1}) = H_x(t_{k+1})P_{xx}(t_{k+1}) + H_\theta(t_{k+1})P_{x\theta}(t_{k+1}) + H_\theta(t_{k+1})P_{\theta\theta}(t_{k+1}) \]  

(II.55.f)

\[ K(t_{k+1}) = [P_{xx}(t_{k+1}) + P_{x\theta}(t_{k+1})]^{-1} \]  

(II.55.g)

\[ L(t_{k+1}) = [P_{x\theta}(t_{k+1}) + P_{\theta\theta}(t_{k+1})]^{-1} \]  

(II.55.h)

\[ \hat{x} = \hat{x} + K(t_{k+1})[z(t_{k+1}) - h(\hat{x}, \hat{\theta})] \]  

(II.55.i)
\[
\hat{\theta}^+ = \hat{\theta}^- + L(t_{k+1})[z(t_{k+1}) - h(\hat{x}^-, \hat{\theta}^-)] \quad \text{(II.55.j)}
\]
\[
P^+_{xx} = [I - K(t_{k+1})H_x(t_{k+1})]P^-_{xx} - K(t_{k+1})H_\theta(t_{k+1})P^-_{x\theta} \quad \text{(II.55.k)}
\]
\[
P^+_{x\theta} = [I - K(t_{k+1})H_x(t_{k+1})]P^-_{x\theta} - K(t_{k+1})H_\theta(t_{k+1})P^-_{\theta \theta} \quad \text{(II.55.l)}
\]
\[
P^+_{\theta \theta} = [I - L(t_{k+1})H_\theta(t_{k+1})]P^-_{\theta \theta} - L(t_{k+1})H_x(t_{k+1})P^-_{x\theta} \quad \text{(II.55.m)}
\]

Eq. (II.55.k) has an alternative form, which is
\[
P^+_{xx} = [I - K(t_{k+1})H(t_{k+1})]P^-_{xx} [I - K(t_{k+1})H(t_{k+1})]^T - K(t_{k+1})H_\theta(t_{k+1})P^-_{x\theta} \quad \text{(II.55.n)}
\]

It can be verified that substitution of (II.55.g) into (II.55.n) leads to (II.55.k).

\[
S(t_{k+1}) \text{ is the covariance matrix of the innovations vector defined by}
\]
\[
v(t_k) = z(t_k) - h(\hat{x}^-, \hat{\theta}^-) \quad \text{(II.56)}
\]

The system of equations (II.55) summarizes the Extended Kalman filter (EKF) algorithm for parameter estimation of a general dynamic system, which could also be nonlinear. In the particular case where the parameters are linear in the data, as in the linear regression model, the linear filtering approach leads to the Kalman filter algorithm which, for the linear regression model, takes the same form as the RLS algorithms of Section (II.3.1).

As will be discussed in Section II.5, the EKF algorithm does not, on a theoretical basis, guarantee convergence in the parameter estimation problem. In addition, a need to concurrently estimate the noise covariance matrices Q and R (or have good a-priori estimates of them) could be another disadvantage of nonlinear filtering approaches in general. However, these approaches have been used extensively for parameter estimation.

Even within the nonlinear filtering framework, it is possible to design identification schemes that approach the parameter estimation problem direct-
ly, rather than use state augmentation methods. One such scheme is a Maximum Likelihood type approach, where a likelihood function is defined, the maximization of which yields state estimator equations and parameter estimator equations. There are various possible interpretations of the problem, that lead to different candidate likelihood junctions. These are discussed by Maybeck (1982).

3. Stochastic Approximations Methods. The origins of this approach can be traced to sequential parameter estimation problems in statistical literature in the 1950's. These methods are not optimal in any statistical sense, but yield recursive estimates with certain well defined convergence properties. The motivation for these algorithms stems from the fact that optimal estimation algorithms are very difficult to implement in situations where the statistical characteristics of the parameters are not well defined and driving noise processes are evidently non Gaussian or the dynamics are of too much a nonlinear character for accurate approximation by linearization. In some such circumstances, it is possible to apply stochastic approximation methods to obtain a sequence of estimates that asymptotically approach the "true" value under very weak mathematical assumptions. Most of these algorithms are recursive linear functions of the observations, a feature that makes them easy to implement. A detailed rigorous mathematical exposition on stochastic approximation algorithms, their convergence properties and asymptotic properties can be found in Nevel'son and Has'minskii (1973). For the purposes of the present discussion, only a heuristic treatment is presented.

Consider the solution to an equation of the form

\[ E[L(x,e(t))] = f(x) = 0 \]  (II.57)

where \( E \) is the expectation over \( e(t) \), a sequence of identically distributed random vectors, and \( L() \) is a function (of the unknown \( x \) and of \( e(t) \)) that is either known in form or can be observed. Robbins and Monro (1951) proposed the following scheme to estimate \( x \) recursively.

\[ \hat{x}(t) = \hat{x}(t-1) + \gamma(t)L(\hat{x}(t-1),e(t)) \]  (II.58)

31
where $\gamma(t)$ is a sequence of positive scalars tending to zero. The convergence properties of (II.58) were studied by Robbins and Monro (1951) who proved that $\hat{x}(t)$ converges to the solution of (II.57) under certain assumptions (for instance, $e(t)$ being a sequence of independent random vectors).

The relevance of the equation (II.57) to the system identification context stems from the fact that very often, in the latter, it is required to minimize the expected value of some loss criterion. This is illustrated for the linear regression model below:

$$ y(t) = \phi^T(t)\theta + e(t) \quad \text{(II.59)} $$

we seek the least squares estimate $\hat{\theta}^*$ such that

$$ V(\theta^*) = \min_{\theta \in \Omega} V(\theta) = \min_{\theta \in \Omega} E[y(t) - \phi^T(t)\theta]^2 \quad \text{(II.60)} $$

Analytic minimization of (II.60) leads to the equation

$$ E[\phi(t)[y(t) - \phi^T(t)\theta]] = 0 \quad \text{(II.61)} $$

which is of the form of (II.57) with $\theta$ replacing $x$ and $[\hat{\theta}(t)|y(t)]$ replacing the vector $e(t)$. Applying the scheme of (II.58) to (II.61) yields

$$ \hat{\theta}(t) = \hat{\theta}(t-1) + \gamma(t)\phi(t)[y(t) - \phi^T(t)\theta] \quad \text{(II.62)} $$

This algorithm is used extensively in adaptive signal processing and referred to as the LMS algorithm. The sequence $\gamma(t)$ in (II.62) is the "gain sequence". The choice of $\gamma(t)$ varies from one application to another. For a particular choice of $\gamma(t)$, viz.

$$ \gamma(t) = \left[ \sum_{k=1}^t ||\phi(k)|| \right]^{-1} \quad \text{(II.63)} $$

where $||\phi(k)||$ denotes $\phi^T(k)\phi(k)$, the norm of the vector $\phi(k)$, the following algorithm that bears strong resemblance to the RLS algorithm of Section II.3.1 results
\[ \hat{\theta}(t) = \hat{\theta}(t-1) + L(t) [y(t) - \phi^T(t) \hat{\theta}] \]  

(II.64.a)

\[ L(t) = P^{-1}(t) \cdot \phi(t) \]  

(II.64.b)

\[ P(t) = P(t-1) + ||\hat{\phi}(t)|| \]  

(II.64.c)

\[ \gamma(t) = P^{-1}(t) \]  

(II.64.d)

Central to the derivation of equation (II.61) is the fact that the stationary points of (II.60) were sought. For a general criterion \( V(\theta) \), extending the same idea, we can try to develop an algorithm based on

\[ \frac{d}{d\theta} [V(\theta)] = 0. \]  

(II.65)

If

\[ V(\theta) = E[H(\theta,e(t))] \]  

(II.66)

then (II.65) reduces to (allowing for interchange of expectation and differentiation)

\[ E\left( \frac{\partial H(\lambda)}{\partial \lambda} \right) = 0 \]  

(II.67)

Equations (II.65) and (II.66) are unaltered if a negative sign is introduced on the left-hand side. Thus the Robbins–Monro scheme (II.62) can be viewed as an algorithm to minimize some criterion by making sequential adjustments in the negative gradient direction of that criterion. This is analogous to some of the methods used in optimization literature, known (see Luenberger, 1973) as gradient, descent or first order methods.

It is well known in optimization literature (see Luenberger, 1973) that the class of so-called Newton methods, that derive from quadratic local approximations to general nonlinear functions, have a much faster (second order as against first order of descent methods) convergence rate close to the optimum. In these algorithms, the gradient direction is modified by premultiplying it with the inverse of the second derivative (Hessian) matrix of the criterion function with respect to the independent variables. Extending this
concept to the Robbins-Monro scheme above, for a general criterion function \( V(\hat{\theta}) \), leads to

\[
\hat{\theta}(t) = \hat{\theta}(t-1) + \gamma(t)[V''(\hat{\theta}(t-1))]^{-1}[V'(\hat{\theta}(t-1))]^T
\]  

(II.68)

where \( V'(\hat{\theta}) \) and \( V''(\hat{\theta}) \) are respectively the first and second derivatives of \( V(\hat{\theta}) \) with respect to \( \hat{\theta} \). If \( V(\hat{\theta}) \) is as in (II.60), we can derive, along the lines of (II.68),

\[
\hat{\theta}(t) = \hat{\theta}(t-1) + \gamma(t)R^{-1}(t)\phi(t)[y(t) - \phi^T(t)\hat{\theta}(t-1)]
\]  

(II.69)

where \( [V'(\hat{\theta})]^T \) in (II.68) is replaced by \( \phi(t)[y(t) - \phi^T(t)\hat{\theta}(t-1)] \) and

\[
R(t) = V''(\hat{\theta}(t-1)) = E[\phi(t)\phi^T(t)]
\]  

(II.70)

which can be obtained as the solution to

\[
E[\phi(t)\phi^T(t)] - R(t) = 0
\]  

(II.71)

Applying the Robbins-Monro scheme to (II.71) gives

\[
R(t) = R(t-1) + \gamma(t)[\phi(t)\phi^T(t) - R(t-1)]
\]  

(II.72)

If \( \gamma(t) = 1/t \), then (II.69) and (II.72) together are of the form of the RLS algorithm of Section II.3.1.

Ljung and Soderstrom (1981) refer to (II.60) as a "stochastic gradient" scheme and to (II.69) as a "stochastic Newton" scheme. Since the Newton scheme could possibly diverge far from the optimum, the Hessian is often replaced by some positive definite substitute for minimization problems and such algorithms are referred to as "quasi-Newton" algorithms.

It should be noted at this stage that (II.68) bears a strong resemblance to equation (II.22). The reason for this is that both are based on quadratic approximations to the criterion function. Thus, the stochastic Newton algorithm has a very strong relevance in the context of recursive prediction error methods.
4. Model Reference Approaches and Pseudolinear Regressions  Both these approaches bear striking similarities to each other and the methods previously presented under the heading: Modified Offline Methods. But the motivation for these approaches arose directly from the online situation, for this reason they are classified separately.

The model reference concept (Landau, 1976) is one that is used extensively in adaptive control, where the system is adjusted in real time to produce desired outputs. In the RI context, it involves adjusting the model parameters online to match system outputs. It has the same form as the RLS algorithm (II.13) except that the $\hat{y}$ vector used for generating the prediction $y()$ uses previous predictions, instead of actual observations, as a result of which it is not affected by measurement noise.

The pseudolinear regression methods involve similar modifications of the $\hat{y}$ vector, including past predictions. In problems that are not truly linear regressions, use of the linear regression model, with the $\hat{y}$ vector including certain unknowns that can be estimated (predicted) based on current estimates of $\theta$, leads to pseudolinear regressions (Solo, 1978). One such algorithm that is applied extensively to ARMAX forecasting models, is the Extended Least Squares (ELS) algorithm. This algorithm has striking similarities to the RLS algorithm of (II.13) and the Kalman filter algorithm for linear estimation. Ljung et al. (1975) and Solo (1978) analyzed the convergence properties of this algorithm. Anderson and Moore (1979) give an account of the application of this algorithm to the problem of simultaneous state-parameter estimation. Consider the model

$$x(k+1) = Fx(k) + K\nu(k) + u(k) \quad (II.73)$$

$$z(k+1) = \theta^T x(k+1) \quad (II.74)$$

where $\nu()$ is zero mean white noise and $u()$ is a sequence of vector inputs. If $x()$ and $z()$ are always measurable, the problem of estimating $\theta$ is essentially a linear regression problem. However, if $x()$ is not measurable, extending the least squares algorithm for the linear regression model, based on estimates of $x()$, leads to (see Anderson and Moore, 1979)
\[
\hat{x}(k+1) = F\hat{x}(k) + K\hat{v}(k) + u(k) \quad \text{(II.75.a)}
\]

\[
\hat{v}(k) = z - \hat{\theta}^T(k)\hat{x}(k) \quad \text{(II.75.b)}
\]

\[
\hat{\theta}(k) = \hat{\theta}(k-1) + P(k-1)\hat{x}(k)[z(k) - \hat{\theta}^T(k-1)\hat{x}(k)] \quad \text{(II.75.c)}
\]

\[
P(k+1) = P(k) - \frac{P(k)\hat{x}(k)\hat{x}^T(k)P(k)}{1 + \hat{x}^T(k)P(k)\hat{x}(k)} \quad \text{(II.75.d)}
\]

The similarities of (II.65) to (II.13) are striking.

**D. Unified Approach to Recursive Identification**

From the discussions in Section II.3, it is evident that most recursive identification algorithms (RIA's) involve two vectors and a matrix, i.e. \(\hat{\theta}(t)\) a vector of parameter estimates, \(\hat{\phi}(t)\) a memory vector summarizing past data and a matrix \(P(t)\), which in some cases is a sort of performance criterion, or in general, used to generate an updating direction involved in the updates of parameter estimates. The latter matrix will be referred to as Auxiliary Matrix in the discussion to follow. The algorithms involve after each observation, three operations which can be represented as

Parameter Update: \(\hat{\theta}(t) = U_\theta(\hat{\theta}(t-1),\hat{\phi}(t),P(t),\varepsilon(t))\) \quad \text{(II.76.a)}

Memory Vector Update: \(\hat{\phi}(t) = U_\phi(\hat{\theta}(t-1),\hat{\phi}(t-1),P(t),\varepsilon(t))\) \quad \text{(II.76.b)}

Auxiliary Matrix Update: \(P(t) = U_p(\hat{\theta}(t-1),\hat{\phi}(t),P(t-1),\varepsilon(t))\) \quad \text{(II.76.c)}

where \(\varepsilon(t)\) is an error vector and \(U_\theta, U_\phi, U_p\) are general "updating rules". A formalization of these concepts to develop a general framework for recursive identification follows. The motivating factor for establishing such a framework is the expediency that such a framework would provide in terms of analysis of the behavior and convergence of various algorithms.

Essential to the establishment of a general framework is the establishment of a general model structure, a general criterion function and a general estimator based on the criterion function. The generality of the state-space model (see Schweppe, 1973, Ch. 3) makes it a natural candidate for an appro-
appropriate model structure. The general prediction error criterion function that was used in the discussions in Section II.3.1 is a natural choice for a criterion function, since most RIA's are based on prediction error criteria. The development of an estimator based on the criterion function was possible through analytical solution in some cases (RLS,EKF), but the most general approach to criterion minimization discussed earlier is perhaps the Newton-type approach mentioned in connection with stochastic approximations in Section II.3.3. Armed with these general tools, the rest of the discussion follows (based on Ljung and Soderstrom, 1981).

Consider the general linear (in the input-output data) discrete predictor model

\[
\hat{\phi}(t+1, \theta) = F(\theta)\hat{\phi}(t, \theta) + G(\theta)z(t) \tag{II.77.a}
\]

\[
\hat{y}(t+1, \theta) = H(\theta)\hat{\phi}(t+1, \theta) \tag{II.77.b}
\]

where
- \( \theta \) is a dx1 parameter vector
- \( \hat{\phi}(\cdot) \) is a nx1 memory vector
- \( \hat{y}(\cdot) \) is a px1 (predicted) output vector
- \( y(\cdot) \) is a px1 observations vector
- \( z(\cdot) \) is a mx1 vector, of inputs and observations, \( z = [u \mid y]^T \)

F(\cdot), G(\cdot) and H(\cdot) are respectively nxn, nxm and pxn matrix-valued functions of the parameter \( \theta \).

The prediction error is defined by

\[
\varepsilon(t+1, \theta) = y(t+1) - \hat{y}(t+1, \theta) \tag{II.78}
\]

and a general PE criterion has the form

\[
V(\theta) = E[\ell(t, \theta, \varepsilon(t, \theta))] \tag{II.79}
\]

where the expectation is over \( z_t \) for fixed values of the model parameter vector \( \theta \). For the purpose of this discussion, we will however, assume that \( V(\theta) \) depends on \( \theta \) only through \( \varepsilon(t, \theta) \), i.e..

37
\[ V(\theta) = E[\ell(t, \varepsilon(t, \theta))] \quad (II.80) \]

(A notable exception to this assumption is the maximum likelihood criterion function of Equation (II.53)). For the implementation of the stochastic Newton algorithm (II.68), the first and second derivatives of \( V(\theta) \) with respect to \( \theta \) are required and can be computed as follows:

\[ V'(\theta) = E[\varepsilon_T^\theta(t, \theta)[\ell^\varepsilon_T(t, \varepsilon(t, \theta))]] \quad (II.80.a) \]

where \( \varepsilon_T^\theta() \) denotes the transposed derivative of \( \ell \) with respect to \( \varepsilon() \), and has dimensions px1. \( \varepsilon_\theta() \) denotes the derivative of \( \varepsilon() \) with respect to \( \theta \), and has dimensions dxp.

\[ V''(\theta) = E[\varepsilon_T^\theta(t, \theta)[\ell^\varepsilon(t, \varepsilon(t, \theta))] \varepsilon_\theta(t, \theta)] + \]

\[ E[\varepsilon_T^{\theta \theta}(t, \theta)[\ell^\varepsilon(t, \varepsilon(t, \theta))]] \quad (II.80.b) \]

where it should be noted that \( \varepsilon_\theta(t, \theta) \) is a tensor of dimensions pxspd. The second term should be interpreted as a matrix whose \((i,j)^{th}\) element is given by

\[ E\left\{ \sum_{k=1}^{p} \frac{d^2}{d\theta_i d\theta_j} \varepsilon_k(t, \theta)[\ell^\varepsilon(t, \varepsilon(t, \theta))] \right\} \]

Based on the definition of \( \varepsilon(t, \theta) \) in (II.78), it is evident that

\[ \varepsilon_T^\theta(t, \theta) = -\left[ \frac{dy(t, \theta)}{d\theta} \right]^T = -\psi(t, \theta) \quad (II.81) \]

i.e., this derivative is actually the negative derivative of the predictor with respect to \( \theta \). The properties of this derivative have very important implications for the convergence and efficiency of the algorithm. Intuitively, it is a measure of sensitivity of the predictions to the parameters and its importance can be traced to the fact that a predictor that is insensitive to the parameter values is not of much value in identifying these parameter values.
\( \psi(t, \theta) \) can be found for the linear predictor model (II.77) from a finite dimensional filter as follows:

Define \( n(t, \theta) = [\phi^{(1)}(t, \theta) \quad \ldots \quad \phi^{(d)}(t, \theta)] \), a \( nxd \) matrix whose \( i \)th column is the derivative of the \( \phi \) vector with respect to the \( i \)th element of the parameter vector \( \theta \).

Let

\[
\frac{\partial}{\partial \theta}[F(\theta)\phi + G(\theta)z(t)] = M(\theta, \phi, z), \quad (nxd) \quad (II.82.a)
\]

\[
\frac{\partial}{\partial \theta}[H(\theta)\phi] = D(\theta, \phi) \quad (pxd) \quad (II.82.b)
\]

It should be noted that \( M(\theta, \phi, z) \) is a matrix whose \( i \)th column is given by

\[
[\frac{\partial}{\partial \theta} F(\theta)] \phi(i) + [\frac{\partial}{\partial \theta} G(\theta)]z(i)
\]

and similarly for \( D(\theta, \phi) \).

Then,

\[
n(t+1, \theta) = F(\theta) n(t+1, \theta) + M(\theta, \phi, z) \quad (II.83.a)
\]

\[
\psi^T(t+1, \theta) = H(\theta)n(t+1, \theta) + D(\theta, \phi) \quad (II.83.b)
\]

Equation (II.83) represents the filter for the prediction of the gradient \( \psi() \). For notational convenience, the filters (II.77) and (II.83) will be combined into the following partitioned structure.

\[
\xi(t, \theta) = \begin{bmatrix} \phi(t, \theta) \\ \text{Col } n(t, \theta) \end{bmatrix}
\]

where \( \text{Col } n() \) is the column vector formed by stacking the columns of \( n() \) vertically. Now, (II.77) and (II.83) can be combined as

\[
\xi(t+1, \theta) = A(\theta)\xi(t, \theta) + B(\theta)z(t) \quad (II.84.a)
\]
\[ \begin{bmatrix} y(t/\theta) \\ \text{Col } \psi(t,\hat{\theta}(t-1)) \end{bmatrix} = C(\theta)\xi(t,\theta) \]  

(II.84.b)

where \( A(\theta), B(\theta) \) and \( C(\theta) \) are matrices related to \( F(\theta), G(\theta) \) and \( H(\theta) \). For instance \( A(\theta) \) is a \((d+1)n \times (d+1)n\) square block diagonal matrix with all the \((d+1)\) block diagonal entries containing \( F(\theta) \).

We now note that in the stochastic Newton algorithm (II.82), the quantities \( V'(\theta(t-1)) \) and \( V''(\theta(t-1)) \) appear. To compute these, we need to compute \( \psi(t,\hat{\theta}(t-1)) \) and \( \xi(t,\hat{\theta}(t-1)) \). Since (II.84) is defined for a particular value of \( \theta \), use of (II.84) to compute \( \psi(t,\hat{\theta}(t-1)) \) and \( \xi(t,\hat{\theta}(t-1)) \) requires proceeding from time zero to \( t \), using \( \theta(t-1) \) in place of \( \theta \) in (II.84). Obviously, such a scheme is not truly recursive. We will note also that \( \xi(t,\hat{\theta}(t-1)) \) involves some power of \( A(\theta(t-1)) \) owing to the nature of (II.84.a). As a result, if \( A \) has eigenvalues within the unit circle (the eigenvalues of \( A \) are identical to those of \( F \), \( [A(\hat{\theta}(t-1))]^k \) tends to zero exponentially. Under these circumstances, the following approximations may be introduced. Consider

\[
\xi(t,\hat{\theta}(t-1)) = [A(\hat{\theta}(t-1))]^t \xi(o,\hat{\theta}(t-1)) + \sum_{k=0}^{t-1} [A(\hat{\theta}(t-1))]^{t-k-1} B(\hat{\theta}(t-1)) z(k)
\]

(II.85.a)

An approximate form is

\[
\xi(t) \triangleq \xi(t,\hat{\theta}(t-1)) = \sum_{k=0}^{t-1} [A(\hat{\theta}(t-1))]^{t-k-1} B(\hat{\theta}(t-1)) z(k)
\]

(II.85.b)

\( \xi(t) \) as defined in (II.85) can be computed recursively by

\[
\xi(t) = A(\hat{\theta}(t))\xi(t) + B(\hat{\theta}(t))z(t)
\]

(II.86.a)

similar arguments lead to

\[
\begin{bmatrix} y(t|\hat{\theta}(t-1)) \\ \text{Col } \psi(t,\hat{\theta}(t-1)) \end{bmatrix} \triangleq \begin{bmatrix} y(t) \\ \text{Col } \psi(t) \end{bmatrix} = C(\hat{\theta}(t-1))\xi(t)
\]

(II.86.b)

\[
\xi(t,\hat{\theta}(t-1)) = \xi(t) \triangleq y(t) - \hat{y}(t)
\]

(II.86.c)
The predictor algorithm (II.86) is truly recursive and at each time only \( \xi(t) \), \( \hat{\theta}(t) \) and \( z(t) \) need to be stored. It should be noted that in most algorithms described in Section II.3, the replacement of \( \hat{\theta} \) by \( \hat{\theta}(t-1) \) and the use of a "recursive" predictor like (II.86) needs to be formalized in terms of the above reasoning, and that the stability of the A-matrix was essential to the development of (II.86).

Based on (II.86), the equations (II.80.a) and (II.80.b) can be modified as follows:

\[
\begin{align*}
V_t'(\hat{\theta}(t-1)) &= -E[\psi(t)[\ell_{\epsilon}(t,\xi(t))]] \\
V_t''(\hat{\theta}(t-1)) &= E[\psi(t)[\ell_{\epsilon\epsilon}(t,\xi(t))\psi^T(t)] + E[\psi_{\hat{\theta}}^T(t)[\ell_{\epsilon}(t,\xi(t))]]
\end{align*}
\] (II.87.a)

(II.87.b)

where \( \psi_{\hat{\theta}}(t) \) is equivalent to \( \epsilon_{\hat{\theta}T}(t) \) and the interpretation of the last term in (II.76.b) is similar to that of the last term in (II.80.b). Where analytical expressions exist for the derivatives \( \ell_{\epsilon}() \) and \( \ell_{\epsilon\epsilon}() \), the recursive computation of (II.87.a) and (II.87.b) is straightforward. This will be illustrated by considering a quadratic criterion of the form

\[
\ell(t,\theta,\xi(t,\theta)) = \frac{1}{2} \cdot \epsilon(t,\theta)^T W^{-1} \epsilon(t,\theta)
\] (II.88)

where \( W \) is a positive definite weighting matrix (assumed independent of \( \theta \)).

Using the approximations leading to Eq. (II.87):

\[
\begin{align*}
V_t'(\hat{\theta}(t-1)) &= -E[\psi(t)W^{-1}\epsilon(t)] \\
V_t''(\hat{\theta}(t-1)) &= E[\psi(t)W^{-1}\psi^T(t) + \psi_{\hat{\theta}}^T(t)W^{-1}\epsilon(t)]
\end{align*}
\] (II.89.a)

(II.89.b)

Based on the results quoted at the end of Section II.2.3, an optimal choice for \( W \) is the true covariance matrix of the errors (if there exists a true value of the parameters). Since this matrix is unknown, a reasonable alternative is to replace it with its estimate,

\[
\hat{W}(t) = \frac{1}{t} \cdot \sum_{k=1}^{t} \epsilon(t) \epsilon^T(t)
\] (II.90.a)
which can be computed recursively, based on (see Eqs. (II.71) and (II.72))

\[ \hat{W}(t) = \hat{W}(t-1) + \gamma(t)[\epsilon(t)\epsilon_\theta^T(t) - \hat{W}(t-1)] \]  

(II.90.b)

If \( \hat{L} \) indeed depends explicitly upon \( \theta \), as in the maximum likelihood approach,

\[ V'(t, \theta) = E\left( \frac{d}{d\theta} \hat{L}(t, \theta, \epsilon(t, \theta)) \right) = \]

\[ = E\{\ell_\theta(t, \epsilon(t, \theta)) - \psi(t, \theta)\epsilon_\theta(t, \epsilon(t, \theta))\} \]  

(II.91.a)

and

\[ V''(\theta) = E\{2\theta_\theta(t, \epsilon(t, \theta)) - \ell_\theta(t, \epsilon(t, \theta))\psi_\theta(t, \theta)^T - \psi(t, \theta)\ell_\theta(t, \epsilon(t, \theta)) + \psi(t, \theta)\epsilon_\theta(t, \epsilon(t, \theta))\psi_\theta(t, \theta)^T - \psi(t, \theta)\ell_\theta(t, \epsilon(t, \theta))\} \]  

(II.91.b)

Equations (II.89.b) and (II.91.b) are true Newton directions and can be recursively computed in a manner similar to Eq. (II.72). However, considerable computation may be involved. We may therefore, attempt to replace them by some approximate positive definite alternatives. If the second term on the right hand side of (II.89.b) is neglected (see approximation (iii) following equation (II.26)), we obtain, replacing \( \hat{W}(t) \) by its estimate \( \hat{W}(t) \),

\[ R(t) = \hat{V}''(\hat{\theta}(t-1)) = E\{\psi(t)\hat{W}^{-1}(t)\psi^T(t)\} \]

It is easy to see that \( R(t) \) is indeed guaranteed positive definite. This approximate Newton direction is often referred to as Guass-Newton direction and can be computed recursively by (see eq. II.72):

\[ R(t) = R(t-1) + \gamma(t)[\psi(t)\hat{W}^{-1}(t)\psi^T(t) - R(t-1)] \]  

(II.92)
Similar approximations may be introduced in (II.91.b). Now, the recursive estimation of $\hat{\theta}(t)$ can be carried out along the lines of Equation (II.68).

The summary of the resulting algorithm for recursive identification based on the quadratic criterion (II.88) is:

Prediction Error: $e(t) = y(t) - \hat{y}(t)$  \hspace{1cm} (II.93.a)

Weighting Matrix Update:

$$\hat{W}(t) = \hat{W}(t-1) + \gamma(t) [e(t)e^T(t) - \hat{W}(t-1)]$$  \hspace{1cm} (II.93.b)

Newton Direction Update:

$$R(t) = R(t-1) + \gamma(t) [\psi(t)\hat{W}^{-1}(t)\psi^T(t) - R(t-1)]$$  \hspace{1cm} (II.93.c)

Parameter Update: $\hat{\theta}(t) = \hat{\theta}(t-1) + \gamma(t)R^{-1}(t)\psi(t)\hat{W}^{-1}(t)e(t)$  \hspace{1cm} (II.93.d)

Propagation: $\xi(t+1) = A(\hat{\theta}(t))\xi(t) + B(\hat{\theta}(t))z(t)$  \hspace{1cm} (II.93.e)

Prediction:

$$\begin{bmatrix} \hat{y}(t+1) \\ \hat{c}(t+1) \end{bmatrix} = C(\hat{\theta}(t))\xi(t+1)$$  \hspace{1cm} (II.93.f)

For a more general criterion, (II.93.b-d) need to be modified appropriately ((II.93b) may not be necessary).

Ljung and Soderstrom (1981) refer to the above algorithm as a Recursive Gauss-Newton algorithm. Defining $P(t) = \gamma(t)R^{-1}(t)$, an equivalent algorithm can be obtained that avoids the need to compute the inverses $\hat{W}^{-1}(t)$ and $R^{-1}(t)$. This is done by using the matrix inversion lemma and the resulting form is (Ljung and Soderstrom, 1981):

$$e(t) = y(t) - \hat{y}(t)$$  \hspace{1cm} (II.94.a)

$$\hat{W}(t) = \hat{W}(t-1) + \gamma(t) [e(t)e^T(t) - \hat{W}(t-1)]$$  \hspace{1cm} (II.94.b)

43
\[ S(t) = \psi^T(t)P(t-1)\psi(t) + \lambda(t)\hat{W}(t) \quad (II.94.c) \]

\[ L(t) = P(t-1)\psi(t)S^{-1}(t) \quad (II.94.d) \]

\[ \hat{\theta}(t) = \hat{\theta}(t-1) + L(t)e(t) \quad (II.94.e) \]

\[ P(t) = [P(t-1) - L(t)S(t)L^T(t)]/\lambda(t) \quad (II.94.f) \]

\[ \xi(t+1) = A(\hat{\theta}(t))\xi(t) + B(\hat{\theta}(t))\epsilon(t) \quad (II.94.g) \]

\[ \begin{bmatrix} \gamma(t+1) \\ \phi(t+1) \end{bmatrix} = C(\hat{\theta}(t))\xi(t+1) \quad (II.94.h) \]

where \( \lambda(t) = \gamma(t-1)[1-\gamma(t)]/\gamma(t) \).

In the form (II.94), the recursive Gauss-Newton algorithm for the minimization of the quadratic criterion (II.88), (or a more general criterion, appropriately replacing R(t), P(t) and L(t)) bears a strong resemblance to the RLS algorithm, (II.13). It should be emphasized also that (II.93) and (II.94) are strictly valid only for \( \hat{\theta}(t) \) such that \( A(\hat{\theta}(t)) \) is stable for all \( t \).

The form that (II.94) takes for a particular model and the complexity of the calculations involved depend upon the nature of the predictor. If the predictor is linear in the parameter vector, the model can be cast in the form of a linear regression model. Then \( \psi(t) = \phi(t) \). The resulting algorithm is in fact identical to the RLS algorithm (II.13), allowing for a more general case of vector observations, \( W(t) \) replacing \( \alpha_t \) as a weighting matrix. It is also apparent that the Instrumental Variable methods, (that replace \( \hat{\phi}(t) \), the true gradient for a linear regression model, by some other vector) are in some sense approximate gradient schemes.

In the discrete-time linear situation, the general state-space model is Eq. (II.28). For a fixed value of \( \theta \), the parameter vector, we have (assuming \( \theta \) is time invariant):
\[ x(t+1) = \phi(\theta)x(t) + G(\theta)w_d(t) \]  
\[ y(t) = H(\theta)x(t) + v(t) \]

where \( w_d() \) and \( v() \) are discrete-time white noise sequences with respective covariance matrices \( Q_d() \) and \( R() \). Now, the Kalman filter predictor corresponding to the model (II.95) takes the form

\[ \varepsilon(t,\theta) = y(t) - \hat{y}(t,\theta) \]  
\[ \hat{x}(t+1,\theta) = \Phi(\theta)\hat{x}(t,\theta) + K(t,\theta)\varepsilon(t,\theta) \]  
\[ K(t,\theta) = P(t,\theta)H^T(\theta)H(\theta)P(t,\theta)H^T(\theta) + R(t) \]^{-1}  
\[ P(t,\theta) = \Phi(\theta)P(t-1,\theta)(I - H(\theta)(H^T(\theta)P(t,\theta)H(\theta) + R(t))^{-1}H(\theta) \]  
\[ P(t-1,\theta) = \Phi^T(\theta) + G(\theta)Q_d(t)G^T(\theta) \]

\( \varepsilon(t,\theta) \), the prediction error is also known as the innovation. Now define (with \( x() \) a \( nx1 \) vector, \( \theta \) a \( d \times 1 \) vector, \( y() \) a \( px1 \) vector):

\[ U(t,\theta) = \frac{d}{d\theta} \hat{x}(t,\theta) \] (\( nxd \) matrix) \n
and

\[ D(\theta,x) = \frac{\partial}{\partial \theta} [H(\theta)x] \Big|_{\theta = \hat{\theta}} \] (\( pxd \) matrix)

then,

\[ \psi^T(t,\theta) = \frac{d}{d\theta} \hat{y}(t,\theta) = H(\theta)U(t,\theta) + D(\theta,\hat{x}(t,\theta)) \]

Differentiating (II.96.b) with respect to the parameter vector \( \theta \),

\[ U(t+1,\theta) = \frac{d}{d\theta} [\phi(\theta)\hat{x}(t,\theta) + K(t,\theta)\varepsilon(t,\theta)] \]
Defining \( M(\theta, x) = \frac{3}{\theta} [\phi(\theta)x] \),

we obtain

\[
U(t+1, \theta) = [\phi(\theta) - K(t, \theta)H(\theta)]U(t, \theta) + M(\theta, \hat{x}(t, \theta)) + \\
\frac{\theta}{3\theta} [K(t, \theta)e(t, \theta) - K(t, \theta)D(\theta, \hat{x}(t, \theta))]
\]

(II.101)

The components of the gradient of \( K(t, \theta) \) can be computed by the differentiation of (II.96.c) (and (II.96d)). The details of this differentiation can be found in Ljung and Soderstrom (1981, Appendix 3.B) and are omitted for the purpose of this discussion. Eq. (II.87), is an attempt to approximate the true covariance matrix of the prediction error. In case of the predictor (II.96), however, the covariance matrix of the prediction error is also estimated by the filter (see (II.55.f)). Thus in extending the algorithm (II.94) to the predictor (II.96), we can use \( S(t) \) in place of \( W(t) \). A detailed derivation of the Gauss-Newton algorithm for the predictor model (II.96) can be found in Ljung and Soderstrom (1981). The details are omitted here. In the above reference, an extensive comparison of a Gauss-Newton algorithm applied to (II.96) with the Extended Kalman Filter (EKF) algorithm ((II.55), which also applies to the predictor (II.96), is presented. The EKF algorithm is shown to be an approximate gradient algorithm, in that the second term in the right hand side of equation (II.101) is neglected in the development of the gradient in the EKF algorithm. In effect the coupling between the Kalman gain matrix and the parameters is neglected. While on the one hand, it offers an attractive prospect in terms of computational requirements (especially in on-line situations, since the computation of the derivatives of \( K(\theta) \) with respect to \( \theta \) involves a considerable amount of computation), on the other hand, as Ljung (1979) demonstrates, it is a reason for the possible divergence of parameter estimates generated by the EKF. Westerlund and Tyssö (1980) demonstrated that an approximation to \( \frac{\theta}{3\theta}[K(t, \theta)e(t, \theta)] \) is given by the difference

\[
M(\hat{\theta}(t), \hat{x}(t/t)) - M(\theta(t), \hat{x}(t))
\]

46
where $M()$ is as defined by (II.100) and $\hat{\chi}(t/t)$ is (see Eqs. (II.47) and (II.55.1)) the filtered state vector at time $t$. Even with the use of this approximation, the EKF still is not a true Gauss Newton algorithm.

In conclusion of this section, it should be noted that a Recursive Gauss Newton algorithm for a general Prediction error criterion can be developed based on the concepts in this section. This unified approach exposes the vast similarities between most Recursive Identification algorithms. Central to this similarity is the choice of a "search direction" based on the gradient of the prediction or some approximation thereof.

One aspect that has not been touched upon is the choice of the gain sequence $\gamma(t)$. (In the EKF algorithm, $\gamma(t)$ does not occur and $\lambda(t)$ is assumed equal to 1). This influences the rate of convergence and asymptotic properties of the Gauss Newton algorithm. Discussions on the choice of $\gamma(t)$ can be found in Ljung and Soderstrom (1981) who also demonstrate that if $\gamma(t) = 1/t$, asymptotically, the Gauss Newton algorithm that uses an exact gradient generates parameter estimates that are asymptotically efficient.

While the discussions in this section have focussed on linear predictor models, more general predictor models such as

$$\phi(t,1,0) = f(0, t, \phi(t,0), z(t)) \quad (II.102.a)$$

$$\hat{\chi}(t/0) = h(0, t, \phi(t,0)) \quad (II.102.b)$$

that are nonlinear and time varying can be considered in the same framework, with appropriate reinterpetations of the quantities involved in the general algorithm (II.93) and (II.94). The propagation and prediction steps will be, in particular, altered from (II.93.e and f). For instance, the development of a recursion analogous to (II.86.a) may involve certain ad-hoc approximations (like the replacement of $\theta$ in (II.102) by its current estimate, without rigorous justification). In addition, some form of linearization of (II.102) may be required in developing some of the equations in the estimation algorithm.

47
E. Convergence and Asymptotic Properties of Recursive Identification Algorithms.

It is the purpose of this section to briefly review some of the results related to the properties of the estimates of parameters generated by recursive identification algorithms. There are two ways to approach this issue. The first, is to perform theoretical analyses of the estimate sequence \( \hat{\theta}(t) \), generated conditional on certain assumptions about the data set \( z_t \). Since, in general, the mappings from \( z_t \) are nonlinear and time varying, a rigorous theoretical analysis can be performed only in simple special cases. The second approach, simulation, involves the application of particular algorithms to known systems and data sets (or hypothetical "rigged-up" systems) and evaluate the obtained sequence of estimates \( \hat{\theta}(t) \). However, the results of a simulation study do not have universal implications and merely reflect the characteristics of the particular algorithm applied to a particular system, conditional on particular characteristics of the data set. Simulation studies, could however, prove very useful in identifying potential difficulties in particular applications, especially for complex systems.

In the early 1970's, a plethora of literature was generated in recursive identification, without serious attempts to analyze the properties of the generated sequence of estimates. Achieving a workable recursive identification algorithm was termed a "fiddler's paradise" by Astrom and Eykhoff (1971). Some of the applications were successful, while some works reported partial success, or counter examples to prevalent notions about convergence (e.g. Nelson and Stear, 1976, Ljung et al., 1975). In the area of deterministic methods, the earliest approaches to convergence analysis (e.g. Narendra, and Kudva, 1974, a,b) were based on developing dynamic equations for the errors during the identification processed and analyzing the asymptotic stability of this system of equations. Similar approaches in the stochastic situation were developed by Kushner and Clark (1978), Moore and Ledwich (1980) and Solo (1980). These were based on advanced concepts in probability theory and have been applied only in restrictive situations. A simpler approach developed by Ljung (1977) involves the association of an ordinary differential equation (O.D.E.) with a recursive identification algorithm and studying the stability properties of the associated differential equation. This is commonly referred to as the O.D.E. approach. Ljung (1977) demonstrated that under certain
conditions the Gauss-Newton algorithm, (II.93) or (II.94), has convergence properties identical to offline identification algorithms. A detailed description of the reasoning leading to the above conclusion can be found in Ljung and Soderstrom (1981).

Ljung (1977) and Ljung and Soderstrom (1981) demonstrate that the general Gauss Newton algorithm applied to the state space model (II.95), (viz. one that includes the gradient of the Kalman gain matrix with respect to \( \hat{\theta} \), the parameter vector, in the computation of the gradient of the prediction (see Eq. (II.101)) converges to a value of \( \hat{\theta} \) that corresponds to a local minimum of the criterion function. This is not true in case of the EKF algorithm (II.55) however.

The differential equation associated with the EKF algorithm can be defined, in terms of the processes that (II.55) would generate for some constant value of \( \hat{\theta} \) (i.e., replacing \( \hat{\theta}(t) \) in (II.55) by \( \hat{\theta} \)). Ljung (1979) describes the development of this differential equation for a discrete-time version of (II.55) with \( f() \) and \( h() \) being linear functions. Based on the developed system of O.D.E.'s associated with the EKF algorithm, he demonstrated that incorrect assumptions about the noise covariance matrices \( Q \) and \( R \) lead to biased estimates and that a stability analysis of the algorithm using the O.D.E. approach does not establish convergence of the EKF algorithm. It should be noted, though, that the O.D.E. method provides only a sufficient condition and not a necessary condition for convergence. Thus failure to establish convergence characteristics, using the O.D.E. method cannot be construed as indicative of divergence, unless supported by some other means.

F. Choice of a Recursive Identification Algorithm for Application to Conceptual Models of Watershed Response

The choice of a recursive identification algorithm for a particular application depends upon various factors, such as the model structure, noise characteristics etc. The general nonlinear state space model structure (eq. (II.39)) provides a convenient means for the representation of the dynamics of watershed response. In fact most of the commonly used large scale physical-conceptual models of watershed, referred to in Chapter I, are cast in this form. Owing to the high levels of uncertainty associated with estimates of inputs such as precipitation and the representations of the dynamics, the
parameter estimation problem in watershed modelling clearly merits a stochastic treatment. The stochastic output error approach provides a convenient means of incorporating high noise levels, that are in addition nonstationary, in state-space models of dynamic systems. Such an approach has been used by Sorooshian and Dracup (1980) and Restrepo-Posada and Bras (1982) for parameter estimation of watershed response models, in offline applications. Both these approaches used maximum likelihood criterion functions, which has been cited as a very powerful approach that leads to well behaved response surfaces in the parameter space.

As was noted in Section II.3.2, practical implementation of ML based approaches is extremely complicated, particularly for on-line implementation, a factor that impedes the realization of the well defined asymptotic properties associated with them. However nonlinear filtering theory does provide an attractive framework for recursive identification based on a stochastic output error approach. The Extended Kalman Filter algorithm (II.55) is a simple nonlinear filtering algorithm that is easy to implement, because of its inherently recursive nature, and provides considerable computational simplification over the ML approach. A drawback, though, is that there is no theoretical guarantee on any desirable asymptotic properties of the parameter estimates it generates. On the other hand, it has been used extensively in parameter estimation, with a reasonable amount of success. For these reasons, the Extended Kalman Filter algorithm is used in this study. Furthermore, this study is perhaps the first instance of the application of recursive identification to large-scale watershed models. Thus, the use of a simple algorithm finds justification, in that it could provide valuable information on the prospects offered by recursive identification algorithms in this application. Even if not entirely successful, it could serve as a basis on which to consider the use of more sophisticated algorithms that could possibly lead to improvements and the cost effectiveness of such efforts.

An associated problem with the use of stochastic output error approaches, is the need to estimate noise model parameters (in case of the EKF algorithm (II.55), elements of the matrices Q and R) concurrently, in case they are not accurately known. This is pursued in chapter III, which aims at developing a general and comprehensive approach to the recursive identification of watershed response models.
III. AN APPROACH TO RECURSIVE STATE AND PARAMETER ESTIMATION
OF CONCEPTUAL WATERSHED RESPONSE MODELS

A. Introduction.

As noted in Section II.6, the use of an EKF algorithm for parameter estimation requires the concurrent estimation of the noise covariance matrices Q and R. This problem is widely referred to in literature as adaptive filtering. An obvious approach is to consider the unknown elements of these matrices as parameters to be estimated and use state augmentation schemes to include these parameters in the identification process. The computational requirements of such an approach would however be prohibitively large. An alternative approach is based on the theoretical properties of the prediction error sequence (or residuals sequence) that hold, if the "true" Q and R matrices are used. This can be viewed as a parallel estimation scheme, for which (a) the predictions are the predicted properties of the residuals sequence (assuming the current estimates of Q and R are the true estimates), (b) the observations are the properties of the actual realizations of the residuals sequence, and (c) the estimates of Q and R are updated, so as to achieve agreement between (a) and (b).

Most approaches that have been developed in control theory, apply to steady state filters (see Mehra, 1971, 1976). Owing to the highly nonstationary character of both the dynamics and the uncertainties associated with watershed response models, however, these approaches are not directly applicable. Independent approaches have been developed in the specific context of recursive state estimation for hydrologic forecasting models, that also use the EKF algorithm. Also, owing to the fact that the development in this chapter applies to a simple state estimation problem as well as to a parameter estimation problem based on state augmentation, it is set in the framework of recursive state estimation. This will facilitate comparison with prior approaches in the hydrosciences. It should be noted that accurate knowledge of the noise covariance matrices is essential even in the context of state estimation, since the optimality of the Kalman filter (and Extended Kalman Filter) algorithms holds only if the "true" Q and R matrix are used.

Consider the general nonlinear system dynamics equation
\[
\frac{dx(t)}{dt} = f(x(t), u(t), \theta(t)) + w(t)
\]  

(III.1)

that can be used to describe the flow characteristics in a soil-channel physical system and constitutes the governing system of differential equations in a conceptual hydrologic model. In this equation, \( x(t) \) is the state vector (in hydrologic models, usually the amount of water in storage in each of the reservoirs that make up the model), \( u(t) \) is the vector of inputs (in hydrologic models, usually mean areal precipitation and mean areal evapotranspiration) \( \theta(t) \) is a vector of model parameters (coefficients in the model equations) and \( w(t) \) is an additive error vector accounting for errors in the model structure, inputs and parameters.

Errors in the model structure are associated with inaccuracies in the form of the function \( f(\cdot) \) and the representation of spatially distributed processes (Involved in the catchment response) as spatially lumped ones. Errors in the model input arise because of inaccuracies in the measurement of point precipitation and other meteorological variables (e.g. air temperature), the spatial interpolation of point values for the computation of mean areal values (e.g. Bras and Iturbe, 1985, pg. 372) and discrete measurements of input variables in time (and consequent assumptions that input variables remain constant over certain durations, usually a time step used in the integration of the system equations). Errors in the model parameters arise from inaccuracies in (manual or automatic) parameter estimation procedures.

When conceptual models such as (III.1) are used in real-time hydrologic forecasting, a common practice (Kitanidis and Bras, 1980a,b, Georgakakos, 1986a,b) has been to use a recursive estimator of the system states in order to improve the quality of the forecasts. The most commonly used recursive estimators for this purpose are based on nonlinear filtering theory and are variants of the Extended Kalman Filter (EKF) algorithm, which is summarized below for a general nonlinear time-continuous dynamic system (repeated here for sake of continuity).

**System Model:**

**Dynamics:**

\[
\frac{dx(t)}{dt} = f(x(t), u(t), \theta(t)) + w(t)
\]

(III.1)

**Observations:**

\[
z(t_k) = h(x(t_k), u(t_k), \theta(t_k)) + v(t_k)
\]

(III.2)
Statistical assumptions: \( \{w(t)\} \) - zero mean continuous-time Gaussian white noise process with covariance parameter matrix \( Q(t) \),

\( \{v(t_k)\} \) - zero mean discrete-time Gaussian white noise sequence with covariance matrix \( R(t_k) \),

\( x(o) \) - Gaussian, mean \( \hat{x}(o) \), covariance matrix \( P(o) \),

\( w(t), v(t_k), x(o) \) - are mutually independent for all \( t, t_k \).

Propagation Equations:

State mean:

\[
\frac{dx(t)}{dt} = f(x(t), u(t), \theta(t)) \quad (III.3)
\]

State Covariance:

\[
\frac{dP(t)}{dt} = F_x(t) P(t) + P(t) F_x^T(t) + Q(t) \quad (III.4)
\]

Update Equations:

Gain Matrix:

\[
K(t_k) = P^-(t_k) H_x^T(t_k) \left[ H_x(t_k) P^-(t_k) H_x^T(t_k) + R(t_k) \right]^{-1} \quad (III.5)
\]

State mean Update:

\[
\hat{x}^+(t_k) = \hat{x}^-(t_k) + K(t_k) \left[ z(t_k) - h(\hat{x}^-(t_k), u(t_k), \theta(t_k)) \right] \quad (III.6)
\]

State Covariance Update:
\[ P^+(t_k) = [I - K(t_k)H_x(t_k)] P^-(t_k) \]

\[ [I - K(t_k)H_x(t_k)]^T + K(t_k)R(t_k)K^T(t_k) \]  

(III.7)

where \( F_x(t) = \frac{\partial f}{\partial x} \bigg|_{x(t)} \), \( H_x(t_k) = \frac{\partial h}{\partial x} \bigg|_{x(t_k)} \), the superscript "−" indicates the value of a quantity at the end of the propagation step and just prior to update, and the superscript "+" denotes the value of a quantity after update.

The EKF algorithm is essentially an optimum linear estimator based on a linearization of (III.1) about the most recent estimate, \( \hat{x}(t) \), of the state. If \( Q(t) \), \( R(t_k) \) and system dynamics are known exactly, the gain \( K(t_k) \) in equation (III.5) becomes the optimal gain matrix. In this case, (III.7) reduces to (Gelb, 1974):

\[ P^+(t_k) = [I - K(t_k)H_x(t_k)] P^-(t_k) \]  

(III.8)

If \( Q(t) \) and \( R(t_k) \) are not exactly known, however, the estimates provided by the EKF algorithm are no longer optimal. In absence of satisfactory a-priori knowledge about \( Q(t) \) and \( R(t_k) \), these can be viewed as parameters of the state estimator, that need to be estimated, in addition to the states themselves. Prior approaches to this problem in the hydrosciences include those of Kitanidis and Bras (1978), Todini (1978), Georgakakos (1984) and Fuente and Bras (1985). All of these approaches are based on the statistical properties of the sequence of filter residuals, defined by

\[ \nu(t_k) = z(t_k) - \hat{h}(\hat{x}(t_k), u(t_k), \theta(t_k)) \]  

(III.9)

In general, for a system with \( n \) states, there are \( n^2 \) elements in the \( Q \) matrix and, if there are \( m \) elements in the observations vector, there are \( m^2 \) elements in the \( R \) matrix. Since they are both symmetric, the number of parameters to be estimated is: \( n(n+1)/2 + m(m+1)/2 \). For typical conceptual hydrologic models like the National Weather Service Soil Moisture Accounting model (Peck, 1974), (6 states, 1 observation), estimation of all these parameters (here 22) involves a prohibitive computational burden. A simplification that has commonly been resorted to is to assume a diagonal \( Q \) matrix that is con-
stant in time (Kitanidis and Bras, 1978, Puente and Bras, 1987), and a known R matrix.

If the R matrix is assumed known, then the problem of estimating the parameters of the state estimator reduces to one of estimating the elements of the Q matrix of some assumed form. In the following, an approach to this problem is developed based on an explicit parameterization of \( \bar{w}(t) \), the model error term in equation (III.1), in terms of errors in the input and parameters. In contrast to some of the earlier approaches in the hydrosciences, this approach allows for incorporation of nonstationary uncertainty levels in the estimator and of prior knowledge on parameter and input uncertainty. In addition, an attempt is made to arrive at a parameterization of the Q matrix so that the number of parameters to be estimated is kept at the minimum.

**B. Adaptive Filtering Based on Explicit Parameterizations of Input and Parameter Uncertainty**

1. **Formulation** Model structure errors are difficult to quantify in most cases (especially for hydrologic models) because of the lack of knowledge or lack of data to support more accurate models. In addition, it is expected that structure errors should be associated with a bias rather than with a zero mean random error. (For instance in hydrologic models where a linear interflow response is assumed, whereas the interflow response during periods of high flow is in fact nonlinear, a consistent bias may be observed during high flow periods). For these reasons, in this development, model structure errors are not considered. Owing to the rather high degrees of uncertainty that can be associated with the input estimates, it is perhaps reasonable to assume that structure errors are relatively insignificant.

The inputs may be expressed as the sum of a mean value (either predicted by a model or measured) and a zero mean error term, as follows,

\[
\bar{u}(t) = \hat{\bar{u}}(t) + e_u(t)
\]

(III.10)

The parameters \( \bar{\theta}(t) \) may be expressed in one of the following ways:

\[
\bar{\theta}(t) = \hat{\bar{\theta}} + e_\theta
\]

(III.11)
i.e. $\theta(t) \equiv \theta$, is a time-invariant "true" parameter vector corresponding to the chosen model structure, $\hat{\theta}$ is a time invariant estimate of $\theta$ obtained from a parameter estimation exercise, and $e_\theta$ is a zero mean, random, estimation-error term.

b)

$$\hat{\theta}(t) = \hat{\theta} + e_\theta(t) \quad (III.12)$$

i.e. $\theta(t)$ is assumed to be time varying, $\hat{\theta}$ is a time invariant estimate of $\theta(t)$, based on assuming that $\theta(t)$ is time invariant and $e_\theta(t)$ is a time varying error term accounting for these assumptions. $e_\theta(t)$ can be interpreted as a structural error term accounting for assuming a time invariant $\theta$. In the discussions to follow the expression (III.11) will be used rather than (III.12), based on the fact that structural uncertainty is difficult to quantify.

Similarly the system state may be expressed as

$$x(t) = \hat{x}(t) + e_x(t) \quad (III.13)$$

In the equations (III.11, 12, and 13), the symbol "\~" denotes mean value and $e_u(t)$, $e_\theta$ and $e_x(t)$ represent zero mean errors in the input, parameters and state respectively.

If the system dynamics function $f(\cdot)$ is linearized about $(\hat{x}(t), \hat{u}(t), \hat{\theta})$, we get

$$\frac{dx(t)}{dt} = \hat{f} + \left\{ \frac{\partial f}{\partial x} \right\} (x(t) - \hat{x}(t)) + \left\{ \frac{\partial f}{\partial u} \right\} (u(t) - \hat{u}(t)) + \left\{ \frac{\partial f}{\partial \theta} \right\} (\theta - \hat{\theta}) + e_L(t) \quad (III.14)$$

where

$$\hat{f} = f(\hat{x}(t), \hat{u}(t), \hat{\theta}) \quad (III.15)$$

$$\left\{ \frac{\partial f}{\partial x} \right\} = \left. \frac{\partial f(x(t), u(t), \theta)}{\partial x} \right|_{\hat{x}(t), \hat{u}(t), \hat{\theta}} \quad (III.16)$$
\[
\hat{f}_{\hat{u}} = \frac{\hat{f}(x(t), u(t), \theta)}{\hat{u}} \quad \hat{x}(t), \hat{u}(t), \hat{\theta} \tag{III.17}
\]
\[
\hat{f}_{\hat{\theta}} = \frac{\hat{f}(x(t), u(t), \theta)}{\hat{\theta}} \quad \hat{x}(t), \hat{u}(t), \hat{\theta} \tag{III.18}
\]

and \(e_L(t)\) denotes the linearization errors. If the vectors \(\hat{x}(\cdot), \hat{u}(\cdot)\) and \(\hat{\theta}\) are respectively assumed to have dimensions \(n, k\) and \(l\), \(\{\frac{\partial f}{\partial x}\}\) is a \(nxn\) matrix, \(\{\frac{\partial f}{\partial u}\}\) is a \(nxk\) matrix and \(\{\frac{\partial f}{\partial \theta}\}\) is a \(nxl\) matrix.

Define

\[
F_x(t) = \frac{\hat{f}}{\hat{x}}, F_u(t) = \frac{\hat{f}}{\hat{u}}, F_\theta(t) = \frac{\hat{f}}{\hat{\theta}}
\]

Let

\[
P_{\theta \theta} = E(e_\theta e_\theta^T) \tag{III.19}
\]
\[
P_{xx}(t) = E(e_x(t)e_x^T(t)) \tag{III.20}
\]

where \(E()\) is the expectation operator and superscript "\(T\)" denotes the transpose of a matrix or vector. Then assuming that the linearization error \(e_L(t)\) is negligible in comparison to \(e_u(t)\) and \(e_\theta(t)\), and that the process \(e_u(t)\) is a zero mean white noise process with Covariance parameter matrix \(P_{uu}(t)\), the differential equation for the state error takes the form

\[
\frac{de_x(t)}{dt} = F_x(t)e_x(t) + F_u(t)e_u(t) + F_\theta(t)e_\theta \tag{III.21}
\]

The corresponding differential equation for the state-error covariance matrix \(P_{xx}(t)\) can be shown to take the following form (see derivation in Appendix A).
\[
\frac{dP_{xx}(t)}{dt} = F_x(t)P_{xx}(t) + P_{xx}(t)F_x^T(t) + \alpha_p[F_\theta(t)P_{x\theta}(t) + P_{x\theta}(t)F_\theta^T(t)] \\
+ \alpha_u[F_u(t)P_{uu}(t)F_u^T(t)]
\]  

(III.22)

where

\[
P_{x\theta}(t) = E\{e_x(t)e_\theta^T\}
\]  

(III.23)

and \(\alpha_u\) is a factor with dimensions [time], to account for errors in \(P_{uu}\) and due to temporal averaging of inputs (common in hydrologic models where inputs are assumed constant over time steps of integration of the model equations) and \(\alpha_p\) is a dimensionless factor accounting for the errors in the parameter covariance matrix \(P_{\theta\theta}\) (see Appendix A).

Also derived in Appendix A is the differential equation for the parameter-state cross-covariance matrix \(P_{x\theta}\), which is

\[
\frac{dP_{x\theta}(t)}{dt} = F_x(t)P_{x\theta}(t) + F_\theta(t)P_{\theta\theta}
\]  

(III.24)

The problem of estimation of the state-estimator parameters is now reduced to one of estimating \(\alpha_u\) and \(\alpha_p\). There are many advantages that this approach offers. Firstly, it allows for the incorporation of prior knowledge about input and parameter uncertainty in the model error term, which is indeed consistent with what this term represents physically. As a consequence it allows for a representation of the nonstationarity of \(w(t)\) in a manner that is consistent with the dynamics of the system. In case the prior information about input and parameter uncertainty involve (possibly inaccurate) degree-of-belief estimates, \(\alpha_u\) and \(\alpha_p\) are appropriately estimated to correct for the inaccuracies in these. Secondly, it allows for considerable computational
expediency, since only two parameters need to be estimated. Thirdly, it does not necessitate the simplifying assumption of a time-invariant diagonal $Q$ matrix and in fact captures the correlations between the elements of the error vector in an intuitively very appealing fashion (in that these correlations can be attributed to the fact that the same inputs and parameters enter the dynamic equations for different elements of the state vector).

Of particular significance in hydrologic systems which exhibit a high degree of nonstationarity in their response, is the fact that the non-stationarity of the $Q$ matrix is consistent with the system dynamics. For instance, the system response may be insensitive to certain inputs under certain conditions and the corresponding input uncertainties will not be weighted very much since the $F_u$ matrix would contain "small" terms corresponding to the inputs in question. If a time varying $\hat{\theta}$-matrix is to be used, without an explicit parameterization such as the one used above, simultaneous estimation of $\hat{\theta}$-matrix elements would be required at all times, in addition to state estimation. However, when the above parameterization is used, $\alpha_u$ and $\alpha_p$ estimated using a certain period of record may be used in the future for state estimation. This is another advantage of the parameterization used above. Of course, $\alpha_u$ and $\alpha_p$ may need to be periodically reestimated, to account for long term enhancements in input sensor networks and possibly long-term climatic trends.

2. Adaptive Estimation of State Estimator Parameters The estimation of $\alpha_u$ and $\alpha_p$ may be performed in an off-line or on-line fashion. The criteria used to estimate $\alpha_u$ and $\alpha_p$ depend upon the statistical properties of the sequence of filter residuals defined in Equation (III.9). For the optimum state estimator, this sequence is zero mean discrete-time white noise, with time varying covariance matrix given by
\[ E \left[ v(t_k) v^T(t_k) \right] = H_x(t_k) P_{xx}(t_k) H_x^T(t_k) + R(t_k) \]  

(III.25)

Following the approach of Kitanidis and Bras (1978), an on-line estimation procedure based on a stochastic approximations methodology will now be derived to estimate \( a_u \) and \( a_p \). As noted in Chapter II, stochastic approximations methodology represents a general approach to parameter estimation that is not optimal in any statistical sense, but yields estimates with well defined convergence properties. Consider the following general scalar observations model

\[ z(k) = f(y(k), a) + \xi(k) \]  

(III.26)

where \( y(k) \) is a known vector, \( a \) is the parameter vector to be estimated and \( \xi(k) \) is a random disturbance term. Use of the Robbins–Monro scheme (II.72) results in an updating equation of the form

\[ \hat{a}(k) = \hat{a}(k-1) + \rho(k) [z(k) - f(y(k), \hat{a}(k-1))] \]  

(III.27)

where \( \hat{a}(k) \) is the estimate after the \( k \)th observation is processed and \( \rho(k) \) is a gain vector. Linearizing the function \( f(\cdot) \) about the latest estimate \( \hat{a}(k-1) \), a minimum mean square error linear estimator of \( a \), \( \hat{a}(k) \), can be derived given the observation \( z(k) \), of the form

\[ \hat{a}(k) = \hat{a}(k-1) + \frac{S(k-1) \hat{a}^T(k-1)}{G(k-1) S(k-1) G^T(k-1) + \sigma^2 \xi^2(k)} [z(k) - f(y(k), \hat{a}(k-1))] \]  

(III.28)
where \( S(k-1) \) is the estimation error variance of \( \hat{\alpha}(k-1) \), \( \sigma^2_\xi(k) \) is the variance of \( \xi(k) \) and

\[
\mathbf{g}(k-1) = \frac{\partial f}{\partial \alpha} \bigg|_{\hat{\alpha}(k-1)}
\]

is the matrix of derivatives of the function \( f \) with respect to the parameter vector \( \alpha \) evaluated at \( \hat{\alpha}(k-1) \). The recursive equation for the estimation error variance is

\[
S(k) = S(k-1) - \frac{S(k-1)\mathbf{g}^T(k-1)\mathbf{g}(k-1)S(k-1)}{\mathbf{g}(k-1)S(k-1)\mathbf{g}^T(k-1) + \sigma^2_\xi(k)}
\]  

(III.30)

It should be noted that in Eqs. (III.4) and (III.6), \( \mathbf{g}(k-1)S(k-1)\mathbf{g}^T(k-1) \) is a scalar.

For the problem at hand (viz. that of estimating \( \alpha = [\alpha_u \, \alpha_p]^T \)), \( \nu(t_k)\nu^T(t_k) \) is the observation (observed variance of filter residuals, assuming zero mean filter residuals) and \( [H_x(t_k)P_{xx}^{-1}(t_k)H_x^T(t_k) + R(t_k)] \) is the prediction, replacing the function \( f(*) \) in (III.26). Kitanidis and Bras (1978) show that the corresponding \( \xi(k) \) in (III.26) has variance

\[
\sigma^2_\xi(t_k) = 2[H_x(t_k)P_{xx}^{-1}(t_k)H_x^T(t_k) + R(t_k)]^2 = 2c_o^2
\]

(III.31)

in the scalar observations case. The corresponding derivatives \( \mathbf{g}(*) \) for the problem at hand are

\[
\mathbf{g}(k-1) = H_x(t_k)\frac{\partial P_{xx}^{-1}(t_k)}{\partial \alpha}H_x^T(t_k)
\]

(III.32)
since $R(*)$ and $H_x(*)$ are independent of $\alpha$. It should be noted that in addition to enforcing compatibility between predicted and observed variances of residuals, compatibility of lag-1 and higher autocovariances could be enforced. This leads to more computation, since it requires additional recursions to predict autocovariances.

Thus, in order to implement a recursive scheme for the estimation of $\alpha$, it is necessary to develop equations for the propagation and update of $\frac{\partial P_{xx}(t)}{\partial \alpha}$ which is a tensor of dimensions (n x n x 2). $\frac{\partial P_{xx}(t)}{\partial \alpha}$ and $\frac{\partial P_{xx}(t)}{\partial \alpha}$ are each n x n matrices. The equations for the propagation of these two matrices are derived next.

Consider Equation (III.22). Differentiating with respect to $\alpha_u$ and changing the order of differentiation on the left hand side leads to

$$\frac{d}{dt}(\frac{\partial P_{xx}(t)}{\partial \alpha_u}) = F_x(t) \frac{\partial P_{xx}(t)}{\partial \alpha_u} + \frac{\partial P_{xx}(t)}{\partial \alpha_u} \cdot P_x(t) + P_u(t)P_{uu}(t)F_u(t)$$ (III.33)

Similarly, we can derive

$$\frac{d}{dt}(\frac{\partial P_{xx}(t)}{\partial \alpha_p}) = F_x(t) \frac{\partial P_{xx}(t)}{\partial \alpha_p} + \frac{\partial P_{xx}(t)}{\partial \alpha_p} \cdot F_x(t) + [F_\theta(t)P_{\theta\theta}(t) + P_{\theta\theta}(t)F_\theta(t)]^T$$ (III.34)

Equations (III.33 and 34) (together with (III.24)) constitute the propagation equations for the derivative matrices. Before the update equations for these
are derived, the update equations involved in the state estimator algorithm will be reviewed. The equations for the gain matrix, the state mean update and state covariance update are given by, respectively, (see also (III.5, 6 and 7)),

\[
K(t_k) = \frac{P_{xx}(t_k)}{x_k} H^T(t_k) \left[ H_x(t_k) \frac{P_{xx}(t_k)}{x_k} H_x(t_k) + R(t_k) \right]^{-1}
\]

(III.35)

\[
\hat{x}^+_k = \hat{x}^-_k + K(t_k) [z(t_k) - h(\hat{x}^-_k, \hat{u}(t_k), \hat{\theta})]
\]

(III.36)

\[
P_{xx}^+(t_k) = [I - K(t_k) H_x(t_k)] P_{xx}^-(t_k) [I - K(t_k) H_x(t_k)]^T + K(t_k) R(t_k) K^T(t_k)
\]

(III.37)

In addition, an update equation for \(P_x\theta\) is necessary. This will be derived as follows:

Define

\[
P_{x\theta}^-(t_k) = E\{[x(t_k) - \hat{x}^-_k] \theta^T\}
\]

(III.38)

\[
P_{x\theta}^+(t_k) = E\{[x(t_k) - \hat{x}^+_k] \theta^T\}
\]

(III.39)

Linearization of (III.2) about \(\hat{x}^-_k\), \(\hat{u}(t_k)\) and \(\hat{\theta}\), results in,

\[
z(t_k) = h(\hat{x}^-_k, \hat{u}(t_k), \hat{\theta}) + H_{x_k}(t_k) [x(t_k) - \hat{x}^-_k] + H_{u_k}(t_k) \theta + H_{\theta_k}(t_k) e + v(t_k)
\]

(III.40)

Substituting in Equation (III.36) and simplifying, yields

\[
\hat{x}^+_k = \hat{x}^-_k + K(t_k) [H_x(t_k) [x(t_k) - \hat{x}^-_k] + H_{u_k}(t_k) \theta + H_{\theta_k}(t_k) e + v(t_k)]
\]

(III.41)

where

\[
H_{u_k}(t_k) = \left. \frac{\partial h(x(t_k), u(t_k), \theta)}{\partial u} \right|_{\hat{x}^-_k(t_k), \hat{u}(t_k), \hat{\theta}}
\]

(III.42)
and

$$H_\theta(t_k) = \frac{\partial h(x(t_k), u(t_k), \theta)}{\partial \theta} \big|_{\bar{x}(t_k), \bar{u}(t_k), \bar{v}}$$ (III.43)

substituting $\bar{x}(t_k)$ from (III.41) in (III.39) yields (noting that $e_u(t)$ is a white noise process and $v(t_k)$ is a white noise sequence),

$$P^+_x(t_k) = [I - K(t_k)H_x(t_k)] P^-_x(t_k) - K(t_k)H_\theta(t_k)P_\theta$$ (III.44)

The update equations for the derivative matrices $\frac{\partial P}{\partial u}$ and $\frac{\partial P}{\partial p}$ can be derived, by differentiating equation (III.37) with respect to $a_u$ and $a_p$ respectively. A detailed derivation is presented in Appendix B. The final expression can be written in the following form for the scalar observations case

$$\frac{\partial P^+_{xx}(t_k)}{\partial a} = \frac{\partial P^-_{xx}(t_k)}{\partial a} + (3 - \frac{2r}{S_c}) \frac{h}{S_c} p^-_{xx}(t_k) \frac{\partial P^-_{xx}(t_k)}{\partial a} \frac{\partial P^-_{xx}(t_k)}{\partial a} +$$

$$\frac{h^2}{S_c} \left[ \frac{\partial P^-_{xx}(t_k)}{\partial a} p^-_{xx}(t_k) + p^-_{xx}(t_k) \frac{\partial P^-_{xx}(t_k)}{\partial a} \right]$$

$$\frac{h^6}{S_c^3} \left[ p^-_{xx}(t_k) p^-_{xx}(t_k) \frac{\partial P^-_{xx}(t_k)}{\partial a} \frac{\partial P^-_{xx}(t_k)}{\partial a} p^-_{xx}(t_k) + p^-_{xx}(t_k) p^-_{xx}(t_k) \frac{\partial P^-_{xx}(t_k)}{\partial a} \right] +$$

$$\frac{h^4}{S_c^2} \left[ p^-_{xx}(t_k) p^-_{xx}(t_k) \frac{\partial P^-_{xx}(t_k)}{\partial a} \frac{\partial P^-_{xx}(t_k)}{\partial a} p^-_{xx}(t_k) \right] +$$

64
\[
\frac{\partial P_{xx}^{-}(t_k)}{\partial \alpha} = [I + P_{xx}^{-}(t_k)H_x(t_k)R^{-1}(t_k)H_x(t_k)]^{-1} \frac{\partial P_{xx}^{-}(t_k)}{\partial \alpha}
\]

\[
[I + H_x^T(t_k)R^{-1}(t_k)H_x(t_k)P_{xx}^{-}(t_k)]^{-1}
\]

Along the lines of equations (III.28) through (III.32), the stochastic approximations algorithm for the estimation of \( \alpha = [\alpha_u \alpha_p]^T \) leads to
\[ g(k-1) = H_x(t_k) \frac{\partial p_{xx}(t_k)}{\partial \alpha} H_x^T(t_k) = \begin{bmatrix} \frac{\partial p_{xx}(t_k)}{\partial u} H_x(t_k) \\ \frac{\partial p_{xx}(t_k)}{\partial p} H_x(t_k) \end{bmatrix} \]

(III.48)

\[ \rho(k) = \frac{S(k-1) g^T(k-1)}{g(k-1) S(k-1) g^T(k-1) + \sigma^2_\xi(k)} \]  

(III.49)

for the stochastic approximations gain,

\[ \hat{\alpha}(k) = \hat{\alpha}(k-1) + \rho(k) [v^2(t_k) - S_c(t_k)] \]  

(III.50)

for the recursive estimator of \( \alpha \), and

\[ S(k) = S(k-1) - \frac{S(k-1) g^T(k-1) g(k-1) S(k-1)}{g(k-1) S(k-1) g^T(k-1) + \sigma^2_\xi(k)} \]  

(III.51)

for the estimation error covariance of \( \alpha \).

3. Summary In the foregoing, a general approach has been developed for the estimation of state estimator parameters for conceptual hydrologic models based on a parameterization of model errors in terms of input and parameter errors. A summary of the equations involved in the complete recursive estimation algorithm for the states and state estimator parameters is presented in Appendix C.
Though not of direct relevance, a certain interesting variant of this approach results under the assumption that equation (III.12) holds with \( e_\theta(t) \) being a white-noise Gaussian process. The state covariance equation that results is

\[
\frac{dP_{xx}(t)}{dt} = F_x(t)P_{xx}(t) + P_{xx}(t)F_x^T(t) + \alpha_p [F_\theta(t)P_{\theta\theta}F_\theta^T(t)] + \\
\alpha_u [F_u(t)P_{uu}(t)F_u^T(t)]
\]

(III.52)

There is no need to carry \( P_{x\theta} \) in this case, since it is identically zero, a factor which could reduce the computational needs significantly.

If in equations (III.22) and (III.24) the following approximation is used,

\[
\frac{dP_{x\theta}(t)}{dt} = 0, \quad t \in [t_{k-1}, t_k]
\]

(III.53)

i.e. \( P_{x\theta} \) remains constant during a time step,

\[
P_{x\theta} = \bar{F}_x^{-1}(t)F_\theta(t)P_{\theta\theta}
\]

(III.54)

is the solution to (III.24). Substituting in (III.22) yields

\[
\frac{dP_{xx}(t)}{dt} = F_x(t)P_{xx}(t) + P_{xx}(t)F_x^T(t) + \alpha_p [\bar{F}_x^{-1}(t)F_\theta(t)]P_{\theta\theta}F_\theta^T(t) + \\
F_\theta(t)P_{\theta\theta}[\bar{F}_x^{-1}(t)F_\theta(t)]^T + \alpha_u [F_u(t)P_{uu}(t)F_u^T(t)]
\]

(III.55)
This equation may be used in case the implementation of the propagation update of \( P_{\theta \theta} \) adds significantly to the computational burden. The term \( [F_x^{-1}(t)F_\theta(t)] \) in equation (III.55) can be interpreted as the relative sensitivity of the system dynamics to the parameters as compared to the sensitivity to the states. It can be noted that as \( F_x(\cdot)^* \) "increases", (i.e. the system becomes more sensitive to the states) with \( F_\theta(\cdot)^* \) remaining constant, the state uncertainty is weighted more and the parameter uncertainty is weighted less.

From the implementation point-of-view, the following initial conditions can be used:

\[
\frac{\partial P_{x x}(o)}{\partial a_u} = F_u(o)P_{u u}(o)F_u^T(o) \quad (\text{III.56})
\]

\[
\frac{\partial P_{x x}(o)}{\partial a_p} = F_\theta P_{x \theta}(o) + P_{x \theta}(o)F_\theta^T(o) \quad (\text{III.57})
\]

\[
\alpha_{u}(o) = C \Delta t \quad (\text{III.58})
\]

\[
\alpha_{p}(o) = C' \quad (\text{III.58})
\]

where \( C \) and \( C' \) are appropriately chosen constants and \( \Delta t \) is the time step of integration.

\[
S(o) = \begin{bmatrix}
\alpha_{u}^2(o) & 0 \\
0 & \alpha_{p}^2(o)
\end{bmatrix} \quad (\text{III.59})
\]

The extension of the approach developed in Sections III.2.1 and III.2.2. to the parameter estimation problem is through state-augmentation. The above development holds with the reinterpretation of \( \theta \) as the vector of "known" parameters or those inferrable from other sources of data and the inclusion of the parameters requiring estimation in the state vector \( x \).

It should also be noted that other estimation schemes could be developed within the same framework, with appropriate modifications. For instance, rather than \( \alpha_u \) and \( \alpha_p \), one could estimate some components of the input and parameter covariance matrix. However, if there are too many such parameters, there is a drastic increase in computation (at least about \( n(n+1)/2 \) additional recursions for every additional parameter).
IV. A CASE STUDY: ESTIMATION OF PARAMETERS OF THE HYDROLOGY SUBMODEL OF THE ENHANCED TRICKLE DOWN (ETD) MODEL

A. Introduction: The Enhanced Trickle Down (ETD) Model

The ETD model is a conceptual model designed to evaluate the response of lake-watershed systems to acid deposition. A steady state version of the model was first reported by Stumm et al. (1983), with the name "Trickle-Down model". A time varying version of the model was developed by Schnoor et al. (1984) and an enhanced time varying version by Nikolaidis et al. (1986).

In response to wet and dry loadings of acidity in a watershed, the soils and bedrock undergo weathering reactions to produce alkalinity, which neutralizes the acidity to some extent. The chemical weathering rate in a watershed is determined by its geochemistry and hydrology. The hydrology determines the flow rates and detention times in various zones of the watershed. The ETD model is a compartmentalized model which simulates the hydrology and the transport and reactions of three chemical constituents—alkalinity, sulfate and chloride, in a watershed. It consists of four submodels, the hydrology submodel (corresponding to conservation of water) and three other submodels, one corresponding to the conservation of each of the three chemical constituents listed above.

The terrestrial portion of the watershed is divided into three compartments—the upper zone soil (soil compartment), lower zone soil (unsaturated zone compartment) and the saturated zone (groundwater compartment). This accounts on the average for the fact that the chemical properties of the water flowing through and the predominant chemical reactions in each of these zones are significantly different. There are three additional compartments, the atmosphere compartment, the lake compartment and the snow compartment. The organization of the rest of this chapter is as follows. In Section IV.2, the hydrology submodel of the ETD model is briefly described. In Section IV.3 the data requirements for hydrologic simulation using the ETD model are discussed. In Section IV.4, certain modifications to the form of the equations, required to facilitate use of nonlinear filtering techniques, are discussed. Sources of uncertainty in the data and their incorporation in the parameter estimation framework are discussed in Section IV.5. In Section IV.6, a description of
the study area (the Woods Lake watershed in Adirondack Park, New York) is presented, followed by a description of the available data and discussions on the uncertainty representations used. In Section IV.7, a stepwise procedure for the estimation of parameters of the hydrology submodel is outlined and in Section IV.8 a brief description of the computer programs used in this study is presented.

B. Description of the Hydrology Submodel

The hydrology submodel simulates the flow of water in and out of each compartment. The processes it accounts for are snow accumulation and ablation, interflow, infiltration, overland flow, lake-groundwater seepage flow, frozen ground driven processes, evapotranspiration and streamflow out of the watershed system. It consists of the system of differential equations based on mass balance in each compartment. A schematic representation of the hydrology submodel is presented in Figure 1. The differential equations involved in the model are stated in Table I and a listing of input and output variables and parameters of the model are presented in Table II. Explicit expressions for the flows used in the original version of the model are described in Nikolaidis et al. (1986). The modified form of these equations, used in this study is discussed in Section IV.4.

The state variables of the model are the equivalent depths of water in storage in each compartment. The equivalent depth (m, metres) is defined as the physical volume in storage (m³) divided by the area of the watershed (m²) and is a normalized measure of the storage in each compartment. All flows (m³/day) are also similarly normalized to (m/day), dividing by the watershed area. Such a normalization provides for ease of comparison of dominant flow paths between watersheds of different areal extents and sizes. An extensive description of the model formulation can be found in Nikolaidis et al. (1986). In this discussion, however, only the physical hypotheses that form the basis of various environmental processes are examined. The original version of the model included sublimation at the snow surface as an additional outflow but this process is not included in this study.

1. Snow Compartment The snow compartment extends over both the terrestrial and aquatic portions of the watershed. When snowcover exists, it is
Figure 1. Schematic representation of the hydrology submodel of the ETD model.
Table I
Hydrologic Model Equations

\textbf{IF TEMPC < 0.0}

\textbf{Snow:} \quad \frac{dh(2)}{dt} = Q(2,1) - Q(1,2)

\textbf{Soil:} \quad \frac{dh(3)}{dt} = -Q(5,3) - Q(4,3)

\textbf{Surface Water Body:} \quad \frac{dh(5)}{dt} = Q(5,4) + Q(5,3) + Q(5,6) - Q(1,5) - Q(6,5) - Q\text{OUT}

\textbf{IF TEMPC > 0.0}

\textbf{Snow:} \quad \frac{dh(2)}{dt} = -Q(1,2) - Q(3,2) - Q(5,2)

\textbf{Soil:} \quad \frac{dh(3)}{dt} = Q(3,1) + Q(3,2) - Q(1,3) - Q(4,3) - Q(5,3) - OV\text{LAND}

\textbf{Surface Water Body:} \quad \frac{dh(5)}{dt} = Q(5,1) + Q(5,2) + Q(5,3) + Q(5,4) + OV\text{LAND} + Q(5,6) - Q(6,5) - Q\text{OUT} - Q(1,5)

\textbf{TEMPC INDEPENDENT}

\textbf{Unsaturated Zone:} \quad \frac{dh(4)}{dt} = Q(4,3) - Q(5,4) - Q(6,4) - Q(1,4)

\textbf{Groundwater:} \quad \frac{dh(6)}{dt} = Q(6,4) + Q(6,5) - Q\text{NET} - Q(5,6) - Q(1,6)

\textbf{Note:} In the above equations, h(i) denotes the equivalent water depth in storage in the \textit{i}th compartment and Q(i,j) denotes flow from the \textit{j}th compartment to the \textit{i}th compartment.
### Table II

Listing of Input, Output Variables and Parameters of the Hydrologic Model

#### Input Variables

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Units</th>
<th>Need to Estimate?</th>
</tr>
</thead>
<tbody>
<tr>
<td>PREC</td>
<td>precipitation intensity in mm/day</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEMPC</td>
<td>air temperature in degrees C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QEVAP</td>
<td>pan evaporation in mm/day</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Output Variables

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Units</th>
<th>Need to Estimate?</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAGE</td>
<td>compute lake stage m</td>
<td>meter</td>
<td></td>
</tr>
<tr>
<td>QOUT</td>
<td>computed surface water export from the lake m/day</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Measured Variables

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Units</th>
<th>Need to Estimate?</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAGE</td>
<td>lake stage elevation, m</td>
<td>meter</td>
<td></td>
</tr>
<tr>
<td>WTABLE</td>
<td>water table elevation, m</td>
<td>meter</td>
<td></td>
</tr>
<tr>
<td>QOUT</td>
<td>surface water export from lake, m/day</td>
<td>meter/day</td>
<td></td>
</tr>
</tbody>
</table>

#### Parameters

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Units</th>
<th>Need to Estimate?</th>
</tr>
</thead>
<tbody>
<tr>
<td>AREAT</td>
<td>terrestrial area of watershed</td>
<td>m²</td>
<td>X*</td>
</tr>
<tr>
<td>AREAA</td>
<td>aquatic area of watershed (lake surface area)</td>
<td>m²</td>
<td>X</td>
</tr>
<tr>
<td>AREATT</td>
<td>total area of watershed</td>
<td>m²</td>
<td>X</td>
</tr>
<tr>
<td>β</td>
<td>fraction of snowmelt that goes to surface water</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>κ</td>
<td>convection condensation melt rate for the watershed</td>
<td>in/day/°C</td>
<td>Yes</td>
</tr>
<tr>
<td>n</td>
<td>a melt increasing factor</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>n</td>
<td>approximately = 0.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KPAN(I)</td>
<td>a dimensionless pan coefficient relating evapotranspiration from the compartment I to measured pan evaporation</td>
<td></td>
<td>Yes</td>
</tr>
</tbody>
</table>
KLAT3  a dimensionless lateral flow constant for the soil compartment  Yes

FI_L  limiting value of frost index above which there is no reduction in withdrawal rate  X

C_r  reduction factor for withdrawal rates per °C of frost index below FI_L  X

C_g  bare-ground frost coefficient  X

C_s  reduction in C_g per mm of snow water equivalent  X

C_t  thaw coefficient for water entering the soil  X

HC  daily thaw rate from ground heat °C  X

x  dimensionless exponent in the equation (19) for reduction of withdrawal rates = 8  X

UZSNOM  equivalent depth of water stored in the soil compartment at nominal moisture content  m  X

UZSSAT  equivalent depth of water storage capacity of the soil compartment under saturated conditions  m  X

PORE3  porosity of soil in the soil compartment  X

D3  depth of the upper soil layer or the soil compartment  m  X

KPERC3  a dimensionless vertical percolation constant for soil compartment  Yes

KLAT4  a dimensionless lateral flow constant for the unsaturated zone compartment  Yes

74
<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Unit</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>KPERC4</td>
<td>a dimensionless vertical percolation constant for the unsaturated zone compartment</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>LZSNOM</td>
<td>the equivalent depth of water stored in the unsaturated zone compartment at nominal moisture content</td>
<td>m</td>
<td>X</td>
</tr>
<tr>
<td>LZSSAT</td>
<td>the equivalent depth of water storage capacity of the unsaturated zone compartment under hypothetical saturated conditions</td>
<td>m</td>
<td>X</td>
</tr>
<tr>
<td>PORE4</td>
<td>porosity of the unsaturated zone compartment</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>DEPBR</td>
<td>average depth to bedrock in the watershed</td>
<td>m</td>
<td>X</td>
</tr>
<tr>
<td>FRAX</td>
<td>fraction of evapotranspiration from unsaturated zone</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>ALFI</td>
<td>permeability correction factor for lake seepage</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>ALF2</td>
<td>permeability correction factor for groundwater seepage</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>PORE6</td>
<td>porosity of the groundwater compartment</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>( k_b )</td>
<td>transpiration coefficient</td>
<td>in/day</td>
<td>X</td>
</tr>
<tr>
<td>( p )</td>
<td>monthly percentage of daylight hours</td>
<td>%</td>
<td>X</td>
</tr>
<tr>
<td>( d_1 )</td>
<td>fraction of the lake perimeter over which seepage occurs into the lake</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>DIST</td>
<td>distance between centroids of watershed and lake</td>
<td>m</td>
<td>X</td>
</tr>
<tr>
<td>KHU</td>
<td>horizontal hydraulic conductivity of the soil compartment</td>
<td>m/day</td>
<td>X</td>
</tr>
<tr>
<td>KVU</td>
<td>vertical hydraulic conductivity of the soil compartment</td>
<td>m/day</td>
<td>X</td>
</tr>
<tr>
<td>KHL</td>
<td>horizontal hydraulic conductivity of the unsaturated zone compartment</td>
<td>m/day</td>
<td>X</td>
</tr>
<tr>
<td>Parameter</td>
<td>Description</td>
<td>Unit</td>
<td>X</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------------------------</td>
<td>------------</td>
<td>---</td>
</tr>
<tr>
<td>KVL</td>
<td>vertical hydraulic conductivity of the unsaturated zone compartment</td>
<td>m/day</td>
<td>X</td>
</tr>
<tr>
<td>CBED</td>
<td>effective hydraulic conductivity of surface water body bed</td>
<td>m/day</td>
<td>X</td>
</tr>
<tr>
<td>SLOPE</td>
<td>average surface slope of watershed</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

* X indicates that parameter won't be estimated.
assumed to cover the whole watershed. The inflow to the snow compartment is through precipitation, in the form of snow. Precipitation is designated as snow if the air temperature is less than or equal to 0°C. The outflow from the snow compartment is through snowmelt. Snowmelt is described by empirical equations, one for rain-on-snow periods (based on an approximation of energy balance computation of melt for overcast conditions in forested areas, U.S. Army Corps of Engineers summary report, 1956) and one for dry weather melt. These equations are not strictly process based. The snowmelt is apportioned between the soil and lake compartments, in a fashion that accounts for the areal extents of the aquatic and terrestrial portions of the watershed and the flow of melt water down slopes to the lake, over or within the snowpack.

2. Soil Compartment The soil compartment comprises the upper zone soil (horizons O, A and B). The inflow to the soil compartment is through precipitation over the terrestrial portion of the watershed (during both no-snow and rain-on-snow periods) and percolating snowmelt. The outflows from the soil compartment consist of the lateral flow (interflow) to the lake, vertical percolation to the unsaturated zone compartment, overland flow to the lake in case of supersaturated conditions, and evaporation.

In the original version of the model (Nikolaidis et al., 1986) the interflow and vertical percolation are both computed using an equivalent saturated depth and then applying Darcy's law (a linear withdrawal) for steady saturated flow. However, in this study, a nonlinear (quadratic withdrawal) was assumed for the interflow, since it is a well documented fact (e.g., Amorosco 1963) that interflow response is indeed nonlinear. The flow area for lateral flow is assumed as the average perimeter of the watershed multiplied by the average saturated depth and the hydraulic gradient as the catchment slope. The flow area for vertical percolation is assumed as the area of the terrestrial portion of the watershed. An independent frozen ground model accounts for the retardation of withdrawal rates due to frost conditions.

3. Unsaturated Zone Compartment The unsaturated zone compartment comprises the lower zone soil, below the soil compartment and above the water-table. The inflow to the unsaturated zone is through vertical percolation from the soil compartment. The outflows are to the lake via lateral flow, to the groundwater compartment via deep percolation and to the atmosphere via...
evapotranspiration. The lateral and vertical outflows are computed as in the soil compartment, based on Darcian flow hypotheses. The evapotranspiration is computed by the empirical Blaney-Criddle formula (see Raudkivi, 1979, pg. 128).

4. Lake Compartment The lake compartment comprises the aquatic portion of the watershed. The inflows are from precipitation, snowmelt, lateral flow from the soil and unsaturated zone compartments and lake-groundwater seepage interaction. The lake-groundwater seepage interaction involves flows both in and out of the lake. These are also computed based on Darcian assumptions. The lake-bed permeability is used as the reference permeability. The hydraulic gradient is computed as the absolute difference between the stage and the saturated depth of water above the bedrock, divided by the distance between the centroids of the lake and terrestrial portions of the watershed. The area of flow is computed as the lake perimeter multiplied by the lake stage. The lake is assumed to be recharging over a fixed fraction of its perimeter and discharging over the other fraction. Other outflows from the lake include evaporation and discharge through a stream channel. The stream outflow is determined either by means of a weir-type equation or a stage-discharge rating curve.

5. Groundwater Compartment The groundwater compartment comprises the water table aquifer. Accurate simulation of the flows in and out of this compartment are crucial to the alkalinity submodel, since a dominant portion of the weathering reactions occur here and it has a relatively large detention time. The model however uses a fairly simple representation of seepage interaction, since more complex representations are likely to require more input data than is likely to be available. The other inflow to the groundwater compartment is deep percolation from the unsaturated zone compartment and the other outflow is through evapotranspiration. Groundwater import and export at the catchment boundary can also be included in the model. However, these need to be input to the model and are not computed internally.

C. Data Requirements for the ETD Model

The data requirements for hydrologic simulation of a lake-watershed system using the ETD model are two fold - system related data and input-output data.
The system-related data include the physiographic parameters listed in Table 3 as those not requiring estimation, and relations between lake volume, lake stage and stream outflow from the lake.

The input-data required are time series of daily temperature, evaporation and precipitation, monthly mean temperature and monthly mean percentage of daylight hours. In addition, characteristics of the vegetal cover over the watershed are required to characterize the parameter $k_b$ (see Table II) corresponding to evapotranspiration demand. For parameter estimation purposes, output data required are the measured outflow in the stream, from the lake and if possible, water table elevations. Equivalently, lake stage and water table elevation may be used.

D. Modifications of the Form of the Equations of the ETD Model to Facilitate Applications of Nonlinear Filtering Methods.

Most conceptual hydrologic models involve mathematical descriptions containing what are referred to in hydrologic literature as "threshold-type" parameters that cause the model to operate in different modes depending upon the values of the state variables. In addition, certain provisions need to be made in order to prevent depletion of storages in the various elements (which leads to physically unrealistic negative values for storages) and overflow of the conceptual reservoirs due to supersaturated conditions. As a consequence, the derivatives of the model equation with respect to the state variables and parameters are not continuous functions. In fact, they are likely to exhibit sharp discontinuities and in some cases do not formally exist. As was noted in Chapter II, the nature of the derivatives of the predictor with respect to the parameters is crucial to the estimation algorithm. Also, in case of the EKF algorithm, the derivatives are involved in linearization of the dynamic equations. Thus, it is desirable to have well behaved derivatives. In addition, it is the location of modal transitions that contain valuable information about the threshold type parameters, hence it is of further importance to have well behaved derivatives at these locations. It should be noted that the sharp modality is only in the mathematical representation of the dynamics. In nature, the transition between modes is gradual. Based on this, all the sharp modal transitions were replaced by smoothing functions. Kitanidis and Bras (1978) followed a describing function approach to this problem. In this
study, however, an empirical approach is followed, replacing the sharp discrete functions by approximate continuous functions. Such an approach was previously used by Georgakakos (1986a). Four such functions are used for this purpose. Their choice is discussed in Appendix D. Simulations carried out to evaluate the agreement between the discrete form of the original model and the modified form indicated very close agreement for "real" data. Based on hypothetical data, the performance of the modified form was analyzed during extreme conditions like supersaturation and depletion and it was concluded that the modified form handled these situations very well. The modified form of the model equations is listed in Table III. It should be noted also that the parameterization of the noise matrices involves derivatives with respect to inputs. These are not, however, replaced by continuous functions even though they are discrete, since they are discrete only with respect to inputs and not the state variables. The expressions for the derivatives of the differential equations of the modified form with respect to the state variables, fixed parameters, free parameters and input variables are presented in Appendix E.

E. Sources of Uncertainty in the Data and Their Incorporation in the Parameter Estimation Framework

The uncertainty in the data used for parameter estimation comes from two broad sources: uncertainty in the estimates of fixed model parameters and uncertainty in the estimates of input and output data.

Uncertainties in the fixed model parameters (in this case physiographic parameters) stems from the use of average quantities based on point samples, for parameters such as depths of various soil horizons and permeabilities of soils. On the other hand, it is reasonable to expect that some other quantities like areas are inferred relatively precisely, based on topographic maps. A rigorous approach to the quantification of these uncertainties would involve investigation of the character of the spatial random fields of spatially varying quantities. However, in lieu of this, it is reasonable to use a certain percentage of a quantity as a standard deviation, based on engineering judgment and the features of the particular system.

Uncertainties in estimates of inputs stems from the fact that inputs are spatially and temporally varying processes and the estimates used are daily averages assumed to be uniform over the entire extent of the watershed. In
Table III
Modified Form of The equations for the Flows in the Hydrology Submodel of the ETD Model

Snow Compartment

Q21 = SPREC
0.0

QMELT = k * (1.8 * TEMPC)^n + 1/1000.  TEMPC > 0.0, PREC = 0.0
[(0.007*PREV+0.074)*1.8*TEMPC+0.005]*.254/1000.
TEMPC > 0.0, PREC > 0.0

Q(3,2) = (1-β)*AREAT/AREATT*QMELT*FTHR(H22)
Q(5,2) = (1 - (1-β)*AREAT/AREATT)*QMELT*FTHR(H22)

Soil Compartment

Q(3,1) = WPREC*AREAT/AREATT
Q(1,3) = KPAN3*EVAP*AREAT/AREATT*FINS(h(2))
Q(5,3) = KLAT3*CLAT*KHU*FR*DEP3*(FTHR(H33))^2
Q(4,3) = KPERC3*AREAT/AREATT*KVU*FR*FTHR(H33)*FRET(H44)
OVLAND = FOVL(H33)*UZSSAT/TSPAN

Unsaturated Zone Compartment

QET = k_b*TM*PDAY/100.
Q(1,4) = QET*FRAX*FTHR(H44)
Q(5,4) = KLAT4*CLAT*KHL*DEP4*FTHR(H44)
Q(6,4) = KPERC4*AREAT/AREATT*KVL*FTHR(H44)*FRET(H66)

Surface Water Body (Lake) Compartment

Q(5,1) = WPREC*AREA/AREATT
Q(1,5) = KPAN5*EVAP*AREA/AREATT*FINS(h(2))
Q(5,6) = (1-d_1)*α_2 * CSEEP*HGR*STAGE
Q(6,5) = d_1*α_1 * CSEEP*HGR*STAGE
Q(8,5) = FLAGR2(STAGE), from stage-discharge relationship
Groundwater Compartment

Q(1,6) = QET-Q(1,4)

Note:
1. All quantities not defined in table II are defined below

\[
\begin{align*}
\text{SPREC} &= \{ \\
& \quad \text{PREC}, \quad \text{TEMP} \leq 0.0 \\
& \quad 0.0, \quad \text{TEMP} > 0.0 \\
& \quad 0.0, \quad \text{TEMP} \leq 0.0 \\
\text{WPREC} &= \{ \\
& \quad \text{PREC}, \quad \text{TEMP} > 0.0 \\
\end{align*}
\]

\[
\begin{align*}
H22 &= h(2)/QMELT \\
H33 &= h(3)/UZSSAT \\
H44 &= h(4)/LZSSAT \\
H66 &= h(6)/PORE6/DEP6 \\
CLAT &= \text{SLOPE}(\text{PERIL}+\text{PERIC})/2.0/\text{AREATT} \\
CSEEP &= \text{CBED} \* \text{PERIL}/\text{DIST}/\text{AREATT} \\
HGR &= \left| \text{STAGE} - \frac{h(6)}{PORE6} - \frac{h(4)}{PORE4} \right| \\
\text{STAGE} &= \text{FLAGR}(h(5)), \text{from volume-stage relationship} \\
\text{FR} &= \text{Frost reduction factor}, \\
&= \text{FR1} + (1.-\text{FR1}) \* (1.-H44)^8 \text{ where FR1 is an independently computed Frost index.}
\end{align*}
\]

2. Functions FTHR( ), FOVL( ), FINS( ) and FRET( ) are defined in appendix
addition, the data available may not be from the watershed of interest itself, but from the nearest meteorological stations. This introduces additional uncertainties in the estimates of inputs. Thirdly, the inputs themselves are outputs of some other dynamic processes, and where estimates are generated from other dynamic models, the uncertainties associated with them enter the watershed model. In case of precipitation, data from nearby raingage stations are weighted to obtain a mean areal precipitation. The associated uncertainty can be very high, especially in regions of uneven topography, and even as high as 100% error in estimates is not uncommon. Evaporation and evapotranspiration likely exhibit high variability in space. Evaporation is sometimes estimated by measuring evaporation from pans placed on the site and correlating pan evaporation to basin-scale evaporation based on water balance computations, to obtain pan coefficients. If not, it is estimated by means of semi-empirical equations that involve temperature, wind speed, atmospheric humidity and other meteorological variables. In all these cases, large levels of uncertainties can be expected of the estimates. Evapotranspiration depends upon meteorological variables and on the vegetal cover over the watershed. Uncertainties associated with estimates of evapotranspiration are thus likely to be very high, owing to spatial variability in these factors and the poor knowledge about the evapotranspiration process itself. It can also be expected that the variances associated with higher values of precipitation and evaporation estimates is higher, since the spatial and temporal variability of these processes increases with their magnitudes or intensities.

As an example of estimating uncertainties associated with empirical equations, the evapotranspiration equation used in the ETD model is considered.

\[ QET = k_b \cdot TM \cdot p/100 \]  

(IV.1)

where \( k_b \) is the (vegetation dependent) Blaney-Criddle constant, \( TM \) is the monthly mean temperature and \( p \) is the monthly mean percentage of daylight hours. If the uncertainty in the estimate of \( p \) is considered negligible as compared to that associated with estimates of \( k_b \) and \( TM \), the following scheme may be used to estimate the uncertainty associated with \( QET \).
Linearizing (IV.1), about \( \hat{k}_b, \hat{TM} \), respectively the best estimates of \( k_b \) and \( TM \),

\[
QET = \hat{k}_b \hat{TM} \, p/100 + \hat{k}_b \, p/100 \, (TM - \hat{TM}) + \hat{TM} \, p/100 \, (k_b - \hat{k}_b) \tag{IV.2}
\]

Hence, the variance can be computed as

\[
\text{Var}(QET) = (\hat{TM} \, p/100)^2 \cdot \text{Var}(k_b) + (\hat{k}_b \, p/100)^2 \cdot \text{Var}(TM) \tag{IV.3}
\]

where \( \text{Var}(\cdot) \) represents variance of a quantity. Assuming that the daily mean temperature follows an Autoregressive lag-1 process, (whose statistics can be computed from the temperature time series),

\[
\text{Var}(TM) = \text{Var} \left( \frac{T_1 + T_2 + \cdots + T_{30}}{30} \right) = f(e_T, \text{Var}(T)) \tag{IV.4}
\]

where \( T_i \) is the daily mean temperature of the \( i \)th day of a month and \( \rho_T \) and \( \text{Var}(T) \) are the lag-1 autocorrelation coefficient and variance statistics respectively, computed from the temperature time series. The correlation between \( QET \) and \( T \), the daily temperature can be computed, similarly. It should be noted that Equation (IV.1) was assumed to be exact for some true value of \( k_b \), in the above development. It may be that uncertainties due to incorrect forms of empirical equations may far outweigh uncertainties accounted for by analyses similar to (IV.1 - IV.3).

The measurement noise covariance matrix \( R \) represents the level of confidence in the measured outflows. It is well known (see Sorooshian and Dracup, 1980) that the use of rating curves to estimate discharge leads to higher errors in the estimates during periods of higher discharge. The measured discharges are point estimates in time. Owing to the fact that temporal variability of discharges increases with their magnitudes, it can be expected that higher errors would be incurred from wrong timing of measurements during such periods. For these reasons the choice of the \( R \) matrix should allow for an increase in variance with increase in the magnitude of the measurement.
P. Description of the Study Area: Woods Lake Watershed in Adirondack Park, New York.

The Woods lake watershed is a forested watershed with an areal extent of 2.1 km$^2$. The average area of the lake is about 0.207 km$^2$. The watershed receives an annual precipitation of about 1.2m and the mean annual air temperature is about 5.5°C. The areal average depth to bedrock is about 2.3 m, the average depth of the upper zone soil is about 0.37m and the average catchment slope is about 14 percent.

The data available for this study were provided by the Oak Ridge National Laboratory, under the Direct versus Delayed Response Project (DDRP) at The University of Iowa, sponsored by the U.S. E.P.A. (Environmental Protection Agency). The data was acquired during 1977-84 under the Integrated-Lake Watershed acidification study (Chen et al., 1980). The precipitation, temperature and outflow data were collected on a daily basis at the Woods lake watershed station. Evaporation data was not available and mean areal evaporation was estimated using an empirical equation based on temperature data and other meteorological data (obtained from the nearest first order National Weather Service station at the airport of Burlington, in Vermont state) by Nikolaidis (1986). Outflow data were the only data related to model outputs that were available.

The estimates of input and fixed-parameter uncertainty used in the input and parameter error covariance matrices (see equation (III.13)) are as follows.

A standard deviation of 1 mm/day plus 30 percent of the current estimate was assumed for precipitation and evaporation. This is in accordance with the fact that spatial and temporal variability of these processes are known to increase with their magnitudes. A standard deviation of 1°C was assumed for estimates of the mean daily temperature. The average correlation coefficient between temperature and evaporation was computed based on the time series data of temperature and generated evaporation and found to be 0.55. The variance of the evapotranspiration estimate was computed as described in Section IV.5.

Based on past studies on the Woods lake watershed, four of the fixed parameters were chosen for representation in the fixed parameter covariance matrix. These are DEP4, the depth of the unsaturated zone compartment, DEP6,
the depth of the groundwater compartment, CBED, the lake bed permeability and the lake volume-stage relationship. A standard deviation equal to 25% of their actual values was assumed for DEP4 and DEP6, a 100% standard deviation for CBED and a standard derivation equal to 5 mm plus 10 percent of the difference between the current stage estimate and the stream bed elevation was assumed for the stage estimate.

The measurement noise covariance matrix $R$ for the outflow estimate (in this case a scalar) was assumed based on a standard deviation of 0.1 mm/day (per unit watershed area) plus 10 percent of the current outflow estimate.

**G. A Stepwise Procedure for the Estimation of Parameters of the Hydrology Submodel**

Before the stepwise procedure is outlined certain observations are made regarding the character of the nonstationarities in the system response that motivate the choice of the procedure in question.

The dynamics of the watershed response during periods of snowcover is grossly different from that during periods when there is no snowcover. For instance, during periods of snowcover, there is no evaporation from the soil or lake and, unless there is rain-or-snow, there is no precipitation input to the soil or lake. The dominant input variable during snowcover periods is thus temperature, which also triggers off any dynamic activity only when it is above freezing. When the temperature is below freezing, the dynamics is not influenced by any of the input variables. By contrast, bulk of the dynamic activity in the watershed response during periods of no snowcover is triggered off by precipitation on the watershed. Thus precipitation is the dominant input variable during such periods. In addition, the uncertainty in the estimates of precipitation propagate through the system dynamics instantaneously during summer periods, while during extended periods of subfreezing temperatures, it is not reflected in the system dynamics until the next melt event.

In the light of the way input uncertainty is parameterized in the algorithm presented in Chapter III, (where a single factor $\alpha_u$ multiplies all the error covariance matrix elements), the foregoing discussion suggests that $\alpha_u$ associated with input uncertainty during summer periods would most likely be
different from that during snowcover periods. This suggests that while the algorithm presented in chapter III provides an elegant means of capturing nonstationary uncertainties such as those associated with the increased uncertainty corresponding to increased precipitation uncertainty, it might not prove effective in handling nonstationaries such as those outlined above (which are effectively changes in the model structure).

In addition, since the soil and unsaturated zone storages are not fed by precipitation during periods of snowcover, they are likely to be close to depletion. As a result, the parameters corresponding to their dynamics are likely to be poorly identifiable during this period.

Another important point is that while the structural errors associated with the representations of the dynamics of compartments other than snow are not likely to be very high, the snow compartment, whose dynamics is based on empirical equations, probably involves a high degree of structural uncertainties (it is a well documented fact that the timing of significant flow events triggered off by snowmelt is very poorly predicted by empirical models in general). This poses another problem, because the algorithm developed in Chapter III is based on the assumption that input and parameter uncertainties outweigh structural uncertainties.

Based on the above discussions, it is perhaps judicious to separate the estimation of parameters of the rest of the model from those of the snow compartment, using only periods of no snow cover for the former. It is suggested that the parameters unrelated to the snow compartment be estimated first and then kept fixed while the snow related parameters are estimated using a period of record when snow cover is present.

When the algorithm developed in Chapter III is applied to the parameter estimation problem, the resulting problem is one of simultaneous estimation of model parameters and noise parameters. Experiences in IR literature indicate that (e.g., Ljung, 1978) such algorithms may require long data records to be processed before they converge. In addition, simultaneous estimation of as many as nine parameters (non-snow-related), which likely exhibit considerable mutual interaction, in a highly nonstationary situation may by itself require a long data record to be processed before any convergence occurs. The question arises as to what one does if the available data record for the hydrology
submodel is not long enough. One way out of this problem is to use the same data record repeatedly until some kind of convergence (either to constant values or an identical time varying pattern over each rewind of the data record) is indicated, or otherwise. (It should be noted that there are no theoretical guarantees about the convergence of the algorithm in question).

At first glance, the use of the same period of data repeatedly may seem questionable, based on the innovations property of an optimal linear filter (heuristically, that there is no "information" left in an observation if the filtering is optimal). However, it should be realized that in this situation the filtering is suboptimal, firstly because the noise covariance matrices used are not the true ones, at least in the earlier stages of estimation, and secondly because we are dealing with a nonlinear filter for which the innovations property does not truly hold.

Based on the above discussions, the following stepwise procedure for the estimation of parameters of the hydrology submodel of the ETD model is suggested.

(i) Fix initial values of parameters unrelated to snow, and values of \( \alpha_u \) and \( \alpha_p \).

(ii) Using periods of data record without snowcover, attempt to estimate model parameters, and noise related parameters \( \alpha_u \) and \( \alpha_p \), recursively.

(iii) Upon convergence in step (ii), fix parameters unrelated to snow at converged values. Replace \( P_{\theta \theta} \), cast in terms of topographic parameters, by final parameter covariance matrix of parameters in step (ii) (since their uncertainty is likely to be more significant). Fix initial conditions for snow related parameters and \( \alpha_u, \alpha_p \) corresponding to snow periods.

(iv) Using periods of data with snowcover, attempt to estimate snow related parameters and \( \alpha_u, \alpha_p \) corresponding to snow recursively.

During step (ii), it is essential to monitor the filter performance as an indicator of the progress of the estimation and the optimality of the filtering. This can be achieved by keeping track of the statistical properties of the filter residuals sequence. For nonstationary systems, it is more convenient to analyse the statistical properties of a normalized residuals sequence.
The normalized residual is defined as the residual divided by its standard deviation as predicted by the filter. In case of vector observations, the standard derivation predicted by the filter is the square root of the \( i \text{th} \) diagonal element of the innovations covariance matrix (III.25). In case of scalar observations, it is defined by

\[ \tilde{\nu}(t_k) = \nu(t_k) / S_c(t_k) \]  

(IV.5)

where \( S_c(t_k) \) is as defined in (III.46), \( \nu(t_k) \) is the residual and \( \tilde{\nu}(t_k) \) the normalized residual.

When filtering is optimal, the normalized residuals sequence is a zero mean while noise sequence with variance equal to one. While this may not be satisfied strictly, gross deviations from these values can be construed as indicative of poor filter performance.

It may happen that during step (ii), \( \alpha_u \) and \( \alpha_p \) converge to some values while the model parameters are yet to converge. If the filtering is close to optimal at this stage, one could fix \( \alpha_u \) and \( \alpha_p \) at their respective converged values and let only parameter estimation proceed. On the other hand, it is also possible that the model parameters have converged while \( \alpha_u \) and \( \alpha_p \) are yet to converge. Owing to the mutually interactive nature of the parallel estimation schemes for the model parameters and filter parameters, the evolution of the parameters \( \alpha_u \) and \( \alpha_p \) influence that of the model parameters and vice versa. If the model parameters have converged at a stage where the filtering is far from optimal, it is best to use the final converged values of \( \alpha_u \) and \( \alpha_p \), keep them fixed and repeat the model parameter estimation.

It should be realized that \( \alpha_u \) and \( \alpha_p \) are representative of the noise strengths involved in the simultaneous state and model parameter estimation. Thus, the same values can be expected to apply for all situations, where these noise strengths are similar in character. However, an issue of concern in case of model parameter estimation is the uncertainty associated with initial estimate of these parameters. This is of special significance when multiple optima are likely in the parameter space. While, in case of recursive state estimation only, the effects of the initial estimates is "forgotten" owing to the driving dynamics and is likely insignificant as compared to the noise strengths, this may not be the case for parameter estimation.

89
Since the model is cast in terms of complete physical parameterization of all processes, with the free parameters being non-dimensional linear correction factors in the model equations (see Nikolaidis et al., 1986), perhaps a choice of unity as initial condition is indicated. However, it must be cautioned that the physical hypotheses based on which the model is developed may not apply with the same accuracy for all physical systems and it is best to gain some insight into the system through some experience on model simulations with some parameter values before setting initial conditions (set initial conditions at appropriate orders of magnitude, for instance).

H. Description of the Computer Programs Used in this Study

The cornerstone of the computer programs used in this study is the TSFP (Time-Series, State-space modelling, Forecasting, and Prediction) computer package developed by Georgakakos (1982). This is a general and powerful computer program suitable for state-space modelling of dynamic systems or analysis of time series. It is capable of performing four basic operations. Time series analysis, state space modelling, generation of extended forecasts and searches in two-parameter spaces. Upon option, combinations of these operations can be performed. A detailed description of the program can be found in the above reference. Salient features of this program include a highly accurate integrator based on a fourth-order predictor corrector scheme and its extensive display capabilities. For use in the context of an Extended Kalman Filter based parameter estimation scheme, it requires input-response data to be supplied by the user and in addition, two user supplied subroutines, one containing the differential equations describing the model dynamics and the derivatives of the state equations with respect to the states for linearization, and the other for computing the current observation based on the current states and input variables.

The program in its original version does not allow for the parameterization of model error covariance based on parameter uncertainties (input uncertainty is included, though) or simultaneous estimation of noise model parameters. The program was modified to allow for these, and some additional subroutines were introduced. Further modifications were made to obtain results of computer simulations in a format particularly suited to ease of interpretation in the context of the ETD model. A description of these subroutines is presented in Appendix F.
V. RESULTS AND DISCUSSION

In this chapter, presented are the results of the application of the methodology developed in Chapter IV to the estimation of parameters of the hydrology submodel of the ETD model for the Woods lake watershed. The organization of this chapter is as follows, in Section V.1, the results of the parameter estimation for periods without snowcover are presented, followed by an evaluation of the estimation exercise in Section V.2. In Section V.3, the results of the parameter estimation for the snow model are presented, followed by an evaluation in Section V.4. Finally, Section V.5 reports the results of a verification of the model.

A. Estimation of Parameters for Periods without Snowcover

Two periods of data were used for this purpose. Data from 04/30/1979 to 10/24/1979 and from 04/30/1980 to 08/30/1980. The time series of precipitation and evaporation data corresponding to this period are presented in Figures 2 and 3 respectively. The initial values of the model parameters are listed in Table IV. The initial values of the parameters used were based on a previous study (Nikolaidis, 1987) of the same system. While it may be expected, based on the model formulation, that most of the parameters that are essentially correction factors should be of the order of 1 or at worst 0.1 to 10.0, it was found that (Nikolaidis, 1987) this was not the case. KLAT3, for instance, corrects for the horizontal permeability coefficient of the soil compartment (comprising soil horizons 0 and A). The base value assumed was 30 mm/day, corresponding to horizon A, while horizon 0 had an associated value of about 1500 mm/day. The value of KLAT3 estimated was about 250. The initial standard deviation estimates associated with these values were assumed to be equal to 50 percent of the initial value of the parameter and no correlations between model parameters were assumed. Three runs were made for the simultaneous estimation of filter and model parameters, starting from different initial conditions for $\alpha_u$, $\alpha_p$, $\alpha_u$, and $\alpha_p$ during runs 1, 2 and 3 are respectively presented in Figures 4, 5, and 6. It is clear from these figures that $\alpha_u$ appears to converge faster than $\alpha_p$ in all the runs. Furthermore, in all three runs, $\alpha_u$ converges to a value of about 3.3. Based on the assumption made earlier of a 30% standard deviation associated with precipita-
Figure 2. Precipitation data for the calibration period during periods without snowcover.
a. 04/30/79 to 10/24/79

b. 04/30/80 to 08/30/80

Figure 3. Evaporation data for the calibration period during periods without snowcover.
**Table IV**

Initial and Final Values of Model Parameters and Standard Deviations

<table>
<thead>
<tr>
<th>PARAMETER</th>
<th>INITIAL VALUE</th>
<th>INITIAL STD.DEV.</th>
<th>FINAL VALUE</th>
<th>FINAL STD.DEV.</th>
</tr>
</thead>
<tbody>
<tr>
<td>KLAT3</td>
<td>0.2400E+01</td>
<td>0.1200E+01</td>
<td>0.1260E+02</td>
<td>0.0832E+00</td>
</tr>
<tr>
<td>KPERC3</td>
<td>0.1320E-01</td>
<td>0.0600E-02</td>
<td>0.4500E-01</td>
<td>0.2807E-02</td>
</tr>
<tr>
<td>KPAM3</td>
<td>0.1330E+01</td>
<td>0.7750E+00</td>
<td>0.5000E+01</td>
<td>0.3209E-01</td>
</tr>
<tr>
<td>KLAT4</td>
<td>0.1470E+02</td>
<td>0.7400E+01</td>
<td>0.9073E+01</td>
<td>0.1000E-03</td>
</tr>
<tr>
<td>KPERC9</td>
<td>0.1120E-01</td>
<td>0.5600E-02</td>
<td>0.1000E-03</td>
<td>0.1000E-03</td>
</tr>
<tr>
<td>FRAX</td>
<td>0.4800E-00</td>
<td>0.2400E-00</td>
<td>0.8300E+00</td>
<td>0.1804E-00</td>
</tr>
<tr>
<td>ALF1</td>
<td>0.6000E-00</td>
<td>0.3000E-00</td>
<td>0.1000E-03</td>
<td>0.9309E-03</td>
</tr>
<tr>
<td>ALF2</td>
<td>0.2500E+01</td>
<td>0.1250E+01</td>
<td>0.2070E+01</td>
<td>0.9222E-01</td>
</tr>
<tr>
<td>D1</td>
<td>0.6000E-00</td>
<td>0.3000E-00</td>
<td>0.7371E+00</td>
<td>0.1000E-03</td>
</tr>
<tr>
<td>KPAM5</td>
<td>0.1638E+01</td>
<td>0.8190E+00</td>
<td>0.1386E+01</td>
<td>0.1000E-03</td>
</tr>
<tr>
<td>UZSSAT</td>
<td>0.3848E-01</td>
<td>0.2924E-01</td>
<td>0.0000E-01</td>
<td>0.5317E-02</td>
</tr>
<tr>
<td>Run No.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>( \alpha_u ), initial</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>( \alpha_p ), initial</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>time steps</td>
<td>2700</td>
<td>3900</td>
<td>1500</td>
<td></td>
</tr>
<tr>
<td>( \alpha_u ), final</td>
<td>3.3</td>
<td>3.3</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td>( \alpha_p ), final</td>
<td>not converged</td>
<td>1.6</td>
<td>not converged</td>
<td></td>
</tr>
<tr>
<td>Normalized residual mean</td>
<td>.077</td>
<td>.034</td>
<td>.069</td>
<td></td>
</tr>
<tr>
<td>Normalized residual variance</td>
<td>1.191</td>
<td>0.96</td>
<td>1.203</td>
<td></td>
</tr>
<tr>
<td>Normalized residual lag-1 autocorrelation coefficient</td>
<td>.376</td>
<td>.32</td>
<td>.379</td>
<td></td>
</tr>
</tbody>
</table>
Figure 4. Traces of the parameters $\alpha_p$ and $\alpha_u$ during Run 1.
Figure 5. Traces of the parameters $\alpha_p$ and $\alpha_u$ during Run 2.
Figure 6: Traces of the parameters $a_p$ and $a_u$ during Run 3.
tion estimates, this indicates a standard deviation of about 100%. In all the three runs a reduction in the variance of the normalized residuals (1.6 to 1.7 to begin with) and the lag-one autocorrelation coefficient (0.6 to 0.7 to begin with) were observed. The final values of the standard deviation attained in runs 1, 2 and 3 were respectively 1.2, 0.96 and 1.19. The corresponding lag-one autocorrelation coefficients were 0.376, 0.320 and 0.379. Examination of the normalized residuals sequence during the final pass over the data records indicates similar characteristics in all the three runs. The sequence of normalized residuals for the final pass in run 2 is presented in Figure 7. Model parameter traces during each of the three runs indicated that the model parameters were yet to converge and some of them indicated marked trends, while α_u and α_p had converged. Thus, it was decided to keep α_u and α_p fixed at the final values from run 2 (in view of the superior normalized residuals statistics) and proceed with the parameter estimation. In an attempt to improve the convergence rate, it was decided to restart with initial values of the model parameters equal to those at the end of 1500 time-steps during run 2, and use initial standard deviations for model parameters equal to 50 percent of these parameter values.

Even under these initial conditions, the parameters converged at a rather slow rate. Two of the parameters, KPERC4 and ALFl were driven to very small values. The final values and standard deviations of the parameters are presented in Table IV. A sharp jump in the parameter values is seen around time step number 1320 (see Figures 8-10), that led to a large reduction in variance of parameter estimates. This jump seems to exert a strong influence on the subsequent parameter evolution for most parameters. It occurred at around day no. 249 in the first period of record. A possible explanation for this is that the parameter correlations have grown steadily and reached a level where significant information from the observation can be transmitted to all parameters. In order to further investigate the effects of this event a simple experiment was performed. Starting with arbitrarily chosen initial values of parameters in the vicinity of the final values and the parameter covariance matrix from time step no. 1200, another parameter estimation run was made. Another jump occurred at a much later time step, though, at the day no. 218 in the first period of record. It was found that most of the parameters came close to the final values from the earlier run. A notable exception was
Figure 7. Normalized residuals during final pass over the data in Run 2.
KLAT4. The parameter traces during the estimation are presented in Figures 8 (for the parameters of the soil compartment), 9 (for the parameters of the unsaturated zone compartment) and 10 (for the parameters of the lake and groundwater compartments). Figure 11 presents the parameter traces of KLAT3 and KLAT4 in the experimental run mentioned above. An interesting fact that came to light was that the deterministic simulation with the final values of parameters from the latter run indicated excellent correspondence with field data, around day no. 218, at which the jump occurred during the parameter estimation. On the other hand, the correspondence with field data around day no. 249 had degenerated. The final parameter correlation matrix at the end of the estimation run is presented in Table VI. It is obvious that there is a high degree of correlation between the model parameters. In fact, most of the correlation coefficients seem to be very close to 1. It should be emphasized at this stage that this observation applies to the particular physical system under investigation and to the particular data record. It may be that the application of the same model to a different system does not lead to the same situation. In any case, on the one hand the high correlations between parameters indicate that information from the observations has been passed even to parameters that are poorly correlated with the observation and on the other hand that parameter inference could have been marred by such a high degree of multiparameter interaction. These issues are discussed further in the next section.

B. Evaluation of the Parameter Estimation for Periods with no Snowcover

The bases for the evaluation of the parameter estimation exercise are as follows: the statistics of the normalized residuals sequence during the parameter estimation, parameter traces and convergence, and comparison of deterministic simulations with initial and final values of parameter estimates.

The normalized residuals sequence provides information about potential locations of structural errors, in that high values of normalized residuals would be indicated at these locations. An examination of the normalized residuals plot in Figure 7 indicates that for the most part, the normalized residuals are within the ±2 envelope. This corresponds to a 95% confidence level. The locations of high values of normalized residuals were found to be
Figure 8. Evolution of the parameters of the soil compartment during the parameter estimation run.
Figure 9. Evolution of the parameters of the unsaturated zone compartment during the parameter estimation run.
Figure 10. Evolution of the parameters of the lake and groundwater compartments during the parameter estimation run.
Figure 11. Evolution of the parameters KLAT3 and KLAT4 during the experimental run.
Table VI
Parameter Correlation Matrix at End of Parameter Estimation Run

<table>
<thead>
<tr>
<th>KLAT3</th>
<th>KPERC3</th>
<th>KPAR3</th>
<th>KLAT4</th>
<th>KPERC4</th>
<th>FPAR</th>
<th>ALF1</th>
<th>ALF2</th>
<th>E1</th>
<th>KPAR3</th>
<th>ZE33AT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000000</td>
<td>0.239674</td>
<td>-0.294404</td>
<td>-0.999863</td>
<td>0.991811</td>
<td>-0.972823</td>
<td>0.994613</td>
<td>0.994724</td>
<td>0.97802</td>
<td>0.998022</td>
<td>0.996928</td>
</tr>
<tr>
<td>0.239674</td>
<td>1.000000</td>
<td>-0.991774</td>
<td>-0.976390</td>
<td>0.996390</td>
<td>-0.331669</td>
<td>0.994213</td>
<td>0.842612</td>
<td>0.99439</td>
<td>0.996390</td>
<td>0.383512</td>
</tr>
<tr>
<td>-0.294404</td>
<td>-0.991774</td>
<td>1.000000</td>
<td>-0.991163</td>
<td>0.990228</td>
<td>-0.994147</td>
<td>0.996690</td>
<td>0.992022</td>
<td>0.99022</td>
<td>0.990228</td>
<td>0.992904</td>
</tr>
<tr>
<td>-0.999863</td>
<td>-0.976390</td>
<td>-0.991163</td>
<td>1.000000</td>
<td>0.920100</td>
<td>-0.993724</td>
<td>0.998783</td>
<td>0.990722</td>
<td>0.95790</td>
<td>0.939400</td>
<td>-0.750806</td>
</tr>
<tr>
<td>0.991811</td>
<td>0.996390</td>
<td>0.990228</td>
<td>0.920100</td>
<td>1.000000</td>
<td>0.991464</td>
<td>-0.999967</td>
<td>-0.993802</td>
<td>-0.99810</td>
<td>-0.998100</td>
<td>-0.377218</td>
</tr>
<tr>
<td>-0.972823</td>
<td>-0.994147</td>
<td>-0.992724</td>
<td>0.991464</td>
<td>1.000000</td>
<td>0.994236</td>
<td>0.718300</td>
<td>0.99978</td>
<td>0.99998</td>
<td>0.999984</td>
<td>-0.324674</td>
</tr>
<tr>
<td>0.972813</td>
<td>0.994213</td>
<td>0.994690</td>
<td>0.997813</td>
<td>-0.999967</td>
<td>0.994236</td>
<td>1.000000</td>
<td>-0.991869</td>
<td>-0.992866</td>
<td>-0.992868</td>
<td>-0.994492</td>
</tr>
<tr>
<td>0.996724</td>
<td>0.843612</td>
<td>0.992023</td>
<td>0.991840</td>
<td>-0.993802</td>
<td>0.718300</td>
<td>-0.991869</td>
<td>1.000000</td>
<td>-0.99764</td>
<td>-0.996995</td>
<td>0.603007</td>
</tr>
<tr>
<td>0.990222</td>
<td>0.996390</td>
<td>0.990228</td>
<td>0.927900</td>
<td>-0.990100</td>
<td>0.999984</td>
<td>-0.992866</td>
<td>-0.997443</td>
<td>1.000000</td>
<td>-0.999400</td>
<td>0.808978</td>
</tr>
<tr>
<td>0.990222</td>
<td>0.996390</td>
<td>0.990228</td>
<td>0.927900</td>
<td>-0.990100</td>
<td>0.999984</td>
<td>-0.992866</td>
<td>-0.997443</td>
<td>1.000000</td>
<td>-0.999400</td>
<td>0.808978</td>
</tr>
<tr>
<td>0.9969728</td>
<td>0.843612</td>
<td>0.992023</td>
<td>0.991840</td>
<td>-0.993802</td>
<td>0.718300</td>
<td>-0.991869</td>
<td>1.000000</td>
<td>-0.99764</td>
<td>-0.996995</td>
<td>0.603007</td>
</tr>
</tbody>
</table>
at sharp changes in the response, such as those at day 218 in the first record, and locations of high flow such as days 249-251 in the first record. Day 218 could be a location of potential structural errors (though this cannot be confirmed unless some other structure can be proposed that performs better). This is perhaps because it marks a significant response after an extended dry period, during which some of the nominal storage in the soil may have been depleted and subsequent precipitation would first need to satisfy that deficiency before lateral flow or vertical percolation can begin. It could also be because the smoothing function used provides a better approximation to Darcian conditions close to depletion, which is contrary to physical reality. However, owing to multiparameter interaction and input uncertainty, such a conclusion cannot be drawn with full confidence.

The parameter traces indicate satisfactory convergence for all parameters except KPERC4 and ALF1. A value of .001 was assigned to both these parameters. It should also be noted that during the parameter estimation, none of the parameters were constrained to any bounds apart from requiring positive values (not even FRAX and D1, which physically are meant to lie between 0 and 1). Inspite of this, none of the parameters have converged to physically unrealistic values. In addition, this automatic technique needs little manual interference, to obtain good parameter estimates. This is in contrast to some other manual or semi-manual calibration methods that need the expertise of experienced hydrologists, or model-knowledgeable personnel. The traces of KLAT3 and UZSSAT indicate a regular time varying pattern towards the end of the estimation. Furthermore the features of this variation in KLAT3 and UZSSAT are identical. Examination revealed that the locations where the troughs in the traces appear correspond to the period of the first record between days 210 and 230. This perhaps supports the earlier conjecture about a potential structural error at this location. Intuitively, reduction in KLAT3 and UZSSAT would lead to reduction in lateral flow and hence in flashy response. Thus, in the estimation scheme, these parameters could have been forced down, in order to compensate for the overestimation of lateral flow based on the conditions that apply over the rest of the data record.

Deterministic simulations were performed using the initial and final parameter estimates for both periods of summer record. The simulations obtained using the initial values are presented in Figure 12 while those
obtained using the final values are presented in Figure 13. Comparison of the two figures reveals a marked improvement in the deterministic prediction of the hydrologic response. There is an overprediction of the outflow during the period from day 210 to day 230 in the first period of record. The underprediction of flow during the earlier part of the first record is possibly because of unaccounted late spring snowmelt. The peak around day 250 in the first period of record is captured with perfect timing, but underpredicted. The underprediction could be in part because of the large (daily) time step used in the simulation, since conditions during such extreme events vary significantly in time-periods of the order of few hours. In the second record, there again is an instance of overprediction of response subsequent to an extended low flow period.

Comparison of least squares performance indices between the simulations with the initial and final parameter estimates is presented in Table VII. Again, improvement in these indices is obtained. It should be emphasized at this stage that the parameter estimation procedure by itself does not use a simple least squares criterion and this criterion has been used only as a basis for evaluation of the parameter estimation.

In addition to the above evaluations, the sensitivity of the final parameter estimates and three instances of two-parameter interaction were investigated. Sensitivity analysis was performed by examining the mean squared error as a function of a parameter value by varying it approximately between 50% and 200% of its estimated value. Figure 14 presents a sensitivity analysis of parameters of the soil compartment, Figure 15 that of the parameters of the unsaturated zone compartment, and Figure 16 that of the parameters of the lake and groundwater compartments.

Sensitivity analysis of the parameters of the soil compartment indicate that the response is significantly sensitive to all of them. In addition it appears that all final estimates are reasonably close to local valleys in their respective planes.

Sensitivity analyses of KPERC4 and ALFL indicates that variations of four orders of magnitude in these parameters have an insignificant effect on the mean squared error. It is therefore not surprising that they are poorly identifiable. Sensitivity analyses of the fractional parameters FRAX and DI
Figure 12. Deterministic simulations during periods without snowcover, with initial values of parameters. Observations are signified by the thick solid line.
Figure 13. Deterministic simulations during periods without snowcover, with estimated values of parameters. Observations are signified by the thick solid line.
Table VII  
Comparison of Least Squares Performance Indices  
Before and After Parameter Estimation - Periods Without Snowcover

<table>
<thead>
<tr>
<th>Run</th>
<th>Initial Values</th>
<th>Final Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Summer Period, 1979</td>
<td>1.029x10^{-6}</td>
<td>7.804x10^{-7}</td>
</tr>
<tr>
<td>MSE</td>
<td>Coefficient of efficiency: 0.625</td>
<td>0.726</td>
</tr>
<tr>
<td></td>
<td>Coefficient of determination: 0.634</td>
<td>0.743</td>
</tr>
<tr>
<td>2. Summer Period, 1980</td>
<td>4.87x10^{-7}</td>
<td>4.08x10^{-7}</td>
</tr>
<tr>
<td>MSE</td>
<td>Coefficient of efficiency: 0.401</td>
<td>0.499</td>
</tr>
<tr>
<td></td>
<td>Coefficient of determination: 0.481</td>
<td>0.734</td>
</tr>
</tbody>
</table>

Note: Define \( Q_p(i) \) and \( Q_o(i) \), respectively, as the predicted and observed flows for time \( i \). Also define \( Q_e(i) \) as the estimated flow from the regression line of \( Q_p(i) \) on \( Q_o(i) \). Let

\[
S = \sum_{i=1}^{n} [Q_o(i) - Q_p(i)]^2, \text{ the sum of squared errors.}
\]

\[
S_o = \sum_{i=1}^{n} [Q_o(i) - \overline{Q_o}]^2, \text{ a measure of the variability of flows,}
\]

with \( \overline{Q_o} = \) mean observed flow.

Then, coefficient of efficiency, \( E = \frac{S_o - S}{S_o} \).

Coefficient of determination, \( D = 1 - \frac{\sum_{i=1}^{n} [Q_o(i) - Q_e(i)]^2}{S_o} \).
Figure 14. Sensitivity analysis of the parameters of the soil compartment.
Figure 15. Sensitivity analysis of the parameters of the unsaturated zone compartment.
Figure 16. Sensitivity analysis of the parameters of the lake and groundwater compartments.
indicate an improvement in mean squared error while straying out of 1, their upper bound. This could be partly due to "errors" in other parameters on account of multiparameter interaction, interaction of these parameters with other parameters or call for alternative formulations. It is also apparent that KPAN5 is a sensitive parameter and that its estimate is close to a valley. It appears that the parameter ALF2 is not as sensitive as the parameters of the soil compartment and its reduction below the estimated value seems to improve performance based on mean squared error.

Two-parameter interactions were studied for the following combinations: KLAT3 - UZSSAT, KLAT3 - KLAT4 and D1 - ALF2. The final parameter correlation matrix indicates a high positive correlation between KLAT3 and UZSSAT. This shows up as a clearly discernible ridge with a positive slope in the two parameter plot (Figure 17). It is also apparent that the KLAT3-UZSSAT pair estimated falls within a relatively flat area near the optimum.

The final parameter correlation matrix indicates a high negative correlation between KLAT3 and KLAT4. There is a ridge with a negative slope in the two parameter plot (Figure 18) but the contours appear to become quite concentric close to the optimum. A plot of another criterion (absolute proportional error in the mean) however indicated a sharp ridge, in the vicinity of the optimum, which was at the same location.

The final parameter correlation matrix indicates a high negative correlation between D1 and ALF2. However, there is no ridge in the two parameter plot (Figure 19). Another interesting feature is apparent in this case. Suppose ALF1 were approximately zero or negligible, then D1 and ALF2 appear only as the produce (1.-D1) * ALF2 in the model equations. Thus, only the product (1.-D1) * ALF2 would be identifiable. Indeed, the contours in the D1-ALF2 space indicate a close correspondence with the contours of (1.-D1)*ALF2 = constant, a family of rectangular hyperbolas with origin at (D1 = 1.0, ALF2 = 0.0). This is due to the relative insensitivity with respect to ALF1.

C. Estimation of Parameters of the Snow Model

Two periods of data were used for this purpose. Data from 11/ 29/78 to 04/29/79 and from 10/25/79 to 04/29/80. The time series of precipitation and temperature data for these periods are presented in Figures 20 and 21 respec-
Figure 17. Interaction between the parameters KLAT3 and UZSSAT.
Figure 18. Interaction between the parameters KLAT3 and KLAT4.
PARAMETER SPACE: ORDINATE - ALF2  ABSCISSA - D1

Optimum designated by 0 is equal to 0.72900E-06
Parameter values at optimum: ALF2 = 0.50000E+00  D1 = 0.90000E+00
Contour values:
   Contour 1 = 0.74426E-06  Contour 2 = 0.75952E-06  Contour 3 = 0.77478E-06
   Contour 4 = 0.79004E-06  Contour 5 = 0.80531E-06  Contour 6 = 0.82057E-06
   Contour 7 = 0.83583E-06  Contour 8 = 0.85109E-06  Contour 9 = 0.86635E-06

Figure 19. Interaction between the parameters D1 and ALF2.
Figure 20. Precipitation data for the calibration period during periods with snowcover.

a. 11/29/78 to 04/29/79
b. 10/25/79 to 04/29/80
Figure 21. Evaporation data for the calibration period during periods with snowcover.

a. 11/29/78 to 04/29/79
b. 10/25/79 to 06/29/80
tively. The initial values for the snow model parameters used were 0.2 for BETA and 1.0 for KAPPA with 100% standard deviations on each of them. Initial values used for $\alpha_u$ and $\alpha_p$ were 0.8 and 10.0 respectively, based on results from preliminary runs.

The traces of $\alpha_u$ and $\alpha_p$ during the simultaneous estimation are presented in Figure 22. The traces of the parameters BETA and KAPPA are presented in Figure 23. The model parameters are seen to converge before $\alpha_u$ and $\alpha_p$. While the runs did result in convergence of parameter estimates of $\alpha_u$ and $\alpha_p$ estimates, it was found that not much improvement was attainable in terms of the normalized residual statistics. In some runs, the variance of the normalized residuals began to increase. It seems that this is due to the large structural errors in the simple formulation adopted for snowmelt in the ETM model. It seems to be incapable of capturing the timing of snowmelt events accurately, and consistently tends to underestimate early winter snowmelt. As a result, significant amounts of snow accumulate and extreme melt events are predicted by the model even after the late spring snowmelt as indicated by the field data.

D. Evaluation of the Parameter Estimation of the Snow Model

The normalized residuals statistics during the simultaneous estimation of model parameters and $\alpha_u$ and $\alpha_p$ indicates far from optimal filtering. Typical normalized residual variances computed over a pass on the data record are of the order of 2-3, with a lag -1 autocorrelation coefficient of 0.6 to 0.7. Figure 24 presents a sample plot of normalized residuals during the final pass over the data record.

Comparison of the deterministic simulations obtained using the initial values of the parameters (Figure 25) with that obtained using the final values of parameters (Figure 26) does not indicate much overall improvement. The least squares performance indices computed based on simulations with initial and final parameter estimates are presented in Table VIII. A sensitivity analysis of the values of BETA and KAPPA is presented in Figure 27.

On the whole, it appears that the parameter estimation for the snow model is unsatisfactory, primarily because of large structural errors in the snow model.
Figure 22. Traces of the parameters $\alpha_p$ and $\alpha_u$ during the simultaneous estimation run for periods with snowcover.
Figure 23. Traces of the snow model parameters during the simultaneous estimation run.
Figure 24. Normalized residuals during the final pass over the data in the simultaneous estimation run.
Figure 25. Deterministic simulations during periods with snowcover, with initial values of parameters. Observations are signified by the thick solid line.

Figure 26. Deterministic simulations during periods with snowcover, with final values of parameters. Observations are signified by the thick solid line.
<table>
<thead>
<tr>
<th>PERFORMANCE INDEX</th>
<th>SIMULATION USING INITIAL VALUES</th>
<th>SIMULATION USING FINAL VALUES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean square error</td>
<td>$6.63 \times 10^{-6}$</td>
<td>$4.44 \times 10^{-6}$</td>
</tr>
<tr>
<td>Coefficient of efficiency</td>
<td>-0.088</td>
<td>0.271</td>
</tr>
<tr>
<td>Coefficient of determination</td>
<td>0.507</td>
<td>0.455</td>
</tr>
</tbody>
</table>

126
Figure 27. Sensitivity analysis of the snow model parameters.
E. Verification of the Model

A verification of the model was performed by making a deterministic simulation for the period of record from 09/01/80 to 08/31/81. The field data and simulated outflows are compared in Figure 28. While the simulation appears to be satisfactory during periods of no snowcover, the simulation during periods of snowcover is very poor. In particular, a very significant snowmelt event due to an early winter "Indian summer" is missed out and all the accumulated snowpack is melted only in the spring, causing excessive overestimation of flows.

The least squares performance indices computed were a mean squared error of $1.36 \times 10^{-5} \text{ m}^2/\text{day}^2$, a coefficient of efficiency of $-0.163$, and a coefficient of determination of $0.166$. Least squares performance indices computed based on periods with no snowcover were a mean square error of $1.87 \times 10^{-6} \text{ m}^2/\text{day}^2$, a coefficient of efficiency of $0.337$, and a coefficient of determination of $0.511$. Comparison of the least squares performance indices for the calibration period indicate a drastic degeneration in performance. This is in spite of the fact that the input excitation during the verification period is not significantly different from that during the calibration period.

This affirms the need to use as much data as is available for calibration. It shows the effect of structural errors in the snowmelt model, and the drawbacks of inferring parameters from short periods of record.
Figure 28. Deterministic simulations during the verification period with estimated parameter values.
VI. CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

A comprehensive approach to the problem of recursive state and parameter estimation of conceptual hydrologic models has been proposed. The new approach allows for the incorporation of time-varying uncertainties in the estimation framework, using an explicit parameterization based on input and parameter uncertainties. The resulting uncertainty representation is consistent with the physical processes involved and the dynamics of hydrologic response.

As compared to some of the earlier approaches that use time-invariant uncertainty representations, the new approach while allowing for incorporation of time-varying uncertainties, achieves a significant reduction in the number of parameters related to the uncertainty representation, that need to be estimated. Only two parameters, \( \alpha_u \) and \( \alpha_p \) need to be estimated. It also allows for incorporation in the estimation framework, of prior information about input and parameter uncertainties, in the form of degree-of-belief estimates with \( \alpha_u \) and \( \alpha_p \) being estimated, so as to correct these.

The new approach was applied to the parameter estimation of the hydrology submodel of the ETD model for the Woods Lake watershed in New York state. The results indicated the success of the methodology in the estimation of parameters, as is evident from the significant improvement achieved in the simulation during the calibration period using the estimated parameter values for the periods without snowcover.

However, it also becomes evident that the potential of the methodology is not fully realized in situations where there are significant structural errors in the model, as in the snow model of the ETD model. Though improvement in least squares performance indices was obtained, subsequent to the parameter estimation, the prediction of snowmelt events is far from satisfactory. As is true for other approaches, it is evident that in the context of parameter estimation for conceptual hydrologic models, the success of the new approach is strongly influenced by structural errors in model representations, the quality of input data, and structure-related features that result in a high degree of multiparameter interaction. It is noted, however, that the proposed methodology (unlike previous ones) is capable of identifying periods with structural errors.
While problems of serious structural errors can perhaps be overcome only by developing alternative refined structures, the new approach allows for the incorporation of the modeler's knowledge about quality of the input data, in the parameter-estimation framework. In fact, the parameterization of uncertainty in this approach establishes a strong framework for the analysis of the influence of input data uncertainty on the parameter estimation, in a rather direct fashion. In addition, it establishes a nexus between the problem of input sensor network design and the parameter estimation problem. For instance, analysis of parameter identifiability as related to the input covariance structure can specify optimum input covariance structures for parameter identifiability. Subsequent analyses of the space-time random fields that constitute the input processes could lead to the design of sensor networks that achieves these optimum characteristics for the input estimates.

Using the same uncertainty representations, as in this approach, alternative schemes for parameter estimation could be developed, (offline maximum likelihood schemes, for instance) that could lead to better parameter inference. Such schemes, obviously exploit the state space structure of the dynamic watershed response model to the fullest extent, in contrast to some earlier approaches. One issue that deserves some attention is that of multiparameter interaction. One possible approach to the problem, that applies even in a recursive estimation framework, is to use alternative criterion functions that are based on Ridge Regression concepts. These could lead to better parameter inference, in the pressure of high degree of interaction.

Another area of research that is worthwhile to explore is that of the choice of criteria for the estimation of parameters $\alpha_u$ and $\alpha_p$. In this study, a filter optimality criterion based on covariance matching, was used. The importance of filter optimality in a situation where the filtering is essentially suboptimal is an issue that deserves much attention. An alternative criterion that could be used for estimation of $\alpha_u$ and $\alpha_p$ is to consider a minimization of one-step prediction error covariance considered as a function of $\alpha_u$ and $\alpha_p$. Such a criterion would not enforce filter optimality, but attempt to tailor filter sensitivity so as to improve one-step ahead predictions. Apparently, such a criterion would have important implications in the context of recursive state estimation and flood forecasting. Another possibility is to consider the use of both filter optimality and minimization of one-step prediction error variances simultaneously.
In the specific context of the ETD model and its potential applications, the following observations may be made. Firstly, it is undesirable that the pan coefficients related to the evaporation process be included in the estimation procedure along with other parameters. This is because of high degree of multiparameter interaction involved. Evaporation is a process that removes water from the watershed system as a whole and should ideally be inferred from analyses of water balance of the watershed system as a whole, over longer periods of time. If not, for instance, inaccurate estimation of groundwater discharge to the lake, causing its underestimation, could be compensated for by underestimation of lake evaporation. Alternative schemes for estimation of evaporation reduce the number of parameters to be estimated and reduce multiparameter interaction and prevent erroneous estimation of evaporation rates. This is especially significant, since evapo-concentration (or increase in concentration of chemical constituents in a water body owing to the loss of water by evaporation) is an important component of the chemical response of the system.

Secondly, it may be useful to impose bounds on the parameter values, based on prior knowledge about the physical system. It is, for instance, known that for this physical system, groundwater discharge to the lake far exceeds recharge from the lake. This could be specified as a bound by requiring that \((1-d_1)\alpha_2 > d_1\alpha_1\) always (see Table III).

In regard to the snow model, it is apparent that a simple degree-day approach is quite incapable of reproducing the actual character of snowmelt related dynamics. The snowmelt rate is in general a function of meteorological factors and the heat deficit in the snowpack which in turn, depends on its density and age. A refined approach would be to attempt to develop indices for snow density and age based on cumulative temperature and then relate snowmelt rate to these indices.

It is also clear that improvement can be achieved over the use of quasi-Darcian hypotheses for unsaturated flow. While in this study a quadratic withdrawal rate was used and led to some improvement, a more refined approach would be based on physical relationships between unsaturated hydraulic conductivity and the soil moisture content. Again, this refinement could be applied, both to the soil and unsaturated zone compartments.
In conclusion, the new methodology is a promising approach to the problem of recursive state and parameter estimation of general nonlinear and nonstationary dynamic systems such as watershed systems. It provides a very elegant framework for the analysis of influence of input data uncertainty on the quality of parameter and state estimates and opens up a number of possibilities for further research that could lead to improvements in the capabilities that currently exist for state and parameter estimation of conceptual watershed models.
REFERENCES


Astrom, K.J. and C. Kallstrom, Application of system identification techniques to the determination of ship dynamics, Proc. 3rd IFAC Symposium on identification and system parameter estimation, the Hauge (North-Holland), 1973.


Clarke, R.T., A review of some mathematical models used in hydrology, with observations on their calibration and use, J. Hydrology 19(1), 126, 1970.


Sorooshian, S., and J.A. Dracup, Stochastic parameter estimation procedures for hydrologic rainfall runoff models: Correlated and heterosedastic error cases, Water Resources Research, 16(2), 430, 1980.


Stumm, W., and J.L. Schnoor, Naturwissenschaffen 70, 216 (In German), 1983.


Troutman, B.M., Errors and parameter estimation in precipitation-runoff modelling, 2. Case study, 21(8), 1214, 1985b.


Young, P.C., Regression analysis and process parameter estimations - a cautionary message, simulation, 10, 125, 1968.


APPENDIX A

DERIVATION OF THE STATE ERROR COVARIANCE DIFFERENTIAL EQUATION

Consider the equation

$$\frac{dx(t)}{dt} = \dot{f} + F_x(t)e_x(t) + F_u(t)e_u(t) + F_\theta(t)e_\theta$$  \hspace{0.5cm} (A.1)

which can be obtained from Eq. (II.5) after neglecting linearization errors and redefining $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial u}$, and $\frac{\partial f}{\partial \theta}$, (defined in Eqs. II.7, II.8 and II.9) by $F_x(t)$, $F_u(t)$ and $F_\theta(t)$ respectively.

In equation (A.1) $e(t)$ is assumed to be a continuous, non-stationary Gaussian white noise process, $e$ a zero mean Gaussian random vector. Under these assumptions, $e_x(t)$ would also be a non-stationary Gaussian white noise process, if the linearized system equation (A.1) is valid locally (i.e. $e(t)$ is negligible).

Taking expectations on both sides of equation (A.1), we obtain

$$\frac{dx(t)}{dt} = \dot{f}$$  \hspace{0.5cm} (A.2)

Subtracting equation (A.2) from equation (A.1), we obtain

$$\frac{de_x(t)}{dt} = F_x(t)e_x(t) + F_u(t)e_u(t) + F_\theta(t)e_\theta$$  \hspace{0.5cm} (A.3)

which is the differential equation for the state error and will form the basis for the derivation of the differential equation for the state covariance. The solution to equation (A.3) for the state error $e_x(t)$ takes the form

$$e_x(t) = \Phi(t,t_o)e_x(t_o) + \int_{t_o}^{t} \Phi(t,\tau)F_u(\tau)e_u(\tau)d\tau$$

$$+ \int_{t_o}^{t} \Phi(t,\tau)F_\theta(\tau)e_\theta d\tau$$  \hspace{0.5cm} (A.4)

where $\Phi(t,\tau)$ is the transition matrix associated with the linear vector differential equation (A.3), defined by the solution to the matrix differential equation
\[
\frac{d\hat{\Phi}(t,i)}{dt} = F(t)\hat{\Phi}(t,i),
\]  
(A.5)

with initial conditions \(\hat{\Phi}(t,\tau) = I\), the identity matrix.

The state covariance is defined by

\[
P_{xx}(t) = \mathbb{E}\{e_x(t)e_x^T(t)\},
\]  
(A.6)

where \(\mathbb{E}\{\cdot\}\) is the expectation operator.

An expression for \(e_x^T(t)\) can be obtained by transposing both sides of equation (A.4). The product \(e_x(t)e_x^T(t)\) can be obtained by straightforward multiplication. Upon taking expectations on both sides of the expression for the product, using the properties of the Itô stochastic integral and noting that

\[
\mathbb{E}\{e_u(t)e_x^T(t_o)\}, \mathbb{E}\{e_u(t)e_{\theta}^T\}
\]

and their transposes are identically zero, because \(e_u(t)\) is a white noise process, we get

\[
P_{xx}(t) = \Phi(t,t_o)P_{xx}(t_o)\Phi^T(t,t_o) + \int_{t_o}^{t} \Phi(t,\tau)F_u(\tau)P_{uu}(\tau)F_u^T(\tau)\Phi^T(t,\tau)d\tau +
\]

\[
[\int_{t_o}^{t} \Phi(t,\tau)F_{\theta}(\tau)d\tau]P_{\theta\theta}[\int_{t_o}^{t} F_{\theta}(\tau)\Phi^T(t,\tau)d\tau] + \Phi(t,t_o)P_{x\theta}(t_o)
\]

\[
\int_{t_o}^{t} P_{x\theta}(\tau)\Phi^T(t,\tau)d\tau + \int_{t_o}^{t} \Phi(t,\tau)F_{\theta}(\tau)d\tau]P_{x\theta}(t_o)\Phi^T(t,t_o)
\]  
(A.7)

where \(P_{uu}(t)\) is the covariance parameter matrix of \(e_u(t)\) and \(P_{\theta\theta}(t)\) and \(P_{x\theta}(t)\) are as defined in equations (III.20) and (III.23). Differentiating equation (A.7) with respect to \(t\) and using the Liebnitz rule for the derivatives of the integrals and the properties of the transition matrix \(\Phi(t,\tau)\) (viz. \(\Phi(t,t) = I\), \(\Phi(t,\tau) = F(t)\Phi(t,\tau)\)), yields

\[
\frac{dP_{xx}(t)}{dt} = F_x(t)P_{xx}(t) + P_{xx}(t)F_{x}^T(t) + P_{\theta}(t)[P_{\theta\theta}(t_o)\Phi^T(t,t_o) +
\]

142
\[ P_{\theta \theta} \int_{t_0}^{t} F_{\theta}^T(\tau) \phi(\tau, \tau) d\tau + [\phi(t, t_0)P_{x\theta}(t_0) + \int_{t_0}^{t} \phi(\tau, \tau)F_{\theta}(\tau) d\tau]P_{\theta \theta} \]
\[ + F_u(t)P_{uu}(t)F_u^T(t) \]  
(A.8)

It will now be shown that the expression

\[ \phi(t, t_0)P_{x\theta}(t_0) + \int_{t_0}^{t} \phi(\tau, \tau)F_{\theta}(\tau) d\tau]P_{\theta \theta} \]

occurring in equation (A.8) is simply \( P_{x\theta}(t) \). Consider \( P_{x\theta}(t) = E(e_{x\theta}(t)e_{x\theta}^T(t)) \), an expression for which can be obtained by multiplying equation (A.4) by \( e_{-\theta}(t) \) and taking expectations. This expression turns out to be

\[ P_{x\theta}(t) = \phi(t, t_0)P_{x\theta}(t_0) + \int_{t_0}^{t} \phi(\tau, \tau)F_{\theta}(\tau) d\tau]P_{\theta \theta} \]  
(A.9)

which is the desired expression.

Substituting equation (A.9) in (A.8) yields

\[ \frac{dP_{xx}(t)}{dt} = F_x(t)P_{xx}(t) + P_{xx}(t)F_{x}^T(t) + F_{\theta}(t)P_{x\theta}^T(t) + P_{x\theta}(t)F_{\theta}^T(t) + \]
\[ + F_u(t)P_{uu}(t)F_u^T(t) \]  
(A.10)

\( P_{x\theta}(t) \), for use in (A.10), can be obtained as the solution to the following differential equation (derived from (A.9)) by an approach similar to the one used to derive (A.8) from (A.7)

\[ \frac{dP_{x\theta}(t)}{dt} = F_x(t)P_{x\theta}(t) + F_{\theta}(t)P_{\theta \theta} \]  
(A.11)

It should be noted at this stage that all terms in equation (A.10) must possess dimensions of [state$^2$/time]. While the first four terms on the right hand side can be verified to possess these dimensions, the last term has dimensions [state$^2$/time$^2$]. If \( P_{uu}(t) \) is interpreted as the covariance parameter matrix of the continuous-time nonstationary process \( e_u(t) \). However, if as in most hydrologic models, the inputs are assumed to be constant over each time step used in the integration of the system equations, then \( P_{uu}(t) \) can be interpreted as the covariance matrix of a discrete time process \( e_u(t, t+\Delta t) \), in
which case the fifth term does possess dimensions of \([\text{state}^2/\text{time}]\). In order to allow for a more general interpretation of this term, it can be modified to \(\alpha_u F_u(t) P_{uu}(t) F_u^T(t)\), with \(P_{uu}(t)\) being interpreted as in the first of the two ways mentioned earlier and \(\alpha_u\) being a constant factor with dimensions \([\text{time}]\) to account for the discretization of the continuous time input vector \(u(t)\) to one that is constant over a time step of integration.

To account for the fact that an erroneous \(P_{\theta \theta}\) is used in equation (A.11), a dimensionless factor \(\alpha_p\) will be introduced in all terms involving \(P_{x\theta}(t)\) in equation (A.10). Incorporating \(\alpha_p\) and \(\alpha_u\) in equation (A.10) yields,

\[
\frac{dP_{x x}(t)}{dt} = P_x(t) P_{x x}(t) + P_{x x}(t) F_x^T(t) + \alpha_p [F_{\theta}(t) P_{x \theta}(t) + P_{x \theta}(t) F_{\theta}^T(t)] + \alpha_u [F_u(t) P_{uu}(t) F_u^T(t)]
\]

(A.12)

This is the final form for the state covariance differential equation. The input and parameter errors are accounted for through the last two terms on the right hand side and \(\alpha_u\) and \(\alpha_p\) are free parameters in the state-estimator algorithm that are to be estimated from input-output data.
APPENDIX B
DERIVATION OF THE UPDATE EQUATION FOR
THE DERIVATIVE MATRIX $\frac{\partial P}{\partial \alpha}$

The update equation for the derivative matrix $\frac{\partial P}{\partial \alpha}$ is derived in the following (for a scalar $\alpha$) assuming a scalar observation. The basis of the derivation lies in the equations III.36 and III.37 respectively for the gain matrix and state covariance update. Equation III.35 for the gain matrix is

$$K = P^{-1}H^T(HP^{-1}H^T + R)^{-1}$$  \hspace{1cm} (B.1)

$P^{-1}$ is the state covariance matrix prior to update, $H = H(t_k)$ and all time dependences and subscripts are implicit, but omitted for ease of notation. Defining

$$S_c = (H_x^TP^{-1}H_x^T + R)$$  \hspace{1cm} (B.2)

which is a scalar if the observation is also scalar, (B.1) reduces to

$$K = \frac{1}{S_c}P^{-1}H^T$$  \hspace{1cm} (B.3)

Differentiating (B.3) with respect to $\alpha$ yields

$$\frac{\partial K}{\partial \alpha} = \frac{1}{S_c} \frac{\partial P^{-1}}{\partial \alpha} H^T - \frac{1}{S_c^2} P^{-1}H^T \frac{\partial P}{\partial \alpha} H^T$$  \hspace{1cm} (B.4)

where the second term on the right hand side is $\frac{\partial}{\partial \alpha} \left( \frac{1}{S_c} \right) P^{-1}H^T$. If we now define

$$h^2 = H^TH$$  \hspace{1cm} (B.5)

which is also scalar if the observations are scalar, we get

$$\frac{\partial K}{\partial \alpha} = \frac{1}{S_c} \frac{\partial P^{-1}}{\partial \alpha} H^T - \frac{h^2}{S_c^2} P^{-1} \frac{\partial P}{\partial \alpha} H^T$$  \hspace{1cm} (B.6)

Equation III.37 for the state covariance update is
\[ p^+ = (I-KH)p^-(I-KH)^T + KRK^T \]  
(B.7)

where \( p^+ \) is the updated state covariance matrix. Differentiating (B.7) with respect to \( \alpha \) yields

\[
\frac{\partial p^+}{\partial \alpha} = (I-KH) \frac{\partial p^-}{\partial \alpha} (I-KH)^T - \frac{\partial K}{\partial \alpha} HP^- (I-KH)^T - (I-KH) \frac{\partial p^-}{\partial \alpha} H^T \frac{\partial K}{\partial \alpha}^T + \frac{\partial K}{\partial \alpha} RK^T + KR \frac{\partial K}{\partial \alpha}^T
\]  
(B.8)

Noting that

\[ KH = \frac{1}{S_c} p^- H^T H = \frac{2}{S_c} p^- \]  
(B.9)

The first term on the right hand side of (B.8) reduces to

\[
\frac{\partial p^-}{\partial \alpha} - \frac{2}{S_c} p^- \frac{\partial p^-}{\partial \alpha} + \frac{2}{S_c} \frac{\partial p^-}{\partial \alpha} p^- + \frac{4}{S_c^2} p^- \frac{\partial p^-}{\partial \alpha} p^- p^-
\]  
(B.10)

The second term reduces to

\[
\frac{2}{S_c} \frac{\partial p^-}{\partial \alpha} p^- - \frac{4}{S_c} (p^- \frac{\partial p^-}{\partial \alpha} p^- + \frac{\partial p^-}{\partial \alpha} p^- p^-) + \frac{6}{S_c^2} p^- \frac{\partial p^-}{\partial \alpha} p^- p^- p^-
\]  
(B.11)

The third term is the transpose of the second term. The fourth term reduces to (letting \( r = R, \) scalar)

\[
\frac{4 r}{S_c^2} \frac{\partial p^-}{\partial \alpha} p^- - \frac{4}{S_c^2} \frac{\partial p^-}{\partial \alpha} p^- p^-
\]  
(B.12)

and the fifth term is the transpose of the fourth.

Adding after collecting terms yields the final expression for \( \frac{\partial p^+}{\partial \alpha} \) which is

\[
\frac{\partial p^+}{\partial \alpha} = \frac{\partial p^-}{\partial \alpha} + \left(3 - \frac{2r}{S_c}\right) \frac{4}{S_c^2} p^- \frac{\partial p^-}{\partial \alpha} p^- + \frac{2}{S_c} \left[ \frac{r}{S_c} - 2 \right] \frac{\partial p^-}{\partial \alpha} p^- + \frac{\partial p^-}{\partial \alpha} p^- - \frac{6}{S_c^3} [p^- \frac{\partial p^-}{\partial \alpha} p^- + p^- \frac{\partial p^-}{\partial \alpha} p^- p^-] + \frac{4}{S_c^2} \left[ \frac{\partial p^-}{\partial \alpha} p^- p^- + p^- p^- \frac{\partial p^-}{\partial \alpha} \right]
\]  
(B.13)
In case of vector observations, the update equation for the state covariance can be derived as follows:

Equation (B.7) can be rewritten as

$$P^+ = -P^{-H}KT - KHP^{-1}K(HP^{-1}HR)K^{T} \quad (B.14)$$

Substituting from (B.1) in (B.7) leads to (see equation I.8 in text)

$$P^+ = P^- - P^{-H}(HP^{-1}HR)^{-1}HP^- \quad (B.15)$$

Differentiating (B.15) (noting that for any nonsingular square matrix $A$, $\frac{dA^{-1}}{dt} = -A^{-1}\frac{dA}{dt}A^{-1}$) with respect to $\alpha$ gives

$$\frac{\partial P^+}{\partial \alpha} = 2P^- - 2P^-H^T(HP^{-1}HR)^{-1}HP^- - P^-H^T(HP^{-1}HR)^{-1}H \frac{\partial P^-}{\partial \alpha}$$

$$+ P^-H^T(HP^{-1}HR)^{-1}H \frac{\partial P^-}{\partial \alpha}H^T(HP^{-1}HR)^{-1}HP^- \quad (B.16)$$

A general alternative form of the update equation can be derived based on the following alternative expression for the updated state covariance matrix. (See equation 6.2.5. in Schwepe, 1973). The equation III.35 is derived from the following equation by the application of the matrix inversion lemma

$$P^+ = [(P^-)^{-1} + H^T R^{-1}H]^{-1} \quad (B.17)$$

Differentiating B.17 with respect to $\alpha$, (using $\frac{dA^{-1}}{dt} = -A^{-1}\frac{dA}{dt}A^{-1}$), and simplifying,

$$\frac{\partial P^+}{\partial \alpha} = [(P^-)^{-1} + H^T R^{-1}H]^{-1}(P^-)^{-1} \frac{\partial P^-}{\partial \alpha} (P^-)^{-1}[(P^-)^{-1} + H^T R^{-1}H]^{-1} \quad (B.18)$$

Now, using the result $A^{-1}B^{-1} = (BA)^{-1}$, where $A$ and $B$ are two square matrices with identical dimensions, (B.18) can be simplified to

$$\frac{\partial P^+}{\partial \alpha} = [I + P^-H^T R^{-1}H]^{-1} \frac{\partial P^-}{\partial \alpha} [I + H^T R^{-1}HP]^{-1} \quad (B.19)$$
It should be noted at this stage that provided $\mathbf{R}^{-1}$ always exists (which is a reasonable assumption to make), both inverses in (B.19) exist. This is due to the fact that, $\mathbf{R}$ being a nonsingular covariance matrix, $\mathbf{R}^{-1}$ is positive-definite, whence the quadratic form $\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$ is greater than zero. $\mathbf{P}^{-1}$, also a covariance matrix is positive semidefinite, so that $\mathbf{P}^{-1} H^T \mathbf{R}^{-1} \mathbf{H}$ and $\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{P}$ are positive-semidefinite. The identity matrix is always positive definite. Thus the sums $[\mathbf{I} + \mathbf{P}^{-1} H^T \mathbf{R}^{-1} \mathbf{H}]$ and $[\mathbf{I} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{P}]$ are positive definite, consequently their determinants are positive and their inverses exist.

In the case of conceptual hydrologic models, when the matrix $\mathbf{H}$ tends to be sparse and $\mathbf{R}$ relatively smaller than $\mathbf{H} \mathbf{P}^{-1} \mathbf{H}^T$, experience has indicated that forms (B.13) and (B.16) are very sensitive to numerical errors that result from having elements of the $\mathbf{P}$ matrix varying over a few orders of magnitude and the "larger" elements not being reflected in the product $\mathbf{H} \mathbf{P}^{-1} \mathbf{H}^T$. The form (B.19) however, has proved to be very robust with respect to numerical errors, owing mainly to the presence of "$\mathbf{I}$" in the sum.
APPENDIX C

SUMMARY OF THE EQUATIONS INVOLVED IN
THE RECURSIVE ALGORITHM FOR STATE ESTIMATION
AND ESTIMATION OF PARAMETERS OF THE STATE ESTIMATOR

Propagation:

State mean: (equation III.3)
\[
\frac{d\hat{x}(t)}{dt} = f(\hat{x}(t), \hat{u}(t), \hat{\theta})
\]  
(C.1)

State covariance: (equation III.22)
\[
\frac{dP_{xx}(t)}{dt} = F_x(t)P_{xx}(t)X(t) + P_{xx}(t)F_x^T(t) + \alpha_p [F_\theta(t)P_{x\theta}(t) + P_{x\theta}(t)F_\theta^T(t)] + \alpha_u [F_u(t)P_{uu}(t)F_u^T(t)]
\]  
(C.2)

State-parameter cross covariance: (equation III.24)
\[
\frac{dP_{x\theta}(t)}{dt} = F_x(t)P_{x\theta}(t) + F_\theta(t)P_{\theta\theta}(t)
\]  
(C.3)

Derivative matrix \(\frac{\partial P_{xx}}{\partial u}\): (equation III.33)
\[
\frac{d}{dt} \left( \frac{\partial P_{xx}(t)}{\partial u} \right) = F_x(t) \frac{\partial P_{xx}(t)}{\partial u} + \frac{\partial P_{xx}(t)}{\partial u} F_x^T(t) + F_u(t) P_{uu}(t) F_u^T(t)
\]  
(C.4)

Derivative matrix \(\frac{\partial P_{xx}}{\partial p}\): (equation III.34)
\[
\frac{d}{dt} \left( \frac{\partial P_{xx}(t)}{\partial p} \right) = F_x(t) \frac{\partial P_{xx}(t)}{\partial p} + \frac{\partial P_{xx}(t)}{\partial p} F_x^T(t) + F_\theta(t) P_{x\theta}(t) + P_{x\theta}(t) F_\theta^T(t)
\]  
(C.5)

UPDATE:

Gain matrix: (equation III.35)

149
\[ K(t_k) = P^{-}_{xx}(t_k)H_x^T(t_k)H_x(t_k)P^{-}_{xx}(t_k)H_x^T(t_k) + R(t_k) \]^{-1} \\
= \frac{1}{S_c(t_k)} P^{-}_{xx}(t_k)H_x^T(t_k) \\
(C.6) 

State mean update: (equation III.36) \\
\[ \hat{x}^+(t_k) = \hat{x}^-(t_k) + K(t_k)[z(t_k) - h(\hat{x}^-(t_k), u(t_k), \theta)] \] 

State covariance update: (equation III.37) \\
\[ P^+_{xx}(t_k) = [I - K(t_k)H_x(t_k)] P^{-}_{xx}(t_k)[I - K(t_k)H_x(t_k)]^T + K(t_k)R(t_k)K(t_k)^T \] 

State-parameter cross covariance update: (equation III.44) \\
\[ P^{+}_{x\theta}(t_k) = [I - K(t_k)H_x(t_k)] P^{-}_{x\theta}(t_k) - K(t_k)H_{\theta}(t_k)P_{\theta\theta} \] 

Derivative matrices update \( \frac{\partial P}{\partial \alpha} \): (equation III.45.a or III.45.b) \\
\[ \frac{\partial P^+_{xx}(t_k)}{\partial \alpha} = \frac{\partial P^{-}_{xx}(t_k)}{\partial \alpha} + (3 - \frac{2r}{S_c}) \frac{h}{S_c} \frac{4}{S_c^2} P^{-}_{xx}(t_k) \frac{\partial P^{-}_{xx}(t_k)}{\partial \alpha} P^{-}_{xx}(t_k) + \] \\
\[ \frac{h}{S_c} \frac{r}{S_c} - 2 \left[ \frac{\partial P^{-}_{xx}(t_k)}{\partial \alpha} P^{-}_{xx}(t_k) + P^{-}_{xx}(t_k) \frac{\partial P^{-}_{xx}(t_k)}{\partial \alpha} \right] = \] \\
\[ \frac{h}{S_c} \left[ P^{-}_{xx}(t_k)P^{-}_{xx}(t_k) \frac{\partial P^{-}_{xx}(t_k)}{\partial \alpha} P^{-}_{xx}(t_k) + P^{-}_{xx}(t_k) \frac{\partial P^{-}_{xx}(t_k)}{\partial \alpha} \right] + \] \\
\[ \frac{h}{S_c} \left[ \frac{\partial P^{-}_{xx}(t_k)}{\partial \alpha} P^{-}_{xx}(t_k) + P^{-}_{xx}(t_k) \frac{\partial P^{-}_{xx}(t_k)}{\partial \alpha} \right] \] 

(C.9.a.)
where

\[
S_c = [H_x(t_k) p_{xx}(t_k) H_x^T(t_k) + R(t_k)], \text{ scalar}
\]

\[
h^2 = H_x^T(t_k) H_x(t_k), \text{ scalar}
\]

\[
r^\Delta = R \quad \text{, scalar}
\]

\[
\frac{\partial p^+(t_k)}{\partial a} = \left[ I + p_{xx}(t_k) H_x^T(t_k) R^{-1}(t_k) H_x(t_k) \right]^{-1} \frac{\partial p^-(t_k)}{\partial a}.
\]

\[
[I + H_x^T(t_k) R^{-1}(t_k) H_x(t_k) p_{xx}(t_k)]^{-1}
\]

(C.9.b)

Stochastic approximations gain: (equation III.49)

\[
\hat{g}(k) = \frac{S(k-1)g_T(k-1)}{g(k-1) S(k-1)g_T(k-1) + \sigma_\xi^2(k)}
\]

where \( g(k-1) = H_x(t_k) \frac{\partial p^-(t_k)}{\partial a} H_x^T(t_k) \), \( \sigma_\xi^2(k) = 2S_c^2 \)

Estimator-parameters update: (equation III.50)

\[
\hat{a}(k) = \hat{a}(k-1) + \hat{g}(k) [\nu^2(t_k) - S_c(t_k)]
\]

Estimation error covariance update: (equation III.51)

\[
S(k) = S(k-1) - \frac{S(k-1)g_T(k-1)g(k-1)S(k-1)}{g(k-1)S(k-1)g_T(k-1) + \sigma_\xi^2(k)}
\]

(C.12)
APPENDIX D
SMOOTHING FUNCTIONS USED TO REPLACE THE MODAL
CHARACTER OF THE EQUATIONS OF THE ETD MODEL

A. A Smoothing Function to Replace Threshold-Type Behavior at Saturation.

Function: \( FTHR(x) = \tan^{-1}(x) \), where \( x \) is a nondimensional storage equal
to the ratio of the actual to saturated storage. Let \( h \) denote the storage and
\( s \) the saturated storage.

Consider a withdrawal of the type

\[
\frac{dh}{dt} = \begin{cases} 
-kx & x < 1.0 \\
-k & x > 1.0 
\end{cases} \tag{D.1}
\]

A smoothing function is sought such that it approximates equation D.1,
and in addition its behavior near \( x = 0.0 \), should be such that \( x = 0.0 \) is a
stable equilibrium point of (D.1), so that self-corrective action is intro-
duced when there is a tendency to negative values.

The smoothing function chosen has the form

\[
\frac{dh}{dt} = -k \tan^{-1}(x) \quad \forall x \tag{D.2}
\]

Figure 29 depicts the modality and the smoothing function for comparison.
This function is a relatively simple function. One disadvantage of this
function is that it implies closer correspondence to the Darcian hypothesis at
lower saturation ratios, rather than the opposite which is more realistic.

B. A Smoothing Function for the Overland Flow Representation.

Function: \( FOVL(x) = 0.5003(x-1) + (23.2(x-1))^2 - 0.00036(x-1) 
+ 4.6 \times 10^{-8} \sqrt{9.62} \]

This is the equation of a hyperbola with \( y = 0 \) and \( y = x-1 \) as its assympotes
and vertex at \((0.0, 1.0)\). It was obtained by the transormation of the equa-
tion of a hyperbola at the origin with an angle of 135° between assympotes.

The overland flow function is of the form

152
Figure 29. Smoothing function for threshold type behaviour of saturation.
\[ \text{OVLAND} = \begin{cases} 0.0 & x < 1.0 \\ \frac{(x-1.0)}{\Delta t} & x \geq 1.0 \end{cases} \]

where \( x \) is the saturation ratio and \( \Delta t \) is the time step of integration. (FOVL(\( x \)) above is to be divided by \( \Delta t \) for \( \Delta t \) other than 1.0).

Figure 30 depicts the modality and the smoothing function for comparison.

C. A Smoothing Function to Retard Percolation when Lower Compartments Approach Saturation

Function: \( FRET(x) = (1-x^4) \), \( x < 1.0 \)

where \( x \) is the saturation ratio in the lower compartment. Let \( h \) denote the storage, \( s \) the saturation storage and let \( x_u = h/s \) consider the percolation equation

\[ \frac{dh}{dt} = -k x_u \quad x < 1.0 \]

with the provision that if \( x > 1.0 \), \( \frac{dh}{dt} \) is adjusted so that \( x = 1.0 \) (see Nikolaidis et al. 1986).

The smoothing function chosen has the form

\[ \frac{dh}{dt} = -k x_u (1-x)^4 \quad \forall x \]

Figure 31 depicts the modality and the smoothing function for comparison. It is obvious that with this representation, if for some reason (another inflow to the lower compartment) \( x > 1.0 \), the lower compartment spills into the upper. If there is no other inflow to the lower compartment, other than percolation, there is no possibility of \( x > 1.0 \) occurring and the percolation rate gradually reduces to zero as the lower zone saturation deficiency becomes very small. This structure is in fact used in other hydrologic models (e.g. NWSRFS-SMA model), based on the hypothesis that there is indeed reduction in percolation rate with saturation deficiency.

D. A Smoothing Function to Eliminate Evaporation from Soil and Lake Compartments in Presence of Snowcover.

Function: \( FINS(h_2) = \exp(1(10^{-4}h_2)^2) \)

where \( h_2 \) is the depth of snowcover.
Figure 30. Smoothing function for overflow representation.
Figure 31. Smoothing function for the retardation of percolation.
Consider the evaporation equation

\[ \text{EVAPORATION} = \begin{cases} \text{computed evaporation if } h_2 = 0.0 \\ 0.0 \quad \text{if } h_2 > 70.0 \end{cases} \]

A smoothing function that approximates this is

\[ \text{EVAPORATION} = \text{COMPUTED EVAPORATION} \cdot \exp\left(-\left(10^{-4}h_2\right)^2\right)-Vh_2. \]

Figure 32 compares the modallity and the smoothing function. In terms of physical reality, this smoothing function in part accounts for possible partial snowcover over the watershed for lower snowdepths, when the evaporation rate is not completely reduced.
Figure 32. Smoothing function of the elimination of evaporation from the lake and soil in presence of snowcover.
APPENDIX E
DERIVATIVES OF THE ETD MODEL EQUATIONS
WITH RESPECT TO STATE VARIABLES, FREE
PARAMETERS, FIXED PARAMETERS AND INPUT VARIABLES

The ETD model equations are presented in Tables I and III. In this appendix
the derivatives of the model equations with respect to the state variables,
free parameter, fixed parameters and input variables are derived. Only deri-
vatives that are not identically zero are presented.

STATE VARIABLES: h(2), h(3), h(4), h(5), h(6)

FREE PARAMETERS: k, β, KLA3, KPERC3, KPAN3, KLA4, KPERC4, FRAX,
\[\alpha_1, \alpha_2, d_1, KPAN5, UZSSAT\]

FIXED PARAMETERS: CBED, DEP4, DEP6, FLAGR-FLAGR2 (volume-stage-discharge
relationship)

INPUT VARIABLES: TEMPC, PREC, EVAP, QET

Define

\[f_i = \frac{dh(i)}{dt}, i = 2, \ldots 5\]

\[DFTHR(x) = \frac{\partial FTHR(x)}{\partial x}\]

\[DFOVL(x) = \frac{\partial FOVL(x)}{\partial x}\]

\[DFRET(x) = \frac{\partial FRET(x)}{\partial x}\]

\[DFINS(x) = \frac{\partial FINS(x)}{\partial x}\]

\[FINS2 = FINS(h(2))\]

\[FIHR2 = FTHR(H22), FTHR3 = FTHR(H33), FTHR4 = FTHR(H44)\]
FOVL3 = FOVL(H33)

FRET4 = FRET(H44), FRET6 = FRET(H66)

DFINS2 = DFINS(h(2))

DFTHR2 = DFTHR(H22)

DFTHR3 = \( \frac{\partial FTHR3}{\partial h(3)} = DFTHR(H33)/UZSSAT \)

DFTHR4 = \( \frac{\partial FTHR4}{\partial h(3)} = DFTHR(H44)/LZSSAT \)

DFOVL3 = DFOVL(H33)

DFRET4 = \( \frac{\partial FRET4}{\partial h(4)} = DFRET(H44)/LZSSAT \)

DFRET6 = \( \frac{\partial FRET6}{\partial h(6)} = DFRET(H66)/DEP6/PORE6 \)

Snow Compartment:

If QMELT > 0.0

\[ \frac{\partial f_2}{\partial h(2)} = - QMELT*DFTHR2*1/QMELT = - DFTHR2 \]

\[ \frac{\partial f_2}{\partial QMELT} = - FTHR2 + DFTHR2*h(2)/QMELT^2 = - FTHR2 + DFTHR2*H22 \]

\[ \frac{\partial f_2}{\partial k} = \frac{\partial f_2}{\partial QMELT} * \frac{\partial QMELT}{\partial k} \]

\[ \frac{\partial f_2}{\partial TEMPC} = \frac{\partial f_2}{\partial QMELT} * \frac{\partial QMELT}{\partial TEMPC} \]

\[ \frac{\partial f_2}{\partial PREC} = \frac{\partial f_2}{\partial QMELT} * \frac{\partial QMELT}{\partial PREC} \]

where
\[ \frac{\partial QMELT}{\partial k} = \begin{cases} (1.8 \times \text{TEMP}C)^{n+1}/1000. & \text{PREC} = 0.0 \\ 0.0 & \text{PREC} > 0.0 \end{cases} \]

\[ \frac{\partial QMELT}{\partial \text{PREC}} = \begin{cases} k 	imes 1.8^{n+1} \times (n+1) \times \text{TEMP}C^n/1000. & \text{PREC} = 0.0 \\ (0.007 \times \text{PREC} + 0.074) \times 1.8 \times 0.254/10000. & \text{PREC} > 0.0 \end{cases} \]

\[ \frac{\partial QMELT}{\partial \text{PREC}} = \begin{cases} 0.0 & \text{PREC} = 0.0 \\ 0.007 \times 1.8 \times \text{TEMP}C \times 0.254/1000. & \text{PREC} > 0.0 \end{cases} \]

\[ \text{IF QMELT = 0.0} \]

\[ \frac{\partial f_2}{\partial h(2)} = 0.0 \]

\[ \frac{\partial f_2}{\partial \text{QMELT}} = 0.0 \]

\[ \frac{\partial f_2}{\partial k} = 0.0 \]

\[ \frac{\partial f_2}{\partial \text{TEMP}C} = 0.0 \]

\[ \frac{\partial f_2}{\partial \text{PREC}} = \begin{cases} 1.0 & \text{TEMP}C < 0.0 \\ 0.0 & \text{TEMP}C > 0.0 \end{cases} \]

**Soil Compartment**

\[ \frac{\partial f_3}{\partial h(2)} = -(1-\beta) \times \text{PER} \times \frac{\partial f_2}{\partial h(2)} - \text{KPAN3} \times \text{EVAP} \times \text{PER} \times \text{DFINS2} \]

\[ \frac{\partial f_3}{\partial h(3)} = -(\text{KLAT3} \times \text{CLAT} \times \text{KHU} \times \text{FR} \times \text{DEP3}^2 \times \text{FTHR3} + \text{KPERC3} \times \text{PER} \times \text{KVU} \times \text{FR} \times \text{FRET4} + \text{KPAN3} \times \text{EVAP} \times \text{PER} \times \text{FINS2}) \times \text{DFTHR3} - \text{DFOVL3}/\text{TSPAN} \]

\[ \frac{\partial f_3}{\partial h(4)} = -(\text{KLATA3} \times \text{CLAT} \times \text{KHU} \times \text{DEP3} \times \text{FTHR3}^2 + \text{KPERC3} \times \text{PER} \times \text{KVU} \times \text{FTHR3} \times \text{FRET4}) \times \frac{\partial \text{FR}}{\partial h(4)} - \text{KPERC3} \times \text{PER} \times \text{KVU} \times \text{FR} \times \text{FTHR3} \times \text{DFRET4} \]

\[ \text{where} \]

\[ \frac{\partial \text{FR}}{\partial h(4)} = (1-\text{FR1}) \times -8.0 \times (1.0 - 44)^7 / \text{LZSSAT} \]

161
\[ \frac{\partial f_3}{\partial k} = -(1-\beta) \cdot \text{CPR} \cdot \frac{\partial f_2}{\partial k} \]
\[ \frac{\partial f_3}{\partial B} = -\text{QMLE} \cdot \text{FTHR} \cdot \text{CPR} \]
\[ \frac{\partial f_3}{\partial KLAT} = -\text{CLAT} \cdot \text{KHUD} \cdot \text{FR} \cdot \text{DEP} \cdot \text{FTHR}^2 \]
\[ \frac{\partial f_3}{\partial KPERC} = -\text{CPR} \cdot \text{KVU} \cdot \text{FR} \cdot \text{FTHR} \cdot \text{FRET} \]
\[ \frac{\partial f_3}{\partial KPA} = -\text{CPR} \cdot \text{EVAP} \cdot \text{FTHR} \]
\[ \frac{\partial f_3}{\partial VZSAT} = -\left( \text{KLAT} \cdot \text{CLAT} \cdot \text{KHUD} \cdot \text{FR} \cdot \text{DEP} \cdot \text{FTHR} \cdot \text{FRET} + \text{KPERC} \cdot \text{CPR} \cdot \text{KVU} \cdot \text{FR} \cdot \text{FRET} + \text{KPA} \right) \cdot \frac{\partial \text{FTHR}}{\partial \text{VZSAT}} - \left( \text{UZSAT} \cdot \frac{\partial \text{FOVL}}{\partial \text{UZSAT}} + \text{FOVL} \right) / \text{TSPAN} \]

where \[ \frac{\partial \text{FTHR}}{\partial \text{UZSAT}} = \text{DFTHR} \cdot -\frac{\text{H}(3)}{\text{UZSAT}} = -\text{DFTHR} \cdot \text{H3} \]
\[ \frac{\partial \text{FOVL}}{\partial \text{UZSAT}} = \text{DFOVL} \cdot -\frac{\text{H}(3)}{\text{UZSAT}} = -\text{DFOVL} \cdot \text{H3} / \text{UZSAT} \]

\[ \frac{\partial f_3}{\partial \text{DEP}} = -\left( \text{KLAT} \cdot \text{CLAT} \cdot \text{KHUD} \cdot \text{DEP} + \text{KPERC} \cdot \text{CPR} \cdot \text{KVU} \cdot \text{PRET} + \text{KPA} \right) \cdot \frac{\partial \text{FR}}{\partial \text{DEP}} - \text{KPERC} \cdot \text{CPR} \cdot \text{KVU} \cdot \text{FR} \cdot \frac{\partial \text{FRET}}{\partial \text{DEP}} \]

where \[ \frac{\partial \text{FR}}{\partial \text{DEP}} = \left( 1.0 - \text{FRL} \right) - 8.0 \cdot \left( 1.0 - \text{H44} \right)^7 - \text{h}(4) / \text{LZSAT}^2 \cdot \text{POR} \]
\[ \frac{\partial \text{FRET}}{\partial \text{DEP}} = -\text{DFRET} \cdot (\text{H44})^h(4) / \text{LZSAT}^2 \cdot \text{POR} \]

\[ \frac{\partial f_3}{\partial \text{TEMP}} = -(1-\beta) \cdot \text{CPR} \cdot \frac{\partial f_2}{\partial \text{TEMP}} \]

\[ \frac{\partial f_3}{\partial \text{PREC}} = \begin{cases} 0.0 & \text{TEMP < 0.0 OR PREC = 0.0} \\ -(1-\beta) \cdot \text{CPR} \cdot \frac{\partial f_2}{\partial \text{PREC}} + \text{CPR} & \text{TEMP > 0.0, PREC > 0.0} \end{cases} \]

\[ \frac{\partial f_3}{\partial \text{EVAP}} = -\text{KPA} \cdot \text{CPR} \cdot \text{FTHR} \cdot \text{FINS} \]

162
Unsaturated Zone Compartment:

\[
\frac{\partial f_4}{\partial h(3)} = K_{PERC3}*C_{PER}*K_{UV}*F_{R}*F_{RET4}*D_{FTHR3}
\]

\[
\frac{\partial f_4}{\partial h(4)} = (K_{PERC3}*C_{PER}*K_{UV}*F_{THR3})*(F_{R}*D_{FRET4} + F_{RET4} + \frac{\partial F_R}{\partial h(4)})
- (K_{LAT4}*C_{LAT}*K_{HL}*D_{P4} + K_{PERC4}*C_{PER}*K_{VL}*F_{RET6} + Q_{ET}*F_{RAX})*D_{FTHR4}
\]

\[
\frac{\partial f_4}{\partial h(6)} = - K_{PERC4}*C_{PER}*K_{VL}*F_{THR4}*D_{FRET6}
\]

\[
\frac{\partial f_4}{\partial K_{PERC3}} = C_{PER}*K_{UV}*F_{R}*F_{THR3}*F_{RET4}
\]

\[
\frac{\partial f_4}{\partial K_{LAT4}} = - C_{LAT}*K_{HL}*D_{P4}*F_{THR4}
\]

\[
\frac{\partial f_4}{\partial E_{PERC4}} = - C_{PER}*K_{VL}*F_{RET6}*F_{THR4}
\]

\[
\frac{\partial f_4}{\partial F_{RAX}} = - Q_{ET}*F_{THR4}
\]

\[
\frac{\partial f_4}{\partial U_{ZSSAT}} = K_{PERC3}*C_{PER}*K_{UV}*F_{R} + \frac{\partial F_{THR3}}{\partial U_{ZSSAT}}*F_{RET4}
\]

\[
\frac{\partial f_4}{\partial D_{P4}} = - \frac{\partial f_3}{\partial D_{P4}} - C_{LAT}*K_{LAT4}*K_{HL}*F_{THR4} - (Q_{ET}*F_{RAX} + K_{PERC3}*C_{PER} + K_{VL}*F_{RET6} + C_{LAT}*K_{LAT4}*K_{HL}*D_{P4})*\frac{\partial F_{THR4}}{\partial D_{P4}}
\]

where \[
\frac{\partial F_{THR4}}{\partial D_{P4}} = - D_{FTHR}(H44)*h(4)/L_{ZSSAT}^2*P_{ORE4}
\]

\[
\frac{\partial f_4}{\partial D_{P6}} = - K_{PERC3}*C_{PER}*K_{VL}*F_{THR4} + \frac{\partial F_{RET6}}{\partial D_{P6}}
\]
where \( \frac{\partial \text{DFRET6}}{\partial \text{DEP6}} = - \text{DFRET(H66)}/\text{DEP6}^2/\text{PORE6} \)

\( \frac{\partial \text{f_4}}{\partial \text{QET}} = - \text{FRAX*FTHR4} \)

Lake Compartment:

\( \frac{\partial \text{f_5}}{\partial \text{h(2)}} = -(1-(1-\beta)*\text{CPER}) \cdot \frac{\partial \text{f_2}}{\partial \text{h(2)}} - \text{KPAN5*EVAP*(1-\text{CPER})*DFINS2} \)

\( \frac{\partial \text{f_5}}{\partial \text{h(3)}} = \text{KLAT3*CLAT*KHU*FR*DEP3*2*FTHR3*DFTHR3 + DFOVL3/TSPAN} \)

\( \frac{\partial \text{f_5}}{\partial \text{h(4)}} = \text{KLAT3*CLAT*KHU*DEP3*FTHR3^2} * \frac{\partial \text{FR}}{\partial \text{DEP3}} + \text{KLAT4*CLAT*KHL*DEP4*DFTHR4} + \frac{\partial \text{HGR}}{\partial \text{h(4)}} \cdot ((1-d_1\alpha_1-d_1\alpha_2)*\text{CSEEP*STAGE}}\)

\( \frac{\partial \text{f_5}}{\partial \text{h(5)}} = ((1-d_1) \cdot \alpha_2-d_1 \cdot \alpha_1) \cdot \text{CSEEP*(HGR \frac{\partial \text{STAGE}}{\partial \text{h(5)}} + \text{STAGE*SGN*}} \frac{\partial \text{STAGE}}{\partial \text{h(5)}} - \frac{\partial \text{FLAGR2(STAGE)}}{\partial \text{h(5)}} \)

where \( \frac{\partial \text{STAGE}}{\partial \text{h(5)}} = \frac{\partial \text{FLAGR(h(5))*AREATT)}}{\partial \text{h(5)}} \)

\( \text{SGN} = \begin{cases} 1.0 & \text{if STAGE} > (\text{h(6)}/\text{PORE6} + \text{h(4)}/\text{PORE4}) \\ -1.0 & \text{if STAGE} < (\text{h(6)}/\text{PORE6}/\text{h(4)}/\text{PORE4}) \end{cases} \)

\( \frac{\partial \text{HGR}}{\partial \text{h(4)}} = - \text{SGN/PORE4} \)

\( \frac{\partial \text{f_5}}{\partial \text{h(6)}} = ((1-d_1) \cdot \alpha_2-d_1 \cdot \alpha_1) \cdot \text{CSEEP*STAGE*(-SGN/PORE6)} \)

\( \frac{\partial \text{f_5}}{\partial \text{k}} = -(1-(1-\beta)*\text{CPER})*\frac{\partial \text{f_2}}{\partial \text{k}} \)

\( \frac{\partial \text{f_5}}{\partial \text{b}} = - \frac{\partial \text{f_3}}{\partial \text{b}} \)

\( \frac{\partial \text{f_5}}{\partial \text{KLAT3}} = \text{CLAT*KHU*FR*DEP3*FTHR3^2} \)

\( \frac{\partial \text{f_5}}{\partial \text{KLAT4}} = \text{CLAT*KHL*DEP4*FTHR4} \)
\[ \frac{\partial f_5}{\partial \alpha_1} = -d_1 \cdot \text{CSEEP} \cdot \text{STAGE} \cdot \text{HGR} \]

\[ \frac{\partial f_5}{\partial \alpha_2} = (1 - d_1) \cdot \text{CSEEP} \cdot \text{STAGE} \cdot \text{HGR} \]

\[ \frac{\partial f_5}{\partial d_1} = - (\alpha_2 + \alpha_1) \cdot \text{CSEEP} \cdot \text{STAGE} \cdot \text{HGR} \]

\[ \frac{\partial f_5}{\partial \text{UZSSAT}} = \text{KLAT3} \cdot \text{CLAT} \cdot \text{KHU} \cdot \text{FR} \cdot \text{DEP3} \cdot 2 \cdot \text{FTHR3} \cdot \frac{\partial \text{FTHR3}}{\partial \text{UZSSAT}} + \]

\[ (\text{UZSSAT} \cdot \frac{\partial \text{FOVL3}}{\partial \text{UZSSAT}} + \text{FOVL3}) / \text{TSPAN} \]

\[ \frac{\partial f_5}{\partial \text{CBED}} = ((1 - d_1) \alpha_2 - d_1 \alpha_1) \cdot \text{STAGE} \cdot \text{HGR} \cdot \text{PERIL} / \text{DIST} / \text{AREATT} \]

\[ \frac{\partial f_5}{\partial \text{DEP4}} = \text{KLAT4} \cdot \text{CLAT} \cdot \text{KHL} \cdot \text{FTHR4} + \text{KLAT4} \cdot \text{CLAT} \cdot \text{KHL} \cdot \text{DEP4} \cdot \frac{\partial \text{FTHR4}}{\partial \text{DEP4}} \]

\[ \frac{\partial f_5}{\partial \text{STAGE}} \text{ is computed as follows:} \]

\[ \frac{\partial f_5}{\partial \text{STAGE}} = \frac{\partial f_5}{\partial \text{FLAGR2(STAGE)}} \cdot \frac{\partial \text{FLAGR2(STAGE)}}{\partial \text{STAGE}} \]

\[ = -1 \cdot \frac{\partial \text{FLAGR2(STAGE)}}{\partial \text{STAGE}} \]

\[ \frac{\partial f_5}{\partial \text{TEMPC}} = - (1 - (1 - \beta) \cdot \text{CPER}) \cdot \frac{\partial f_2}{\partial \text{TEMPC}} \]

\[ \frac{\partial f_5}{\partial \text{PREC}} = \begin{cases} 0.0 & \text{TEMPC} < 0.0 \text{ OR PREC} = 0.0 \\ - (1 - (1 - \beta) \cdot \text{CPER}) \cdot \frac{\partial f_2}{\partial \text{PREC}} + (1 - \text{CPER}) \text{ TEMPC} > 0.0, \text{ PREC} > 0.0 \end{cases} \]

\[ \frac{\partial f_5}{\partial \text{EVAP}} = - \text{KPAN5} \cdot (1 - \text{CPER}) \cdot \text{FINS2} \]

**Groundwater Compartment**

\[ \frac{\partial f_5}{\partial \text{h(4)}} = \text{KPERC4} \cdot \text{CPER} \cdot \text{KVL} \cdot \text{FRET6} \cdot \frac{\partial \text{FTHR4}}{\partial \text{h(4)}} - ((1 - d_1) \alpha_2 - d_1 \cdot \alpha_1) \]

\[ \text{CSEEP} \cdot \text{STAGE} \cdot \frac{\partial \text{HGR}}{\partial \text{h(4)}} \]

163
\[ \frac{3f_6}{3h(5)} = - ((1-d_1)*a_2 - d_1*a_1) + CSEEP*(STAGE*SGN* \frac{3STAGE}{3h(5)} + HGR* \frac{3STAGE}{3h(5)}) \]

\[ \frac{3f_6}{3h(6)} = - \frac{3f_5}{3h(6)} - \frac{3f_4}{3h(6)} \]

\[ \frac{3f_6}{3KPERC4} = - \frac{3f_4}{3KPERC4} \]

\[ \frac{3f_6}{3FRAX} = - \frac{3f_4}{3FRAX} \]

\[ \frac{3f_6}{3\alpha_1} = - \frac{3f_5}{3\alpha_1} \]

\[ \frac{3f_6}{3\alpha_2} = - \frac{3f_5}{3\alpha_2} \]

\[ \frac{3f_6}{3d_1} = - \frac{3f_5}{3d_1} \]

\[ \frac{3QET}{3QET} = - \frac{3QET}{3QET} \]

\[ \frac{3f_6}{3DEP4} = KPERC4*CPER*KVL*FRET6* \frac{3FTHR4}{3DEP4} \]

\[ \frac{3f_6}{3DEP6} = - \frac{3f_4}{3DEP6} \]

\[ \frac{3f_6}{3CBED} = - \frac{3f_5}{3CBED} \]
APPENDIX F

BRIEF DESCRIPTION OF THE SUBROUTINES INVOLVED
IN THE ADAPTIVE ESTIMATION ALGORITHM IN APPENDIX C

In this appendix, the subroutines used in the computer program, that are
related to the adaptive estimation of the parameters u and v, are briefly
described. In addition to some subroutines added to the existing TSFP ver-

tion, some of the already existing subroutines were modified. These are also
described in this appendix. A more extensive description of all subroutines,
with their listing is presented by Rajaram (1987).

Subroutine STOIN

This subroutine reads input of data related to the adaptive estimation. The
input data read are input covariance matrix related data, parameter covar-
iance matrix related data, initial conditions for parameter-state cross covar-
iance, initial values for u, v, and their covariance matrix and integration
schemes desired during the adaptive estimation. It is called by the TSFP
subroutine PMSC, on option.

This subroutine allows for inclusion of proportionality (direct or in-
verse) of standard deviations of variables to their respective values. The
integration schemes that are allowed for are

1) Complete coupled integration using a predictor corrector routine, with
state vector (x), state covariance (P_x), state-parameter cross
covariance (P_xv) and derivatives of state covariance.

ii) Retention of only state vector and state covariance in the coupled
integration using the predictor-corrector scheme and a Runge-Kutta
scheme for P_xv and the derivative matrices, with initial conditions
from the beginning of the time step and decoupled from the rest of
the system.

iii) Same as (i), with P_xv not integrated, but a steady state assumption
used (see eq. III.55) with a provision for computation of a general-
ized inverse of P_x (see eq. III.65) if desired.

167
Subroutine STOUP

This subroutine is called from the TSFP subroutine PXDUPD and updates the derivative matrices \( \frac{\partial \mathbf{F}}{\partial \mathbf{a}} \), \( \mathbf{a}_u \) and \( \mathbf{a}_p \), and their covariance matrix, after each observation is processed. In effect, it implements the equations (C.9) to (C.12).

Subroutine UPPXT

This subroutine is called from the TSFP subroutine UPDT. It updates the state parameter cross-covariance based on eq. (C.8).

Subroutine RUNSET

This subroutine is called from the TSFP subroutine FLOWS, at the beginning of every time step and uses the initial values of all quantities to manage the integration \( P_{x\theta} \) and \( \frac{\partial P_{xx}}{\partial \mathbf{a}} \) matrices using a Runge-Kutta scheme. The integration itself is accomplished by the subroutine RUNGE, to be described below.

Subroutine RUNGE

This subroutine integrates a system of ordinary differential equations using a fourth order Runge-Kutta scheme, using Gill's method.

In addition the above subroutine introduced, the following TSFP subroutines were modified.

Subroutine PMSC

This subroutine reads the control parameters for the TSFP program. An additional control parameter "IQEST" is introduced. For IQEST < 0, the program runs in its original mode. For IQEST = 1, the corresponding option is the use of a Q matrix parameterized based on input and parameter uncertainty, but \( \mathbf{a}_u \) and \( \mathbf{a}_p \) kept fixed. For IQEST > 1, adaptive estimation of \( \mathbf{a}_u \) and \( \mathbf{a}_p \) is performed. For IQEST > 0, the subroutine PMSC calls the subroutine STOIN. When IQEST = 1, the updating based on subroutine STOUP is not performed.
**Subroutine VRPR**

The modified version of this TSFP subroutine which specifies the differential equations for the state covariance matrix propagation is STOVPR. This subroutine allows for (when IQEST > 0) incorporation of input and parameter uncertainty in the Q matrix, and the specification of the differential equations for the propagation of $P_{xθ}$ and $\frac{∂x}{∂θ}$, in case they are integrated coupled with the TSFP array Y.

**Subroutine PRDUPD**

The modified version of this TSFP subroutine which manages the prediction and update of state and state covariance matrices is PRDSTO. This subroutine allows for the inclusion of $P_{xθ}$ and $\frac{∂x}{∂θ}$ in the prediction and update. In addition, it calls the subroutine STOUP to perform the updates related to the adaptive estimation of $α_u$ and $α_p$.