# Time varying parameter models for catchments with land use change: the importance of model structure

S. Pathiraja[1], D.Anghileri[2], P. Burlando[2], A. Sharma[1], L. Marshall[1], H. Moradkhani[3]

[1]Water Research Centre

School of Civil and Environmental Engineering

University of New South Wales

Sydney, NSW

AUSTRALIA

Email: s.pathiraja@unsw.edu.au

[2]Institute of Environmental Engineering

ETH Zurich

Zurich

SWITZERLAND

[3]Department of Civil and Environmental Engineering

Portland State University

Portland, Oregon

USA

# Abstract

Rapid population and economic growth in South-East-Asia has been accompanied by extensive land use change with consequent impacts on catchment hydrology. Modelling methodologies capable of handling changing land use conditions are therefore becoming ever more important, and are receiving increasing attention from hydrologists. A recently developed Data Assimilation based framework that allows model parameters to vary through time in response to signals of change in observations is considered for a medium sized catchment (2880 $km^2$) in Northern Vietnam experiencing substantial but gradual land cover change. We investigate the efficacy of the method as well as the importance of the chosen model structure in ensuring the success of time varying parameter methods. The framework was utilized with two conceptual models (HBV and HyMOD) that gave good quality streamflow predictions during pre-change conditions. Although both time varying parameter models gave improved streamflow predictions under changed conditions compared to the time invariant parameter model, persistent biases for low flows were apparent in the HyMOD case. It was found that HyMOD was not suited to representing the modified baseflow conditions, resulting in extreme and unrealistic time varying parameter estimates. This work shows that the chosen model can be critical for ensuring the time varying parameter framework successfully models streamflow under changing land cover conditions. It can also be used to determine whether land cover changes (and not just meteorological factors) contribute to the observed hydrologic changes in retrospective studies where the lack of a paired control catchment precludes such an assessment.

# 1. Introduction

Population and economic growth in South-East Asia has led to significant land use change, with rapid deforestation occurring largely for agricultural purposes [*Kummer and Turner*, 1994]. Forest cover in the Greater Mekong Sub-region (comprising Myanmar, Thailand, Cambodia, Laos, Vietnam, and South China) has decreased from about 73% in 1973 to about 51% in 2009 [*WWF*, 2013]. Vietnam in particular has had the second highest rate of deforestation of primary forest in the world, based on estimates from the Forest Resource Assessment by the United Nations Food and Agriculture Organization [*FAO*, 2005]. Such extensive land use change has the potential to significantly alter catchment hydrology (in terms of both quantity and quality), with its effects sometimes not immediate but occurring gradually over a lengthy period of time. Recent estimates from satellite measurements indicate that rapid deforestation continues in the region, although at lower rates [e.g. *Kim et al.,* 2015]. Persistent land use change necessitates modelling methodologies that are capable of providing accurate hydrologic forecasts and predictions, despite non-stationarity in catchment processes. This is also particularly relevant for water resource management which requires reliable estimates of water availability, both in terms of volume and timing, to properly allocate the resource between different water uses and to prevent flood damages. Vietnam has built many reservoirs in the last decades and more are planned because they are considered to be fundamentally important for electricity production, flood control, water supply and irrigation, ultimately contributing to the development of the country [*Giuliani et al*., 2016].

The literature on land-use change and its impacts on catchment hydrology is extensive, with studies examining the effects of 1) conversion to agricultural land-use [*Thanapakpawin et al*, 2007; *Warburton et al.,* 2012]; 2) deforestation [*Costa et al.,* 2003; *Coe et al*, 2011]; 3) afforestation [e.g. *Yang et al., 2012*; *Brown et al*, 2013] and urbanization [*Bhaduri et al.,* 2001; *Rose & Peters*, 2001]. Fewer studies have examined how traditional modelling approaches must be modified to handle

46 non-stationary conditions, or how modelling methods can be used to assess impacts of land use
47 change. Split sample calibration has been used frequently to retrospectively examine changes to
48 model parameters due to land use or climatic change [*Seibert & McDonnell, 2010; Coron et al., 2012;*
49 *McIntyre & Marshall*, 2010; *Legesse et al*, 2003]. Several other studies have employed scenario
50 modelling, whereby hydrologic models are parameterized to represent different possible future land
51 use conditions [e.g. *Niu & Sivakumar*, 2013; *Elfert & Borman*, 2010]. A related approach involves
52 combining land use change forecast models with hydrologic models [e.g. *Wijesekara et al.,* 2012].
53 However the aforementioned approaches are unsuited to short-term predictive modelling or
54 hydrologic forecasting in dynamic catchments, as the predicted land use change may not reflect
55 actual changes. A potentially more suitable approach in such a setting is to allow model parameters
56 to vary in time, rather than assuming a constant optimal value or stationary probability distribution.
57 Many existing methods utilising such a framework require some *apriori* knowledge of the land use
58 change in order to inform variations in model parameters (see for instance *Efstratiadis,* 2015; *Brown*
59 *et al.,* 2006; and *Westra et al.,* 2014). Recent efforts have examined the potential for time varying
60 models to automatically adapt to changing conditions using information contained in hydrologic
61 observations and sequential Data Assimilation, without requiring explicit knowledge of the changes
62 [see for example *Taver et al.,* 2015, *Pathiraja et al.,* 2016a&b]. Such approaches can objectively
63 modify model parameters in response to signals of change in observations in real time, whilst
64 simultaneously providing uncertainty estimates of parameters and streamflow predictions. They can
65 also be used to determine whether land cover changes (and not solely meteorological factors)
66 contribute to observed changes in streamflow dynamics in retrospective studies where the lack of a
67 paired control catchment precludes such an assessment.
68
69 *Pathiraja et al.* [2016a] presented an Ensemble Kalman Filter based algorithm (the so-called Locally
70 Linear Dual EnKF) to estimate time variations in model parameters. The method sequentially
71 assimilates observations into a numerical model in real time to generate improved estimates of

72 model states, fluxes and parameters at a given time based on their respective uncertainties. Its
73 purpose is to infer changes to catchment properties (e.g. land cover change) from hydrologic
74 observations, without prior knowledge of such changes, at the time scale of the observation
75 frequency. It can therefore be used for various applications: 1) to retrospectively estimate time
76 variations in model parameters; 2) for short-term predictive modelling (e.g. flood forecasting); and 3)
77 for on-line/real time water resource management (e.g. determining releases from reservoirs in
78 catchments with changing land cover conditions). When used for prediction, states and parameters
79 are updated sequentially using all available observations up until the current time. These updated
80 states and parameters are then used along with the prior parameter generating model to produce
81 hydrologic predictions over a short time horizon. The efficacy of the method was demonstrated in
82 *Pathiraja et al.* [2016b] through an application to small experimental catchments (< 350 ha) with
83 drastic land cover changes and strong signals of change in streamflow observations.
84
85 Here we investigate two issues that are pertinent to modelling in realistic catchments with land use
86 change. Firstly, we investigate the efficacy of the time varying parameter method for sparsely
87 observed, larger catchments with gradual and spatially complex land use change. Several authors
88 have demonstrated that impacts of land use change on the hydrologic response are dependent on
89 many factors including the type and rate of land cover conversion as well the spatial pattern of
90 different land uses within the catchment [*Dwarakish & Ganasri*, 2015; *Warburton et al.*, 2012]. In
91 such situations, the effects of unresolved spatial heterogeneities in model inputs (e.g. rainfall) and
92 the relatively less pronounced changes in land surface conditions make time varying parameter
93 detection and accurate hydrologic prediction more difficult. The second objective is to examine the
94 role of the hydrologic model in determining the ability of the time varying parameter framework to
95 provide high quality predictions in changing conditions. Often there may be several candidate
96 hydrologic models (with time invariant parameters) that have similar predictive performance for a
97 catchment when calibrated and validated over a time series of static land cover conditions. This

work examines whether all such candidate models in time varying parameter mode are also capable of providing accurate predictions under changing conditions.

These issues are investigated for the Nammuc catchment (2880 $km^2$) in Northern Vietnam which has experienced deforestation largely due to increasing agricultural development. It serves as an ideal test catchment to study the efficacy of the time varying parameter algorithm due to its size, spatially complex pattern of land use changes, and lack of information on the precise timing of such changes. Land cover change is estimated to have occurred at varying rates, with cropland accounting for roughly 23% between 1981 and 1994, and 52% by 2000. We also consider two conceptual hydrologic models (given the availability of point rainfall, temperature, and streamflow data) to address the second objective. Both models demonstrate similar performance in representing streamflow at the outlet during the pre-change period. The effect of the model structure on hydrologic predictions from the time varying parameter models is studied.

The remainder of this paper is structured as follows. Details of the study catchment and the impact of land cover change are analysed in Section 2. Section 3 summarizes the experimental setup including the hydrological models and the time varying parameter estimation method used. Results are provided in Section 4, along with an analysis of whether the time varying model structures reflect the observed catchment dynamics. Finally, we conclude with a summary of the main outcomes of the study as well as proposed future work.

## 2. The Nammuc Catchment

The Nammuc catchment (2880 $km^2$) is located in the Red River Basin, the second largest drainage basin in Vietnam which also drains parts of China and Laos. The local climate is tropical monsoon dominated with distinct wet (May to October) and dry (November to April) seasons. The wet season tends to have high temperatures (on average 27 to 29 °C) due to south-south easterly winds that

bring humid air masses. Conversely, during the dry season, circulation patterns reverse carrying cooler dry air masses to the basin (leading to average temperatures of 16 to 21°C). Streamflow response is consequently monsoon driven, with high flows occurring between June and October (generally peaking in July/August) and low flows in the December to May period (Vu, 1993). Average annual rainfall at Nammuc varies between 1300 and 2000 mm (on average 1600 mm). A summary of catchment properties is provided in Table 1**.**

## 2.1.Data & Land Cover Change

Figure 1 shows the available land cover information for the Nammuc catchment. The first land cover map refers to the period 1981-1994 and was obtained by the Vietnamese Forest Inventory and Planning Institute (http://fipi.vn/Home-en.htm). The second land cover map refers to year 2000 and was obtained from the FAO Global Land Cover database (http://www.fao.org/geonetwork/srv/en/metadata.show?id=12749&currTab=simple). A comparison of the two maps shows a reduction in forest cover in favor of cropland; Evergreen Leaf decreases from about 60% to 30% whilst cropland increases from about 23% to 52%. The change in land cover is patchy, although mostly concentrated in the northern part of the catchment. Because of the scant information available, it is not easy to identify the precise time period of these changes. Based on the available land cover map information and the changes to observed runoff (see Section 2.2), we posit that a period of rapid extensive deforestation occurred in early-1990s.

Daily point rainfall data is available at four precipitation stations surrounding the catchment (Dien Bien, Tuan Giao, Quynh Nhai and Nammuc, see Figure 1). Catchment averaged rainfall was developed as a weighted sum of the four stations with weights determined by Thiessen Polygons. Daily mean temperature was calculated in a similar fashion using temperature records from the 2 closest gauges (Lai Chau and Quynh Nhai, see Figure 1). This was used to estimate Potential Evapotranspiration through the empirical temperature-latitude based Hamon PET method [*Hamon*,

148 1961]. Daily rainfall, temperature and streamflow data was provided by the Vietnamese Institute of
149 Water Resources Planning.

## 150 2.2.Impact of Land Cover Change on Streamflow

151 An examination of the observed streamflow and rainfall records shows that distinct changes to the
152 hydrologic regime are evident after the mid-1990s. The annual runoff coefficient varies between 0.4
153 and 0.6 prior to 1994, after which it increases to between 0.6 and 0.8 until 2004 (see Figure 2a).
154 However, increases to annual yields are driven mostly by changes to baseflow volume. This is
155 evident in Figure 2a, which shows that the increase in the annual direct runoff coefficient
156 $\left(\frac{runoff-baseflow}{rainfall}\right)$ is less than the increase in the total runoff coefficient (roughly 0.1 increase
157 compared to 0.2 respectively). Baseflow was estimated using the two parameter recursive baseflow
158 filter of *Eckhardt* [2005], with on-line updating of baseflow estimates to match low flows. A small
159 increase in the Annual Baseflow Index $\left(\frac{baseflow}{runoff}\right)$ is apparent also, from about 0.32 on average in the
160 period 1970 to 1982 to 0.39 on average after 1994 (Figure 2b). This indicates that the annual
161 increases to baseflow volume exceed the increases to direct runoff volume. Similar changes were
162 found by *Wang et al.* [2012] who analyzed records in the entire Da River basin which drains the
163 largest river in the Red River catchment.
164
165 At a seasonal time scale, it is apparent that both wet and dry season flows exhibit temporal
166 variations. We utilized the Moving Average Shifting Horizon (MASH) [*Anghileri et al.,* 2014] and
167 Mann-Kendall test to assess seasonal trends in observed streamflow, precipitation, and temperature
168 data. The MASH tool can be used to qualitatively assess inter-annual variations in the seasonal
169 pattern of a variable. It works by calculating a statistic of the data (e.g. mean) over the same block of
170 days in consecutive years. A steady increase in baseflow is again apparent (see February to April in
171 Figure 2c), as well as increases to wet season flows (see June to September in Figure 2c). Mann-
172 Kendall test (with significance level equal to 5%) on annual and monthly streamflow time series

shows increasing trends in almost all months, i.e., from October to July. No concurrent increases are apparent in rainfall (see Figure 2d). Also the Mann-Kendall test applied to precipitation time series does not show any statistically significant trend, except a decrease in September for Nammuc and Quynh Nhai station and an increase in July for Dien Bien station. Temperature variations are not evident from the MASH analysis (not shown) and no significant trend can be detected by applying the Mann-Kendall test. These results indicate that changes in streamflow dynamics are likely due to land use change rather than climatic impacts.

# 3. Experimental Setup

## 3.1.Hydrologic Models

Conceptual lumped models were adopted due to the availability of point rather than distributed hydro-meteorological data of sufficient length. We considered the HyMOD [*Boyle*, 2001] and Hydrologiska Byrans Vattenbalansavdelning (HBV) [*Bergstrom et al.,* 1995] models. They differ mainly in the way components of the response flow are separated (HBV has near surface flow, interflow, and baseflow components whilst HyMOD has a quickflow and slow flow component only) and how these flows are routed. A schematic of the models is shown in Figure 3.

In the HyMOD model, spatial variations in catchment soil storage capacity are represented by a Pareto distribution with shape parameter $b$ and maximum point soil storage depth $c_{max}$. Excess rainfall ($V$) is partitioned into three cascading tanks representing quick flow and a single slow flow store through the splitting parameter $\alpha$. Outflow from these linear routing tanks is controlled by parameters $k_q$ (for the quick flow stores) and $k_s$ (for the slow flow store). The model has a total of 5 states and 5 parameters.

196 In the HBV model, input to the soil store is represented by a power-law function (see Figure 3, note
197 the snow store is neglected for this study). Excess rainfall enters a shallow layer store which
198 generates: 1) near surface flow ($q_0$) whenever the shallow store state ($stw1$) is above a threshold
199 ($hl1$) and 2) interflow ($q_1$) by a linear routing mechanism controlled by the $K1$ parameter.
200 Percolation from the shallow layer store to the deep layer store (controlled by $perc$ parameter) then
201 leads to the generation of baseflow also via linear routing (controlled by the $K2$ parameter). Finally, a
202 triangular weighting function of base length $Maxbas$ is used to route the sum of all three flow
203 components. There are a total of 9 parameters and 3 states.
204
205 The Shuffled Complex Evolution Algorithm (SCE-UA) [*Duan et al.*, 1993] and the Borg Evolutionary
206 Algorithm [*Hadka & Reed*, 2013] were used to calibrate the models to pre-change conditions (1973
207 to 1979). The period 1973 to 1979 was selected for calibration as it was expected to have minimal
208 land cover changes, and also to ensure sufficient data availability for the assimilation period. Both
209 models had very similar performance in terms of reproducing observed runoff (an NSE of 0.75 and
210 0.77 for HyMOD and HBV respectively). HBV was slightly better at reproducing low flows whilst
211 HyMOD was slightly better at mid-range flows (see Table 2).

212 ## 3.2. Time Varying Parameter Estimation

213 A Data Assimilation based framework for estimating time varying parameters was presented in
214 *Pathiraja et al.* [2016a]. The approach relies on an Ensemble Kalman Filter (EnKF) [*Evensen*,1994] to
215 perform sequential joint state and parameter updating. EnKFs were developed to extend the
216 applicability of the celebrated Kalman Filter [*Kalman,* 1960] to non-linear systems, although they
217 provide a sub-optimal update as only the mean and covariance are considered in generating the
218 posterior. However they have been used with much success in many hydrologic applications [see for
219 example *Reichle et al.,* 2002; *Gu et al.*, 2005; *Komma et al.,* 2008; *Sun et al.,* 2009; *Xu et al.,* 2016].
220 EnKFs offer a practical alternative to Sequential Monte Carlo/Particle Filter methods that propagate

the full probability density through time, but suffer from several implementation issues even in moderate dimensional systems. The Locally Linear Dual EnKF method of *Pathiraja et al.* [2016a] works by sequentially proposing parameters, updating these using the Ensemble Kalman filter and available observations, and subsequently using these updated parameters to propose and update model states. An approach for proposing parameters in the time varying setting was also presented, for cases where no prior knowledge of parameter variations is available. The method was verified against multiple synthetic case studies as well as for 2 small experimental catchments experiencing controlled land use change [*Pathiraja et al.,* 2016a and *Pathiraja et al.,* 2016b]. The algorithm is summarised below, for full details refer to *Pathiraja et al.* [2016a].

### 3.2.1. Locally Linear Dual EnKF

Suppose a dynamical system can be described by a vector of states $\boldsymbol{x}_t$ and outputs $\boldsymbol{y}_t$ and a vector of associated model parameters $\boldsymbol{\theta}_t$ at any given time *t.* The uncertain system states and parameters are represented by an ensemble of states $\{\boldsymbol{x}_t^i\}_{i=1:n}$ and parameters $\{\boldsymbol{\theta}_t^i\}_{i=1:n}$ each with *n* members. The prior state and parameter distributions $\{\boldsymbol{x}_t^{i-}\}_{i=1:n}$ and $\{\boldsymbol{\theta}_t^{i-}\}_{i=1:n}$ respectively represent our prior knowledge of the system, usually derived as the output from a numerical model. Suppose also that the system outputs are observed $(\boldsymbol{y}_t^o)$ but that there is also some uncertainty associated with these observations. The purpose of the data assimilation algorithm (here the EnKF) is to combine the prior estimates with measurements, based on their respective uncertainties, to obtain an improved estimate of the system states and parameters. A single cycle of the Locally Linear Dual EnKF procedure for a given time *t* is undertaken as follows. Note in the following, the overbar notation is used to indicate the ensemble mean.

1. ***Propose a prior parameter ensemble.*** This involves generating a parameter ensemble using prior knowledge. In this case, our prior knowledge comes from the updated parameter ensemble from the previous time ($\boldsymbol{\theta}_{t-1}^{i+}$) and how it has changed over recent time steps. The

assumed parameter dynamics is a Gaussian random walk with time varying mean and variance, given by:

$$\boldsymbol{\theta}_t^{i-} \sim N\left(\boldsymbol{\theta}_{t-1}^{i+} + \boldsymbol{m}_t, s^2\boldsymbol{\Sigma}_{t-1}^{\theta}\right) \; for\; i = 1{:}n \tag{1}$$

$$\boldsymbol{\Sigma}_{t-1}^{\theta} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\boldsymbol{\theta}_{t-1}^{i+} - \overline{\boldsymbol{\theta}_{t-1}^{+}}\right)\left(\boldsymbol{\theta}_{t-1}^{i+} - \overline{\boldsymbol{\theta}_{t-1}^{+}}\right)^{\mathrm{T}} \tag{2}$$

where $\boldsymbol{\Sigma}_{t-1}^{\theta}$ is the sample covariance matrix of the updated parameter ensemble at time $t-1$; $\overline{\boldsymbol{\theta}_{t-1}^{+}}$ indicates the ensemble mean of the updated parameters at time $t-1$; $(\,)^{\mathrm{T}}$ represents the transpose operator; and $s^2$ is a tuning parameter. The prior ensemble mean is determined as the linear extrapolation of the updated ensemble means from the previous two time steps, i.e.:

$$\boldsymbol{m}_t[k] = \begin{cases} \boldsymbol{m}_{t-1}[k], & |\boldsymbol{m}_{t-1}[k]| \le m_{max} \\ \boldsymbol{m}_{t-2}[k], & |\boldsymbol{m}_{t-1}[k]| > m_{max} \end{cases} \tag{3}$$

$$\boldsymbol{m}_{t-1} = \frac{\overline{\boldsymbol{\theta}_{t-1}^{+}} - \overline{\boldsymbol{\theta}_{t-2}^{+}}}{\Delta t} \tag{4}$$

$$\boldsymbol{m}_{t-2} = \frac{\overline{\boldsymbol{\theta}_{t-2}^{+}} - \overline{\boldsymbol{\theta}_{t-3}^{+}}}{\Delta t} \tag{5}$$

where $\boldsymbol{m}_t[k]$ indicates the kth component of the vector $\boldsymbol{m}_t$. Note that the extrapolation is forced to be less than a pre-defined maximum rate of change $m_{max}$ to minimise overfitting and avoid parameter drift due to isolated large updates.

2. ***Consider observation and forcing uncertainty.*** This is done by perturbing measurements of forcings and system outputs with random noise sampled from a distribution representing the uncertainty in those measurements. The result is an ensemble of forcings ($\boldsymbol{u}_t^i$) and observations ($\boldsymbol{y}_t^i$) each with *n* members. For example, if random errors in measurements of system outputs (herein also referred as observations) are characterized by a zero mean Gaussian distribution, the ensemble of observations is given by:

$$\boldsymbol{y}_t^i \sim N\left(\boldsymbol{y}_t^o, \boldsymbol{\Sigma}_t^{y^o y^o}\right) \; for\; i = 1{:}n \tag{6}$$

where $\boldsymbol{y}_t^o$ is the recorded measurement at time $t$ and $\boldsymbol{\Sigma}_t^{y^o y^o}$ is the error covariance matrix of the measurements.

3. ***Generate simulations using prior parameters***. The prior parameters from Step 1, $\boldsymbol{\theta}_t^{i-}$ and updated states from the previous time, $\boldsymbol{x}_{t-1}^{i+}$ are forced through the model equations to generate an ensemble of model simulations of states ($\widehat{\boldsymbol{x}}_t^i$) and outputs ($\widehat{\boldsymbol{y}}_t^i$):

$$\widehat{\boldsymbol{x}}_t^i = f\left(\boldsymbol{x}_{t-1}^{i+}, \boldsymbol{\theta}_t^{i-}, \boldsymbol{u}_t^i\right) \; for\; i = 1{:}n \tag{7}$$

$$\widehat{\boldsymbol{y}}_t^i = h\left(\widehat{\boldsymbol{x}}_t^i, \boldsymbol{\theta}_t^{i-}\right) \; for\; i = 1{:}n \tag{8}$$

4. ***Perform the Kalman update of parameters.*** Parameters are updated using the Kalman update equation and the prior parameter and simulated output ensemble from Step 1 and 3:

$$\boldsymbol{\theta}_t^{i+} = \boldsymbol{\theta}_t^{i-} + \mathbf{K}_t^{\theta}\left(\boldsymbol{y}_t^i - \widehat{\boldsymbol{y}}_t^i\right) \; for\; i = 1{:}n \tag{9}$$

$$\mathbf{K}_t^{\theta} = \boldsymbol{\Sigma}_t^{\theta\hat{y}}\left[\boldsymbol{\Sigma}_t^{\hat{y}\hat{y}} + \boldsymbol{\Sigma}_t^{y^o y^o}\right]^{-1} \tag{10}$$

where $\boldsymbol{\Sigma}_t^{\theta\hat{y}}$ is a matrix of the sample cross covariance between errors in parameters $\boldsymbol{\theta}_t^{i-}$ and simulated output $\widehat{\boldsymbol{y}}_t^i$ ; and $\boldsymbol{\Sigma}_t^{\hat{y}\hat{y}}$ is the sample error covariance matrix of the simulated output:

$$\boldsymbol{\Sigma}_t^{\theta\hat{y}} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\boldsymbol{\theta}_t^{i-} - \overline{\boldsymbol{\theta}_t^{-}}\right)\left(\widehat{\boldsymbol{y}}_t^i - \overline{\widehat{\boldsymbol{y}}_t}\right)^{\mathrm{T}} \tag{11}$$

$$\boldsymbol{\Sigma}_t^{\hat{y}\hat{y}} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\widehat{\boldsymbol{y}}_t^i - \overline{\widehat{\boldsymbol{y}}_t}\right)\left(\widehat{\boldsymbol{y}}_t^i - \overline{\widehat{\boldsymbol{y}}_t}\right)^{\mathrm{T}} \tag{12}$$

5. ***Generate simulations using updated parameters.*** Step 3 is repeated with the updated parameter ensemble $\boldsymbol{\theta}_t^{i+}$ to generate the prior ensemble of model simulations of states ($\boldsymbol{x}_t^{i-}$) and outputs ($\widetilde{\boldsymbol{y}}_t^i$):

$$\boldsymbol{x}_t^{i-} = f\left(\boldsymbol{x}_{t-1}^{i+}, \boldsymbol{\theta}_t^{i+}, \boldsymbol{u}_t^i\right) \; for\; i = 1{:}n \tag{13}$$

$$\widetilde{\boldsymbol{y}}_t^i = h\left(\boldsymbol{x}_t^{i-}, \boldsymbol{\theta}_t^{i+}\right) \; for\; i = 1{:}n \tag{14}$$

6. ***Perform the Kalman update of states and outputs***. Use the Kalman update equation for correlated measurement and process noise (equations 15 to 18) and the simulated state ($\boldsymbol{x}_t^{i-}$) and output ($\widetilde{\boldsymbol{y}}_t^i$) ensembles from Step 5 to update them. Since the measurements have already been used to generate $\widetilde{\boldsymbol{y}}_t^i$, the errors in model simulations and measurements are now correlated. The standard Kalman update equation (as in the form of equations 9 and

10) can no longer be used as it relies on the assumption that errors in measurements and model simulations are independent.

$$\boldsymbol{x}_t^{i+} = \boldsymbol{x}_t^{i-} + \mathbf{K}_t^{x}\left(\boldsymbol{y}_t^{i} - \widetilde{\boldsymbol{y}}_t^{i}\right) \; for\; i = 1{:}\,n \tag{15}$$

$$\mathbf{K}_t^{x} = \left[\boldsymbol{\Sigma}_t^{x\tilde{y}} + \boldsymbol{\Sigma}_t^{\varepsilon_x y^o}\right]\left[\boldsymbol{\Sigma}_t^{\tilde{y}\tilde{y}} + \boldsymbol{\Sigma}_t^{\varepsilon_{\tilde{y}} y^o} + \left(\boldsymbol{\Sigma}_t^{\varepsilon_{\tilde{y}} y^o}\right)^{\mathrm{T}} + \boldsymbol{\Sigma}_t^{y^o y^o}\right]^{-1} \tag{16}$$

$$\boldsymbol{\varepsilon}_{x_t}^{\;i} = \boldsymbol{x}_t^{i-} - \widehat{\boldsymbol{x}}_t^{i} \tag{17}$$

$$\boldsymbol{\varepsilon}_{\tilde{y}_t}^{\;i} = \widetilde{\boldsymbol{y}}_t^{i} - \widehat{\boldsymbol{y}}_t^{i} \tag{18}$$

where $\boldsymbol{\Sigma}_t^{x\tilde{y}}$ is a matrix of the sample cross covariance between simulated states $\left\{\boldsymbol{x}_t^{i-}\right\}_{i=1:n}$ and outputs $\left\{\widetilde{\boldsymbol{y}}_t^{i}\right\}_{i=1:n}$ from Step 5; $\boldsymbol{\Sigma}_t^{\varepsilon_x y^o}$ represents the sample covariance between $\left\{\boldsymbol{\varepsilon}_{x_t}^{\;i}\right\}_{i=1:n}$ and the observations; and $\boldsymbol{\Sigma}_t^{\varepsilon_{\tilde{y}} y^o}$ represents the sample covariance between the $\left\{\boldsymbol{\varepsilon}_{\tilde{y}_t}^{\;i}\right\}_{i=1:n}$ and the observations.

The above algorithm specifies the updating of states and parameters at any given time, based on available observations. This allows one to retrospectively estimate time variations in model parameters, as well as provide one time step ahead forecasts of states & outputs (as per equations 7 and 8). Forecasts at longer time horizons would be made by generating prior parameters and states as detailed in Steps 1 to 3, although the local linear extrapolations are only valid close to the current time point.

### 3.2.2. Application to the Nammuc Catchment

Joint state and parameter estimation was undertaken for the Nammuc Catchment over the period 1975 to 2004 by assimilating streamflow observations into the HyMOD and HBV models at a daily time step. Given the fairly low parameter dimensionality of HyMOD, all model parameters were allowed to vary in time whilst for HBV the $lp$ and $Maxbas$ parameters (see Figure 3) were held fixed ($lp$ = 1 and $Maxbas$ = 1 day). This was based on the results of Variance Based Sensitivity Analysis or Sobol method [see for example *Saltelli et al.,* 2008] implemented through the SAFE toolbox [*Pianosi*

298 *et al.,* 2015] which found these to be the least sensitive and least important in defining variations to
299 catchment hydrology (see Table 3). Note that although the $hl1$ parameter was found to have low
300 sensitivity, it was retained as a time varying parameter due to its conceptual importance in
301 separating interflow and near surface flow (refer Figure 3).
302
303 Unbiased normally distributed ensembles of the parameters and states are required to initialise the
304 LL Dual EnKF. Initial parameter ensembles were generated by sampling from a Gaussian distribution
305 with mean equal to the calibrated parameters over the pre-change period and variance estimated
306 from parameter sets with similar objective function values. Parameter sets with similar objective
307 function values were obtained when using different starting points to the optimization algorithm
308 during the model calibration stage. Initial state ensembles were also sampled from normal
309 distributions with mean equal to the simulated state at the end of the calibration period. An
310 ensemble size of 100 members was adopted and assumed sufficiently large based on the findings of
311 *Moradkhani et al.* [2005] and *Aksoy et al.* [2006]. Due to the stochastic-dynamic nature of the
312 method, ensemble statistics were calculated over 20 separate realisations of the LL Dual EnKF. The
313 prior parameter generating method described in Step 1 of Section 3.2 requires specification of the
314 tuning parameter $s^2$ to define the variance of the perturbations. This was tuned by selecting the $s^2$
315 value that optimized the quality of forecast streamflow over the calibration period. Forecast quality
316 was assessed using the logarithmic score (LS) [*Good,* 1952] of background streamflow predictions
317 ($\tilde{y}_t^i$) using updated parameters (equation 14), which was averaged over the calibration period of
318 length *T*:

$$\overline{LS} = \sum_{t=1}^{T} LS_t \quad (19)$$

$$LS_t = \log\left(f(y = y_t^o)\right) \quad (20)$$

319 where $f(y)$ is the probability density function of the background streamflow predictions
320 (represented by the empirical pdf of the sample points $\{\tilde{y}_t^i\}_{i=1:n}$); and $y_t^o$ is the measurement of the

system outputs. The $s^2$ value that gave the largest $\overline{LS}$ was adopted for the assimilation period. The maximum allowable daily rate of change in the ensemble mean was based on assuming a linear rate of change within the entire feasible parameter space over a three year period.

As detailed in Section 3.2, observation and forcing uncertainty is considered by perturbing measurements with random noise. Here streamflow errors were assumed to be zero-mean normally distributed (truncated to ensure positivity) and heteroscedastic. The variance is defined as a proportion of the observed streamflow, to reflect the fact that larger flows tend to have greater errors than low flows:

$$y_t^i \sim TN(y_t^o, d.y_t^o) \;\; for\; i = 1{:}n \qquad (21)$$

where TN indicates the truncated normal distribution to ensure positive flows and $d$ = 0.1. A multiplier of 0.1 was chosen based on estimates adopted for similar gauges in hydrologic DA studies [e.g. *Clark et al.,* 2008; *Weerts & Serafy,* 2006; *Xie et al.,* 2014].

Several studies have noted that a major source of rainfall uncertainty arises from scaling point rainfall to the catchment scale [*Villarini & Krajewski*, 2008; *McMillan et al.*, 2011] and that multiplicative errors models are suited to describing such errors [e.g. *Kavetski et al.*, 2006]. Rainfall uncertainties were therefore described using unbiased, lognormally distributed multipliers:

$$P_t^i = P_t.M^i \qquad (22)$$

$$M^i \sim LN(m, v) \text{ and } X^i = \log\left(M^i\right) \sim N(\mu, \sigma^2) \quad for\; i = 1{:}n \qquad (23)$$

where $P_t$ is the measured rainfall at time $t$; $m$ and $v$ are the mean and variance of the lognormally distributed rainfall multipliers $M$ respectively; and $\mu$ and $\sigma^2$ are the mean and variance of the normally distributed logarithm of the rainfall multipliers $M$. For unbiased perturbations, we let $m$ = 1. The variance of the rainfall multipliers ($v$) was estimated by considering upper and lower bound error estimates in the Thiessen weights assigned to the four rainfall stations (see Section 2.1 for calculation of catchment averaged rainfall, $P_t$). The resulting upper and lower bound catchment

averaged rainfall data were then used to estimate error parameters due to spatial variation in rainfall:

$$v = e^{(2\mu + \sigma^2)}.\left(e^{\sigma^2} - 1\right) \quad (24)$$

$$\sigma^2 = \widehat{\sigma^2} = var\left(\log\left[\frac{P_{upper,10}}{P_{lower,10}}\right]\right) \quad (25)$$

$$\mu = \log(m) - \frac{\sigma^2}{2} = -\frac{\sigma^2}{2} \quad (26)$$

where $P_{upper,10}$ indicates catchment averaged rainfall data estimated using the upper bound Thiessen weights with daily depth greater than 10mm (similar for $P_{lower,10}$) . $\widehat{\sigma^2}$ A 10mm rainfall depth threshold was chosen to avoid large rainfall fractions due to small rainfall depths. $\widehat{\sigma^2}$ was found to be 0.05 in this case study. Similarly, we assume the dominant source of uncertainty in temperature data arises from spatial variation. Differences in temperature records at Lai Chau and Quynh Nhai (only available gauges with temperature records) were analysed and found to be approximately normally distributed with sample mean 0.2 deg C and variance of 1.4 deg C. A perturbed temperature ensemble was then generated according to equation 27:

$$T_t^i \sim TN\left(T_t^{avg}, 1.4\right) \;\; for \;\; i = 1{:}n \quad (27)$$

where $T_t^{avg}$ represents catchment averaged temperature data (see Section 2.1). Note that perturbations were taken to be unbiased (zero mean) as the sample mean of the differences in the temperature records was close to zero. The same perturbed input and observation sequences were used for the HyMOD and HBV runs for the sake of comparison. A summary of the values adopted for the various components of the Locally Linear Dual EnKF for each model is provided in Table 4 and Table 5.

# 4. Results and Discussion

Temporal variations in the estimated parameter distributions from the LL Dual EnKF are evident for both models. In the case of the HBV model, changes at an inter-annual time scale are evident for the

363 $perc$ and $\beta$ (see Figure 4). The decrease in the $\beta$ parameter means that a greater proportion of
364 rainfall is converted to runoff (i.e. more water entering the shallow layer storage). Additionally, the
365 increase in the $perc$ parameter means that a greater volume of water is made available for baseflow
366 generation. These changes correspond with the observed increase in the annual runoff coefficient
367 (Figure 2) and increase in baseflow volume (as discussed in Section 2.2). Similar parameter
368 adjustments are seen for HyMOD, at least at a qualitative level (see Figure 5). The sharp increase in
369 the $b$ parameter during the post-change period means that a greater volume of water is available for
370 routing (as larger $b$ values mean that a smaller proportion of the catchment has deep soil storage
371 capacity) and the downward inter-annual trend in $\alpha$ means that a greater portion of excess runoff is
372 routed through the baseflow store. Intra-annual variations in updated model parameters for both
373 HyMOD and HBV are also apparent (refer Figure 4 and Figure 5). This is due to the inability of a
374 single parameter distribution to accurately model both wet and dry season flows, an issue that is
375 commonly encountered when modelling large heterogeneous catchments experiencing significant
376 spatial variation in rainfall. Such variations were not observed when using the time varying
377 parameter framework for small deforested catchments (< 350ha) [see *Pathiraja et al.,* 2016b]. The
378 comparatively less clear parameter changes for the Nammuc catchment are due to a combination of
379 the increased difficulty in accurately modelling the hydrologic response (even in pre-change
380 conditions) and due to the relatively more subtle and gradual changes to land cover. Nonetheless,
381 the method is shown to generate a temporally varying structure that is conceptually representative
382 of the observed changes.
383
384 Despite the overall correspondence between changes to model parameters and observed
385 streamflow, a closer examination shows that the hydrologic model structure is critical in determining
386 whether the time varying parameter models accurately reflect changes in all aspects of the
387 hydrologic response (not just total streamflow). In order to examine the impact of parameter
388 variations on the model dynamics, we generated model simulations with the time varying parameter

389 ensemble from the LL Dual EnKF, but without state updating (hereafter referred to as TVP-HBV and
390 TVP-HyMOD). Streamflow predictions from the LL Dual EnKF (i.e. with state and parameter updating)
391 for both the HyMOD and HBV are generally of similar quality and superior to those from the
392 respective time invariant parameter models, although a slight bias in baseflow predictions from
393 HyMOD is evident (see for example Figure 6). However, differences in predictions from TVP-HBV and
394 TVP-HyMOD are more striking due to the lack of state updating. Figure 7 shows annual statistics of
395 simulated streamflow from the TVP-HBV and TVP-HyMOD models and observed runoff. The TVP-
396 HBV gives direct runoff and baseflow predictions that are consistent with runoff observations,
397 meaning that the parameter adjustments reflect the observed changes in the runoff response. This
398 however is not the case for the TVP-HyMOD. The annual runoff coefficient and annual direct runoff
399 coefficient are severely under-estimated in the post-change period by the TVP-HyMOD, whilst the
400 Annual Baseflow Index has an increasing trend of magnitude far greater than observed (Figure 7c).
401 All three quantities on the other hand are well represented by the TVP-HBV (Figure 7).
402
403 Similar conclusions can be drawn from Figure 8, which shows the results of a Moving Average
404 Shifting Horizon (MASH) analysis (see Section 2.2) on total and direct runoff (observed and
405 simulated). Observed increases in January to April flows (see Figure 8a) and wet season direct flows
406 (July to September) (see Figure 8e) are well represented by the TVP-HBV but not TVP-HyMOD. The
407 reason for these differences between the two models lies in their structure. In joint state-parameter
408 updating using HyMOD, underestimated runoff predictions during recession periods lead to
409 adjustments to the $k_s$ and $\alpha$ parameters to increase baseflow depth. Unlike HBV, HyMOD has no
410 continuous supply of water to the routing stores (i.e. the quick flow and slow flow stores) during
411 recession periods (which typically have extended periods of no rainfall, so that $V$ in Figure 3 is zero).
412 This means that $k_s$ and $\alpha$ are updated to extreme values to compensate for the volumetric shortfall.
413 HBV on the other hand has a continuous percolation of water into the deep layer store even during
414 periods of no rain (so long as the shallow water store is non-empty). In summary, the HyMOD model

structure prevents the parameters from being updated to values that realistically reflect the observed changes to catchment dynamics.

Having established that the TVP-HBV provided a good representation of the observed streamflow dynamics, we used a modelling approach to determine whether the observed changes were solely driven by forcings and which (if any) components of runoff were also affected by land use change. A resampled rainfall and temperature time series was generated by sampling the data without replacement across years for each day (for instance rainfall and temperature for 1st January 1990 is found by randomly sampling from all records on 1st January). This maintains the intra-annual (e.g. seasonal) variability but destroys any inter-annual trends in the meteorological data. Streamflow simulations were then generated using this resampled meteorological sequence as inputs to the TVP-HBV (i.e. without state updating). If the resulting streamflow simulations do not reproduce the observed changes to streamflow dynamics, then this indicates that changes to meteorological forcings are the main contributor. However, if it is able to at least partially (or fully) reproduce the observed streamflow changes, this means that land cover changes are impacting catchment hydrology (but potentially in addition to forcing changes, due to the presence of ecosystem feedbacks). Figure 8d&h show the results of a MASH undertaken on the resulting simulations of total and direct runoff using the resampled forcing time series and TVP-HBV model. Observed increases in baseflow during the January – April period (see Figure 8a) and increases in direct runoff in the June – September period (see Figure 8e) are reproduced. The magnitude of increase in direct runoff in July is slightly lower, indicating the potential for some climatic influences also. This is consistent with findings from the Mann-Kendall test which identified a statistically significant increase in July rainfall (see Section 2.2). Overall however, these results lend further weight to the conclusion that land cover change has impacted the hydrologic regime of the Nammuc catchment. These results also demonstrate that parameter changes correspond to actual changes in catchment hydrology, and are

not just random fluctuations that reproduce the observed streamflow statistics only when the observed forcing time series is used.

# 5. Conclusions

As our anthropogenic footprint expands, it will become increasingly important to develop modelling methodologies that are capable of handling dynamic catchment conditions. Previous work proposed the use of models whose parameters vary with time in response to signals of change in observations. The so-called Locally Linear Dual EnKF time varying parameter estimation algorithm [*Pathiraja et al.,* 2016a] was applied to 2 sets of small (< 350 ha) paired experimental catchments with deforestation occurring under experimental conditions (rapid clearing of 100% and 50% of land surface) [*Pathiraja et al.,* 2016b]. Here we demonstrate the efficacy of the method for a larger catchment experiencing more realistic land cover change, whilst also investigating the importance of the chosen model structure in ensuring the success of time varying parameter methods. We also demonstrate that the time varying parameter framework can be used in a retrospective fashion to determine whether land cover changes (and not just meteorological factors) contribute to the observed hydrologic changes.

Experiments were undertaken on the Nammuc catchment (2880 $km^2$) in Vietnam, which experienced a relatively gradual conversion from forest to cropland over a number of years (cropland increased from roughly 23% of the catchment between 1981 and 1994 to 52% by 2000). Changes to the hydrologic regime after the mid-1990s were detected and attributed mostly to an increase in baseflow volume. Application of the LL Dual EnKF with two conceptual models (HBV and HyMOD) showed that the time varying parameter framework with state updating improved streamflow prediction in post-change conditions compared to the time invariant parameter case. However, baseflow predictions from the LL Dual EnKF with HBV were generally superior to the HyMOD case which tended to have a slight negative bias. It was found that the structure (i.e. model equations) of HyMOD was unsuited to representing the modified baseflow conditions, resulting in extreme and

unrealistic time varying parameter estimates. This work shows that the chosen model is critical for ensuring the time varying parameter framework successfully models streamflow in unknown future land cover conditions. Appropriate model selection can be a difficult task due to the significant uncertainty associated with future land use change, and can be even more problematic when multiple models have similar performance in pre-change conditions (as was the case in this study). One possible way to ensure success of the time varying parameter approach is to use physically based models whose fundamental equations more closely model physical processes (for instance, modelling sub-surface flow using Richard's equation with hydraulic conductivity allowed to vary with time). The drawback of such approaches is that they are generally data intensive, both in generating model simulations (i.e. detailed inputs) and specifying parameters. Another possibility is to combine time varying parameter framework with multi-model approaches.

## 6. Acknowledgements

This study was funded by the Australian Research Council as part of the Discovery Project DP140102394. Dr Marshall is additionally supported through a Future Fellowship FT120100269.

The data used in this paper were collected under the project IMRR (Integrated and sustainable water Management of Red Thai Binh Rivers System in changing climate), funded by the Italian Ministry of Foreign Affairs (Delibera n. 142 del 8 Novembre 2010). We greatly acknowledge Dr. Andrea Castelletti for provision of data and for discussions on this work.

Data utilized in this study can be made available from the authors upon request.

# 7. References


Aksoy, A., Zhang, F., Nielsen-Gammon, J. (2006). Ensemble-Based Simultaneous State and Parameter Estimation in a Two-Dimensional Sea-Breeze Model. *Monthly Weather Review*, 134, 2951–2970.

Anghileri, D., Pianosi, F., & Soncini-Sessa, R. (2014). Trend detection in seasonal data: From hydrology to water resources. *Journal of Hydrology*, *511*, 171–179. http://doi.org/10.1016/j.jhydrol.2014.01.022

Bergström, S. 1995. The HBV model. In: Singh, V.P. (Ed.) *Computer Models of Watershed Hydrology*. Water Resources Publications, Highlands Ranch, CO., pp. 443-476.

Bhaduri, B. B., Minner, M., Tatalovich, S., Member, A., & Harbor, J. (2001). Long-term hydrologic impact of urbanization: A tale of two models. *Journal of Water Resources Planning and Management*, *127*(February), 13–19.

Boyle, D. (2001). Multicriteria calibration of hydrological models, *Ph.D. dissertation*, Univ. of Ariz., Tucson.

Brown, A. E., Mcmahon, T. A., Podger, G. M., & Zhang, L. (2006). A methodology to predict the impact of changes in forest cover on flow duration curves, *CSIRO Land and Water Science Report 8/06*.

Brown, A. E., Western, A. W., McMahon, T. a., & Zhang, L. (2013). Impact of forest cover changes on annual streamflow and flow duration curves. *Journal of Hydrology*, *483*, 39–50. http://doi.org/10.1016/j.jhydrol.2012.12.031

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., … Uddstrom, M. J. (2008). Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. *Advances in Water Resources*, *31*(10), 1309–1324. http://doi.org/10.1016/j.advwatres.2008.06.005

Coe, M. T., Latrubesse, E. M., Ferreira, M. E., & Amsler, M. L. (2011). The effects of deforestation and climate variability on the streamflow of the Araguaia River, Brazil. *Biogeochemistry*, *105*(1–3), 119–131. http://doi.org/10.1007/s10533-011-9582-2

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., & Hendrickx, F. (2012). Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments. *Water Resources Research,* 48(5), 1–17. doi:10.1029/2011WR011721

Costa, M. H., Botta, A., & Cardille, J. A. (2003). Effects of large-scale changes in land cover on the discharge of the Tocantins River, Southeastern Amazonia. *Journal of Hydrology*, *283*(1–4), 206–217. http://doi.org/10.1016/S0022-1694(03)00267-1

Duan, Q. Y., Gupta, V. K., & Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications*, *76*(3), 501–521. doi:10.1007/BF00939380

Dwarakish, G. S., & Ganasri, B. P. (2015). Impact of land use change on hydrological systems: A review of current modeling approaches. *Cogent Geoscience*, *1*(1), 1115691–1115691. http://doi.org/10.1080/23312041.2015.1115691

523 Eckhardt, K. (2005). How to construct recursive digital filters for baseflow separation. *Hydrological*
524 *Processes*, *19*(2), 507–515. http://doi.org/10.1002/hyp.5675

525 Efstratiadis, A., Nalbantis, I., & Koutsoyiannis, D. (2015). Hydrological modelling of temporally-varying
526 catchments: facets of change and the value of information. *Hydrological Sciences Journal*, *60*(7–
527 8), 1438–1461. http://doi.org/10.1080/02626667.2014.982123

528 Elfert, S., & Bormann, H. (2010). Simulated impact of past and possible future land use changes on
529 the hydrological response of the Northern German lowland "Hunte" catchment. *Journal of*
530 *Hydrology*, 383, 245–255. http://dx.doi.org/10.1016/j.jhydrol.2009.12.040

531 Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using
532 Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, *99*(C5).
533 http://doi.org/10.1029/94JC00572

534 FAO (2005). Global Forest Resources Assessment 2005 (FRA 2005)

535 Good, I.J. (1952). Rational Decisions. *Journal of the Royal Statistical Society.* B 14: 107–114.
536
537 Gu, Y., and D. S. Oliver (2005), History matching of the PUNQ-S3 reservoir model using the ensemble
538 Kalman filter, *SPE J.*, *10*(2), 217–224, doi:10.2118/89942-PA.
539
540 Giuliani, M., Anghileri, D., Castelletti, A., Vu, P. N., & Soncini-Sessa, R. (2016). Large storage
541 operations under climate change: expanding uncertainties and evolving tradeoffs.
542 *Environmental Research Letters*, 11(3), 035009.
543
544 Hadka, D., Reed, P., (2013). Borg: an auto–adaptive many–objective evolutionary computing
545 framework. *Evol. Comput*. 21 (2), 231–259.

546 Hamon, W. (1961). Estimating potential evapotranspiration. *Transactions of the American Society of*
547 *Civil Engineers*, 128(1), pp.324-337.

548 Kalman, R.E. (1960) A new approach to linear filtering and prediction problems, Transactions of the
549 ASME – *Journal of Basic Engineering*, Series D, 82 (1960), 35–45.

550 Kavetski, D., Kuczera, G., & Franks, S. W. (2006). Bayesian analysis of input uncertainty in hydrological
551 modeling: 1. Theory. *Water Resources Research*, *42*(3), n/a–n/a. doi:10.1029/2005WR004368
552
553 Kim, D.-H., J. O. Sexton, and J. R. Townshend (2015). Accelerated deforestation in the humid tropics
554 from the 1990s to the 2000s, *Geophysical Research Letters*, 42, 3495–3501, doi:10.1002/
555 2014GL062777.

556 Komma, J., G. Blöschl, and C. Reszler (2008), Soil moisture updating by Ensemble Kalman Filtering in
557 real-time flood forecasting, *J. Hydrol.*, *357*(3–4), 228–242, doi:10.1016/j.jhydrol.2008.05.020.
558
559 Kummer, D., and Turner, B. (1994). The Human Causes of Deforestation in Southeast Asia. *BioScience,*
560 *44*(5), 323-328. doi:10.2307/1312382
561
562 Legesse, D., Vallet-Coulomb, C., & Gasse, F. (2003). Hydrological response of a catchment to climate
563 and land-use changes in Tropical Africa: case study South Central Ethiopia. *Journal of Hydrology*,
564 *275*(1-2), 67–85. doi:10.1016/S0022-1694(03)00019-2

McIntyre, N., & Marshall, M. (2010). Identification of rural land management signals in runoff response. *Hydrological Processes*, *24*(24), 3521–3534. doi:10.1002/hyp.7774

McMillan, H., Jackson, B., Clark, M., Kavetski, D., & Woods, R. (2011). Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models. *Journal of Hydrology*, *400*(1-2), 83–94. doi:10.1016/j.jhydrol.2011.01.026

Moradkhani, H., Sorooshian, S., Gupta, H. V., & Houser, P. R. (2005). Dual state–parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources*, *28*(2), 135–147. http://doi.org/10.1016/j.advwatres.2004.09.002

Niu, J., & Sivakumar, B. (2013). Study of runoff response to land use change in the East River basin in South China. *Stochastic Environmental Research and Risk Assessment*. doi:10.1007/s00477-013-0690-5

Pathiraja, S., Marshall, L., Sharma, A., & Moradkhani, H. (2016a). Hydrologic modeling in dynamic catchments: A data assimilation approach. *Water Resources Research*, *52*, 3350–3372. http://doi.org/10.1002/2015WR017192

Pathiraja, S., Marshall, L., Sharma, a., & Moradkhani, H. (2016b). Detecting non-stationary hydrologic model parameters in a paired catchment system using data assimilation. *Advances in Water Resources*, *94*, 103–119. http://doi.org/10.1016/j.advwatres.2016.04.021

Pianosi, F., Sarrazin, F., Wagener, T. A Matlab toolbox for Global Sensitivity Analysis, *Environmental Modelling & Software*, 70, 80-85, http://dx.doi.org/10.1016/j.envsoft.2015.04.009.

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi (2002), Hydrologic Data Assimilation with the Ensemble Kalman Filter, *Am. Meteorol. Soc. - Mon. Weather Rev.*, *130*(1), 103–114, doi:10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2.

Rose, S., & Peters, N. E. (2001). Effects of urbanization on streamflow in the Atlanta area (Georgia, USA): a comparative hydrological approach. *Hydrological Processes*, *15*(8), 1441–1457. http://doi.org/10.1002/hyp.218

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., (2008). Global Sensitivity Analysis, the Primer. Wiley.

Seibert, J., & McDonnell, J. J. (2010). Land-cover impacts on streamflow: a change-detection modelling approach that incorporates parameter uncertainty. *Hydrological Sciences Journal*, *55*(3), 316–332. doi:10.1080/02626661003683264

Sun, A. Y., A. Morris, and S. Mohanty (2009), Comparison of deterministic ensemble Kalman filters for assimilating hydrogeological data, *Adv. Water Resour.*, *32*(2), 280–292, doi:10.1016/j.advwatres.2008.11.006.

Taver, V., Johannet, a., Borrell-Estupina, V., & Pistre, S. (2015). Feed-forward vs recurrent neural network models for non-stationarity modelling using data assimilation and adaptivity. *Hydrological Sciences Journal*, *60*(7–8), 1242–1265. http://doi.org/10.1080/02626667.2014.967696

605 Thanapakpawin, P., Richey, J., Thomas, D., Rodda, S., Campbell, B., & Logsdon, M. (2007). Effects of
606 landuse change on the hydrologic regime of the Mae Chaem river basin, NW Thailand. *Journal*
607 *of Hydrology*, *334*(1-2), 215–230. doi:10.1016/j.jhydrol.2006.10.012
608
609 Villarini, G., & Krajewski, W. F. (2008). Empirically-based modeling of spatial sampling uncertainties
610 associated with rainfall measurements by rain gauges. *Advances in Water Resources*, *31*(7),
611 1015–1023. doi:10.1016/j.advwatres.2008.04.007
612
613 Vu, V.T., 1993. Evaluation of the impact of deforestation to inflow regime of the Hoa Binh Reservoir
614 in Vietnam, Hydrology of Warm Humid Regions (Proceedings of the Yokohama Symposium,
615 July 1993). IAHS Publ. no. 216
616
617 Wang, J., Ishidaira, H., & Xu, Z. X. (2012). Effects of climate change and human activities on inflow
618 into the Hoabinh Reservoir in the Red River basin. *Procedia Environmental Sciences*, *13*, 1688-
619 1698.

620 Warburton, M. L., Schulze, R. E., & Jewitt, G. P. W. (2012). Hydrological impacts of land use change in
621 three diverse South African catchments. *Journal of Hydrology*, *414–415*, 118–135.
622 http://doi.org/10.1016/j.jhydrol.2011.10.028

623 Weerts, A. H., & El Serafy, G. Y. H. (2006). Particle filtering and ensemble Kalman filtering for state
624 updating with hydrological conceptual rainfall-runoff models. *Water Resources Research*, *42*(9),
625 n/a-n/a. http://doi.org/10.1029/2005WR004093

626 Westra, S.; Thyer, M.; Leonard, M.; Kavetski, D.; Lambert, M. (2014). A strategy for diagnosing and
627 interpreting hydrological model nonstationarity. *Water Resources Research*, 5090–5113.
628 http://doi.org/10.1002/2013WR014719.Received

629 Wijesekara, G. N., Gupta, A., Valeo, C., Hasbani, J. G., Qiao, Y., Delaney, P., & Marceau, D. J. (2012).
630 Assessing the impact of future land-use changes on hydrological processes in the Elbow River
631 watershed in southern Alberta, Canada. *Journal of Hydrology*, *412–413*, 220–232.
632 http://doi.org/10.1016/j.jhydrol.2011.04.018

633 WWF. (2013). Ecosystems in the Greater Mekong: Past trends, current status, possible futures.

634 Xie, X., Meng, S., Liang, S., & Yao, Y. (2014). Improving streamflow predictions at ungauged locations
635 with real-time updating: application of an EnKF-based state-parameter estimation strategy.
636 *Hydrology and Earth System Sciences*, *18*(10), 3923–3936. http://doi.org/10.5194/hess-18-
637 3923-2014

638 Xu, T., and J. . Gomez-Hernandez (2016), Joint identification of contaminant source location, initial
639 release time, and initial solute concentration in an aquifer via ensemble kalman filtering, *Water*
640 *Resour. Res.*, 600–612, doi:10.1002/2015WR018249.

641 Yang, L., Wei, W., Chen, L., & Mo, B. (2012). Response of deep soil moisture to land use and
642 afforestation in the semi-arid Loess Plateau, China. *Journal of Hydrology*, *475*, 111–122.
643 http://doi.org/10.1016/j.jhydrol.2012.09.041

# Tables

| | Pre 1994 | Post 1994 |
|---|---|---|
| ***Land Use*** | | |
| **Evergreen Forest (including evergreen needle and evergreen leaf) (%)** | 77% | 48% |
| **Cropland (%)** | 23% | 52% |
| ***Hydro-Meteorological Properties*** | | |
| **Mean Annual Rainfall (mm)** | 1630 | 1660 |
| **Mean Annual Runoff (mm)** | 838 | 1190 |
| **Mean Annual Runoff Coefficient** | 0.5 | 0.7 |
| **Mean Annual PET (mm)** | 1300 | 1300 |

**Table 1 Study catchment properties**

| | HYMOD | HBV |
|---|---|---|
| NSE [ ] | **0.77** | 0.75 |
| ***Peak flows (q > 5mm/d)*** | | |
| MAE [mm/d] | 3.11 | **2.85** |
| RMSE [mm/d] | **4.55** | 4.72 |
| ***Medium flows (1 mm/d <= q <= 5mm/d)*** | | |
| MAE [mm/d] | **0.66** | 0.80 |
| RMSE [mm/d] | **0.86** | 1.09 |
| ***Low flows (q < 1mm/d)*** | | |
| MAE [mm/d] | 0.35 | **0.20** |
| RMSE [mm/d] | 0.42 | **0.34** |

**Table 2 Model performance in pre-change conditions (1975 – 1979). Bold face numbers correspond to the model with superior performance for the particular metric.**

| | Sensitivity Index |
|---|---|
| *hl1* | 0.10 |
| *lp* | 0.12 |
| *Maxbas* | 0.14 |
| *fcap* | 0.18 |
| *K0* | 0.23 |
| *K2* | 0.23 |
| *K1* | 0.38 |
| *beta* | 0.41 |
| *perc* | 0.47 |

**Table 3 Variance Based Sensitivity Analysis Results for HBV parameters: first order sensitivity index representing the contribution of varying a single parameter to the variance of the model output. Lower values indicate lower sensitivity.**

666

| ***Parameters*** | | | | | | |
|---|---|---|---|---|---|---|
| | **Description** | **Units** | **Initial Sampling Distribution** | **Feasible Range** | $s^2$ | **Max allowable daily rate of change (LL)** |
| $\beta$ | Soil Moisture exponent | [ ] | *N*(2, 0.1) | 0 – 7 | 0.003 | $1.8 \times 10^{-3}$ |
| $fcap$ | Maximum soil moisture store depth | [mm] | *N*(467, 10) | 10 – 2000 | 0.003 | 0.4 |
| $hl1$ | Threshold for generation of near surface flow | [mm] | *N*(120, 10) | 0 – 400 | 0.003 | 0.1 |
| $K0$ | Near Surface Flow Routing Coefficient | [ ] | *N*(0.3, 0.005) | 0.0625 – 1 | 0.003 | $2 \times 10^{-4}$ |
| $K1$ | Interflow Routing Coefficient | [ ] | $N(0.09, 5 \times 10^{-4})$ | 0.02 – 0.1 | 0.003 | $9 \times 10^{-6}$ |
| $perc$ | Percolation rate | [mm/d] | $N(1.3, 10^{-4})$ | 0 – 3 | 0.003 | $10^{-3}$ |
| $K2$ | Baseflow Routing Coefficient | [ ] | $N(0.01, 10^{-6})$ | $5 \times 10^{-5}$ – 0.02 | 0.003 | $9 \times 10^{-6}$ |
| ***States*** | | | | | | |
| $sowat$ | Soil Moisture Store | [mm] | *N*(0,1) | (0, $fcap$) | | |
| $stw1$ | Shallow Layer Store | [mm] | *N*(0,1) | (0, ∞) | | |
| $stw2$ | Deep Layer Store | [mm] | *N*(0,0.1) | (0, ∞) | | |

667 **Table 4 Locally Linear EnKF inputs for the HBV model case**

668

669

| *Parameters* | | | | | | |
|---|---|---|---|---|---|---|
| | **Description** | **Units** | **Initial Sampling Distribution** | **Feasible Range** | $s^2$ | **Max allowable daily rate of change (LL)** |
| $b$ | Pareto-distributed soil storage shape parameter | [ ] | $N(0.37, 10^{-4})$ | 0 – 0.3 | 0.004 | $3x10^{-4}$ |
| $c_{max}$ | Maximum point soil storage depth | [mm] | $N(651, 10)$ | 300 – 1500 | 0.004 | 0.3 |
| $k_s$ | Surface Runoff Routing Coefficient | [ ] | $N(0.6, 5x10^{-4})$ | 0.55 – 0.99 | 0.018 | $3x10^{-4}$ |
| $k_b$ | Groundwater Routing Coefficient | [ ] | $N(0.04, 5x10^{-4})$ | 0.001 – 0.54 | 0.018 | $4x10^{-5}$ |
| $\alpha$ | Excess Runoff Splitting Parameter | [ ] | $N(0.47, 5x10^{-4})$ | 0.001 – 0.99 | 0.018 | $4x10^{-4}$ |
| *States* | | | | | | |
| $S$ | Soil Store | [mm] | $N(180, 0.1*180)$ | $(0, S_{max} = \frac{bc_{min} + c_{max}}{b+1})$ | | |
| $S_{q1,2,3}$ | Quick Flow Stores | [mm] | $N(0,1)$ | $(0, \infty)$ | | |
| $S_s$ | Slow Flow Store | [mm] | $N(0,1)$ | $(0, \infty)$ | | |

**Table 5 Locally Linear EnKF inputs for the HYMOD model case**

# Figures

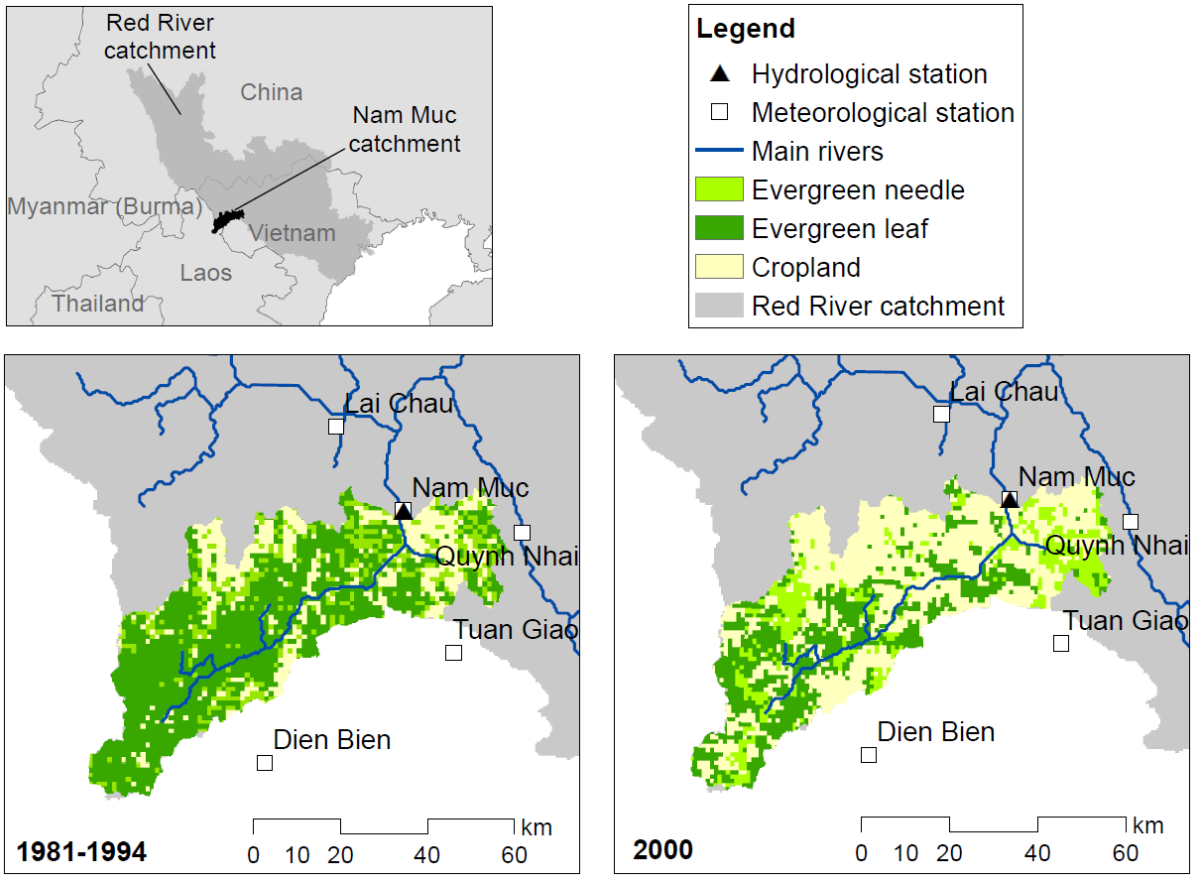



**Figure 1 Study Catchment showing gauges and changes in land cover over time.**

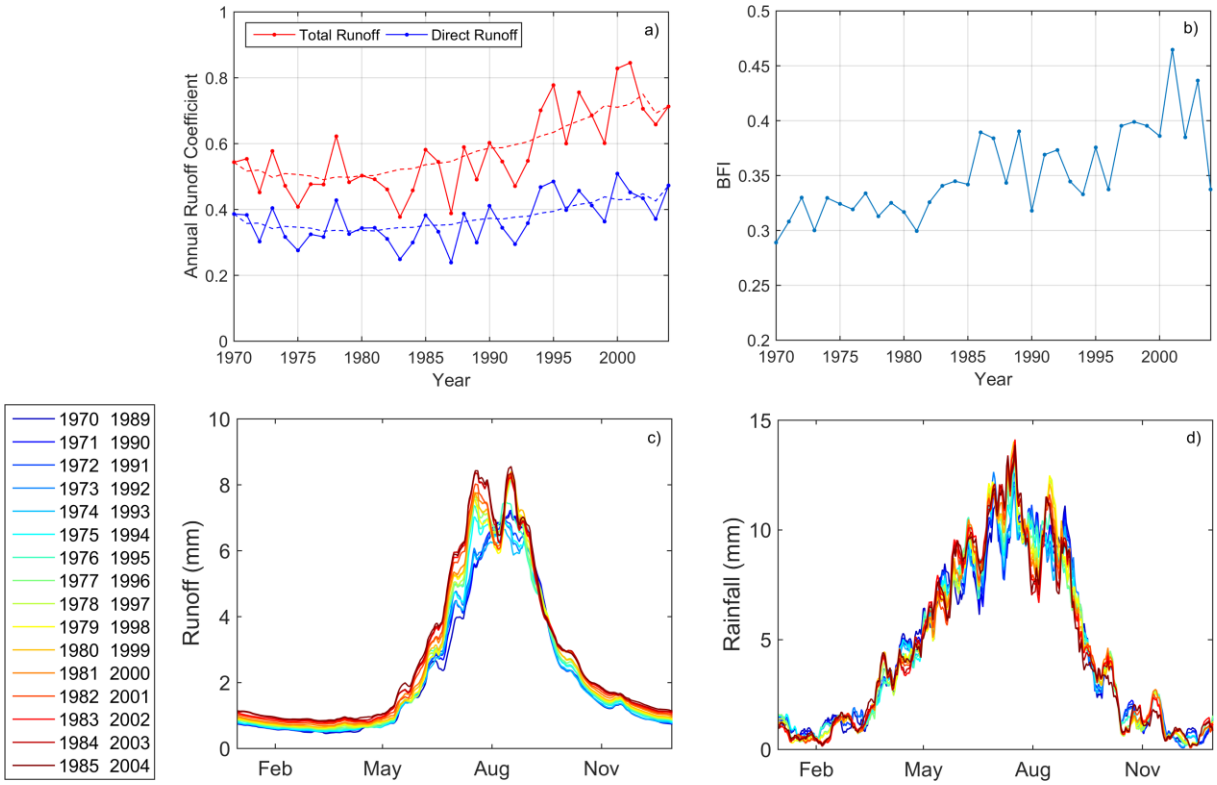



**Figure 2 Impact of land use change on observed streamflow: a) Annual Runoff Coefficient, b) Annual Baseflow Index (BFI), c) Moving Average Shifting Horizon (MASH) results for total observed runoff, d) MASH for observed rainfall.**

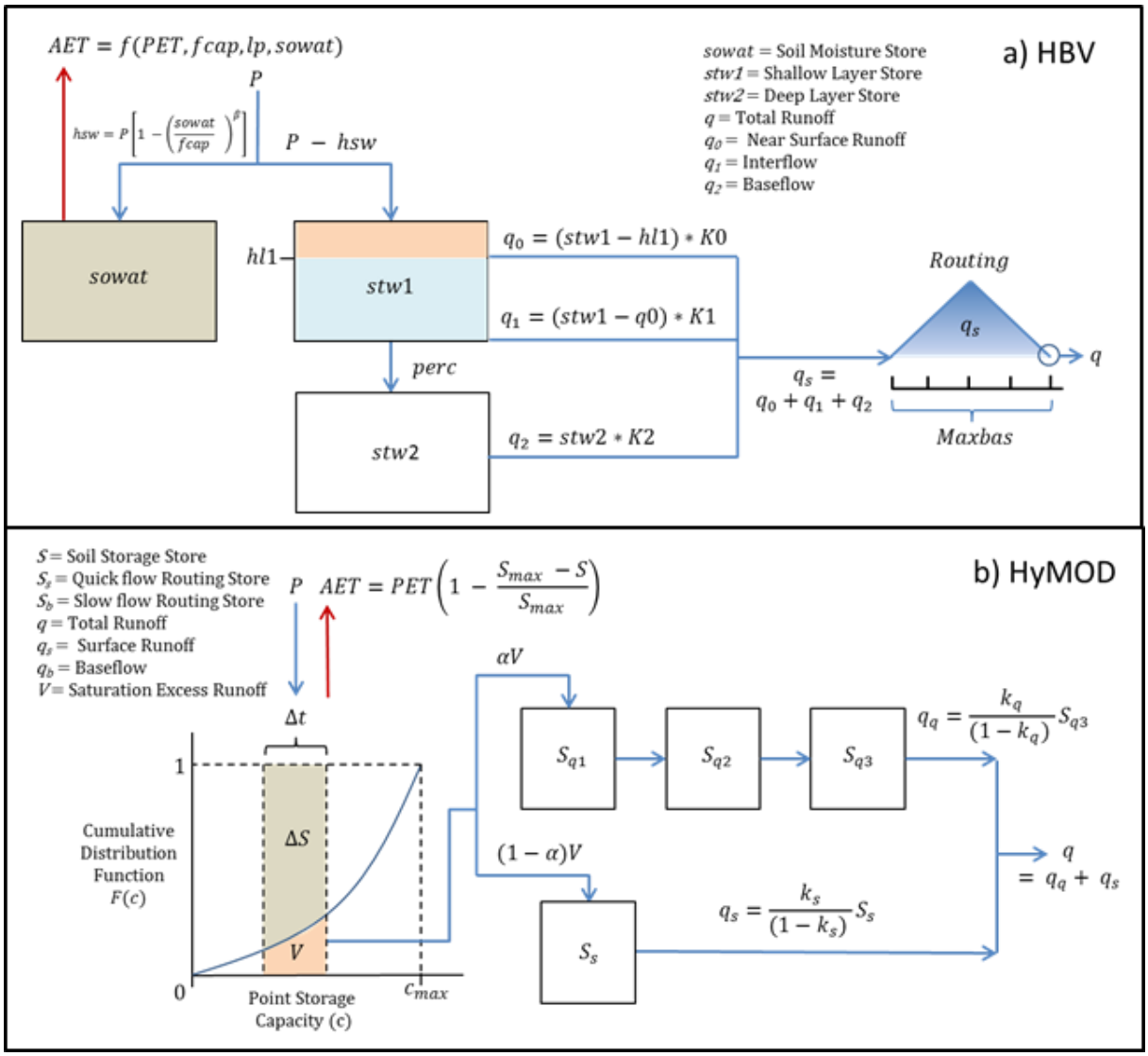



**Figure 3 Schematic of the models used in this study: a) HBV and b) HyMOD**

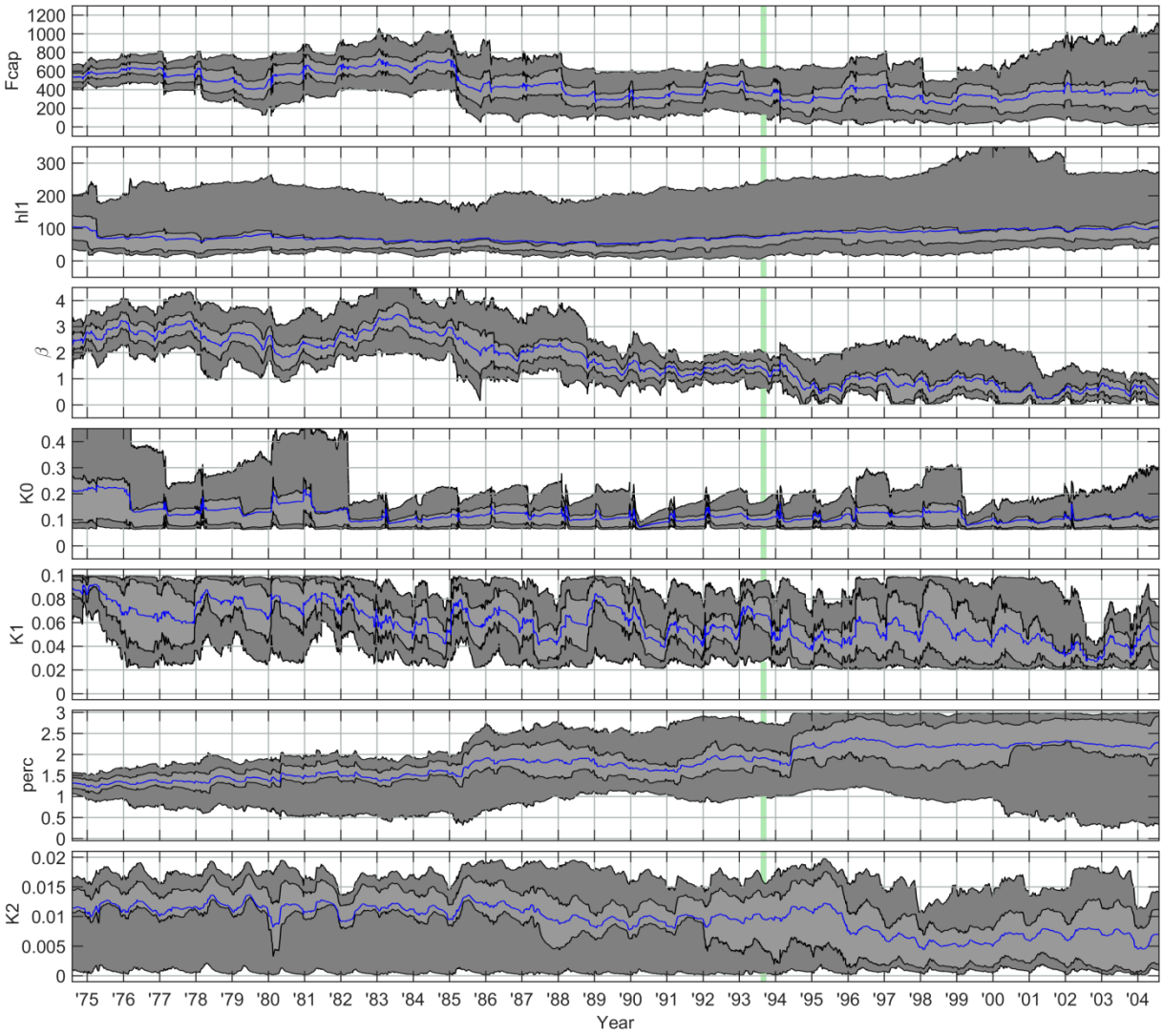



**Figure 4 Parameter Trajectories using the HBV model. The dark grey shaded areas indicate the middle 90% of the ensemble, bounded by the 5th and 95th percentiles. The light grey shaded areas indicate the middle 50% of the ensemble, bounded by the 25th and 75th percentiles. The ensemble mean is indicated by the blue line. The vertical green panel indicates the assumed time period of rapid deforestation.**

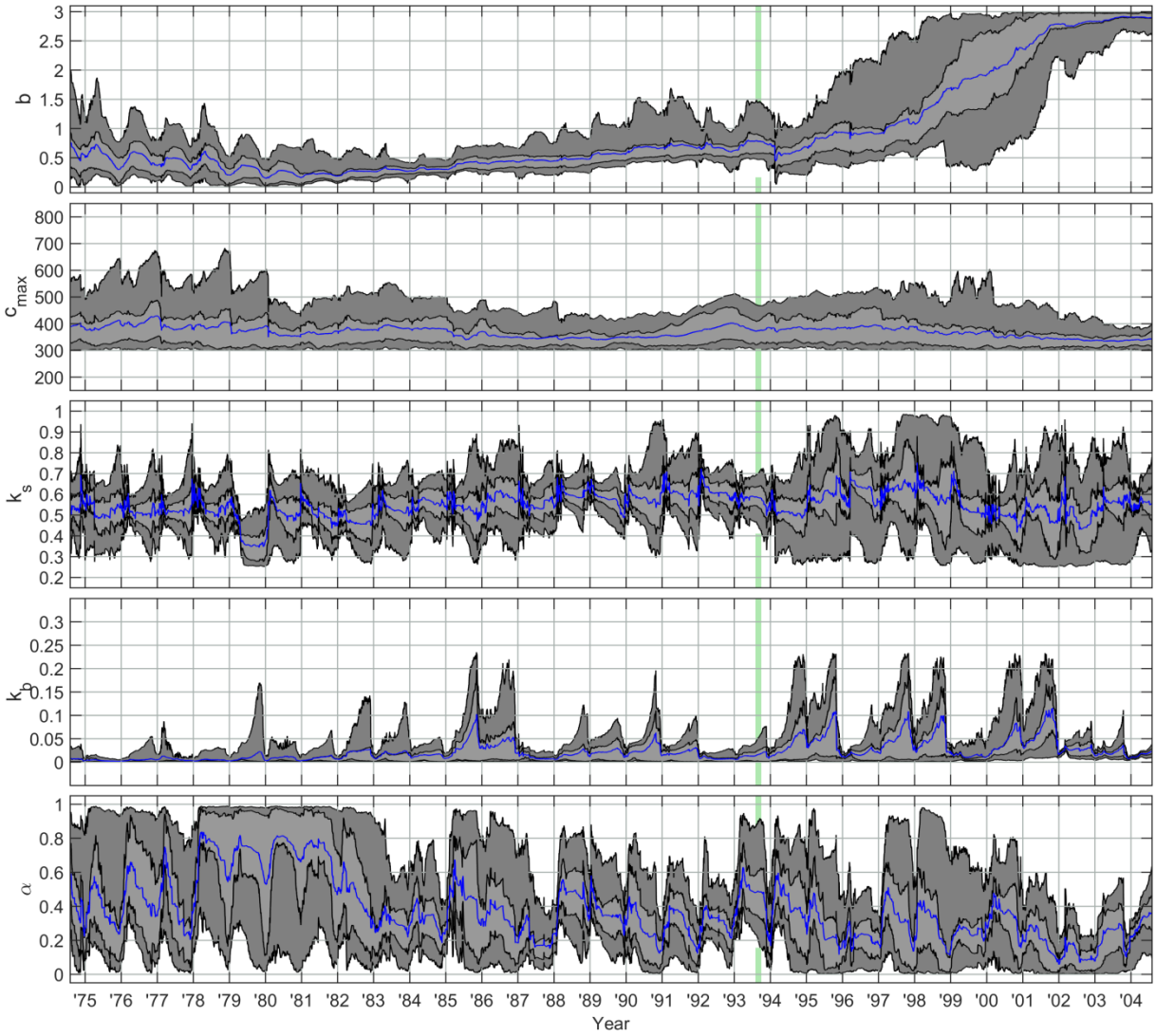
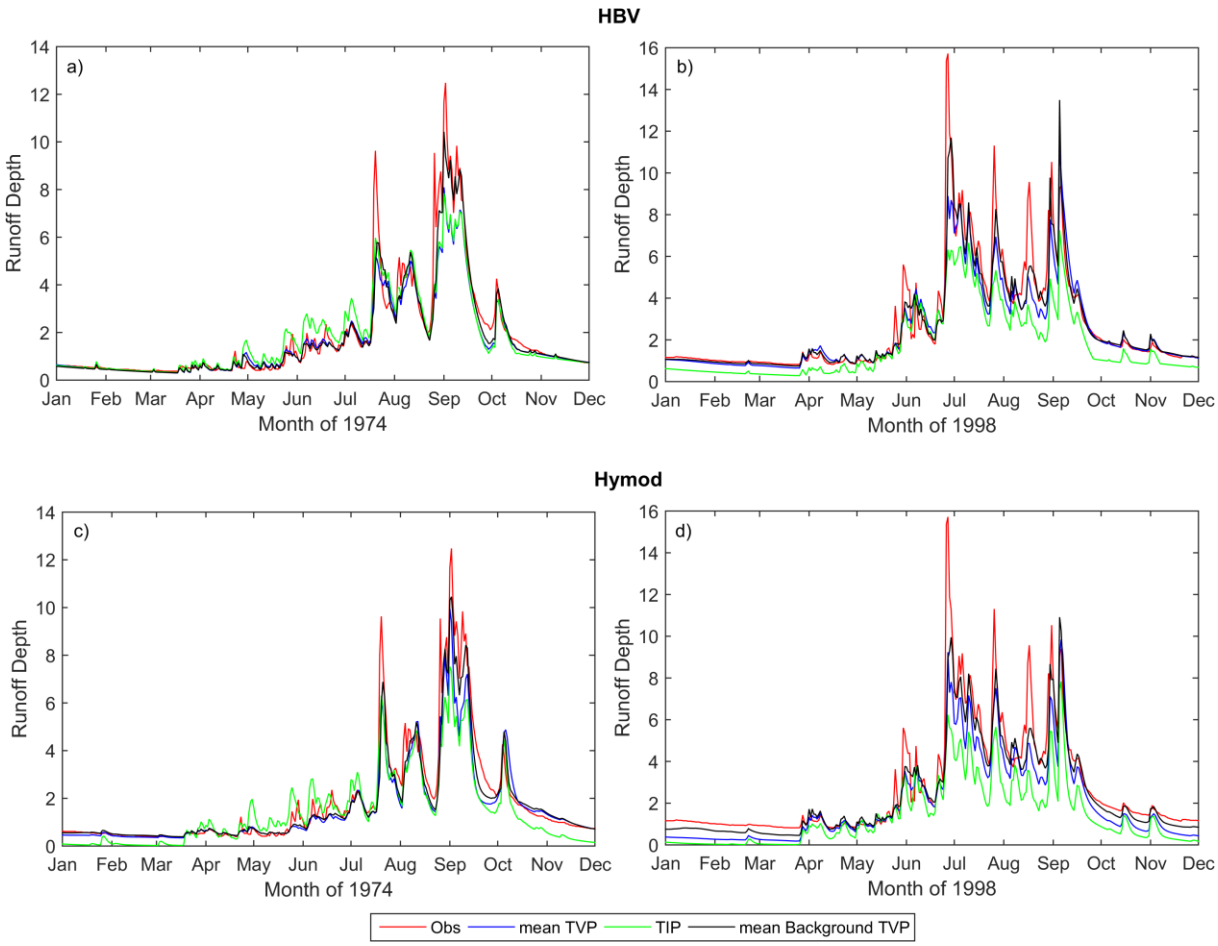



**Figure 5 Parameter Trajectories using the HyMOD model. The dark grey shaded areas indicate the middle 90% of the ensemble, bounded by the 5th and 95th percentiles. The light grey shaded areas indicate the middle 50% of the ensemble, bounded by the 25th and 75th percentiles. The ensemble mean is indicated by the blue line. The vertical green panel indicates the assumed time period of rapid deforestation.**




736

737 **Figure 6 Representative Hydrographs of background streamflow from the LL Dual EnKF (black line),**
738 **Time varying parameter model with no state updating (blue line), time invariant parameter model**
739 **with no DA (green line) and observed streamflow (red line). Results for HBV are shown in the top**
740 **row and HyMOD in the bottom row. A pre-change year (1974) is shown on the left and a post**
741 **change year (1998) on the right.**

742
743

744

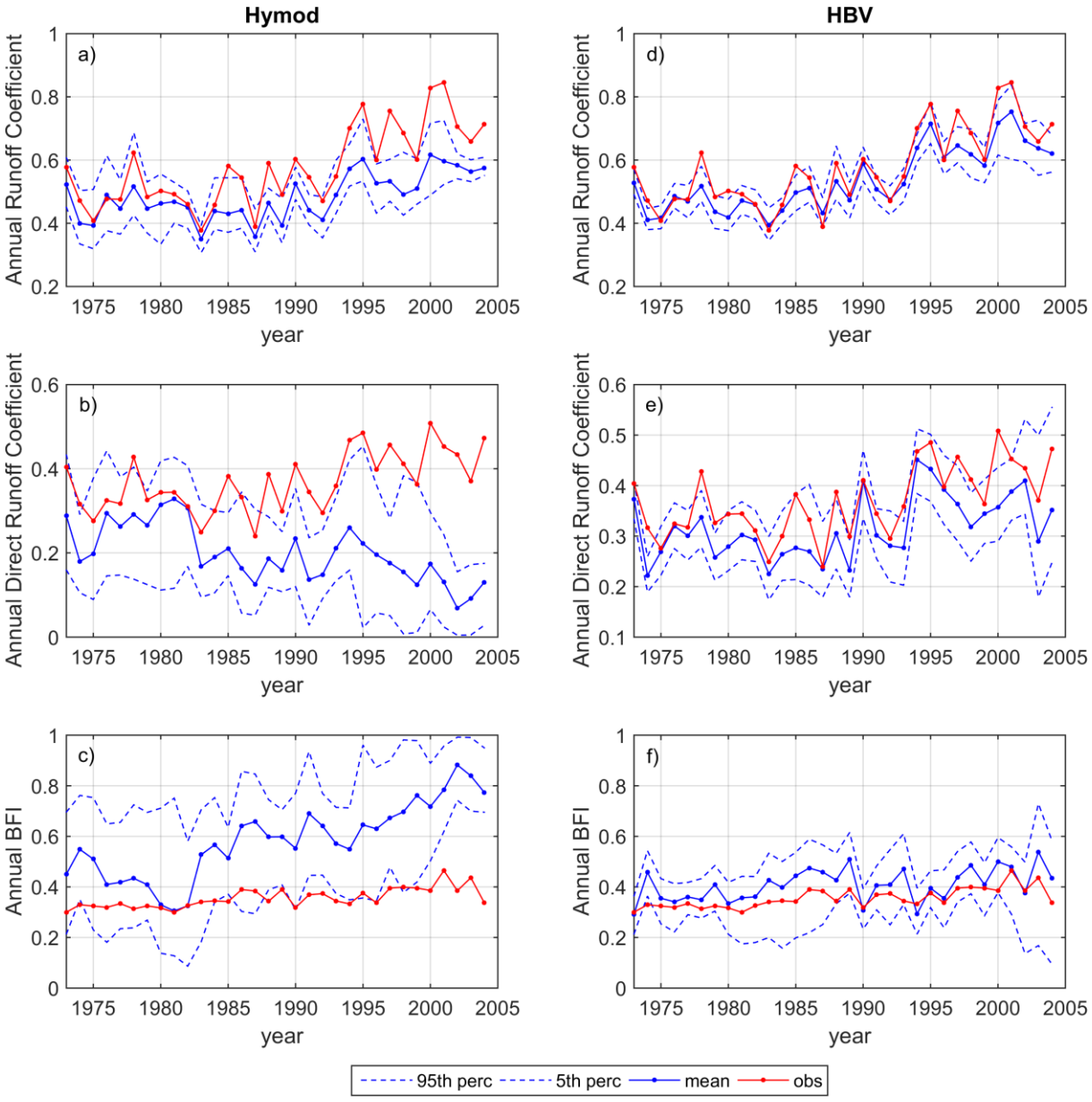



**Figure 7 Influence of time varying parameters on model output (i.e. without state updating) summarized in terms of the Annual Runoff Coefficient (top row), Annual Direct Runoff Coefficient (second row) and Annual Baseflow Index (BFI) (third row). Results for HyMOD are shown in the first column, HBV are shown in the second column.**

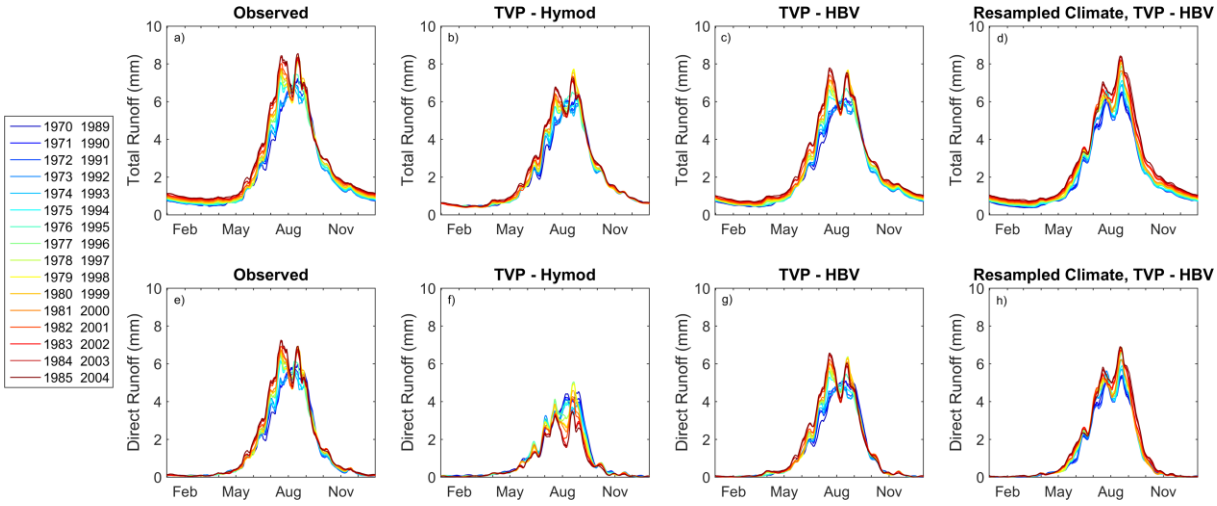



**Figure 8 Moving Average Shifting Horizon (MASH) results for observed streamflow (first column), simulated streamflow from time varying parameter model (without state DA) for HYMOD (2nd column), HBV (third column), resampled climate HBV (fourth column). These are split into total runoff (first row) and direct runoff or surface runoff (2nd row).**