METEOROLOGICAL APPLICATIONS
Meteorol. Appl. **16**: 57–64 (2009)
Published online 26 February 2009 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/met.130

Towards the provision of site specific flood warnings using wireless sensor networks

Paul J. Smith,[a]* Danny Hughes,[b] Keith J. Beven,[a] Philip Cross,[a] Wlodek Tych,[a] Geoff Coulson[b] and Gordon Blair[b]

[a] *Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK*
[b] *Computing Department, Lancaster University, Lancaster, LA1 4WA, UK*

ABSTRACT: An important aspect of flood risk management is the issuing of timely flood alerts. The spatial, as well as temporal, scale of these warnings is important. In many situations efficient risk management may be aided by the provision of local flood predictions at a high spatial resolution. Examples of such situations include issuing warnings for small groups of outlying houses or key infrastructure locations such as power sub-stations. In this paper a methodology for providing automated, detailed and location specific warnings which are computed 'on-site' is presented. Copyright © 2009 Royal Meteorological Society

KEY WORDS flood warning; wireless networks; kalman filter; DBM

Received 10 October 2008; Revised 15 December 2008; Accepted 9 January 2009

1. Introduction

An ability to predict local patterns of inundation and water surface elevations resulting from high flows is a critical step within any integrated modelling framework used to calculate flood warnings or assess flood risk. Despite the very wide application of hydraulic models for the assessment of flood risk, an examination of the detailed results reveals that the predictions often do not match inundation observations very well in some parts of the modelled domain (Aronica *et al*., 1998; Pappenberger *et al*., 2005, 2007).

This is partly a result of numerical approximations within such models, but more critically because of inadequate data on the geometry and connectivity of the channel and floodplain; difficulty in estimating effective roughness coefficients; and uncertainty in the initial and boundary conditions (including poorly known upstream and lateral inflows at flood levels). These issues, often combined with a lack of adequate calibration and validation data, means that a single hydraulic model may not be the most suitable tool for making accurate flood forecasts everywhere on the flood plain. In addition, such models, with their very many distributed state variables, are not easily made adaptive which can be an important factor in improving the accuracy of forecasts in the *next* event. This paper outlines an alternative methodology for providing local predictions at specific points at risk of flooding.

The methodology presented is based upon overlaying and complementing existing observational networks with an adaptive system, called the GridStix platform, consisting of nodes (GridStix) communicating using wireless technology. Each GridStix can interface with one or more sensors allowing the spatial resolution of an observation network to be increased significantly in areas of interest at a relatively low cost. The technology behind the GridStix platform is outlined in Section 2. This highlights how the adaptive nature of the GridStix platform can be used to increase the usefulness, robustness and manageability of the monitoring system.

Each GridStix has autonomous computational power. This can be used to provide local flood warnings. The ability to do this is dependent upon the availability of input data from upstream (or from rainfall measurements) and the use of an accurate and adaptive flood forecasting model which estimates the uncertainty in its predictions yet has a low computational burden. A suitable family of parsimonious time series models is outlined in Section 4. The effectiveness of these models is demonstrated using historic data from the River Dee catchment.

2. The GridStix Platform

The GridStix wireless sensor node platform is a hardware and software platform designed for environmental monitoring applications. As described in previous work (Smith *et al*., 2008), the GridStix sensor node platform uses two computational cores. A *low power* core operates during normal environmental conditions in order to maximise node lifetime while a *high performance* core

* Correspondence to: Paul J. Smith, Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK.
E-mail: p.j.smith@lancs.ac.uk

operates during flood events to provide improved performance and resilience.

Section 2.1 discusses the GridStix hardware platform, which provides support for connection to sensor devices, wireless networking and the local execution of on-site prediction models. Section 2.2 then presents the modular software solution that is used to support GridStix functionality. Finally, Section 3 discusses briefly the deployment of the GridStix platform and how the data gathered may be accessed.

2.1. GridStix Hardware

The performance and power requirements of event-centric environmental monitoring are dichotomous. On the one hand, as deployment sites are often remote and difficult to access, sensor nodes must minimise power consumption in order to maximise battery lifetime. On the other hand, during extreme events, maximising battery life may be less important then increasing the frequency of observations or exploiting sensors with higher performance requirements (e.g. digital imaging equipment). To meet these conflicting requirements, the GridStix use two computational cores.

Specifically, each GridStix consists of a SmartDust 2135 mote (Dust Networks, 2006a) and a Waysmall GumStix compact computer (see http://www.gumstix.com/).

Each Dust 2135 mote consists of a low powered microcontroller and radio. The motes form part of a low powered networking solution from Dust Networks (Dust Networks, 2006a). The motes communicate using 802.15.4 radio, operating in the 2.4 GHz range, which is capable of achieving ranges of up to 2 km in free space and approximately 600 m at the average height of field deployments.

The network of Dust motes is formed using a Time Synchronized Mesh Protocol (TSMP) (Dust Networks, 2006b). The use of a time-synchronised network protocol means that motes only listen for and broadcast in their allocated time-slots. This significantly reduces power consumption by minimising radio activity time. However, as motes must wait for a time-slot before transmitting, communication incurs higher latency than Carrier Sense Medium Access (CSMA) based protocols such as those based on B-MAC (Polastre *et al.*, 2004). Fortunately, the transmission of sensor data is latency-tolerant at such a fine-grained level in most practical flood management situations.

GridStix can be added to the TSMP network on the fly and if a connection is lost (e.g. due to node failure) they will reconfigure dynamically and make use of any available alternative routes to the manager. Reconnection to the network is a relatively power-hungry operation since TSMP nodes joining the network must listen to an entire sequence, or 'frame', of time slots to pick up the timing sequence. However, since GridStix joins are rare this is not problematic.

While the Dust motes provide an adequate base for monitoring during normal environmental conditions, adequate computational power for flood forecasting and the potential for high bandwidth communication are provided by the GumStix mote. A GumStix (so called since it is roughly the size of a pack of chewing gum) is a compact computer equipped with a 400 MHz Intel XScale PXA255 CPU, 64 MB of RAM and 16 MB of flash memory. These hardware resources support the execution of a standard Linux kernel and Java virtual machine, making a GumStix an inherently flexible platform for execution of the software element, albeit of limited computational power. It is, however, sufficiently powerful to support the local flood forecasting model used here, including adaptation in real-time. Further to this the GumStix may be equipped with an Ethernet port along with 802.11b/g (where b and g refer to the two versions of the IEEE 802.11 standard for wireless local area networking), Bluetooth and GSM hardware. This provides a number of options for high bandwidth communication. Specifically, in the field GumStix are equipped with 802.11 b which provides a high performance communication substrate though at a much higher power cost than the Dust motes.

GumStix consume a significant amount of power (up to 0.3 W) even when idle and, thus, to maximise battery life, must be used sparingly (Hughes *et al.*, 2008). This is achieved by allowing the Dust mote to act as an actuator for the GumStix allowing the GridStix to switch between three operational modes. These three modes, in order of increasing power usage, are:

- Mode 1: In this mode the Dust mote is used to relay low bandwidth data for example readings from pressure and conductivity sensors. The GumStix is not actuated;
- Mode 2: As with Mode 1 the Dust mote is used to relay low bandwidth data. However the GumStix is actuated to provide data processing. This mode is best suited for supporting on-site flood modelling outlined in Section 4;
- Mode 3: In this mode the GumStix is constantly actuated providing 802.11b networking. This mode is best suited relaying complex measurements such as digital images from a camera installed at a site of interest.

A small subset of the GridStix implements special purpose gateway functionality. These nodes are equipped with a specialized Dust 'manager' mote and GSM gateway hardware. The functionality of these nodes is discussed in more detail in Section 3.1.

2.2. GridStix Software

The software element of the GridStix platform performs two tasks: managing the individual GridStix and controlling the networking of the GridStix nodes (including the site gateway). These two tasks need to be performed in a co-ordinated fashion. For example, to execute its flood forecasting model a GridStix needs to be managed such

DOI: 10.1002/met

that it is in an appropriate mode (Mode 2 or 3) and that a suitable model is present. Alongside this, the network of GridStix needs to be co-ordinated so that appropriate data are made available to the GridStix in a timely fashion.

The GridStix software platform is implemented using the GridKit middleware (Grace *et al.*, 2004). GridKit is based on the OpenCOM (Coulson *et al.*, 2008) component model, which hosts the Open Overlays (Grace *et al.*, 2008) component framework that is responsible for co-ordinating networking and the SensorNode component framework which manages the scheduling of sensor readings and hosts the flood prediction models. OpenCOM supports run-time reconfiguration and provides safe support for the switching between, or deployment of, new networking technologies and models at runtime. Triggers for reconfiguration may be internal (e.g. swapping out a flood model that is providing inaccurate results) or external (e.g. switching to high-performance networking (Mode 3) when a flood is predicted).

2.2.1. *The Overlays Component Framework*

The Open Overlays CF offers rich support for virtual, overlay-based, networking of the GridStix.

Each overlay is a component which consists of three sub-components:

- A **Control** sub-component manages cooperation between GridStix to build and maintain an overlay-specific network topology;
- A **Forwarding** sub-component determines how the GridStix in the overlay route messages over the topology;
- A **State** sub-component is maintained for each overlay, e.g. containing a record of neighbouring GridStix.

As already noted, the Overlay component framework may be reconfigured autonomously at run time, allowing networking functionality to respond to changes in the environment and adapt to suit the demands of the GridStix platform.

To support the specific demands of the flood monitoring and prediction application, a number of specialised overlay implementations have been developed. These include a low bandwidth spanning-tree implementation based upon Dijkstra's algorithm (Dijkstra, 1959), a more robust decentralised spanning tree algorithm based upon Bellman-Ford (Coulson *et al.*, 2008) and specialised overlay components that allow the Dust motes to be used as a network interface card for the GumStix as reported in Smith *et al.* (2008).

3. Using the GridStix platform

The technology present in the GridStix platform allows for a wide variety of potential deployment configurations. In this section of the paper we discuss briefly the software tools available for the management of the deployed network and also those available for helping decide upon network configuration.

3.1. Interacting with the Sensor Network Deployment

In order to ensure redundancy, sensor data are logged at multiple levels. On each node deployed in the field a small cache of sensor data is maintained to deal with temporary disruptions caused by node failures. At the gateway a powered node maintains a master log of all observed sensor data and this is replicated off-site.

Data from the wireless sensor network may be accessed in three basic ways. Firstly, a dial-in GSM server at the powered gateway allows logs to be retrieved in a bandwidth-efficient manner using simple graphical interface software. Alternatively, the full management API is exposed over the same connection *via* XMLRPC, though this is bandwidth prohibitive over a GSM connection.

Finally, the presence of Bluetooth and 802.11 networking on GumStix allows for the possibility of direct interaction with consumer electronics devices (for example disseminating flood warnings directly to local residents' mobile phones).

The standard site interface software allows for the configuration of common parameters such as the frequency of data logging and the number of factors to be logged. These include not only sensor data, but also data on node battery life, link reliability and node-uptime which can be used to judge the health of the installation and debug the sensor network deployment.

3.2. Deciding upon suitable networking strategies

The three main networking technologies available for the GridStix nodes, 802.15.4 (2.4 GHz) radios, 802.11 b and GSM each operate at different spatial resolutions for differing power costs. Both the 802.15.4 radios and 802.11 b are affected by local topography, with the 802.11 b gaining greater range at the cost of increased power usage. The GSM communication technology is the most inefficient in terms of power usage but is largely independent of local topography. A cost-efficient and robust network requires the successful intermixing of these technologies.

To aid in determining the correct mix of technologies a network optimisation tool has been developed. Given a pre-specified list of locations for deployment of the sensors it performs two functions. Firstly, it evaluates the spatial coverage of the communication technologies for the GridStix located at these sites. Following this a series of candidate sites for relay points can be given (or automatically generated) and a genetic algorithm used to optimise the location of relay nodes.

For example, Figure 1 illustrates a current test site for the GridStix platform. At this test site GridStix nodes are deployed around the confluence of the Rivers Alyn and Dee, which lies between Chester and Wrexham on the England-Wales border. This area suffers from regular flooding and the GridStix are deployed both to offer local

DOI: 10.1002/met

forecasting of water levels and to provide information for the development of better hydraulic models.

The shaded areas in Figure 1 indicating the communication coverage of the 802.15.4 radios show direct communication between the GridStix sensor nodes is not possible. However, the optimised location of GridStix bridging nodes indicates that the majority of the sensors can be easily integrated into one network. Those that cannot, located further upstream on the River Alyn around Rossett, are connected to the gateway using GSM technology.

4. Flood Forecasting Models

The modular nature of the GridStix platform software outlined in Section 2.2 means a variety of models could be utilised on the GridStix for flood forecasting purposes. This section outlines the implementation of a family of parsimonious time series models, termed Data Based Mechanistic (DBM) models, that have been shown to be effective in flood forecasting (Lees *et al.*, 1994; Young, 2002; Romanowicz *et al.*, 2006). Further details and a fuller discussion of DBM modelling can be found in Young (2006) or Young *et al.* (2004). The remainder of this section focuses the prediction of a series of water levels observed at the Environment Agency gauging station located in Farndon which forms the upstream boundary of the test site shown in Figure 1.

Deployment of a DBM model for real-time flood forecasting at Farndon consists of two steps; off line estimation of the DBM model from historic data followed by on line forecasting utilising data assimilation. The first of these steps is not performed on the GridStix but remotely and is discussed in Section 4.1. The resulting models can then be deployed on the GridStix as software components. Section 4.2 focuses upon the data assimilation and on line forecasting, which is performed on the GridStix.

4.1. Identification of a DBM model

The water level observed at Farndon is influenced by the transport of water from upstream but also tidal forcing. Observed water levels from an upstream location above the tidally influenced reach are used to represent the forcing by water from upstream. A second input series based on the observed water levels in the estuary is used to represent the tidal forcing.

In keeping with past DBM studies (Young, 2002; Romanowicz *et al.*, 2006) a model is formulated to predict the vector $\mathbf{y} = (y_1, \ldots, y_T)$, indexed by time, where y_t is deviation of the water level at Farndon at time t above the minimum water level observed. The series of water levels observed at the upstream location (the Environment Agency gauging station at Manley Hall) is transformed in a similar fashion around its own

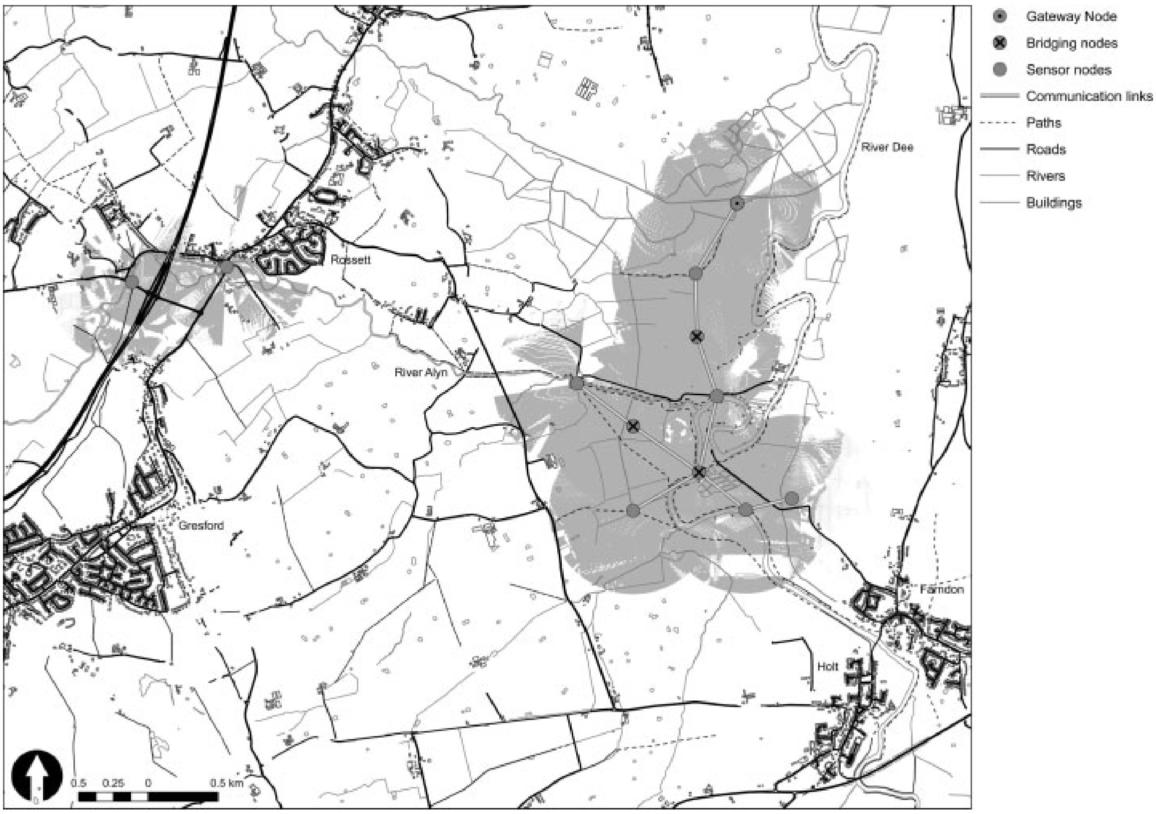

Figure 1. A map showing the test deployment of the GridStix at the confluence of the River Dee (flowing from south to north) and the River Alyn (flowing from east to west) located between Chester and Wrexham (UK). Shaded areas represent the coverage of the 2.4 GHz radios allowing for topography. The optimised bridging nodes lie in locations that allow for completed connectivity of the system to the gateway node. GSM technology is used to connect the two nodes to the east of the map to the gateway. (©Crown Copyright/OS 1 : 10000 Colour Raster 2008. An Ordnance Survey/EDINA supplied service.).

minima to give $\mathbf{u}_1 = (u_{1,1}, \ldots u_{1,\mathrm{T}})$. The second input series $\mathbf{u}_2 = (u_{2,1}, \ldots u_{2,\mathrm{T}})$, used to represent the tidal forcing, are the deviations of the estuary water level above the height of a controlling weir downstream of the study site, which acts the boundary between the estuary and river. That is $u_{2,t}$ equals zero if the estuary water level is below the weir height. For all the series data were available from the start of January 2000 until the end of December 2007 at a 15 min time step. The first two years of data were taken as a calibration period with the remainder being used for validation.

Initial analysis of the system proceeded by fitting the linear continuous time transfer function:

$$y_t = \frac{\sum_{i=1}^{2}(b_{i,0} + \cdots + b_{i,m_i}s^{m_i})u_{i,t-\tau_i}}{1 + a_1 s + \cdots + a_n s^n} + \varepsilon_t \tag{1}$$

(where ε_t is a stochastic disturbance) to the observed data using the methods discussed in Young and Garnier (2006). In doing this a number of model structures, determined by the quintet of positive integers $(n, m_1, m_1, \tau_1, \tau_2)$, were trialled and an optimum choice for the lags (τ_1, τ_2) made using the robust model selection criteria presented in Young (1984).

Analysis of the residuals of this initial model suggested that the relationship between the inputs and the observed water level was non-linear and that the behaviour of the system changes significantly when the river goes out of bank. The nonlinear nature of the relationship was investigated using state dependent parameter estimation (Young *et al.*, 2001) to consider the discrete time model:

$$\begin{aligned} y_t &= a(y_t)y_{t-1} + b_1(y_{t-\tau_1})u_{1,t-\tau_1} \\ &\quad + b_2(y_{t-\tau_2})u_{2,t-\tau_2} + \eta_t \end{aligned} \tag{2}$$

with η_t a stochastic disturbance and, for example, the notation $a(y_{t-1})$ indicating the value of the parameter a is believed to vary with the value of the state y_{t-1}.

The state dependent parameter analysis indicated that all the parameters had significant state dependency. To proceed further with the modelling a number of pragmatic choices are made. Firstly, the model structure:

$$\begin{aligned} y_t &= a(y_{t-\kappa})y_{t-1} + b_1(y_{t-\tau_1})u_{1,t-\tau_1} \\ &\quad + b_2(y_{t-\tau_2})u_{2,t-\tau_2} + \eta_t \end{aligned} \tag{3}$$

is presumed a reasonable representation of the system. This is in keeping with past studies (Young, 2003; Romanowicz *et al.*, 2006) and supported by the initial continuous time modelling. The state dependencies of the parameters (a, b_1, b_2) are presumed to be represented by polynomials of unknown order (r, s_1, s_2) respectively. That is:

$$a(y_{t-\kappa}) = \alpha_0 + \alpha_1 y_{t-\kappa} + \ldots + \alpha_r y_{t-\kappa}^r$$

and for $i = 1, 2$

$$b_i(y_{t-\tau_i}) = \beta_{i,0} + \beta_{i,1}y_{t-\tau_i} + \ldots + \beta_{i,s_i}y_{t-\tau_i}^{s_i}.$$

The integer time lag κ in the formulation for a is introduced to ensure that the polynomials giving parameter values can be evaluated using observed value of $\mathbf{y}$, thus avoiding feeding back predictions into the model. This allows the parameters of the polynomials to be estimated using an instrumental variable algorithm similar to the basic algorithm presented in Young (1984). As with the initial continuous time transfer function estimation a range of model structures, given by $(r, s_1, s_2, \kappa, \tau_1, \tau_2)$ were considered. The resulting model, selected based on robust selection criteria (Young, 1984) and consideration of the resulting polynomials, is given by $(r, s_1, s_2, \kappa, \tau_1, \tau_2) = (3, 5, 3, 9, 18, 7)$.

4.2. Data Assimilation and forecasting

This section outlines how the model determined by the off-line process outlined in Section 4.1 can be cast in a state space form. This allows the use of the Kalman Filter to assimilate observations and produce forecasts with appropriate error bounds.

4.2.1. *State Space Model*

There are several methods of constructing an appropriate state space model from Equation (3) (Young, 1998, 2002). The minimal state space representation consists of a single state representing the unobserved true value of y_t. Instead of using this a two state representation is used. Each component of state vector $\mathbf{x}_t = [x_{t,1}x_{t,2}]^{\mathrm{T}}$ relates to one of the two inputs. This allows the noise variance matrix of the Kalman Filter (see Section 4.2.2) to be used to capture the different levels of noise on each path of the model.

Take $\mathbf{F}_t$ and $\mathbf{G}_t$ to be time index matrices defined as:

$$\mathbf{F}_t = \begin{bmatrix} a(y_{t-\kappa}) & 0 \\ 0 & a(y_{t-\kappa}) \end{bmatrix}$$

and

$$\mathbf{G}_t = \begin{bmatrix} b_1(y_{t-\tau_1}) & 0 \\ 0 & b_2(y_{t-\tau_2}) \end{bmatrix}.$$

Let $\mathbf{u}'_t = [u_{1,t-\tau_1}u_{2,t-\tau_2}]^{\mathrm{T}}$ and $\mathbf{h} = [1\ 1]$. These, along, with realisations of stochastic disturbances ζ_t and ξ_t, allow, the state space model to be expressed as:

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{F}_{t+1}\mathbf{x}_t + \mathbf{G}_{t+1}\mathbf{u}'_{t+1} + \zeta_{t+1} \\ y_{t+1} &= \mathbf{h}^T\mathbf{x}_{t+1} + \xi_{t+1}. \end{aligned}$$

4.2.2. *Kalman Filter*

To apply the Kalman Filter a number of assumptions need to be made. Firstly ξ_{t+1} and ζ_{t+1} are presumed to be random draws from independent, zero mean symmetric distribution with variances σ^2 and $\mathbf{Q}\sigma^2$ respectively. The matrix $\mathbf{Q}$ is commonly referred to as the noise variance ratio (NVR) matrix. Further to this presume that the distribution of the unknown $\mathbf{x}_t$ can be summarised by its first two moments: the mean $\hat{\mathbf{x}}_t$ and variance $\mathbf{P}_t\sigma^2$.

The Kalman Filter can now be used for updating the summary of the state distribution as observation become available. Using the subscript $t+1|t$ to indicate the prediction at time $t+1$ given all the information up to time t, the *a priori* prediction equations are:

$$\hat{\mathbf{x}}_{t+1|t} = \mathbf{F}_{t+1}\hat{\mathbf{x}}_t + \mathbf{G}_{t+1}\mathbf{u}'_{t+1} \quad (4)$$

$$\mathbf{P}_{t+1|t} = \mathbf{F}_{t+1}\mathbf{P}_t\mathbf{F}^{\mathrm{T}}_{t+1} + \mathbf{Q}. \quad (5)$$

Following the observation of y_{t+1} the *a posteriori* correction is computed using:

$$\mathbf{K}_{t+1} = \mathbf{P}_{t+1|t}\mathbf{h}(1 + \mathbf{h}^{\mathrm{T}}\mathbf{P}_{t+1|t}\mathbf{h})^{-1}$$

$$\hat{\mathbf{x}}_{t+1} = \hat{\mathbf{x}}_{t+1|t} + \mathbf{K}_{t+1}(y_t - \mathbf{h}^{\mathrm{T}}\hat{\mathbf{x}}_{t+1|t}) \quad (6)$$

$$\mathbf{P}_{t+1} = \mathbf{P}_{t+1|t} - \mathbf{K}_{t+1}\mathbf{h}^{\mathrm{T}}\mathbf{P}_{t+1|t} \quad (7)$$

The manner in which the state evolution matrix $\mathbf{F}_t$ alters with time indicates that the variance of the state and hence any resulting predictions is heteroscedastic. However, for simplicity in this initial study none of the state dependent formulations for σ^2 presented in Young (2002), Smith *et al.* (2008) and Romanowicz *et al.* (2006) were implemented. In fact, treating σ^2 as a constant may also reflect the observational errors on water level measurements more realistically since, unlike flow, the error in the observation is unlikely to alter with river stage. Further methods of incorporating heteroscedasticity into the model using a state dependent NVR matrix are currently under investigation. However, as the results in Section 4.3 show, they are not needed in this case.

4.2.3. *Prediction Error Decomposition*

The Kalman Filter introduces two variance terms (σ^2 and $\mathbf{Q}$) which require optimising for given forecast horizons. Presuming that the prediction errors are normally distributed such an optimisation, for the n-step ahead forecast, can be readily achieved in a maximum likelihood framework using Prediction Error Decomposition (Schweppe, 1965).

For the n-step ahead prediction the innovations are given by $e_{t+n} = y_{t+n} - \mathbf{h}^{\mathrm{T}}\hat{\mathbf{x}}_{t+n|t}$ and the prediction variance by $\sigma^2(1 + \mathbf{h}^{\mathrm{T}}\mathbf{P}_{t+n|t}\mathbf{h})$. The values of $\hat{\mathbf{x}}_{t+n|t}$ and $\mathbf{P}_{t+n|t}$ can be computed by repeated application of the *a priori* steps of the Kalman Filter given by Equations (4) and (5).

The concentrated log likelihood of observing $\mathbf{y}$, assuming the presence of appropriate $\mathbf{u}'_t$, is:

$$l(\mathbf{Q}) = K - \frac{\mathrm{T}-n}{2}\log(\hat{\sigma}^2) - \frac{1}{2}\sum_{t=0}^{\mathrm{T}-n}\log(1 + \mathbf{h}^{\mathrm{T}}\mathbf{P}_{t+n|t}\mathbf{h}) \quad (8)$$

where:

$$\hat{\sigma}^2 = \frac{1}{\mathrm{T}-n}\sum_{t=0}^{\mathrm{T}-n}\frac{\mathrm{e}^2_{t+n}}{(1 + \mathbf{h}^{\mathrm{T}}\mathbf{P}_{t+n|t}\mathbf{h})}$$

is the maximum likelihood estimate of σ^2 and K is a constant with respect to σ^2 and $\mathbf{Q}$. The value of $l(\mathbf{Q})$ can be optimised numerically to give maximum likelihood parameter estimates for σ^2 and $\mathbf{Q}$.

4.3. Example Results

The model constructed in Section 4.1 was embedded into the Kalman Filter. Here, results are shown for predicting 3 h ($n = 12$ time steps) ahead. While a 3h lead time would normally be too short for decision making about issuing a flood warning, here we are concerned with the potential for local direct forecasts for which 3h lead time might be more than adequate. To do this requires the evaluation of $a(y_{t-\kappa})$ and $b_2(y_{y-\tau_2})$ at time steps for which the observed values of y_t are not available. Instead at these points the parameter vectors are evaluated using expected value of y_t given by the most recent forecast. Also, in the absence of a suitable predictive model for the estuary water level, the observed values are used even when they have not (theoretically) been taken.

Figure 2 shows the results for a flood event during the calibration data. Figures 3 and 4 show similar plots for two flood events observed during the validation period. Visually the results of the model indicate that it appears a reasonably reliable predictor of the water level at this lead time. However, certain characteristics, such as that displayed during the predictions for the 29 October 2000 (Figure 2), indicate that further refinement of the tidal interactions may be required.

The water level at Farndon is used to issue flood alerts for the Lower Dee Valley which runs downstream to Chester. Three level of alert are issued which in increasing order of severity are flood watch, flood warning and

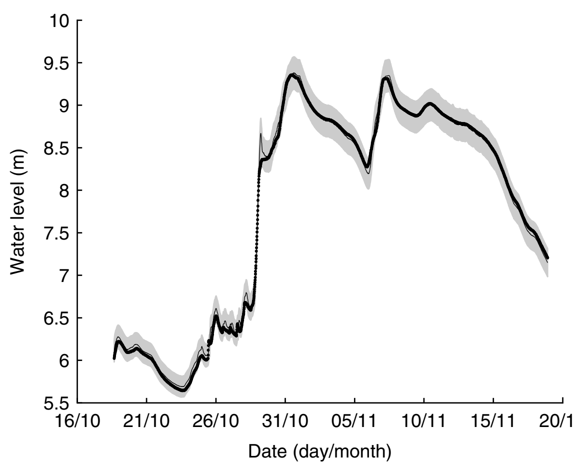

Figure 2. Summary plot of the 3-h ahead predictions for a flood event during October 2000, part of the calibration period. The shaded area represents the 95% prediction interval, the solid line the expected value in prediction and the points the observations.

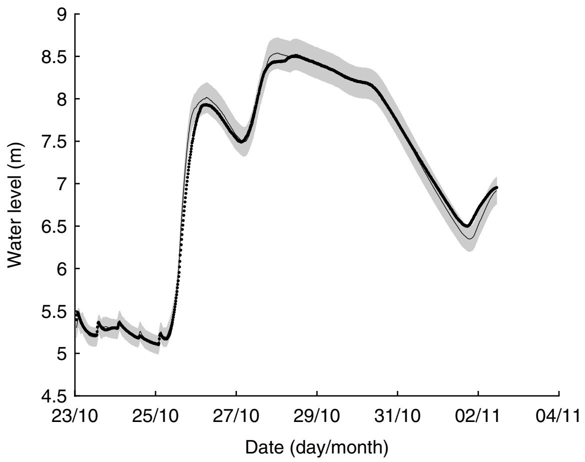

Figure 3. Summary plot of the three hour ahead predictions for a flood event during October 2002, part of the validation period. The shaded area represents the 95% prediction interval, the solid line the expected value in prediction and the points the observations.

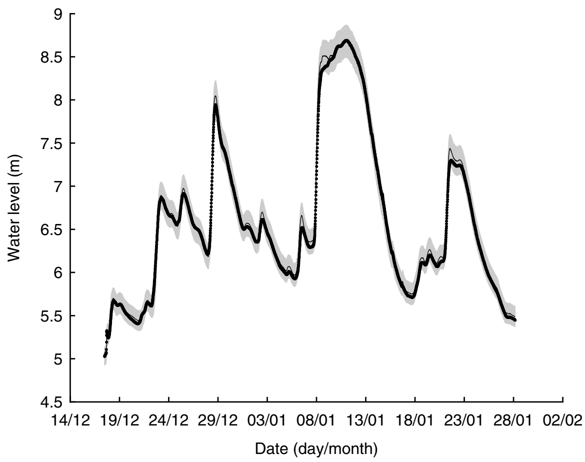

Figure 4. Summary plot of the three hour ahead predictions for a flood event during January 2005, part of the validation period. The shaded area represents the 95% prediction interval, the solid line the expected value in prediction and the points the observations.

severe flood warning. Table I highlights the high correspondence between when both the observed water level and the expected value of the forecast are above the specified alert heights. This is further evidence that the model may be a useful tool for applied flood forecasting.

5. Conclusions

This paper outlines a methodology for providing location specific flood forecasts. The solution outlined is based on an adaptive network of sensors, the GridStix platform (Sections 2 and 3), which brings benefits through its autonomous, flexible, nature and robust networking.

A family of flood forecasting models suitable for deployment on the GridStix platform is outlined in Section 4. This is shown to be an effective forecasting tool. At the site used in the example further work is required to improve the representation of the tidal response of the system as well as integration with further forecasting models to increase the lead time.

Table I. Summary of 3-h lead time forecasts against flood alert levels for the Lower Dee Valley. Correctly identified time steps are those where both the observed and expect value of the model prediction are above the alert threshold. False positives are those time steps where the expected model prediction is above the alert level yet the observation is below. All results are expressed as percentage of the number of time steps above the alert level.

Alert Level	Correctly identified time steps (%)	False Positives (%)
Flood Watch	99.0	1.5
Flood Warning	98.9	1.7
Severe Flood Warning	91.5	2.9

Acknowledgements

The authors wish to acknowledge the ongoing support of the Welsh Environment Agency in implementing the trial deployment. They also provided the data used in this study. Discussions with Peter Young have been helpful in the formulation of the models used. Two of the authors PJS and DH are currently funded through the NERC FREE program.

References

Aronica G, Hankin B, Beven K. 1998. Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data. *Advances in Water Resources* **22**(4): 349–365.

Coulson G, Blair G, Grace P, Taiani F, Joolia A, Lee K, Ueyama J, Sivaharan T. 2008. A generic component model for building systems software. *ACM Transactions on Computer Systems* **26**(1): 4–45 online.

Dijkstra E. 1959. A note on two problems in connection with graphs. *Numerische Mathematik* **1**: 269–271.

Dust Networks. 2006a. Smartmesh XT 2135 mote data sheet, http://www.dustnetworks.com/docs/M2135.pdf.

Dust Networks. 2006b. Technical overview of TSMP, white paper, http://www.dustnetworks.com/docs/TSMP_Whitepaper.pdf.

Grace P, Coulson G, Blair G, Mathy L, Yeung WK, Cai W, Duce D, Cooper C. 2004. Pluggable overlay networks for grid computing. On The Move To Meaningful Internet Systems 2004: Coopls, Doa and Odbase, Pt 2, Proceedings 3291, 1463–1481.

Grace P, Hughes D, Porter B, Blair G, Coulson G, Taiani F. 2008. Experiences with open overlays: A middleware approach to network heterogeneity. In *Proceedings of the European Conference on Computer Systems (EuroSys'08)*, Glasgow, UK.

Hughes D, Daude M, Coulson G, Blair G. 2008. Managing heterogeneous data flows in wireless sensor networks using a 'split personality' mote platform. In *Proceedings of the 2nd International Workshop on SensorWebs, Databases and Mining in Networked Sensing Systems (SWDMNSS 2008)*, Turku.

Lees MJ, Young PC, Ferguson S, Beven KJ, Burns J. 1994. An adaptive flood warning scheme for the River Nith at Dumfries. In *2nd International Conference on River Flood Hydraulics*, White WR, Watts J (eds).: Chichester, UK; New York.

Pappenberger F, Beven K, Frodsham K, Romanowicz R, Matgen P. 2007. Grasping the unavoidable subjectivity in calibration of flood inundation models: A vulnerability weighted approach. *Journal of Hydrology* **333**(2–4): 275–287.

Pappenberger F, Beven K, Horritt M, Blazkova S. 2005. Uncertainty in the calibration of effective roughness parameters in HES-RAS using

DOI: 10.1002/met

inundation and downstream level observations. *Journal of Hydrology* **302**(1–4): 46–69.

Polastre J, Hil J, Culler D. 2004. Versatile low power media access for wireless sensor networks. In *Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems (Sensys'04)*, Baltimore, Maryland, US; 95–107.

Romanowicz RJ, Young PC, Beven KJ. 2006. Data assimilation and adaptive forecasting of water levels in the River Severn catchment, United Kingdom. *Water Resources Research* **42**(6): W06407.

Schweppe FC. 1965. Evaluation of likelihood functions for gaussian signals. *IEEE Transactions on Information Theory* **11**(1): 61–70.

Smith P, Beven K, Tych W, Hughes D, Coulson G, Blair G. 2008. The provision of site specific flood warnings using wireless sensor networks. In *Proceeding of FLOODRisk2008*, Oxford, UK.

Young PC. 1984. *Recursive Estimation and Time Series Analysis: An Introduction*. Springer Verlag: Berlin.

Young PC. 1998. Data-based mechanistic modelling of environmental, ecological, economic and engineering systems. *Environmental Modelling & Software* **13**(2): 105–122.

Young PC. 2002. Advances in real-time flood forecasting. *Philosophical Transactions of the Royal Society of London Series A: Mathematical Physical and Engineering Sciences* **360**(1796): 1433–1450.

Young PC. 2003. Top-down and data-based mechanistic modelling of rainfall-flow dynamics at the catchment scale. *Hydrological Processes* **17**(11): 2195–2217.

Young PC. 2006. The data-based mechanistic approach to the modelling, forecasting and control of environmental systems. *Annual Reviews in Control* **30**(2): 169–182.

Young PC, Chotai A, Beven KJ. 2004. Data-based mechanistic modelling and the simplification of environmental systems. In *Environmental Modelling: Finding Simplicity in Complexity*, Wainwright J, Mullgan M (eds). Wiley: Chichester, UK; New York 371–388.

Young PC, Garnier H. 2006. Identification and estimation of continuous-time, data-based mechanistic (DBM) models for environmental systems. *Environmental Modelling & Software* **21**(8): 1055–1072.

Young PC, McKenna P, Bruun J. 2001. Identification of non-linear stochastic systems by state dependent parameter estimation. *International Journal of Control* **74**(18): 1837–1857.