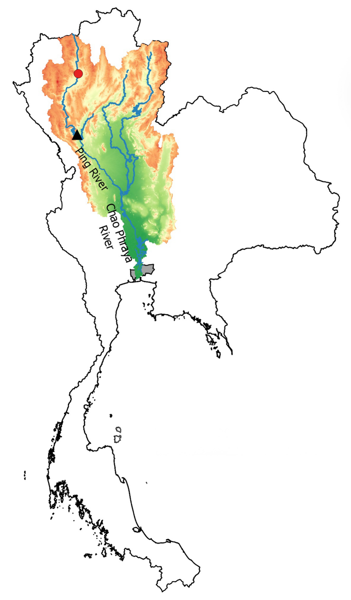

A linear dynamical systems algorithm for streamflow reconstruction reveals history of regime shifts in northern Thailand

Hung T.T. Nguyen and Stefano Galelli

Pillar of Engineering Systems and Design, Singapore University of Technology and Design

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

European Geosciences Union General Assembly 2018

Paper No. EGU2018-459

Key points

- The linear dynamic model has higher accuracy than conventional linear regression.
- The state variable reveals regime-like behaviour in the catchment history.
- The model can generate stochastic replicates of both streamflow and catchment state.

1. Introduction

Streamflow reconstruction is the study of catchment in the distant past, using statistical techniques to reconstruct streamflow from climate proxies (e.g., tree-rings). Since its inception in the 1970s, streamflow reconstruction has brought fought insights that were unattainable with instrumental records, such as better understanding of extreme events and long term streamflow variability.

Most reconstruction studies use principal component linear regression, which establishes an empirical relationship between climate proxies u and streamflow y via the regression equation

$$y_t = \alpha + \beta u_t + \varepsilon_t \qquad (1)$$

where α, β are regression parameters and $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. Equation (1) neglects catchment dynamics and its effects on streamflow generation; therefore, it may not fully capture important phenomena such as long-range dependence, complex flood generation mechanisms, or clustering of extreme events.

An alternative is to use water-balance-based methods. This approach, however, requires extensive hydrologic data that may be unavailable in developing countries.

Research Questions

- How do we account for catchment dynamics without requiring more data?
- Will a dynamic streamflow reconstruction be more accurate?
- What insights can we gain with a dynamic model?

We propose a linear dynamical systems approach to answer these questions.

2. Case Study

We reconstruct 406 years of streamflow (1600–2005) for station P1 on the Ping River, Thailand. The Ping is a main tributary of the Chao Phraya Basin, home to 22 million people, including 8 million in Bangkok. The Ping River supplies water to the Bhumibol Reservoir, the largest reservoir in the basin, with an active capacity of 9.6 billion m³.

Our paleoclimate proxy is the Monsoon Asia Drought Atlas (MADA) (Cook *et al*, 2010), a gridded time series of the Palmer's Drought Severity Index (PDSI) reconstructed over Asia from a rich tree-ring network. A similar gridded PDSI dataset was shown by Ho et al (2016) to be a reliable paleoclimate proxy. Figure 1 shows the MADA grid points used for reconstruction and its correlation with streamflow at P1.

We benchmark our reconstruction with a typical backward stepwise principal component linear regression similar to Woodhouse *et al* (2006).

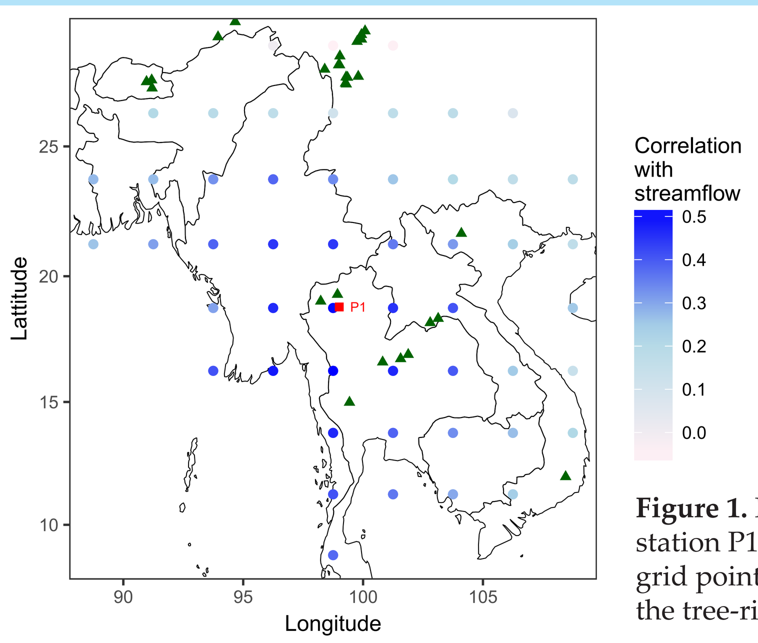

Figure 1. Map showing the study region with station P1 (red triangle) and nearby MADA grid points (coloured circles), together with the tree-ring sites used in the MADA.

3. Methodology

3.1. Linear dynamical systems

We model the catchment as a linear dynamical system (LDS) governed by equations (2) and (3):

$$x_{t+1} = Ax_t + Bu_t + w_t \qquad (2)$$
$$y_t = Cx_t + Du_t + v_t \qquad (3)$$
$$w_t \sim \mathcal{N}(0, Q)$$
$$v_t \sim \mathcal{N}(0, R)$$

Here, y and u are streamflow and climate proxy, as before. We introduce the hidden system state x, which indicates the catchment's flow regime (i.e., whether it is wet or dry). Observe that linear regression is a subclass of the LDS model.

The system parameters $\theta = (A, B, C, D, Q, R)$ and state are learned with the Expectation-Maximization algorithm (Figure 2). The E-step fixes the system parameters and estimates the hidden states; the M-step fixes the state and estimates the parameters. EM iterates between E- and M-steps until convergence (Cheng and Sabes, 2006).

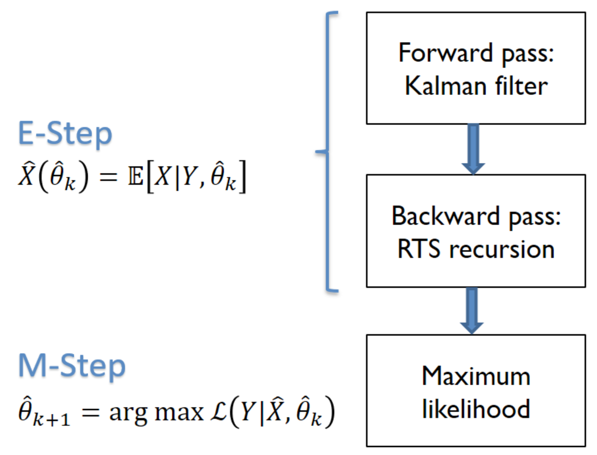

Figure 2. The EM algorithm for linear dynamical systems

3.2. Simultaneous learning-reconstruction

Typically, a paleoreconstruction problem is solved in two phases: learning and reconstruction. Learning involves building a regression model for the instrumental period. Reconstruction involves feeding the paleo period's input into the regression model to obtain the paleo period's streamflow (Figure 3a). This approach does not work with our LDS model. Because of the system dynamics, one must propagate the system state from past to present, which may cause a mismatch of system state where the two periods intersect (Figure 3b). To overcome this problem, we eliminate the paleo-instrumental delineation and perform learning-reconstruction simultaneously (Figure 3c). We achieve this by modifying the EM algorithm (see details in Nguyen and Galelli, 2018).

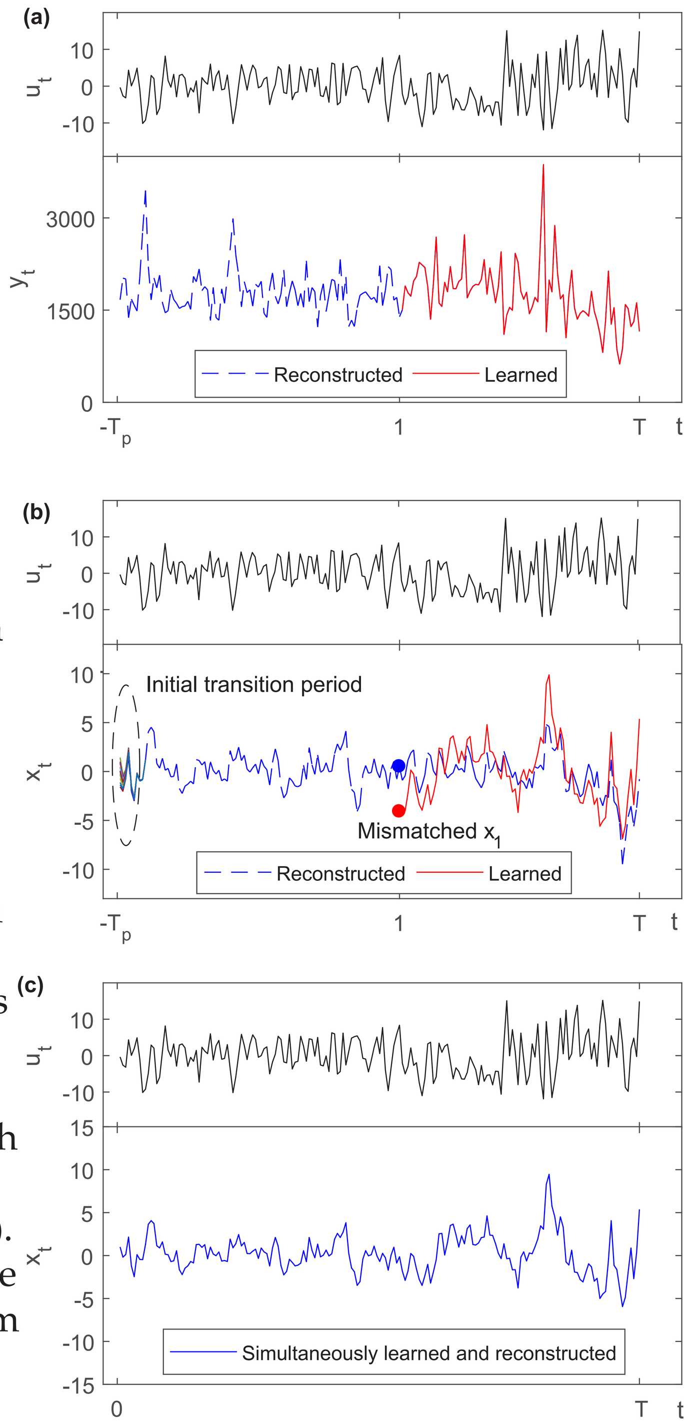

Figure 3. Motivation for simultaneous learning-reconstruction

4. Results and Discussion

4.1. Model performance

LDS performed much better than the benchmark (Figure 4). Linear regression tended to overestimate streamflow when the catchment was dry and underestimate it when the catchment was wet while the LDS model matched observations better. This shows that information about the catchment state is beneficial.

The cross-validated performance scores of LDS are 45–497% better than linear regression (Figure 5), e.g., R^2 increased from 0.54 to 0.82 and coefficient of efficiency from 0.12 to 0.74.

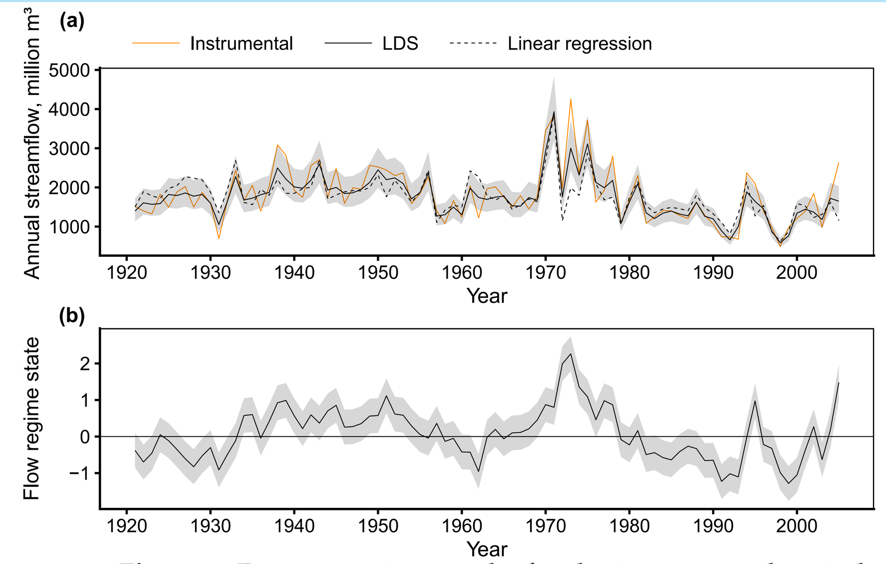

Figure 4. Reconstruction results for the instrumental period

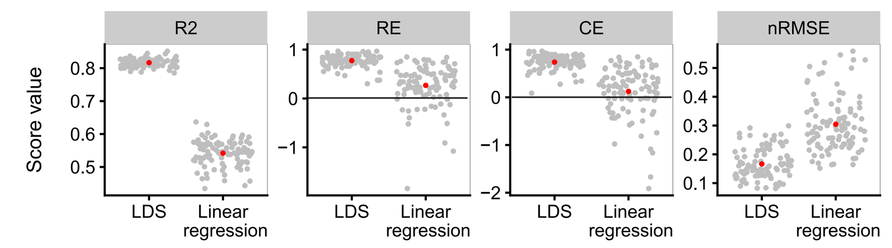

Figure 5. Model performance in leave-10%-out cross validation

4.2. Full reconstruction

LDS reveals a drier history of the Ping River than does linear regression (Figure 6a). The reconstructed flow regime shows different patterns of regime shift over time (Figure 6b). Notably, the last century contains both the wettest period (including the wettest year) and the driest year. Our reconstruction is in agreement with the MADA with regards to the Asian megadroughts, but also complements it with information on local droughts (Cook *et al*, 2010).

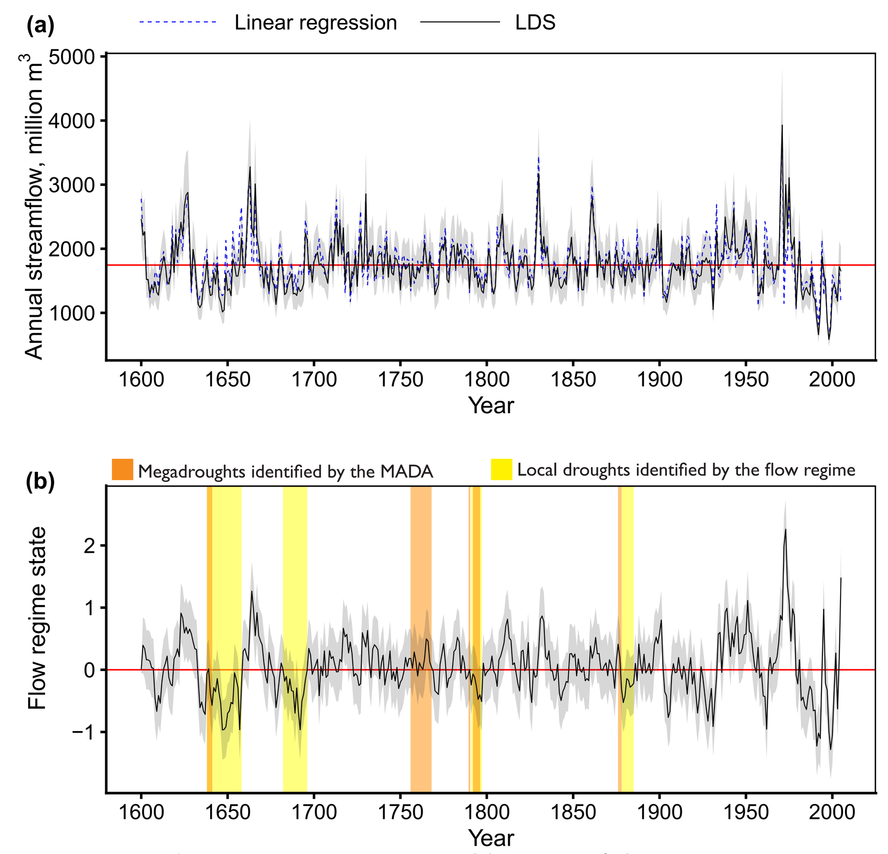

Figure 6. A reconstructed history of the Ping River

4.3. Stochastic streamflow

Stochastic streamflow replicates are generated by sampling w_t, v_t and ε_t with historical climate input.

Figure 7 shows that LDS is a better stochastic streamflow generator. In linear regression, climate input only accounts for 54% of streamflow variability; hence, the noise process generates large and unrealistic deviations from the mean. In LDS, input and catchment state account for 82% streamflow variation; thus, the LDS replicates resembles the reconstruction more closely.

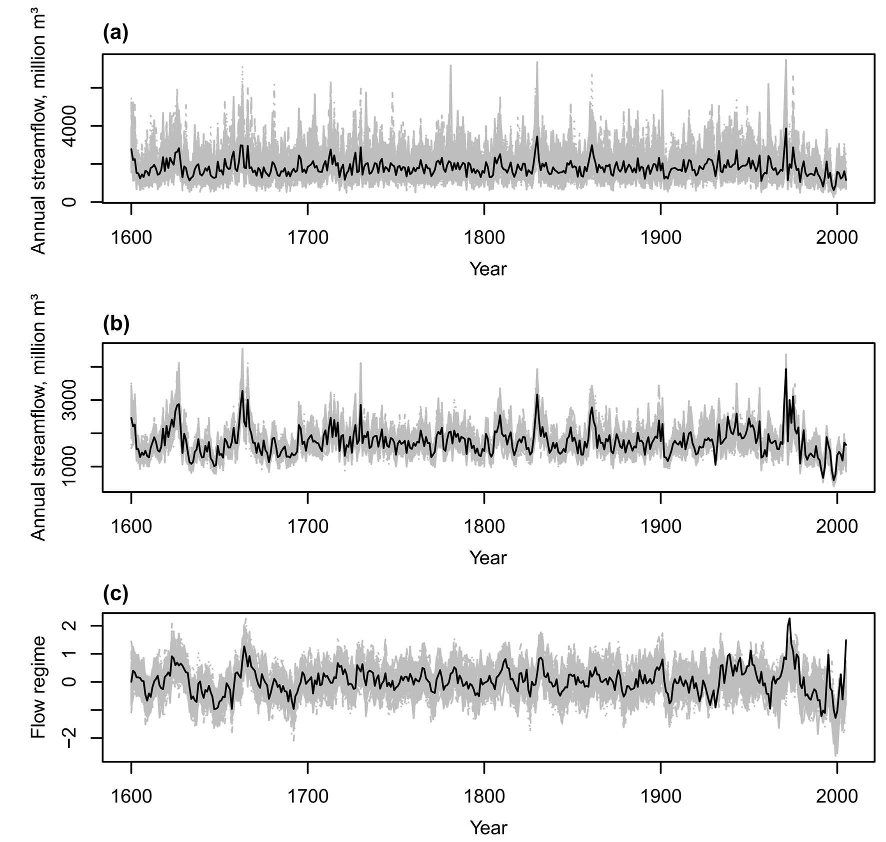

Figure 7. Stochastic streamflow replicates

5. Conclusions

We contribute a technique of applying the linear dynamical systems model and EM learning to streamflow reconstruction. Results reveal a history of regime shifts in the catchment with decadal to bi-decadal droughts and pluvials in the paleo period, while highlighting that the instrumental period contains the wettest period and the driest year.

The model scores are notably higher than the conventional linear regression (45–497% improvement), suggesting that it is important to account for catchment dynamics. The model's success is attributed to its two key advantages: it estimates the trajectory of the catchment state during the paleo and instrumental periods, and it accounts for the effect of both catchment state and climate proxies on the streamflow generation process.

The LDS model also has several desirable features: (i) the reconstructed trajectory of the state variable provides more insights about the catchment's history than the reconstructed streamflow alone, (ii) the learning algorithm is computationally efficient, and (iii) the model can be readily used as a stochastic streamflow generator.

The LDS model can replace linear regression in future streamflow reconstruction studies. Most importantly, the model's regime state, not available in conventional methods, may add value to downstream water resources management. Through the findings in this work, not only has the values of streamflow reconstruction been strengthened, but its potential applications have also been widened.

Acknowledgements

We thank Joost Buurman for his initial suggestions on the Ping River case study; Robert Wasson for his insightful suggestions and for providing additional information and data on paleo floods; Brendan Buckley and Edward Cook for their valuable insights on the Monsoon Asia Drought Atlas (MADA). Hung Nguyen is supported by the SUTD President's Graduate Fellowship.

References

Nguyen, H. T. T., & Galelli, S. (2018). A Linear Dynamical Systems Approach to Streamflow Reconstruction Reveals History of Regime Shifts in Northern Thailand. *Water Resources Research*, 1–21.

Cook, E. R., Anchukaitis, K. J., Buckley, B. M., D'Arrigo, R. D., Jacoby, G. C., & Wright, W. E. (2010). Asian Monsoon Failure and Megadrought During the Last Millennium. *Science, 328*(5977), 486–489.

Cheng, S., & Sabes, P. N. (2006). Modeling Sensorimotor Learning with Linear Dynamical Systems. *Neural Computation, 18*(4), 760–793.

Ho, M., Lall, U., & Cook, E. R. (2016). Can a paleodrought record be used to reconstruct streamflow?: A case study for the Missouri River Basin. *Water Resources Research, 52*(7), 5195–5212.

Woodhouse, C. A., Gray, S. T., & Meko, D. M. (2006). Updated streamflow reconstructions for the Upper Colorado River Basin. *Water Resources Research, 42*(5), 1–16.

Resilient Water Systems Group
Stefano Galelli: stefano_galelli@sutd.edu.sg
people.sutd.edu.sg/~stefano_galelli
Hung Nguyen: tanthaihung_nguyen@mymail.sutd.edu.sg
peole.sutd.edu.sg/~ntthung