Model Discrepancy in the Saturated Path Hydrology Model: Initial Analysis

Tom Fricker
University of Sheffield

25 January 2010

1 Introduction

In this report we present some very basic results regarding model discrepancy in the Saturated Path Hydrology Model (logSPM, Kuczera et al., 2006). The purpose is to investigate whether this model is likely to be a suitable case study on which to implement and test the ideas about learning model discrepancy in dynamic models set out in Wilkinson et al. (2009).

2 Model discrepancy in data assimilation

Data assimilation is the task of inferring the trajectory of a system using a combination of noisy observations and the output of a mathematical model. The mathematical model is an attempt to describe the dynamics of the system, but inevitably the output will be an imperfect representation of the true system state. Traditional approaches to data assimilation treat the difference between the model output and reality (the *model discrepancy* or *model error*) as white noise. Wilkinson et al. (2009) consider the alternative approach of modelling the model discrepancy as a state dependent term in order to improve the predictive distribution of the model. They consider the discrete-time data assimilation problem, in which the system dynamics are described by

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t) + \delta(\mathbf{x}_t, \mathbf{u}_t), \quad t = 0, ..., T. \tag{1}$$

Here, $\mathbf{x}_t$ is the true system state, $\mathbf{u}_t$ is a set of forcing inputs, and $f(x_t, u_t)$ is the mathematical model that predicts the state at time $t+1$ given the state at time t. The model discrepancy is $\delta(\mathbf{x}_t, \mathbf{u}_t)$.

3 The Saturated Path Hydrology Model

We consider the Saturated Path Hydrology Model (logSPM, Kuczera et al., 2006). This is a simple model that simulates the movement of water from a rainfall catchment area through the ground and out to a river. The model comprises three compartments: soil, groundwater and river. The features of the model are as follows:

- The state variable at a given time t is the volume of water in each compartment, denoted $\mathbf{x}_t^T = (h_s(t), h_{gw}(t), h_r(t))$.
- The forcing inputs are rainfall, $rain(t)$, and the evapotranspiration potential, $pet(t)$.
- There are 7 unknown calibration parameters denoted $\boldsymbol{\theta}$.
- There are 3 initial conditions, $\mathbf{x}_0^T = (h_s(0), h_{gw}(0), h_r(0))$, which are also considered unknown.
- We have $N_t = 1827$ consecutive daily observations recorded in the Abercrombie river catchment in Australia. The data consists of the forcing inputs $rain$ and pet, and Q_r^{obs}, the volume of water discharged from the system. The forcing input data is complete, but there are 410 missing values of Q_r^{obs}.
- The deterministic relationship between (true) discharge Q_r and the state variables is assumed to be

$$Q_r(t) = f_Q A_w k_r h_r(t) \tag{2}$$

  where k_r is an element of $\boldsymbol{\theta}$, A_w is the area of catchment (it is 2770km2 for the Abercrombie data), and f_Q is an additional 'modification parameter' used to tune the model.

Further details of the model are given in Kuczera et al. (2006) and MUCM (2007).

The model has previously been calibrated using the Abercrombie data by Peter Reichert (see the slides attached in MUCM (2007)). We denote the MAP values of the 3 initial conditions and 7 calibration parameters found in this calibration as $\hat{\mathbf{x}}(0)^T$ and $\hat{\boldsymbol{\theta}}$. We ran the model with $\hat{\mathbf{x}}(0)^T$ and $\hat{\boldsymbol{\theta}}$ to obtain model output data $\{\mathbf{x}^{mod}(t) : t = 0, ..., N_t\}$. We used this to estimate the modification parameter f_q in equation (2) by minimising the mean squared error:

$$\hat{f}_q = \text{argmin} \frac{1}{N_t} \sum_{t=1}^{N_t} [Q_r^{obs}(t) - f_Q A_w \hat{k}_r h_r^{mod}(t)]^2. \tag{3}$$

For the following analyses we assume that the uncertain parameters to be fixed at $\hat{\mathbf{x}}(0)^T$, $\hat{\boldsymbol{\theta}}$ and $\hat{f}_q$.

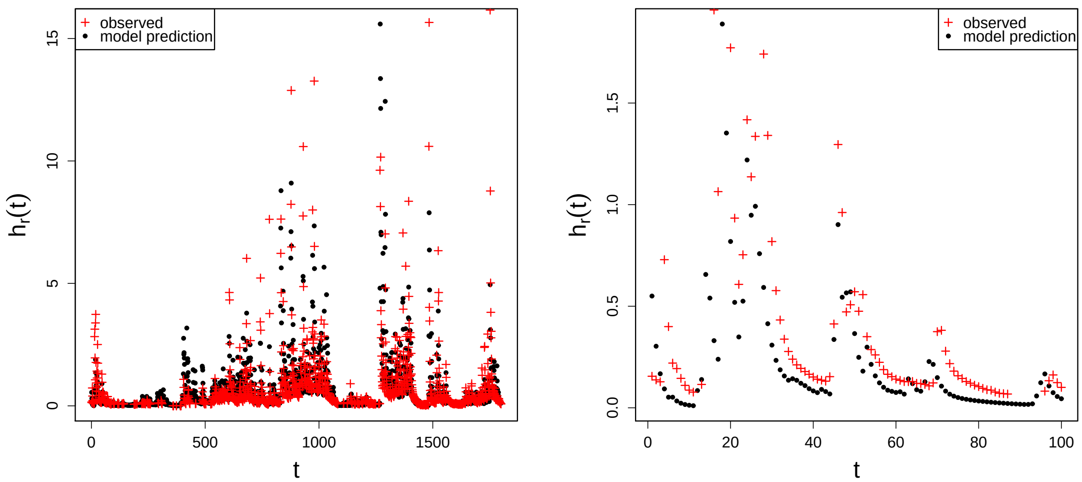

Figure 1: Observations and model predictions of the system discharge Q_r, plotted against time.

Figure 1 compares the observations of Q_r with the series of Q_r obtained from the model. We see that the fit is reasonable, suggesting that the model run with the fixed values $\hat{\mathbf{x}}(0)^T$, $\hat{\boldsymbol{\theta}}$ and $\hat{f}_q$ is an adequate description of the real system.

4 Visual inspection of model discrepancy

Following Wilkinson et al. (2009), we consider the system dynamics to be described by

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{z}_t) + \delta(\mathbf{x}_t, \mathbf{z}_t), \tag{4}$$

where $f(.) = (f_s(.), f_{gw}(.), f_r(.))$ is the deterministic model, $\delta(.)$ is the state dependent model discrepancy.

Interest is in using the data to learn about $\delta(.)$. This cannot be done directly because we never know the true values of the state variables $\mathbf{x}_t$. We do, however, have observations which give us estimates of the $h_r(t)$ state variable:

$$h_r^{obs}(t) = \frac{Q_r^{obs}(t)}{\hat{f}_Q A_w \hat{k}_r}. \tag{5}$$

We can therefore estimate the h_r component of the discrepancy as

$$\hat{\delta}_r(\mathbf{x}_t, \mathbf{z}_t) = h_r^{obs}(t+1) - h_r^{mod}(t+1), \tag{6}$$

where $h_r^{mod}(t)$ is the result of running the model forward from the fixed initial conditions until time t. Note that we are not performing any data assimilation here, so $h_r^{mod}(t)$ is the model prediction taken from the series without any state corrections being made.

We wish to model the discrepancy as a function of the state variable. We therefore examine plots of $\hat{\delta}_r(\mathbf{x}_t, \mathbf{z}_t)$ against estimates of the state variables provided by observations and by the model. Figure 2 shows $\hat{\delta}_r(\mathbf{x}_t, \mathbf{z}_t)$ plotted against the observed state variable $h_r^{obs}(t)$, and figure 3 shows $\hat{\delta}_r(\mathbf{x}_t, \mathbf{z}_t)$ plotted against the predicted state variable $h_r^{mod}(t)$. Neither indicates any discernable relationship between discrepancy and $h_r(t)$. Figures 4 and 5 show $\hat{\delta}_r(\mathbf{x}_t, \mathbf{z}_t)$ plotted against the model predictions $h_{gw}^{mod}(t)$ and $h_s^{mod}(t)$ respectively (note that we do not observe h_{gw} and h_s so there is no analogue to figure 2 for these state variables). Again, there is no evidence of a relationship between discrepancy and $h_s(t)$. On the other hand, the closeup plot in figure 4 shows some clear lines of points, suggesting that we may be able to model discrepancy as a function of h_{gw}.

We note that the discrepancy is also a function of the forcing inputs, so we plot $\hat{\delta}_r(\mathbf{x}_t, \mathbf{z}_t)$ against the observed $rain(t)$ and $pet(t)$. We see no evidence of a relationship here.

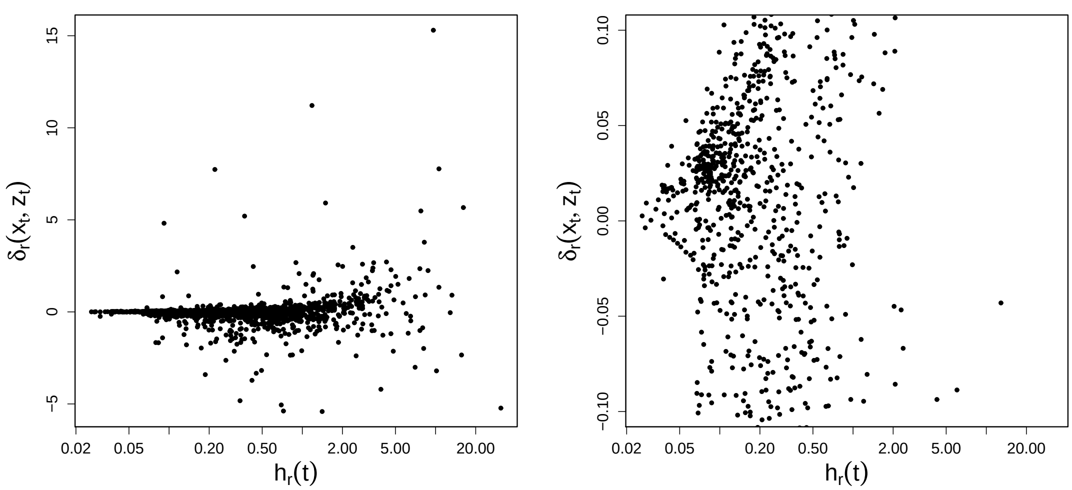

Figure 2: The discrepancy between observed and predicted $h_r(t+1)$, plotted against observed $h_r(t)$. The right plot is a closeup of the left plot.

5 Conclusions

The simple analysis here suggests that we may struggle to model the discrepancy as a smooth function of the state variables and the forcing inputs. The only discernable relationship we have found is between discrepancy and h_{gw}. However, the analysis we have performed here is somewhat crude. A major limitation is that we have not performed any data assimilation, so the estimates of the state variables that we have used are likely to be poor since we do not correct the model at any stage.

The plots we have presented here can be compared to the upper plot of figure 3 in Wilkinson

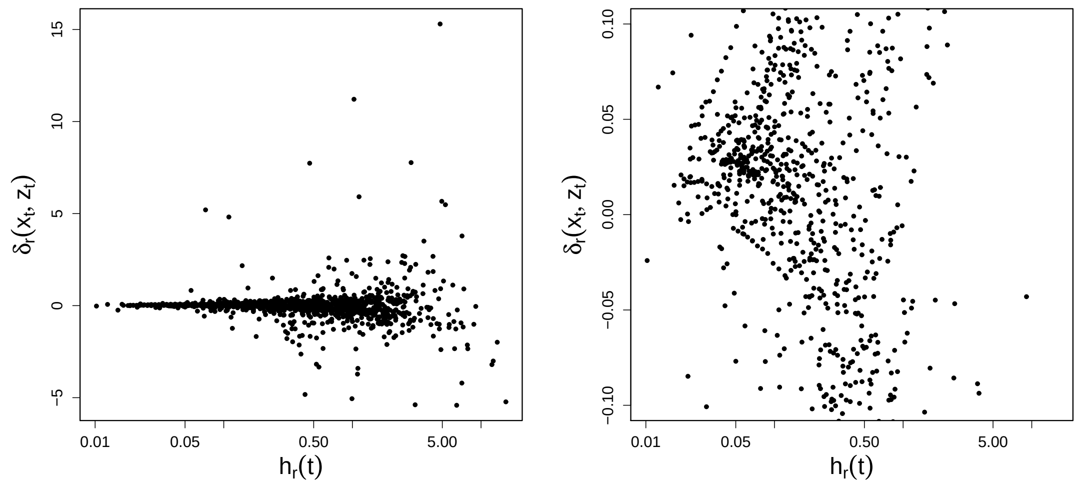

Figure 3: The discrepancy between observed and predicted $h_r(t+1)$, plotted against predicted $h_r(t)$. The right plot is a closeup of the left plot.

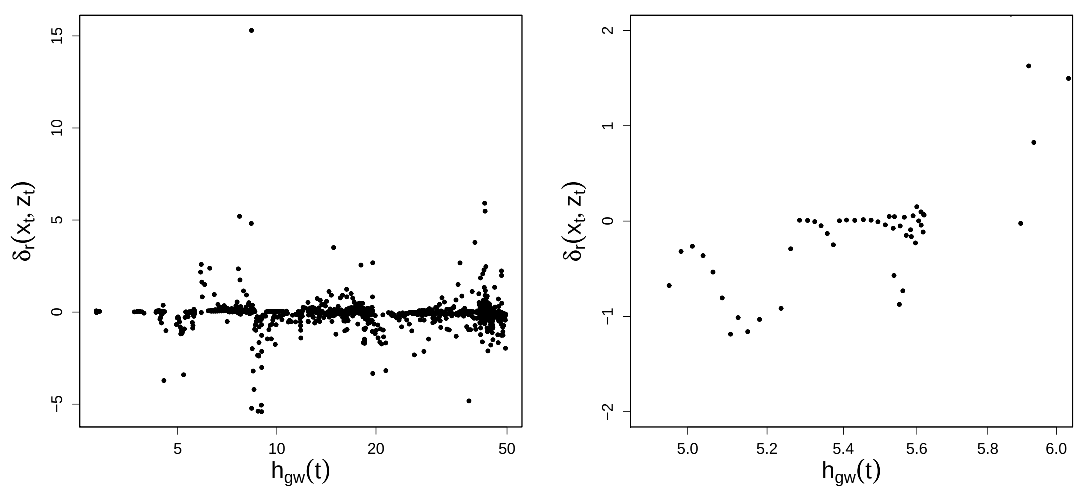

Figure 4: The discrepancy between observed and predicted $h_r(t+1)$, plotted against predicted $h_{gw}(t)$. The right plot is a closeup of the left plot.

et al. (2009), which shows the discrepancy based on non-corrected state variables in a toy example. There is little evidence in that plot of a discrepancy-state relationship even though we know it exists. In that example the model discrepancy in not revealed until the trajectory is corrected to the true state at every time step (see the lower plot of figure 3 in Wilkinson et al. (2009)). This shows that our analyses should not in themselves be taken to show that the logSPM model is unsuitable to test methods for learning model discrepancy. We may find that when a more sophisticated approach is taken with the logSPM model (performing data assimilation, for example), we will start to see more structure in the model discrepancy.

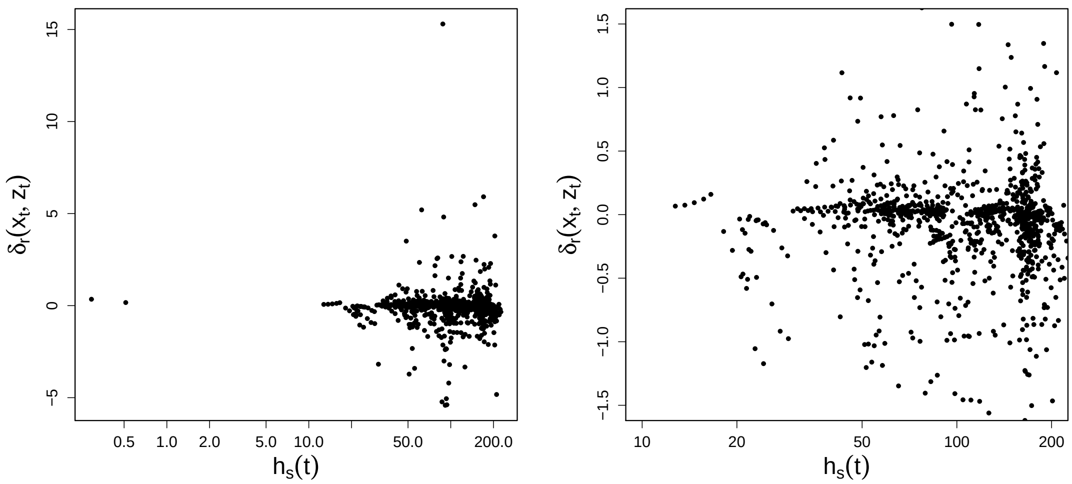

Figure 5: The discrepancy between observed and predicted $h_r(t+1)$, plotted against predicted $h_s(t)$. The right plot is a closeup of the left plot.

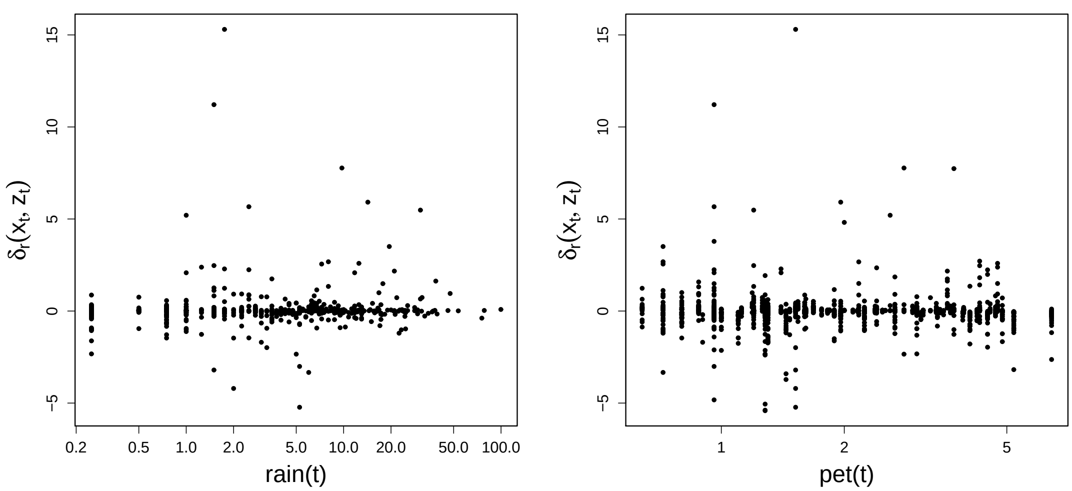

Figure 6: The discrepancy between observed and predicted $h_r(t+1)$, plotted against forcing inputs $rain(t)$ (left plot) and $pet(t)$ (right plot).

References

Kuczera, G., Kavetski, D., Franks, S. and Thyer, M. (2006). Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters, *Journal of Hydrology*, **331 (1-2)**: 161–177.

MUCM (2007). Saturated path hydrology model (logspm) description, http://wiki.aston.ac.uk/cgi-bin/view/MUCM/MUCMPrivate/LogspmModel.

Wilkinson, R., Oakley, J. and O'Hagan, A. (2009). Estimating model error in data assimilation, *Unpublished technical report*.