University of Reading

School of Mathematics, Meteorology & Physics

Flood Prediction and Uncertainty

by

Evangelia-Maria Giannakopoulou

August 18, 2008

This dissertation is submitted to the Department of Mathematics and Meteorology in partial fulfilment of the requirements for the degree of Master of Science

Abstract

A grid-based approach to fluvial flood modelling has been investigated in this dissertation. A spatially-distributed hydrological model can simulate flow on an area-wide basis and a runoff production is used to estimate river flows with a simple kinematic

Contents

List of Figures

List of Tables

List of Symbols

Latin

a	Parameter that depends on topography.
a (Superscript)	Analysis value.
b (Subscript)	Sub-surface pathways of water movement.
c	Kinematic wave speed.
$c_{\max}$	Upper limit of storage capacity.
E	Evaporation.
$\mathbf{e}$	Model error.
e (Subscript)	Ensemble value.
f (Superscript)	Forecast value.
$\overline{g}$	Average topographic gradient.
$g_{\max}$	Upper limit of gradient.
H	Nonlinear observation operator.
$\mathbf{H}$	Linear observation operator.
$\mathbf{I}$	Identity matrix.
i (Subscript)	Ensemble member.
$\mathbf{K}$	Kalman gain matrix.
k (Subscript)	Position in discrete time.
L	Length scale.
l	Dimension of rainfall input vector.
l (Subscript)	Flow over land.
M	Nonlinear dynamical model.
$\mathbf{M}$	Linear dynamical model.
m	Number of nonzero singular values in SVD.
N	Ensemble size.
$\mathbf{N}$	Matrix that relates the rainfall input $\mathbf{u}$ to the state vector $\mathbf{x}$.

n

List of Abbreviations

CEH	Centre for Ecology and Hydrology
EKF	Extended Kalman Filter
EnKF	Ensemble Kalman Filter
EnSRF	Ensemble Square Root Filter
ETKF	Ensemble Transform Kalman Filter
KF	Kalman Filter
NWP	Numerical Weather Prediction
ODE	Ordinary Differential Equation

Chapter 1

Introduction

1.1 Background

Two main categories of flood forecasting models have developed in the last decades, 'lumped conceptual models' and 'physically based distributed models' (Tingsanchali, 1974). In this thesis, we focus on distributed hydrological models, such as the Grid-to-Grid flow model by Moore *et al.*, (2006), where a vital issue is the spatial discretization since stream flow data are integrated over catchment areas. Distributed flood models have the ability to take into account changes in the landscape such as topography and land-use and provide spatially and temporally distributed output variables (Moore *et al.*, 2006).

The main sources of uncertainty in flood modelling are initialization errors, (rainfall) input errors and forecast model errors (Leahy *et al.*, 2007). Initialization errors can be reduced by implementing data assimilation methods and well known examples, which are often used in practice, are the Kalman Filter (KF) and its generalizations, such as Ensemble Kalman Filter (EnKF) techniques (Koster *et al.*, 2004). An ensemble approach has been developed to try and deal with rainfall uncertainty, by using ensemble rainfall forecasts as an input to an ensemble flood model. Generally, ensemble flood forecasting is becoming more popular, using ensemble rainfall inputs from Numerical Weather Prediction (NWP) forecasts (Roberts, 2005). The Ensemble Kalman Filter is a natural candidate for initializing ensemble flood models, however, unlike the standard Kalman Filter; it has not been developed for situations where inputs play a significant role (Reichle *et al.*, 2002).

The basic idea of the Ensemble Kalman Filter (EnKF) is to use a statistical sample of state estimates instead of a single estimate. The mean of this ensemble sample represents the 'best' state estimate, while the variance provides a measure of the spread of the ensemble errors (Leahy *et al*., 2007). Also, with the use of a statistical sample in EnKF algorithm we calculate the error covariance matrix from this ensemble instead of maintaining a separate covariance matrix and that leads in a better representation of nonlinearity and is less expensive than the Extended Kalman Filter (Evensen, 2003). Finally, another benefit of the EnKF comes from the calculation of the Kalman gain matrix for all statistical members which decreases the fixed cost of the additional ensemble members (Leahy *et al*., 2007).

1.2 Goals

The goals of this thesis are

- To design and implement a simplified one dimensional (1-D) distributed flow model, based on some of the ideas from the distributed Grid-to-Grid model (Moore *et al*., 2006 and Bell *et al*., 2007).

- To implement an Ensemble Square Root Filter (EnSRF), (Livings *et al*., 2008); the Ensemble Transform Kalman Filter (ETKF), (Bishop *et al*., 2001) in conjunction with this simplified 1-D distributed flow model.

- To modify the ETKF for use with rainfall inputs.

- To investigate the effects of ensemble size and observation frequency on the behaviour of the forecast - assimilation dynamical system.

1.3 Principal Results

A simplified 1-D distributed flow model is selected for implementation in Chapter 4. It is found to be useful to follow a related to the Grid-to-Grid routing scheme that described in Chapter 2 and to assume periodic boundary conditions for reasons of simplicity. Experiments with this simplified 1-D distributed flood model in Chapter 4 (Section 4.6) show that low order numerical schemes, such as the upwind scheme (first order accurate in time and space) used to integrate the simple kinematic wave equation (4.1) of the flow model in Section 4.1, tend to have numerical diffusion.

The Ensemble Transform Kalman Filter (ETKF) using rainfall inputs and the simple flood model, which described in Chapter 4, is selected for implementation in Chapter 5. Experiments with the ETKF, in Chapter 6, show that the usage of a simplified low dimensional distributed flow model and the sequential nature of the ETKF may lead to filter convergence. In view of the experimental results in Chapter 6, we expect that the assimilation results might be quite different when obtained on the basis of a more active assimilation model than the one we use in this research. Such a model will be if we increase the dimension of the state space, the number of days we run the model and the size of ensemble members.

1.4 Outline

In Chapter 2 two main points are selected for discussion. Firstly, we present an overview of flood forecast models, focusing on the Grid-to-Grid flow model (physically based distributed model), and then consider the application of these models to extreme flood conditions. The second focus is on sources of uncertainty in flood modelling.

Chapter 2

Flood Forecasting

With the incidence of severe weather and flooding on the increase around the world, there is a need to improve flood forecasting and warning (Dehotin & Braud, 2008). Floods cause physical damage, loss of basic sanitation that leads to disease, economic hardship due to rebuilding costs and food shortages. They are also the most frequent and costly natural disasters in terms of human hardship and economic loss (Perry, 2000). By improving flood forecasts it becomes possible to take mitigating actions in advance of the

this flood forecast model is the fact that has minimal data requirements, since only a small number of spatial parameters are needed.

A number of studies (Michaud & Sorooshian, 1994) compared simple distributed and lumped conceptual models. Several distributed flood models use algorithms similar to those of conceptual lumped models for runoff production, but in many cases methods have been devised to estimate the spatial variability of model parameters within a basin. In lumped models, model parameters are rela

respectively. A runoff production scheme operates within each grid square and the resulting runoffs considered in the model are surface flow due to precipitation excess and subsurface flow. We consider a simple formulation for the grid-to-grid flow routing which is a model for water movement over the whole terrain (land and river) and the sub-surface. The generated runoffs are translated from cell to cell using a routing scheme

where q is the resulting runoff, P is the depth of precipitation, E is the evaporation, $S_{\max}$ is the maximum water storage capacity for each grid-square and S_0 is the initial depth of water in storage. The details of how the maximum water storage capacity, $S_{\max}$,

2.2.3 Grid-to-Grid Flow Routing

In the Grid-to-Grid flow model the runoff production is routed by using the simple kinematic wave equation:

$$\frac{\partial q}{\partial t} + c\frac{\partial q}{\partial x} = c(u + R) \quad \textbf{(2.3)}$$

where q is the channel flow, c is the kinematic wave speed, u is the lateral inflow per unit length of river and R is the return flow.

The flood model is applied separately in two different layers, as figure 2.3 shows. In this case there are two pathway,ayper yt, on thsu facse fca1(,)5(t)-65(())0.f5a) (y) and f.3(cse(sFlo())-7.3.( )5.1Assu cshjeFmie,i an sub-surfmiea

yr

components of river flow (Bell *et al.*, 2007). The subscripts l and r denote the flow over land and over river pathways respectively and the subscript b denotes the sub-surface pathways of water movement.

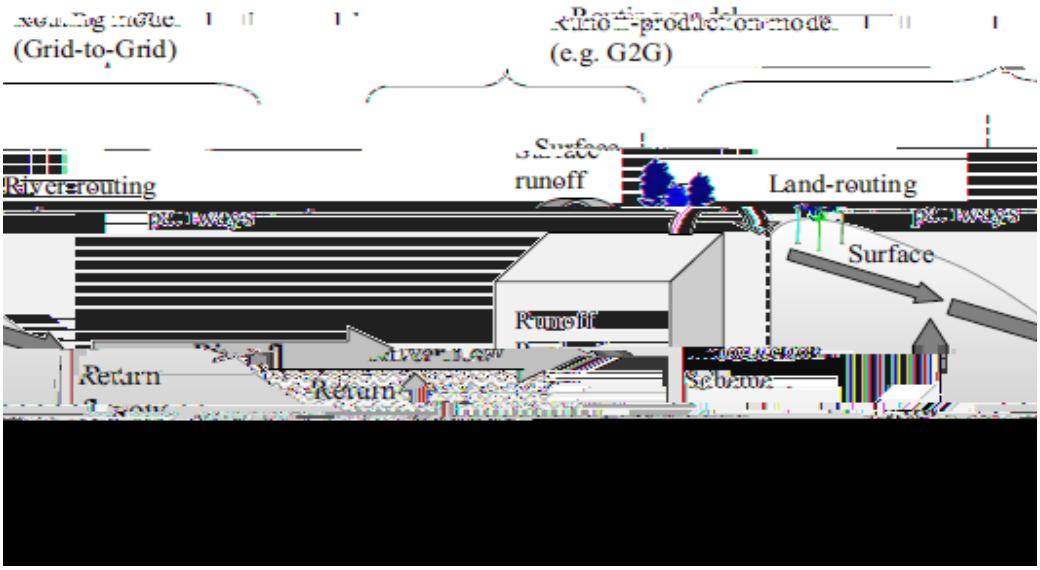


Figure 2.3 Schematic of the Grid-to-Grid model structure from Bell *et al.*, 2007.

To approximate the four partial differential equations by finite differences, we divide a chosen model domain by a set of lines parallel to x-axis and t-axis to form a grid or a mesh. We shall assume that the sets of lines are equally spaces and the line spacings are equal to Δx and Δt such that n and k denote positions in discrete space and time and the crossing points are given by $(x_n \quad n, x, t_n \quad k\Delta t)$.

k and R_k^n is the return flow of the n^{th} space at time k. Equation (2.4) thus represents flow out of the n^{th} space at time k, q_k^n, as a linear weighted combination of the flow out of the reach at the previous time $k-1$, q_{k-1}^n, the inflow to the reach from upstream at the previous time $k-1$, q_{k-1}^{n-1}, and the total lateral inflow along the reach at the same time k, where the total lateral inflow is given by the sum of lateral inflow u_k^n and return flow R_k^n (Moore *et al*., 2006).

However, this scheme is not consistent (we believe there is a typographical error in the paper of Moore *et al.*, 2006). Hence, instead we analyse a similar to the following difference scheme,

$$q_k^n = (1-\theta)q_{k-1}^n + \theta q_{k-1}^{n-1} + \theta\Delta t(u_k^n + R_k^n) \qquad \textbf{(2.5)}$$

which is stable and accurate, simple and quick to run. In the finite difference scheme (2.5), $\theta = c\dfrac{\Delta t}{\Delta x}$ is the dimensionless wave speed and for stability we require $0 < \theta < 1$. But the most useful result of that selection is the fact that the scheme allows for different values of the dimensionless wave speed, θ, for different pathway (surface or sub-surface) and surface type (land or river) combinations, since θ depends on the different values of the kinetic wave speed c. In Chapter 4, which is about the implementation of a simplified 1-D distributed flow model using similar to the grid-to-grid routing scheme, we give an analytic description of how we determine the dimensionless wave speed θ and the order of accuracy of the finite difference scheme.

2.2.4 Parameterization

In the grid-to-grid flood model the flow-routing and the return flow are parameterized as water depths by Moore *et al*., (2006) as follows:

- The routing is given by: $q_k^n = \kappa S_k^n$, where κ is a parameter that depends on soil, geology, terrain and land cover and S_k^n is the depth of water in store over the grid square of the n^{th} reach at time k.
- The return flow is given by: $R_k^n = rS_k^n$, where r is the return flow fraction. Since the return flow fraction is proportional to the depth of the water of the sub-surface store, can take values between zero and one. In this case S_k^n represents the depth

Figure 2.4 highlights where the main source of errors in the flood forecasting procedure. These uncertainties are discussed in detail in the following Sections, rainfall input uncertainty (Section 2.3.1), model uncertainty (Section 2.3.2) and output uncertainty (Section 2.3.3).

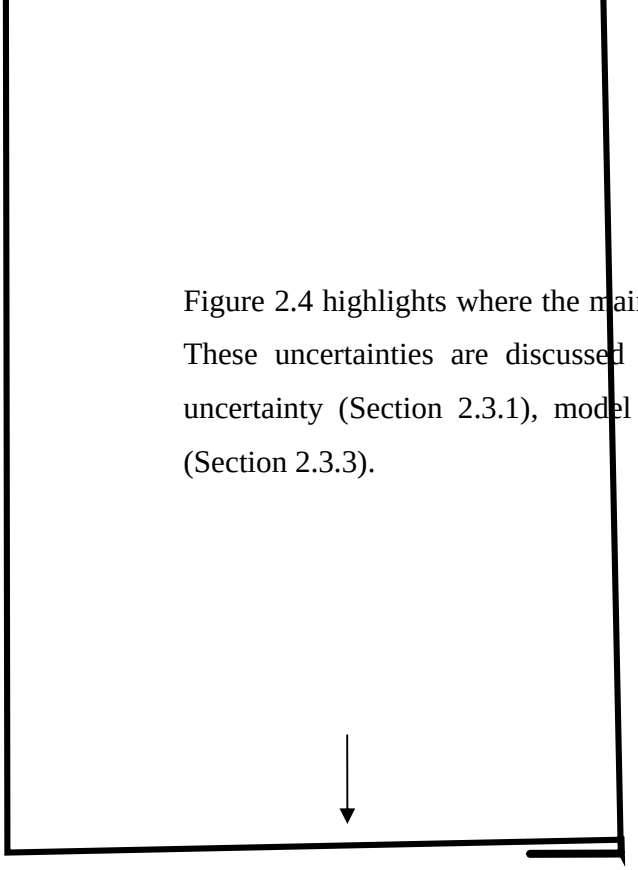


Figure 2.4 Error framework for rainfall-runoff models used in flood forecasting after Leahy *et al.*, 2007.

2.3.1 Input Uncertainty of Rainfall

The main source of uncertainty for models of both distributed and lumped forms is the rainfall input (Leahy *et al.*, 2007). For the distributed rainfall-runoff model, the main input is precipitation and the model output is basin flow. Hence, errors in rainfall measurement lead for example to inaccurate values of water in store and this is one of the situations we need to improve. There are important uncertainties even when precipitation

input to a flood forecasting model is based on recorded rainfall, since radar methods can observe large areas but they do not directly measure rainfall (Leahy *et al*., 2007).

2.3.2 Model Uncertainty

We are not able to entirely model every process of the ‘real’ world, especially if that has to do with flood forecasting. Any flood forecasting model is a gross simplification of reality (Leahy *et al*., 2007). Since we want to achieve a model that works, we make assumptions which lead to errors. These errors however are not resolved with more data and therefore remain constant through an event (Leahy *et al*., 2007).

Errors will also be introduced due to model parameters. In practice, in flood forecasting models the model parameters are used to account errors such as errors in the volume and the distribution of precipitation (Leahy *et al*., 2007). Most of the model parameters have a physical meaning and are determined by the spatial distribution of topography, soil and land cover. However, in flood forecasting parameter errors tend to decrease with time, since more recorded and previous runoff data are available to calibrate the model parameters (Leahy *et al*

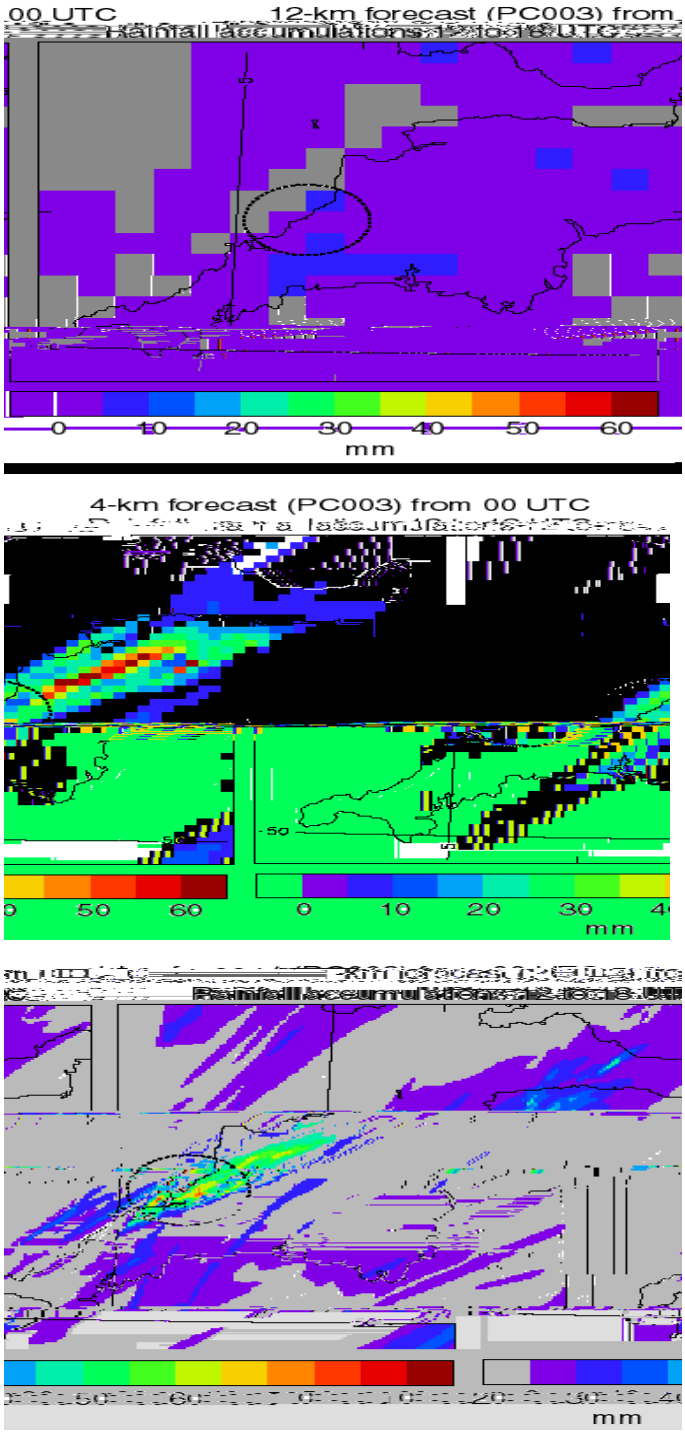


Figure 2.5 Rainfall accumulations over SW England during the period 12 to 18UTC on 16th August 2004 from 12-km, 4-km and 1-km grid-space forecast models starting from 00UTC by May *et al.*, 2004.

2.5 Summary

This Chapter has demonstrated the need to improve the flood forecasting models (lumped conceptual and physically based distributed models) and warning, since floods cause loss of life, human suffering and economic hardship due to rebuilding costs. The distributed Grid-to-Grid model (Section 2.2) show promise in providing an integrated approach to modelling for any location and since there is uncertainty associated with rainfall forecasts whatever the resolution of the flood forecasting model; the Grid-to-Grid model needs improvement.

The main sources of uncertainty in flood modelling are divided in three categories by Leahy *et al.*, (2007): input uncertainty of rainfall (Section 2.3.1), model uncertainty (Section 2.3.2) and output uncertainty (Section 2.3.3). An ensemble approach has been developed to try and deal with rainfall uncertainty, by using ensemble rainfall forecasts as an input to an ensemble flood model. It seems natural to combine this approach with an ensemble data assimilation system. These ideas are discussed in Chapter 3. Finally, in

capable of providing ensembles of rainfall forecasts (Moore *et al*., 2005). This technique developed within the Gandolf and Nimrod systems (rainfall advection nowcasting systems) to produce an ensemble of advection forecasts in which small-scale features are replaced by random noise as the forecast progress (Roberts, 2005). This, thus, will give the additional information of an ensemble of precipitation predictions and provide a probabilistic forecast approach (Roberts.

foundation in Kalman filtering theory, which we will describe in subsequent Sections. Note that the Ensemble Kalman Filter is a natural candidate for use in ensemble flood models such as those we described above with rainfall inputs; however, unlike the standard Kalman Filter, it has not been developed for situations where inputs play a significant role. Hence, we will describe the ensemble filter from this literature in this Chapter and develop it in Chapter 5 to include inputs.

3.3 The Kalman Filter

This Section introduces some notation and gives some desired properties for the Kalman Filter. This is an established sequential data assimilation technique which is characterized by alternate forecast and analysis steps. Generally, in the forecast step a previous state estimate is evolved forward in time to give a forecast state at the time of the latest observations. In the analysis step these observations are used to update the forecast state and to determine the state of the dynamical system by giving an improved state estimate called the 'analysis' (Welch & Bishop, 2006). For a detailed treatment see Welch & Bishop, (2006).

We assume a state vector $\mathbf{x}$ of size n that describes the state of the forecast model. In particular, the true state of the system at time t_k will be denoted by $\mathbf{x}^t(t_k)$. The analysis at this time (denoted with the superscript a) and the forecast (denoted with the superscript f) are given by $\mathbf{x}^a(t_k)$ and $\mathbf{x}^f(t_k)$ respectively and are of size n. The observation vector, of size p, at time t_k will be denoted by $\mathbf{y}(t_k)$.

We shall assume that we use random variables to model errors in the flood forecasting model and in observations. We denote these errors $\mathbf{e}^f = \mathbf{x}^f - \mathbf{x}^t$ and $\mathbf{e}^a = \mathbf{x}^a - \mathbf{x}^t$ for the forecast and analysis, respectively. We assume that these forecasts and analyses are unbiased so that $\langle \mathbf{e} \rangle = 0$ and $\langle \mathbf{e} \rangle = 0$

components and are given by $\mathbf{P}^f = \left\langle (\mathbf{x}^f - \mathbf{x}^t)(\mathbf{x}^f - \mathbf{x}^t)^T \right\rangle$ and $\mathbf{P}^a = \left\langle (\mathbf{x}^a - \mathbf{x}^t)(\mathbf{x}^a - \mathbf{x}^t)^T \right\rangle$ (Welch & Bishop, 2006). Note that the errors and means are all for a single time t_k.

3.3.1 The Kalman Filter Algorithm

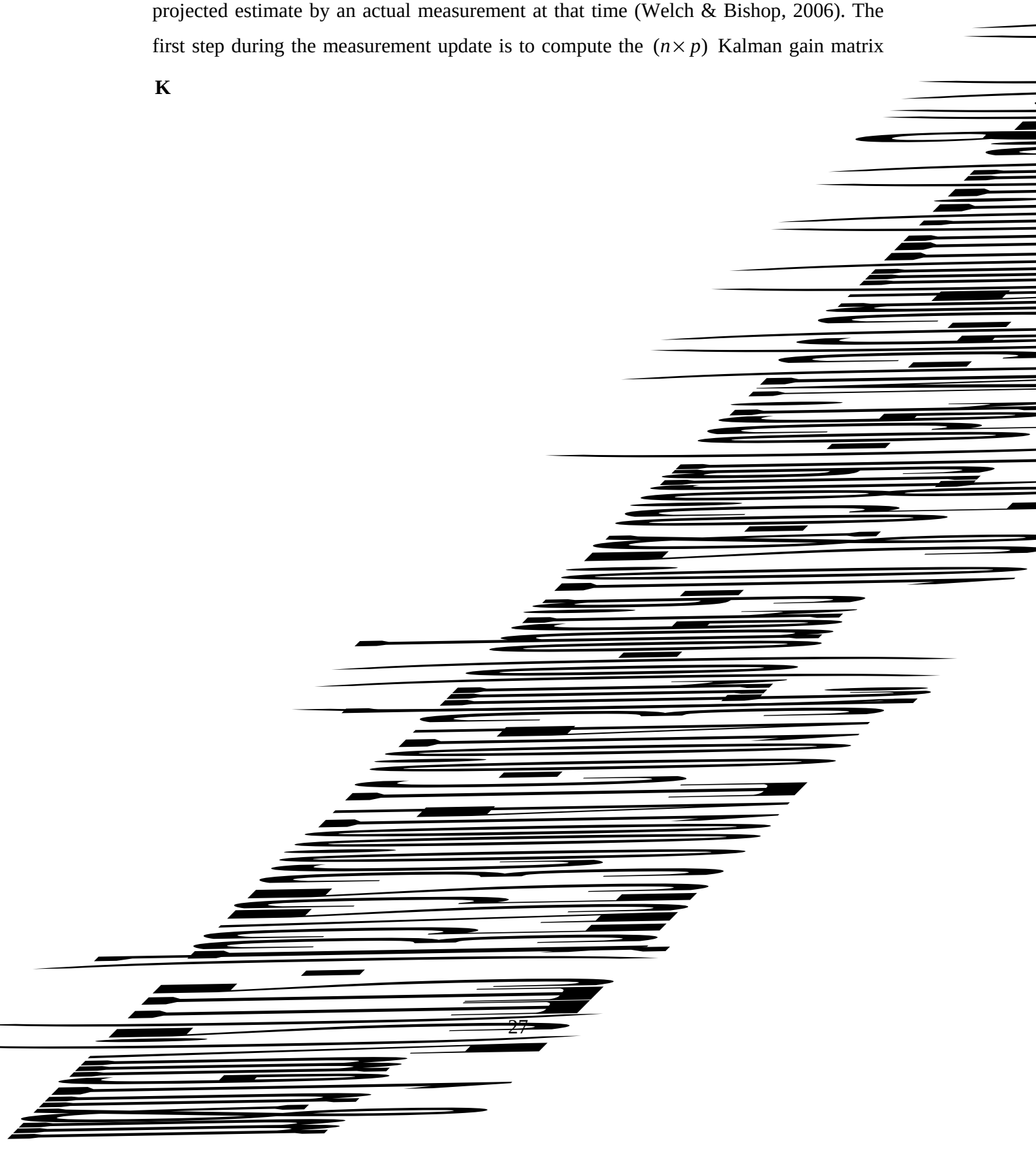
The Kalman Filter (Gelb, 1974) was developed for linear dynamic systems and provided a means of explicitly taking account of input, model and output errors (Srikanthan *et al*., 2007). In this Section, we follow the description of the Kalman Filter algorithm in Welch & Bishop, (2006). We consider the general problem of trying to estimate the state vector $\mathbf{x}$ of a discrete-time controlled process (indicates that the problem is done in steps rather than continuously). The true state of the system at the current time t_k, satisfies

$$(\quad) = \quad {}^{t}(\quad _{k-1}) + \quad (\quad _{k-1}) + \quad (\quad _{k-1})$$

interpolation from model grid to the location of an observation. Also, (t_k) is a Gaussian, random, unbiased and uncorrelated observation noise at the same time t_k with mean zero and known covariance matrix $\mathbf{R}$ (the observation error covariance matrix, a $(p \times p)$

2006). Finally, matrix $\mathbf{P}^a(t_{k-1})$, in equation (3.4), is the 'state error covariance matrix' which is a $(n \times n)$ matrix and describes the random errors in the 'initial guess'.

It is worth noting that the measurement update equations (3.5), (3.6) and (3.7) alter the projected estimate by an actual measurement at that time (Welch & Bishop, 2006). The first step during the measurement update is to compute the $(n \times p)$ Kalman gain matrix **K**



3.4 The Ensemble Kalman Filter

In the last decade the Ensemble Kalman Filter (EnKF) and its derivatives have been used extensively in real time flow forecasting, especially with the *Probability Distributed Model* (Moore *et al*

state from the previous time step is also used for the approximation of the probability function of the actual state (Srikanthan *et al*., 2007). The light blue ellipses represent the model state prediction with uncertainty and the pink ellipses denote the measurement uncertainty. Finally, the Ensemble Kalman Filter combines the forecast with measurements and then the updated state estimate, associated with uncertainty, is shown

3.4.1 Notation

3.4.2 The Forecast Step

The EnKF moves sequentially from one measurement time to the next and is divided into

3.4.3 The Analysis Step

We assume in the beginning an observation $\mathbf{y}$ of dimension p, and an observation operator $\mathbf{H}$ which maps the state vector to the observation vector. We introduce an ensemble of forecast observation $\mathbf{y}_i^f = \mathbf{H}\mathbf{x}_i^f$, where $\mathbf{y}_i^f$ represents that observation if $\mathbf{x}_i^f$

$$\overline{\mathbf{x}^a} = \overline{\mathbf{x}^f} + \mathbf{K}_e(\mathbf{y} - \overline{\mathbf{y}^f}), \tag{3.19}$$

and the analysis ensemble covariance matrix, using equations (3.14) and (3.17), will be defined by

$$\begin{aligned}\mathbf{P}_e^a &= (\mathbf{I} - \mathbf{K}_e\mathbf{H})\mathbf{P}_e^f \\ &= (\mathbf{I} - \mathbf{X}^f(\mathbf{Y}^f)^T\mathbf{S}^{-1}\mathbf{H})\mathbf{X}^f(\mathbf{X}^f)^T\end{aligned}$$

with $\mathbf{T}$ an $(N \times N)$ matrix which we want to satisfy

$$\mathbf{T}\mathbf{T}^T = \mathbf{I} - (\mathbf{Y}^f)^T \mathbf{S}^{-1} \mathbf{Y}^f . \qquad \textbf{(3.23)}$$

Thus, using equations (3.22), (3.23) and (3.17), the analysis ensemble covariance matrix

$$\begin{aligned} \mathbf{P}^a &= \mathbf{X}^a (\mathbf{X}^a)^T \\ &= (\mathbf{X}^f \mathbf{T})(\mathbf{X}^f \mathbf{T})^T \\ &= \mathbf{X}^f (\mathbf{I} - (\mathbf{Y}^f)^T \mathbf{S}^{-1} \mathbf{Y}^f)(\mathbf{X}^f)^T \\ &= (\mathbf{X}^f - \mathbf{K}_e \mathbf{Y}^f)(\mathbf{X}^f)^T , \end{aligned} \qquad \textbf{(3.24)}$$

by multiplying the left hand side of the equality with $\mathbf{I} + (\mathbf{Y}^f)^T \mathbf{R}^{-1} \mathbf{Y}^f$ and using the definition (3.18) of matrix $\mathbf{S} = \mathbf{Y}^f (\mathbf{Y}^f)^T + \mathbf{R}$. In this case is easier to compute the $(N \times N)$ matrix $(\mathbf{Y}^f)^T \mathbf{R}^{-1} \mathbf{Y}^f$, since $\mathbf{R}$

$$\mathbf{X}^a = \mathbf{X}^f \mathbf{T}$$

$$\mathbf{X}^f \mathbf{U} \; \mathbf{I} \qquad)^{\frac{1}{2}} \mathbf{U}^T$$

4.2 Analytic Solution

In this Section we present the analytic solution of equation (4.1) for specific functions of the initial condition, the precipitation P and the evaporation E. Our choice of initial condition for this particular problem is given by

$$\begin{aligned} q(x,0) &= f(x) \\ &= 1+\sin(x), \end{aligned} \tag{4.4}$$

for $0 \le x \le 2\pi$. Then, we distribute precipitation and evaporation over the hours of day and we assume that the precipitation P is changing over time with the following function

$$P(t) = 1+\sin(wt+\delta), \tag{4.5}$$

which is a time varying function with w the frequency of rainfall and δ the phase. The above assumption for precipitation P is not very realistic, but the simple function we chose is useful for the calculation of the analytic solution, to enable us to validate the code. We assume that time $t=0$ corresponds to midnight on the first day ($t=1$ is midnight on the second day) and each time unit corresponds to 24 hours. The first panel

that case, we chose the phase 5 /4

constant along the characteristics and will be points along the $t = 0$ axis in the $x - t$ plane; on the other hand the new variable s will vary along the characteristic line. If, we use the form

$$q(x,t) = q(x(s),t(s)),$$

where $(x(s),t(s))$ is a characteristic line we have, using the chain rule, that

$$\frac{d}{ds}q(x(s),t(s)) = \frac{\partial q}{\partial x}\frac{dx}{ds} + \frac{\partial q}{\partial t}\frac{dt}{ds}. \qquad \textbf{(4.7)}$$

The left hand side of the PDE (4.1) is given if we set $\frac{dx}{ds} = c$ and $\frac{dt}{ds} = 1$; the right hand side of equation (4.1) is given if we set $\frac{dq}{ds} = a(P - E)$. Thus, tn Tm0 06 19574.4 m2(P)T9u6415echb30 (

si

$$\frac{dq}{ds} = a(1+\sin(wt+\delta) - (0.05+\beta\sin(zt+\gamma)). \qquad \textbf{(4.10)}$$

By integrating this equation, where $t = s$, we obtain

$$q(x,t) = a(1-0.05)t - \frac{a}{w}\cos(wt+\delta) + \frac{\beta a}{z}\cos(zt+\gamma) + g(x_0). \qquad \textbf{(4.11)}$$

To determine the function

4.3 Numerical Implementation

To approximate the partial differential equation (4.1) by finite differences, we divide a chosen model domain by a set of lines parallel to x-axis and t-axis to form a grid. We assume that the sets of lines are equally spaced and the line spacings are equal to x

$$\frac{q_k^n - q_{k-1}^n}{\Delta t} + c\frac{q_{k-1}^n - q_{k-1}^{n-1}}{\Delta x} = a(P_k^n - E_k^n), \qquad \textbf{(4.15)}$$

and finally, from the above equation, we obtain the following finite difference representation for the approximate values at time k:

$$q_k^n = (1-\theta)q_{k-1}^n + \theta q_{k-1}^{n-1} + a\Delta t(P_k^n - E_k^n), \qquad \textbf{(4.16)}$$

where θ is the dimensionless wave speed equal to $\theta = c\frac{\Delta t}{\Delta x}$, P_k^n and E_k^n represent the precipitation and the evaporation of the n^{th} space at time k, respectively.

4.4 Accuracy of the finite difference scheme

To determine the order of accuracy of equation (4.15) we use the truncation error which defined by

$$\tau_k^n = \frac{q_k^n - q_{k-1}^n}{\Delta t} + c\frac{q_{k-1}^n - q_{k-1}^{n-1}}{\Delta x} - a(P_k^n - E_k^n). \qquad \textbf{(4.17)}$$

Assuming that the channel flow q, the precipitation P and the evaporation E are smooth functions of space and time, we expand in Taylor series about (x_n, t_{k-1}) in powers of Δx and Δt:

$$q_{k-1}^n = q(n\Delta x, (k-1)\Delta t) = q(x_n, t_k - \Delta t) = q(x_n, t_{k-1}) \equiv q \qquad \textbf{(4.18)}$$

$$q_{k-1}^{n-1} = q((n-1)\Delta x,(k-1)\Delta t) = q(x_n - \Delta x, t_k - \Delta t) = q(x_n - \Delta x, t_{k-1})$$

$$= q(x_n, t_{k-1}) - \Delta x q_x(x_n, t_{k-1}) + \frac{\Delta x^2}{2!} q_{xx}(x_n, t_{k-1}) - \frac{\Delta x^3}{3!} q_{xxx}(x_n, t_{k-1}) + \mathrm{O}(\Delta x^4)$$

$$= q - \Delta x q_x + \frac{\Delta x^2}{2} q_{xx} - \frac{\Delta x^3}{6} q_{xxx} + \mathrm{O}(\Delta x^4) \qquad \textbf{(4.20)}$$

$$P_k^n = P(n\Delta x, k\Delta t) = P(x_n, t_k - \Delta t + \Delta t) = P(x_n, t_{k-1} + \Delta t)$$

$$= P(x_n, t_{k-1}) + \Delta t P_t(x_n, t_{k-1}) + \frac{\Delta t^2}{2!} P_{tt}(x_n, t_{k-1}) + \frac{\Delta t^3}{3!} P_{ttt}(x_n, t_{k-1}) + \mathrm{O}(\Delta t^4)$$

$$= P + \Delta t P_t + \frac{\Delta t^2}{2} P_{tt} + \frac{\Delta t^3}{6} P_{ttt} + \mathrm{O}(\Delta t^4) \qquad \textbf{(4.21)}$$

$$E_k^n = E(n\Delta x, k\Delta t) = E(x_n, t_k - \Delta t + \Delta t) = E(x_n, t_{k-1} + \Delta t)$$

$$= E(x_n, t_{k-1}) + \Delta t E_t(x_n, t_{k-1}) + \frac{\Delta t^2}{2!} E_{tt}(x_n, t_{k-1}) + \frac{\Delta t^3}{3!} E_{ttt}(x_n, t_{k-1}) + \mathrm{O}(\Delta t^4)$$

$$= E + \Delta t E_t + \frac{\Delta t^2}{2} E_{tt} + \frac{\Delta t^3}{6} E_{ttt} + \mathrm{O}(\Delta t^4) \qquad \textbf{(4.22)}$$

We can make the substitution of equations (4.18), (4.19), (4.20), (4.21) and (4.22) into the truncation error τ_k^n (4.17) and cancel:

$$\tau_k^n = \frac{1}{\Delta t}\left[q + \Delta t q_t + \frac{\Delta t^2}{2} q_{tt} + \frac{\Delta t^3}{6} q_{ttt} + \mathrm{O}(\Delta t^4) - q \right.$$

$$+ \frac{c}{\Delta x}\left[q - \left[q - \Delta x q_x + \frac{\Delta x^2}{2} q_{xx} - \frac{\Delta x^3}{6} q_{xxx} + \mathrm{O}(\Delta x^4) \right] \right.$$

$$- a\left[P + \Delta t P_t + \frac{\Delta t^2}{2} P_{tt} + \frac{\Delta t^3}{6} P_{ttt} + \mathrm{O}(\Delta t^4) \right.$$

$$\left. \left. - \left[E + \Delta t E_t + \frac{\Delta t^2}{2} E_{tt} + \frac{\Delta t^3}{6} E_{ttt} + \mathrm{O}(\Delta t^4) \right] \right] \right]$$

Assuming that channel flow q, precipitation P and evaporation E satisfy the PDE (4.1) and using the relationship $q_t + cq_x = a(P - E)$, we have that the final structure of the truncation error τ_k^n is revealed more clearly by the following:

$$\tau_k^n = \frac{\Delta t}{2} q_{tt} - c\frac{\Delta x}{2} q_{xx} - a\Delta t[P_t - E_t] + \mathrm{O}(\Delta t^2) + \mathrm{O}(\Delta x^2), \qquad \textbf{(4.23)}$$

where we conclude that

$$\tau_k^n = \mathrm{O}(\Delta t) + \mathrm{O}(\Delta x), \qquad \textbf{(4.24)}$$

which shows that the scheme is first order accurate in time and first order accurate in space.

4.5 Stability of the finite difference scheme

To analyse the stability of the finite difference scheme (4.15), we are going to represent the approximated solution at some particular time step by a finite Fourier series and examine the stability of each individual component. We start the Von Neumann's stability analysis (Wesseling, 1996) by assuming that the approximate solution q_k^n is given by a Fourier mode as follows:

$$q_k^n = (A_k)e^{ip(n\Delta x)},$$

where A is the mode amplification factor and p is the mode wave number. Then, we substitute the above expression into the equation (4.15) and for reasons of simplicity we assume that the right hand side of equation (4.15) is equal to zero. Note that this may have implications for the applicability of the stability analysis to the numerical scheme (4.15). This will be discussed further in Section 4.6. Hence, we obtain that:

$$\frac{1}{\Delta}\Big(\quad {}^{\Delta x 50\theta} \quad {}_{k-1}{}^{ipn\Delta x} \Big) + \frac{}{\Delta}\Big(\quad {}_{k-1}{}^{ipn\Delta x} - \quad {}_{k-1}{}^{ip(n-1)\Delta x} \Big) = 0$$

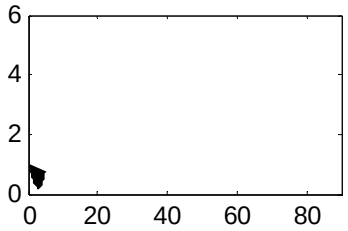
4.6 Validation of the

parameters for application and hence that makes our model simple and quick to run. By

In this figure we give analytic and approximate solutions of equation (4.1) by running the flood model for different values of Δx. In each panel, along the x-axis we plot the time (days) and in y-axis the river flow (volume/time). The first panel in figure 4.2 shows the analytic solution (4.12) obtained as we described in Section 4.2. The second panel (a.) gives the numerical solution of the finite difference scheme (4.16) in a periodic domain with period 2π, for state space dimension equal to 40. The third (b.) and the fourth (c.) panel illustrate again the approximate solution but for different number of space grid points; 150 and 450 grids respectively. These are examples of the evolution of the flow model subject to a small number (40), to a medium number (150) and a large number (450) of space grid points, since we want to show the problems that we face with the numerical diffusion for small state space dimension. Note that the river flow values are taken at a specific grid point. However, the behaviour is qualitatively similar at all grid points.

These experiments have revealed two main points. Firstly, we are able to see that the river flow (and the error also) increases with time. This is to be expected since the rain-rate is twenty times larger than the evaporation rate. The most important result of this experiment in figure 4.2 has to do with the numerical diffusion. As we can see in the second panel (a.) for number of grid points equal to 40 the approximate solution behaves differently from the analytic solution. The amplitude of the river flow oscillations decreases. This is a consequence of the use of our numerical scheme, which has a diffusive character (Morton & Mayers, 2005). In Section 4.3, we discussed that time and space are divided into a discrete grid and the simple kinematic wave equation (4.1) is discretized into finite difference equation (4.15), which in general is more diffusive than the original differential equation (4.12). Consequently, the approximate solution behaves differently from the analytic solution, since the simulated system depends on the type of discretization that is used, which is the upwind scheme. This scheme is first order accurate in time and space and that is one of the reasons that cause numerical diffusion (Morton & Mayers, 2005) especially if we run the flood model for small number of space grid points and for more than 30 days. Usually, higher order numerical methods tend to have less numerical diffusion than low order numerical schemes, such as the upwind scheme that we use in these experiments. As noted by Morton & Mayers, (2005), one of

the approaches that is useful to manage the numerical diffusion is to be careful to have sufficiently many spatial grid points. It is clear from figure 4.2 (b.) and (c.) that we indeed observe better results if we increase the number of grid points.



4.6.2 Validation of the flow model

Our basic measure of estimation error is the difference between the exact solution and the

4.7 Summary

In this Chapter, we described our new simplified 1-D distributed flow model and its numerical implementation. In the presentation of the methodology of the flood model main point was the assumption of periodic boundary conditions (not very realistic). We showed also that the numerical scheme (upwind scheme), which used to discretize the

$$(\mathbf{Y}^f)^T \mathbf{R}^{-1} \mathbf{Y}^f = (\hat{\mathbf{Y}}^f)^T \hat{\mathbf{Y}}^f . \tag{5.2}$$

Using the Singular Value Decomposition (SVD) by Livings, (2005) we have that

$$(\hat{\mathbf{Y}}^f)^T = \mathbf{U} \quad \mathbf{V}^T , \tag{5.3}$$

where $\mathbf{U}$ is the $(N \times N)$ orthogonal matrix, is the $(N \times p)$ diagonal matrix and $\mathbf{V}$ is the $(p \times p)$ orthogonal matrix. Note that $\mathbf{U}$ matrix in equation (5.3) is in fact the same $\mathbf{U}$ matrix in equation (3.26). The singular value matrix, , and the diagonal matrix of eigenvalues, , are related by = ^{T}.

The updated ensemble perturbation matrix is then

$$\begin{aligned} \mathbf{X}^a &= \mathbf{X}^f \mathbf{T} \\ &= \mathbf{X}^f \mathbf{U}(\mathbf{I} + \quad ^T)^{-\frac{1}{2}} \mathbf{U}^T . \end{aligned} \tag{5.4}$$

This choice of matrix $\mathbf{T}$ ensures that the filter is unbiased in the sense of Livings *et al*., (2008). Using equation (3.17) and the SVD (5.3) we conclude to the following expression of the Kalman gain

$$\begin{aligned} \mathbf{K}_e &= \mathbf{X}^f (\mathbf{Y}^f)^T (\mathbf{Y}^f (\mathbf{Y}^f)^T + \mathbf{R})^{-1} \\ &= \mathbf{X} \ (\hat{\mathbf{Y}} \ ) \ (\hat{\mathbf{Y}} \ (\hat{\mathbf{Y}} \ ) \ + \mathbf{I})^{-1} \mathbf{R}^{-\frac{1}{2}} \end{aligned}$$

from right to left and then update the ensemble mean by $\overline{\mathbf{x}^a} = \overline{\mathbf{x}^f} + \mathbf{X}^f \mathbf{U}\mathbf{z}$. With this process we avoid storing the Kalman gain $\mathbf{K}_e$ and we only need to store a vector at each stage of building up $\mathbf{z}$ (Livings, 2005).

The experimental results by implementing the ETKF algorithm in MATLAB are given in detail in Chapter 6. Note that the SVD (5.3) is performed using the standard MATLAB svd function which uses the LAPACK routine DGESVD (Anderson *et al*

Case 1: Rainfall input $\mathbf{u}$ is perfectly known.

In this case there is only one input $\mathbf{u}$ and the only change to the algorithm described in Section 3.3.1 is to take account of the inputs in the forecast step as in equation (3.3). Considering a nonlinear dynamical model M in the state forecast step the ensemble is propagated forward in time using the following nonlinear model:

$$\mathbf{x}_i^f(t_k) = M(\mathbf{x}_i^a(t_{k-1})) + \mathbf{N}\mathbf{u}(t_{k-1}) + \ _i(t_{k-1}), \qquad \textbf{(5.8)}$$

for $1 \le i \le N$. In this relationship $\mathbf{x}_i^f(t_k)$ represents the ith ensemble forecasted state at time t_k, $\mathbf{x}_i^a(t_{k-1})$ is the ith updated ensemble state at the previous time t_{k-1}, $\mathbf{u}(t_{k-1})$ is the precipitation input at the previous time t_{k-1} and $_i(t_{k-1})$ is a Gaussian pseudo-random model noise at the previous time t_{k-1} which for our implementation we assume that is equal to zero. Matrix $\mathbf{N}$, in equation (5.8), is defined based on knowledge of the process given in Section 3.3.1. Note that a more complex (nonlinear) relationship between the state and input $\mathbf{u}$ is possible, and could be dealt with in the same way. The case described here is the case we have done our experiments on.

Then, the ensemble mean, the ensemble perturbation matrix and the ensemble covariance matrix are given by equations (3.12), (3.13) and (3.14) from Section 3.4.2. We assume, after, as in Section 3.4.3 an observation $\mathbf{y}$ of dimension p j and an observation

and then update the ensemble mean by $\overline{\mathbf{x}^a} = \overline{\mathbf{x}^f} + \mathbf{X}^f \mathbf{U}\mathbf{z}$. The above equations can be generalized for both linear and nonlinear observation operators (Livings *et al*., 2008).

Case 2: Rainfall input $\mathbf{u}$ is uncertain and can be treated as a random variable.

5.3 Summary

The purpose of this Chapter was to provide a presentation of the ETKF algorithm

Chapter 6

Experimental Results

This Chapter presents the results of experiments with an ETKF implemented as described in Chapter 5 using observations of the simplified 1-D distributed flow model as implemented in Chapter 4. The experiments with the ETKF differ in the analysis step, the ensemble size, the number of observations and whether we assume perfect or imperfect background.

6.1 Experiments with the ETKF

In this Section we present the experimental results with the ETKF implemented as described in Chapter 5. The experiments were carried out using our MATLAB code, where the filtering part was written by Livings, (2005), but changes and additions have been made for this thesis. Note that the filter code was validated in Livings, (2005) experiments.

Before we start giving some experimental results we need to describe briefly the filter procedure. In the beginning, of the filter code, we declare the global variables; model, precipitation and evaporation parameters, as defined in Chapter 4 (Table 4.2) for the implementation of the simplified 1-D distributed flow model. We set the initial conditions equal to (4.4), which differ for each ensemble member, and we generate the true state using the finite difference scheme (4.16) we used to discretize the simple kinematic wave equation (4.1) in the flood model in Chapter 4. We generate, then, the observations using the truth. We assume perfect (without observation noise) or imperfect (with observation noise) observations which in our experiments differ in time and space.

Particularly, we run the filter code assuming imperfect observations and we generate the ensemble of observations at each update time by giving noise with zero mean and covariance equal to the observation error covariance matrix $\mathbf{R}$. An initial ensemble of state estimates is then chosen. This ensemble of vectors is drawn from a normal distribution with a given covariance matrix and the true initial state. We assume in our experiments perfect or imperfect background, where for the first case the ensemble is translated so that the ensemble mean coincides exactly with the true initial state. These sample points are then propagated through the system. The approximation of forecast state error covariance matrix is made by propagating the ensemble of model states using the updated state from the previous time step. Note that there is no model noise in the ensemble forecasts as well as in the truth forecast. Finally, we have to mention that rainfall inputs (single inputs) and evaporation are specified a priori inside the flow model.

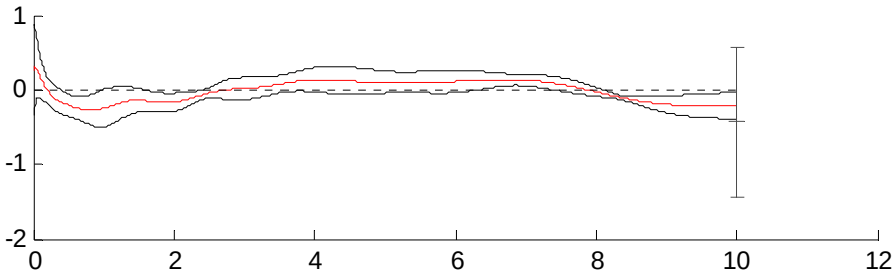
Figures 6.1 and 6.2 show the experimental results by running the ETKF for 10 days over 0.01-time interval and a state space of dimension n=150 (this choice ameliorates numerical diffusion in the model, see Section 4.6). The main point of these figures is to compare the results, if we assume that the estimates are derived with the ETKF using ensemble members $N = 4$ and we run the ETKF for different number of imperfect observations over time. The ensemble size is rather small, although that is more like operations. Examples assuming ensemble size $N = 27$ are given in Appendix B. Specifically, in figures 6.1 and 6.2 we give a sequence of plots where in each graph we simulate 1, 5, 10 and 20 distinct measurements (imperfect observations) that have errors normally distributed around zero with standard deviation 1.0; in figure 6.1 we assume perfect background and in figure 6.2 imperfect. All coordinates are observed in space but in the 1st panel of figures observation is taken at the last time grid point (the last day), in the 2nd panel the measurements are taken every 200 time steps, in the 3rd panel of figures every 100 time steps and finally in the 4th panel observations are taken every 50 time steps up to the 10th day. Figures 6.1 and 6.2, generally, illustrate the difference between the filter and the truth. Specifically, each panel in these figures represents the first component of the output of the ETKF plotted relative to the truth. The value of the true state is indicated by the dotted line at zero. The three solid lines show ensemble mean (red line) and ensemble mean $\pm$ ensemble standard deviation. Finally, in each panel the

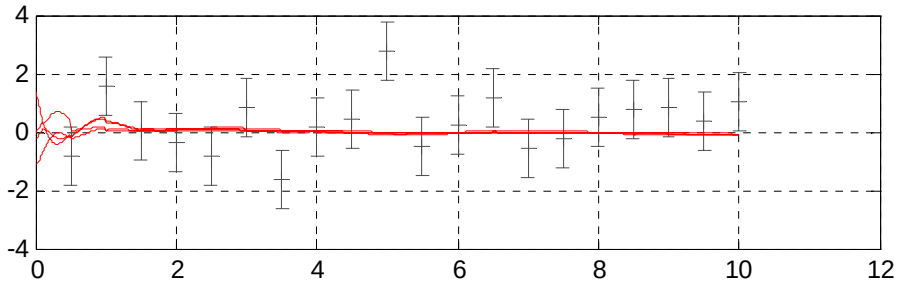
first component of assimilated observations are given as + with error bars indicating the observation standard deviation.

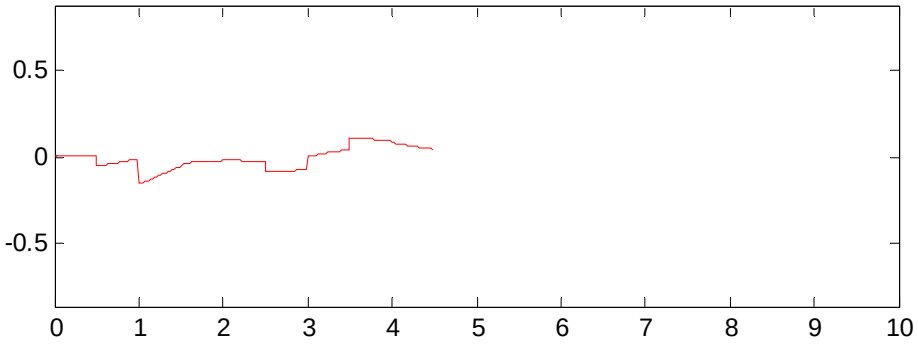
deviation. Note, finally, that the choice of the ensemble size play significant role and we have to point out that for larger number of ensemble members we observe 'better' results. These are illustrated in Appendix B.

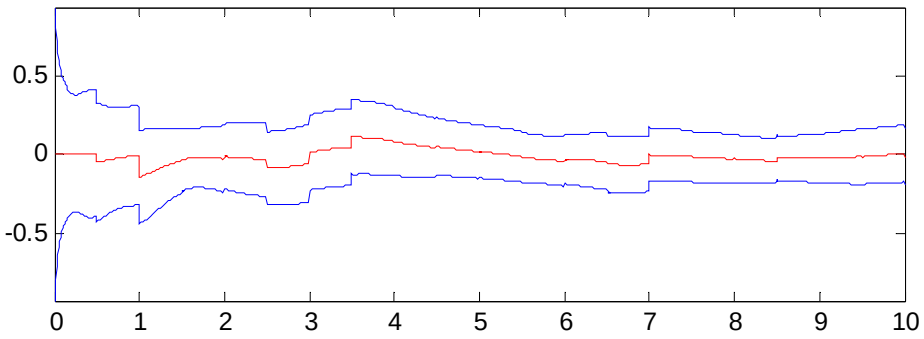
Generally, the use of small ensemble sizes often leads to less accurate results (Livings *et al*., 2008). It is useful to consider how the ETKF estimates converge to the true state as the ensemble size increases. Hence, for that purposes, figure 6.3 illustrates individual ensemble members with the assumption of imperfect observations. For the experimental results in figure 6.3 we run the ETKF for 10 days over 0.01-time interval and a state space of dimension $n=150$, since for small number of space grid points we face problems (numerical diffusion) with the implemented flood model. Each panel of figure 6.3 is for 4, 10, 50 and 100 ensemble members respectively which are represented as red lines. In these experiments, we observe all the components in space and we assume 20 measurements over time plotted as error bars of the 20^{th} coordinate. In figure 6.3, we assume perfect background and we use the same observations for each ensemble. The case of imperfect background is illustrated in Appendix B. An important feature of the these experimental results is the fact that the ensemble spread is getting smaller after the first observation; and that is clear in all panels of figure 6.3. Finally, we are able to observe that after the first couple of observations are assimilated, the ETKF estimates converge to the true state, as we expected. That becomes more obvious when the ensemble size is increased, as in the last three panels of figure 6.3.

It is very important to be mentioned that after several trials of the filter we chose again, for all the experiments, to run the simplified 1-D distributed flow model and hence the ETKF for value of parameter a equal to 0.05 (the parameter that was chosen in our research to depend on soil, geology, land cover etc.). With that value we allow to rainfall and evaporation to have the significant role in the filter, but at the same time we keep a balance as it concerns the increasing values of river flow during the last experimental days. Note that in Chapter 4, we made the assumption that is raining 20 times more than we have evaporation. Hence, it was expected, by increasing the value of the parameter (increasing also the forecast errors) to observe filter divergence and by decreasing the value (decreasing the forecast errors) to obtain stable filter behaviour, whether we assume perfect or imperfect background (the plots are not shown).


1
0
-1
-2
0
2
4
6
8
10
12


4
2
0
-2
-4
0
2
4
6
8
10
12


0.5
0
-0.5
0
1
2
3
4
5
6
7
8
9
10


0.5
0
-0.5
0
1
2
3
4
5
6
7
8
9
10

The ETKF has been performed within contexts of the simplified 1-D distributed flow model, which is a low dimensional model and offers the obvious advantage of a much reduced dimensionality n of the state vector $\mathbf{x}$ (in terms of calculations in the implemented filter). In this implementation of the ETKF we make use of a state space of dimension $n = 150$ (such a small number), for computational reasons. This is an important limitation as reviewed in Ehrendorfer, (2007) and the possibility of generalizing the experimental results may lead

observation frequency on the behaviour of the forecast-assimilation dynamical system. Hence, it is important to be mentioned that the use of small ensemble sizes often led to less accurate results, since we observed that the ETKF estimates converge to the true state as the ensemble size increases. It was clear also, from the experiments, the fact that if more observations were assimilated, during the days we run the filter, the true state was often inside the band defined by the ensemble mean ± ensemble standard deviation, indicating filter convergence. Moreover, the experiments have revealed and the following main points. The first feature pointed in the fact that the choice of perfect or imperfect background may affect the filter behaviour. The selection of perfect background usually led in stable filter behaviour and for the selection of imperfect background we had to take into consideration the fact that the problem was a random sampling problem, where sometimes the samples do not lie within the band of ensemble mean ± ensemble standard deviation. The second feature was the fact that although we observed stable filter behaviour, the usage of the simplified low-dimensional (1-D) distributed flow model and the sequential nature of the ETKF may lead to this filter convergence. Finally, the choice of the values of parameter a which controls the behaviour of the rainfall inputs in our research may lead to different results than the one we observed. However, parameter, space, time, ensemble size and background limitations can result in the accuracy of the filter but we still can have good results.

Chapter 7

Conclusions

7.1 Summary and Discussion

The purposes of this dissertation were to design and implement a simplified one dimensional (1-D) distributed flow model, using a similar to the Grid-to-Grid routing scheme (Moore *et al*., 2006) and to modify an Ensemble Kalman Filter (EnKF), the Ensemble Transform Kalman Filter (ETKF), for use with (rainfall) inputs and the simplified 1-D distributed flow model.

In Chapter 2, we introduced an overview of flood forecast models, focusing on the distributed Grid-to-Grid flow model (Moore *et al*., 2006), and we represented briefly the sources of uncertainty in flood modelling which divided in three categories by Leahy *et al*., (2007): input uncertainty of rainfall, model uncertainty and output uncertainty. An ensemble approach has been developed to try and deal with rainfall uncertainty, by using ensemble rainfall forecasts as an input to an ensemble flood model. It seems natural to combine this approach with an ensemble data assimilation system and these ideas were discussed in Chapter 3.

Our fundamental issue in Chapter 3 related to the description of data assimilation techniques, such as the Kalman (presented in Welch & Bishop, 2006) and Ensemble Kalman Filter methods which are valuable in flood forecasting. The Kalman Filter (KF) developed for linear dynamic systems and provided a means of explicitly taking account of input, model and output uncertainties. For nonlinear dynamic systems EnKF techniques (presented in Evensen, 2003) provided an alternative method of estimating these uncertainties by the use of an ensemble of state estimates instead of a single state estimate and without maintaining a separate error covariance matrix. These data

assimilation methods are useful in flood forecasting, since the use of real-time flood models requires attention to uncertainty estimation and model initialization (i.e. state estimation); problems which can be solved using these techniques. It is important to be mentioned, that the EnKF was not originally designed to take into account (rainfall) inputs and hence the algorithms described without that assumption.

In Chapter 4, the key idea was to introduce a new simplified 1-D distributed flood model, as a subject for experiments, assuming periodic boundary conditions. The assumption of periodic boundary conditions was not very realistic, since few rivers have a loop shape. However, we have chosen these conditions to make our flow model easier to implement numerically. We gave an an

Chapter 5 was about the presentation of an implemented EnKF; the ETKF using the MATLAB code written for Livings, (2005), through modifications and additions have been made for this thesis. The purpose of this Chapter was to provide a complete interpretation of the ETKF and to present new ideas that used in the implemented ETKF. The key idea was to modify the ETKF for use with inputs and a simple flood model, as described in Chapter 4. For our implementation we used single inputs and the possibility of working with an ensemble of inputs may be an area for further investigation.

Finally, in Chapter 6 we gave an analytic presentation of the experimental results using the ETKF implementation of Chapter 5. We gave explanations of the features we observed from the experiments and of the problems that were encountered with the implemented ETKF. We observed that the filter behaviour depends on the assumption of

7.2 Further work

Three main areas may be identified for further investigation, as it concern the

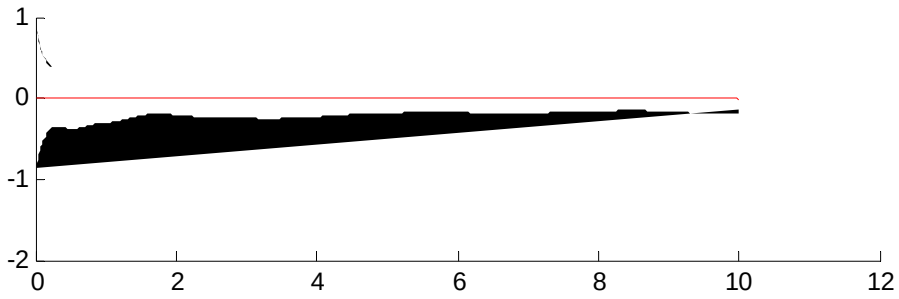
Appendix A

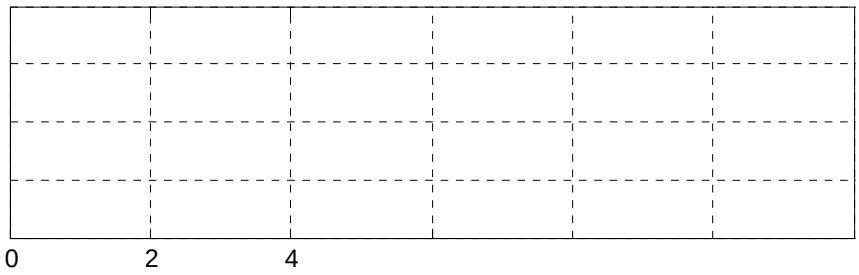
Graphs of the simplified 1-D distributed flow

Appendix B

Graphs of the ETKF

This Appendix is a supplement to Chapter 6. Figures B.1, B.2 and B.3 correspond to figures 6.1, 6.2 and 6.3 of Chapter 6 and represent the experimental results by running the ETKF for ensemble size $N = 27$ for figures B.1 and B.2 and for imperfect background for figure B.3. They have been placed in Appendix because they mainly used for comparison of the results observed in Chapter 6. By increasing the ensemble size we observed 'better' results; filter convergence for both assumptions of perfect and imperfect background in figures B.1 and B.2. Assuming imperfect background in figure B.3 we observed that after the fourth observation is assimilated, the ETKF estimates converge to the true state. The ETKF was run making all the assumptions as in Chapter 6 for each figure.


1
0
-1
-2
0
2
4
6
8
10
12


0
2
4

Bibliography

Anderson, E., Bai, Z., Bischop, C., Blackford, S., Demmel, J., Dongarra, J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A. & Sorensen, D. (1999). *LAPACK User's Guide*. SIAM, 3rd Edition. Available from: www.netlib.org/lapack/lug/lapack_lug.html.

Barlow, R. J. (1989). Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences, John Wiley & Sons Ltd., Chichester.

Bell, V. A., Kay, A. L., Jones, R. G. & Moore, R. J. (2007). Development of a high resolution grid-based river flow model for use with regional climate model output. *Hydrology & Earth System Sciences*, **11**(1), 532-549.

Bishop, C. H., Etherton, B. J. & Majumdar, S. H. (2001). Adaptive sampling with the ensemble transform Kalman Filter, part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420-436.

Daley, R. & Menard, R. (1993). Spectral characteristics of Kalman filter systems for atmospheric data assimilation. *Mon. Wea. Rev.*, **121**, 1554 - 1565.

Dehotin, J. & Braud, I. (2008). Which spatial discretization for distributed hydrological

Gelb, A. (1974). Editor. *Applied Optimal Estimation*. The M.I.T. Press.

Kalnay, E. (2002). *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge, UK: Cambridge University Press.

Koster, T., Sefary, G. E., Heemink, A. W., Boogaard, H. F. P. & Mynett, A. M. (2004). Input correction in rainfall-runoff models using Ensemble Kalman Filtering (EnKF). In: Lee, J. H. W. & Lam, K. M. (Eds.) Proc. 4th Int. Symp. On Environmental Hydraulics, Hong Kong, 15-18 December 2004, Leiden Proc, Balkema, 1991 - 1996. Available from:

Michaud, J. & Sorooshian, S. (1994). Comparison of simple versus complex distributed runoff models on a midsized semiarid watershed. *Water Resources Research*, **30**(3), 593 -605.

Moore, R. J. (2007). The PDM rainfall-runoff model. *Hydrology & Earth System Sciences*, **11**(1), 483 - 499.

Moore, R. J., Bell, V. A. & Jones, D.

Wang, X., Bishop, C. H. & Julier S. J. (2004). Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble? *Mon. Wea. Rev*., **132**, 1590 - 1605.

Welch, G. & Bishop, G. (2006). An Introduction to the Kalman Filter. Technical Report