


Article

Associating climatic trends with stochastic modelling of flow sequences

Sandhya Patidar ^{1,*} , Eleanor Tanner ², Bankaru-Swamy Soundharajan ³ and Bhaskar SenGupta ⁴

¹ Institute for Infrastructure and Environment, School of Energy, Geoscience, Infrastructure and Society, Heriot-Watt University, Edinburgh, UK; s.patidar@hw.ac.uk

² Computing Science and Mathematics, University of Stirling, UK; eleanor.tanner@stir.ac.uk

³ Department of Civil Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore 641112, India; b_soundharajan@cb.amrita.edu

⁴ Institute for Infrastructure and Environment, School of Energy, Geoscience, Infrastructure and Society, Heriot-Watt University, Edinburgh, UK; b.sengupta@hw.ac.uk

* Correspondence: s.patidar@hw.ac.uk; Tel.: +44 (0) 131 451 4456

Abstract: Water is essential to all life-forms including various ecological, geological, hydrological, and climatic processes/activities. With changing climate, associated El Niño/Southern Oscillation (ENSO) events appear to stimulate highly uncertain patterns of precipitation (P) and evapotranspiration (EV) processes across the globe. Changes in P and EV patterns are highly sensitive to temperature (T) variation and thus also affecting natural streamflow processes. This paper presents a novel suite of stochastic modelling approaches for associating streamflow sequences with climatic trends. The present work is built upon a stochastic modelling framework (HMM_GP) that integrates a Hidden Markov Model (HMM) with a Generalised Pareto (GP) distribution for simulating synthetic flow sequences. The GP distribution within HMM_GP model is aimed to improve the model's efficiency in effectively simulating extreme events. This paper further investigated the potentials of Generalised Extreme Value Distribution (GEV) coupled with an HMM model within a regression-based scheme for associating impacts of precipitation and evapotranspiration processes on streamflow. The statistical characteristic of the pioneering modelling schematic has been thoroughly assessed for their suitability to generate/predict synthetic river flows sequences for a set of future climatic projections, specifically during the ENSO events. The new modelling schematic can be adapted for a range of applications in the area of hydrology, agriculture and climate change.

Keywords: Stochastic modelling; Climate change; Streamflow; El Niño/Southern Oscillation (ENSO), Extreme events modelling

Nomenclature

EV Evapotranspiration

P Precipitation

T Temperature

ANN Artificial Neural Networks

BBMB Bhakra Beas Management Board

DBN Deep Belief Network

ENSO El Niño/Southern Oscillation

GEV Generalised Extreme Value

GP Generalised Pareto

HMM Hidden Markov Model

IMD Indian Meteorological Department



Citation: Patidar, S.; Tanner, E.; Soundharajan, B.; SenGupta B. Title. *Preprints* **2021**, *1*, 0. <https://doi.org/>

Received:

Accepted:

Published:

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Q	inflow sequences
R	Random
S	Seasonal
SVR	Support vector regression
Tr	Trend
USGS	U.S. Geological Survey
WANN	Wavelet-based Artificial Neural Network

1. Introduction

Rivers are socio-economically valuable assets, providing a range of hydro-ecological services, such as, agriculture, fishing, recreational space, landscape scenery, a healthy environment for society, habitats for a range of marine and terrestrial species. Rivers are considered as a major contributor to the food-water nexus. According to U.S. Geological Survey (USGS), the term streamflow is used to refer the amount of water flowing in a river [1]. Thus, streamflow can be considered as a complex manifestation of interacting and overlapping spatio-temporally distributed physical, environmental, and hydro-morphological processes, such as climate change, sediment transport, etc. [2,3]. Recently, a large volume of research has been conducted to understand the impact of anthropogenic climate change on streamflow and the associated hydro-ecological services. Most of these studies concluded that streamflow is highly sensitive to projected future climate change and extreme events [4–6]. Further, as scientific evidence are growing, it is widely accepted that frequency, intensity, severity, duration, temporal ranges, and spatial extent of extreme events (such as flooding, droughts, heatwaves) will also change considerably in future climate change [7,8].

In this context, the interactions occurring between the tropical Pacific Ocean and Earth's atmospheric system that initiates El Niño/Southern Oscillation (ENSO) events [9–11] are of particular interest. These events last over several months, occur irregularly every few years and impact global climatic patterns [12,13]. There is growing evidence that associates ENSO events with more than average rainfall in South America (specifically in Peru, Ecuador, and Argentina) and drought in South Asia [14,15]. Agriculture is one of the key sectors that could be severely impacted by ENSO events affecting lives in the poorest and most vulnerable communities, specifically in the agricultural dependent economies in south Asian and South American countries. The social and economic impact of ENSO events is expected to be further exacerbated by projected climate change across the globe. Nations worldwide require their government, regulators and statutory consultees to develop, maintain, apply and monitor a national strategy to coordinate actions for managing an efficient Early Warning Early Action System (a new initiative proposed by the Food and Agriculture Organisation of the United Nations) to strengthen the coping capacities of at-risk populations [16].

As a step towards investigating the impacts of ENSO type events on the availability of water resources, land use change and agricultural production we wish to model a suite of synthetic river flows coupled to climatic conditions that can be used as input to a range of conventional software used in hydrological and agricultural applications. For example, to conduct a stochastic assessment of streamflow scenarios within crop simulation tools to assess the probability of crop failure in the face of extreme weather conditions (eg. drought). We build on previous work done by authors (reported elsewhere [17]) on stochastic modelling of synthetic Q . In a novel approach, presented in this paper, we link statistical modelling of Q time series with coincident P and EV data, to provide a direct connection between underlying changes in climatic patterns with water resources available to agriculture and reservoir beneficiaries.

This paper is mainly focused on developing an efficient statistical model that can

predict the Q sequences during ENSO events (which are expected to demonstrate unusual trends similar to extreme event forms) using the changes in the climatic trends. We will focus on the Beas river basin in northern India to provide empirical data for calibrating our statistical models and to understand the implications of changes in the summer monsoon rainfall on the varying inflow of the river Beas into the Pong Dam, a phenomenon that could be intensified during an ENSO event. As a proof of concept, we investigate if we can successfully predict future river flow time series based on trends in climatic conditions and historical flow data during these periods.

2. Background

This section is intended to provide a theoretical background for contextualising proposed research. A literature review covering a brief overview of the impact of climate change on ENSO events along with key statistical/computational modelling approaches developed recently and their application in prediction of streamflow sequences under climatic influences, specifically in the context of extreme events is discussed.

2.1. Impact of climate change on ENSO events

Climate change is a complex process and, as such, climatic projections are associated with a large amount of unquantified uncertainties. Although, ENSO events are widely investigated for influencing extreme weather events such as flooding, drought and tropical cyclone [18–20], there is a limited amount of evidence for directly attributing impacts of climate change on intensifying frequency of extreme ENSO events [21,22]. This is mainly because there are no set rules for classifying ENSO events from moderate to strong ENSO events [21]. Some of the related studies investigated influence of key hydro-climatological variables such as temperature (T), P , and EV on these extreme climatic events and associated uncertainty thoroughly [23]. Also, effects of ENSO on these key hydro-climatic variables (T, P, EV) has been widely investigated [24]. Some widely adapted work that facilitated schematics of summary global maps of regional effects of ENSO event on i) T can be referred elsewhere [25]; and for P can be referred elsewhere [26–28]. A considerable amount of work is done for predicting ENSO events that include both statistical and dynamic models and can be referred elsewhere [29–32].

Within the focus of the paper, effects of ENSO simultaneously on T , P and EV during Monsoon season over the last century for 146 districts across North India are thoroughly investigated in [33]. Results suggest that El Niño years have significant influence on hydro-climatic variables in comparison to La Niño or neutral years and most of the district experienced a significant decreasing trend in T , P and EV variable across the century. When considering adaptation and planning issues related to water resources it is inevitably important to incorporate and quantify these effects and associated uncertainties appropriately. This is essential to ensure a robust and reliable interpretation of model outcomes for optimising confidence in decision-making with future projections.

2.2. Data-driven approaches for predicting streamflow under climatic influence (extreme events)

With growing evidence confirming the phenomenon of global warming, a large amount of research is recently conducted to examine potentials of data-driven modelling approaches in predicting influences of climate change on streamflow. Mathematically, novel nonlinear dynamical system and chaos-based approaches involving phase-space analysis of streamflow for characterization and prediction of runoff dynamics have been explored in several studies [34–38]. Some interesting statistical approaches such as Markov switching time series models [39] and Bayesian approach for neural networks [40] has also been explored for runoff modelling. Applications of stochastic modelling approaches for exploring impacts of climate change on runoff modelling has been demonstrated across a few case-studies [41]. The suitability of computational approaches, such as Artificial Neural Networks (ANN) in rainfall-runoff modelling has been intensively studied in

recent decades [42]. An intensive examination of some of the widely applied Machine learning techniques including wavelet-based artificial neural network (WANN), support vector regression (SVR) and deep belief network (DBN) for multi-step ahead streamflow forecasting has been presented in [43]. In most of the investigation, ANN-based approaches appear to underperform in the estimation of extreme events.

A limited amount of work is done that explore impacts of extreme events either using a block maximum (estimated over a specific length of the period) or threshold-based approaches [44,45]. A model involving block maxima usually utilises the application of GEV distribution whereas those based on threshold values utilises GP distribution [45]. Some of the representative studies are briefly discussed here. A conditional density model, demonstrating the application of extreme value theory for facilitating a parametric modelling base for estimating upper tail of river runoff distribution while incorporating a non-parametric central distribution has been explored in [46]. A non-stationary generalised additive models-based approach for modelling sample extremes has been demonstrated for the estimation of extreme winter temperatures [47]. Potentials for a new criterion based on peak/low flow regime for selection of ANN model has been investigated for improving the forecasting of extreme hydrological events [48].

While most of these studies focused on varying average conditions of climatic variables [49], a very limited number of studies focused on the implications of the extreme events [50,51] that could lead to the most exacerbated impacts on the multifaceted hydro-ecological services associated with river systems [52,53]. The paper aims to bridge some of these gaps by facilitating a data driven modelling framework, which is specifically designed to effectively capture impact of extreme events through the integration of suitable statistical approaches and linking key hydro-climatic variables for simulating multiple realistic alternatives of streamflow sequences. These multiple streamflow sequences are intensively examined for capturing key statistical characteristics and dynamics of original series and thus represents a realistically plausible scenario that can be input into hydrological applications for facilitating a thorough uncertainty analysis of climate impacts and associated uncertainty.

3. Case study and data organisation

3.1. Study Area

Our focus area is the Beas river basin in Himachal Pradesh state, India, where the Beas rises in the western Himalayas eventually flowing westerly into the Pong reservoir, dammed at its western end by the Pong Dam [54–56]. The Pong reservoir stretches to a surface area of 260 km² with a catchment of 12,561 km² [55], managed by the Bhakra Beas Management Board (BBMB) which regulates discharge from the Pong Dam for generating hydroelectric power and providing irrigation to 1.6 Mha of land. Monsoon rainfall between July and September is a major source of water inflow into the reservoir, apart from snow and glacier melt. The historic mean annual runoff (MAR) at dam site is 8485 Mm³ and coefficient of variation of annual runoff is 0.225. Significant higher flows occur during the monsoon season compared to other time periods [46].

3.2. River runoff, precipitation and evapotranspiration data

Historic daily runoff data (Q) measured at the Pong Dam (geographical coordinates: 76° 05E and 32° 01N), was made available by the BBMB for the years 1998 to 2010 (inclusive). Additionally, for the same period daily P and EV (Penman-Monteith method) for the Beas river basin up to the Pong Dam has been gathered from the Indian Meteorological Department (IMD) and gridded TRMM (TRMM 3B42 V7) daily rainfall data [46]. The spatial resolution of TRMM data is $0.25^\circ \times 0.25^\circ$, covering the latitudinal band of 50° N-S. Potential EV were estimated using the Penman-Monteith (P-M) formulation forced with meteorological variables from the NCEP Climate Forecast System Reanalysis (CFSR) data from January 1999 to December 2008. Based on the elevation, Beas river basin has been divided in to three sub-basins, namely upper (5720km²), middle (3440km²) and lower

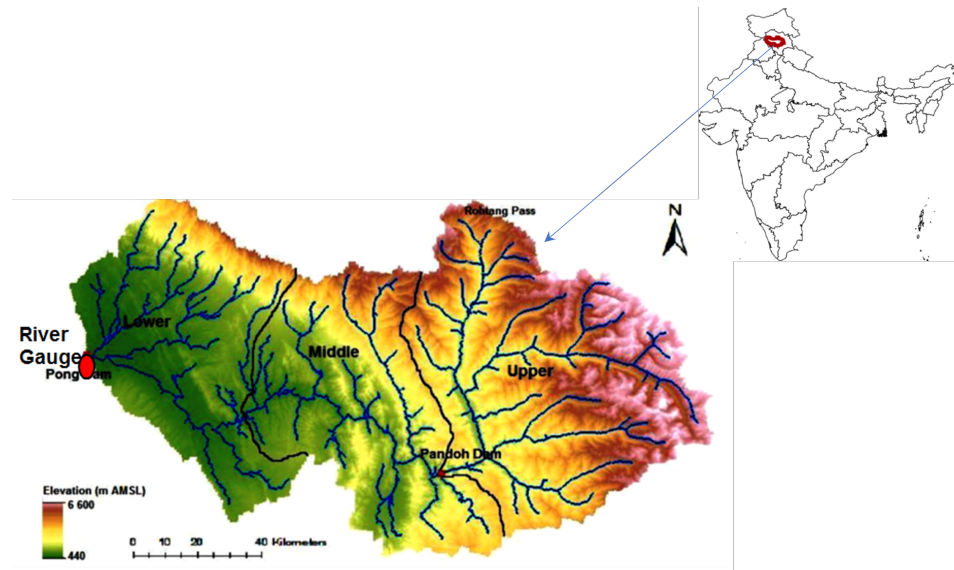


Figure 1. Map of Beas river basin

(3350km²) as shown in Figure 1, and the T , EV and P data are provided in Table 1. Effect of snowmelt on runoff is important. Beas inflow is also influenced by the snowmelt. Further details on snowmelt runoff for Beas basin can be referred elsewhere [57].

4. Research methodology

4.1. Methodological framework of HMM_GP model

The methodological framework of HMM_GP model is consisting of four stages and is illustrated in Figure 2, adapted from [58]. The colour schematics applied in Figure 2 represent the key four stages of HMM_GP modelling framework which are briefly overviews below and can be referred elsewhere [58] for underlying technical details. Gray colour boxes are used for highlighting data pre-processing and post-processing steps, whereas purple boxes indicated start and end of procedure.

Stage 1 (Orange) - Time series decomposition using a robust STL (a Seasonal-Trend decomposition procedure based on Loess) [59]: The STL procedure facilitates a temporal decomposition of observed time-series (O) into three components: i) Long-term trends (Tr); ii) Seasonal movements (S); and iii) Random variations(R).

$$O = Tr + S + R; \text{for additive decomposition} \quad (1)$$

Since, STL procedure is mostly suitable for additive decomposition, input time series are recommended for a log-transformation. The data pre-processing is intended to stabilise variance in non-stationary series and to de-emphasizes the influences of extreme values (outliers). A detailed literature review covering a detailed overview of STL method, along with other closely related mathematical techniques for time series decomposition such as Empirical Mode Decomposition (EDM) can be referred elsewhere [58,60,61].

Stage 2 (Blue) - Fitting of Hidden Markov Model (HMM) to random component: Application of HMM model to random component facilitate simulation of uncertainty / randomness associated with the process. The theoretical structure of HMM model comprises of five components as represented in Table 2. Table 2 also details procedure of fitting HMM model to random component used within the HMM_GP framework.

HMM model fitted to random component is used to simulate n —user specified random components. These synthetic random components are then combined with trend and seasonal components of observed series to construct n synthetic time series corresponding to original series.

Table 1: Summary statistics of the $P(mm/day)$, $EV(mm/day)$ and $T^{\circ}C$ for the Pong sub-catchments

Indices		Minimum	Mean	Maximum
Sub-Catchment: Upper	P	0.00	2.77	96.44
	EV	0.27	1.97	6.31
	$Max.T$	-6.31	10.85	25.75
	$Min.T$	-23.07	-1.42	12.24
Sub-Catchment: Middle	P	0.00	3.76	119.44
	EV	0.53	4.62	9.70
	$Max.T$	6.87	26.12	40.03
	$Min.T$	-1.24	13.40	25.74
Sub-Catchment: Lower	P	0.00	4.12	136.83
	EV	0.73	6.58	15.92
	$Max.T$	10.71	32.78	50.30
	$Min.T$	-3.05	18.45	33.16

Stage 3 (Green) - Fitting GEV or GP distribution: To effectively simulate extreme values a GEV or GP distribution is fitted to extreme values, specifically in the range of 95th to 99.9th percentiles, in observed series. All synthetically simulated series are then processed to resample extreme values from the fitted distribution. This step ensures the extreme limits of the synthetic sequences are not constrained by the observed dataset and the fitted continuous distribution allows to incorporate unseen extreme events in the synthetic series. A technical discussion on appropriateness of selecting a GEV or GP distribution can be referred elsewhere [62].

Stage 4 (Yellow) - Bias Correction: Data transformation procedure involving log-transformation and back transformation often produces biased predictions [63]. To minimise the influence of data transformation procedure, a novel-percentile based bias correction is applied to synthetic series so that synthetic data are not out of synch with actual data.

Following the HMM_GP approach, as detailed above and illustrated in Figure 2, we seek to generate synthetic streamflow sequences for the river inflow data Q , and similarly for the EV and P time series. We aim to both show that a) our synthetic stream flows have the same statistical dynamics and characteristics as the actual dataset, and b) to provide future stream flows to investigate the effect of greater disturbance from ENSO events to the agriculture industry in India.

The underpinning methodology for generation of synthetic time series for Q , EV and P remains same, though procedure need to be adapted to incorporate slight variations noted in different dataset due to difference in the quantitative nature of the data. For example,

- (a) Time series of EV has a considerably small range, [0.73, 15.92], of values, and is not a particularly long-tailed distribution¹. Since time series of EV dataset does not exhibit an overly long-tailed distribution. Therefore, for processing EV dataset need some adaption in Stage 3. Specifically, we choose to fit a GEV distribution rather than a GP type distribution in the observed time series of EV , as detailed in [62].
- (b) Time series of P has a wide range of values [0, 136.83] with many zero values and a long-tailed distribution. For time series of P , as there can be valid zero elements

¹ Long-tailed or heavy-tailed distribution are those with extended tails either in right or left or both directions, due to several values occurring far from the mean or central part of the distribution. Long-tailed distribution are mainly studied within the context of extreme value distribution.

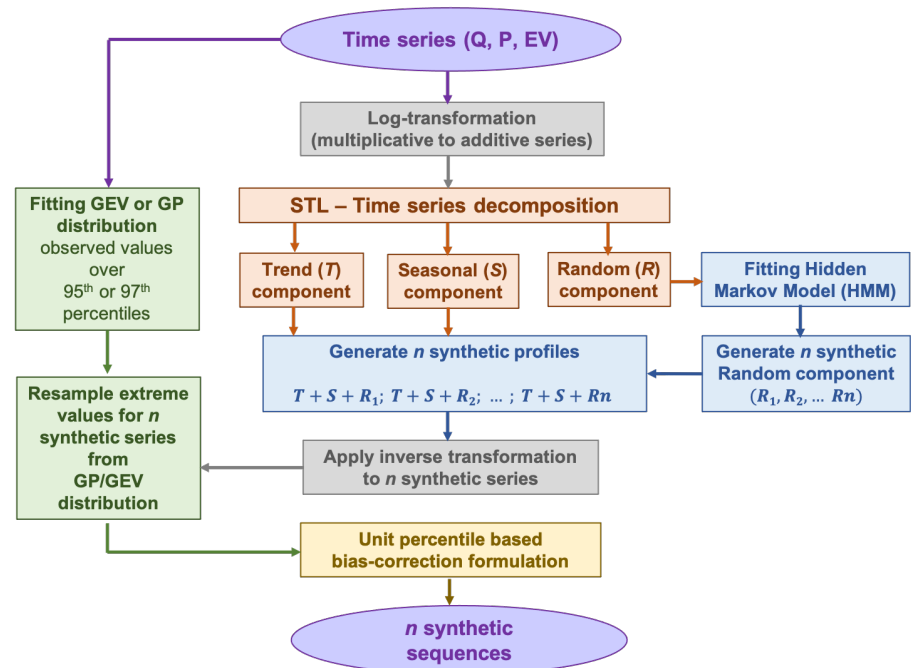


Figure 2. Work flow diagram of HMM_GP model, adapted from 58

of the series, we shift the series by a very small translation, 0.001, so that we can take the \log of the series. As the data is available to 2 decimal places and ranges from 0.01 to 136.83, this small value does not represent a significant change to the data and does not excessively stretch (on the negative side) the range of values that the \log of the series takes (note this is significant when fitting an HMM to the data). Stage 1- 2 are applied and after inverse transforming synthetic series we subtract the small shift value, 0.001, from the synthetic series. In contrast to the EV data, in Stage 3, as the distribution for the P data is very long-tailed, thus, we choose to fit a GP distribution.

- (c) Time series of Q has a very wide range, [17,5915] and a long-tailed distribution. A climatic module is developed to integrated influences of EV and P data in simulated Q sequences, detailed in next subsection.

4.2. Calibrating 'climatic module' for simulating Q sequences

In a novel approach, we seek to link the underlying data trend for Q with underlying climatic conditions, noting that EV , as measured by the Penman-Monteith method, depends on daily mean T , wind speed, relative humidity and solar radiation [16]. We thus investigate fitting a model for the response of the trend for the Q series to changes in the trend for P and EV as follows:

1. We take the \log of the Q time series to convert a multiplicative time series into an additive series.
2. Using the Loess method of time series decomposition from [59] we decompose the $\log(Q)$ series into trend, seasonality and random components.
3. We fit a linear regression model for the response of the decomposed trend of the $\log(Q)$ series to the trends generated by decomposition of $\log(EV)$ and $\log(P)$, as described in Subsection 4.1 above:

$$\log(Q)_{\text{trend}} \sim \log(EV)_{\text{trend}} + \log(P)_{\text{trend}}$$

Table 2: Structural composition of HMM

Set of observed states [O _{State}]	Percentile analysis of random component is conducted to define eleven distinct states, A, B, \dots, K State A - value between 0 th – 10 th percentile; State B - value between 11 th – 20 th percentile; ; State K - value between 95 th – 100 th percentile
State transitional probability matrix [T _{State}]	$[T_{State}]_{11 \times 11} = [p_{ij}; \text{for } i, j \in A, B, \dots, K]$ $= \begin{bmatrix} p_{AA} & p_{AB} & \dots & p_{AK} \\ p_{BA} & p_{BB} & \dots & p_{BK} \\ \dots & \dots & \dots & \dots \\ p_{KA} & p_{KB} & \dots & p_{KK} \end{bmatrix}$
Set of unobserved (hidden) states [U _{State}]	Hidden states corresponding to each of eleven observed state are defined as $[U_{State}] = [U_i; \text{for } i \in A, B, \dots, K]$
Emission probability matrix [E]	Corresponding to each hidden state $U_i \in [U_{State}]$ $[E_i] = [e_i; \text{for } i \in 1, 2, \dots, E_i^{length}]$ where, $E_i^{length} = [(range \text{ of } U_i \times 10) + 1]$
Initial Probability Matrix [I]	Initial probabilities of occurrences of observed states, observed state are defined as $[I] = [I_i; \text{for } i \in A, B, \dots, K]$

Accordingly, we complete the generation of N synthetic series for the Q data complementing the synthetic series generated for P and EV with the following steps:

4. We fit an HMM to the $\log(Q)$ of random component.
5. Using the HMM we generate the prescribed number (N) of synthetic random series for a pre-determined length of time (less than or equal to that of the P and EV length).
6. We recombine the decomposed series by adding the seasonality component for the $\log(Q)$ from Step 2 and the $\log(Q)$ trend fitted by the linear regression model to each of the N synthetic random series from Step 5.
7. Take the exponential of each of the N resultant series from Step 6.
8. Re-sample extreme values in the synthetic series from a GP distribution fitted to the extreme values in the actual Q data, as detailed in Stage 3. As the distribution for the Q data is very long-tailed we choose to fit a GP distribution and apply Stage 4 from Subsection 4.1.

4.3. Application of 'climatic module' for forecasting Q sequences

Further to generating synthetic Q series for a particular time frame using HMM to 'learn' the behaviour of the random component over that time we also investigate how feasible it is to use our synthetic series to predict future Q dependent on climate changes. For this exercise, we propose to use a truncated portion of the Q data to 'learn' the behaviour of the Q data to predict the Q for future times. In this way can compare our predictions with actual river runoff data. We do this with the following steps:

1. We split the time-frame T for our data into two segments T_{learn} and T_{predict} such that the start date in the year for both T_{learn} and T_{predict} is the same (to ensure the correct starting probabilities for our fitted HMM). Note we also assume that $\text{length}(T_{\text{learn}}) \geq \text{length}(T_{\text{predict}})$.
2. For the inflow data in the T_{learn} time-frame we follow the initial steps for generating synthetic inflow series up to fitting an HMM to the random $\log(Q)$ decomposed series, including generating the linear regression model for the $\log(Q)$ trend on time period T_{learn} .

3. Using the HMM fitted for the inflow data in time period T_{learn} generate the prescribed number (N) of synthetic random series for time period T_{predict} .
4. Manufacture the synthetic Q series for time period T_{predict} by adding the predicted random series, the predicted trend for the Q using the linear model whose parameters are fitted from the data in time period T_{learn} and generated using the P and EV trend from time period T_{predict} , and the seasonal inflow component generated by using the annual seasonal component decomposed in time period T_{learn} .
5. Take the exponential of each of the N resultant series.
6. Re-sample extreme values in the synthetic series from a GP distribution fitted to the extreme values in the actual Q data, as detailed in Stage 3 and apply a percentile basis, as detailed in Stage 4.

5. Results

This section is aimed to demonstrate and discuss the key results obtained at various stages of model development (detailed in above section). Key research findings are organised in four subsections discussed below.

5.1. STL decomposition of EV , P and Q

To demonstrate results obtained at Stage 1 of HMM_GP framework, Figures 3, 4 and 5 show the STL decomposition for the log of the time series for EV , P and Q respectively for the years 1998 to 2010. Note that as zero entries may occur in the data for P , before taking logs we add a very small number (10^{-3}) to the P time series. A reverse of this is performed when transforming back from the additive series to the multiplicative series. We set the seasonal decomposition window to a granularity of 1 year as we wish to investigate changes on an annual basis. Both EV and Q show clear annual peaks and troughs, whereas the P , although exhibiting an annual monsoon season, has many days throughout the year when there is either zero or extremely low rainfall. We can see that in general, the trend for $\log(Q)$, $\log(P)$ and $\log(EV)$ exhibits a change in behaviour around the years 2003-2004 where Q shows a change in decreasing trend to increasing; P shows a levelling off after years of increasing, and EV starts a downward trend after a few years of an unchanging trend in general. Therefore we conclude that, in general, as the trend for EV decreases we expect an increase in the trend for Q into the reservoir. We also see that as the trend for P increases to its peak in 2006, then the trend for Q also rises. Please note that, years 2002-03, 2004-05, 2006-07 and 2009-10 were recorded to observe an ENSO events [64].

5.2. Calibration of HMM_GP model for simulating synthetic sequences

We applied HMM_GP methodology to simulate synthetic sequences for EV , P and Q . Figures 6 and 7 show a sample synthetic series generated using the HMM_GP methodology for EV and P respectively for the years 1998 to 2010, and Figure 8 shows a sample synthetic series for inflow (Q) for the years 1998 to 2006. The synthetic time series exhibit good agreement with periodic peaks and troughs of the observed data (Figures 6a, 7a and 8a), although as our modelling is based on probabilities the most extreme values in the synthetic series do not always match the years in the original Q time series when the most extreme events occur. Notably, as EV and Q both have clear annual peaks and troughs the modelled synthetic data show better agreement with annual timings of highs and lows than for P . The percentiles and probability density function for the synthetic series follow the observed time series reliably (Figures 6b&c, 7b&c and 8b&c) and the quantile-quantile plot in Figures 6d, 7d and 8d show good agreement, particularly below the 99th percentile. Figures 6(e,g), 7(e,g) and 8(e,g) show the auto-correlation and partial auto-correlation over a lag time of 1 month for the actual data for EV , P and Q respectively with their counterparts for the synthetic series shown in Figures 6(f,h), 7(f,h) and 8(f,h). There is a good agreement for each of the synthetic series with actual data, and also we note that both EV and Q show a higher level of auto-correlation in general (Figures 6(e,f) and 8(e,f)) when compared with P (Figure 7(e,f)).

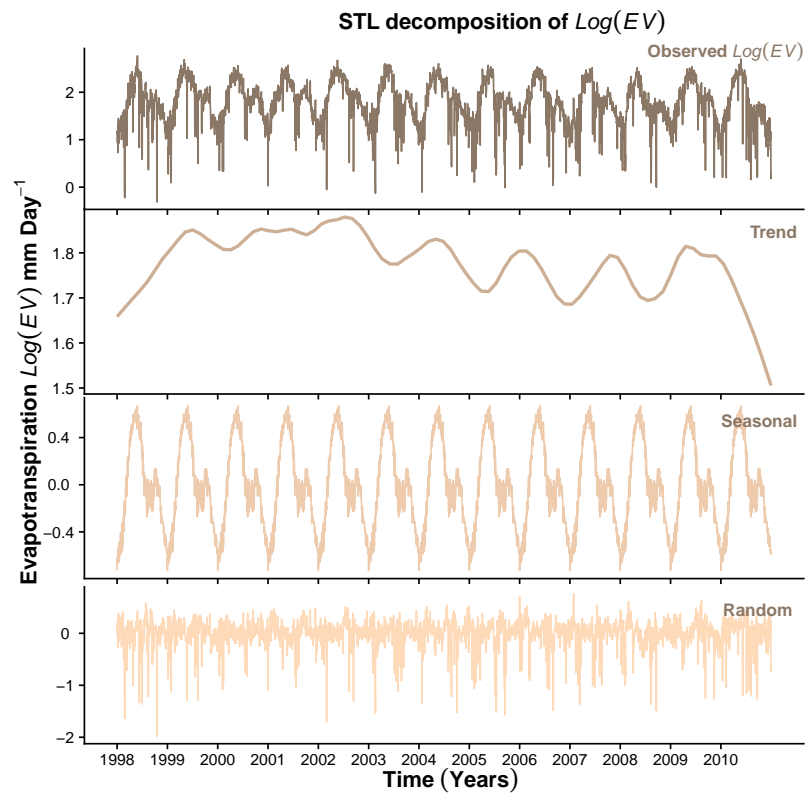


Figure 3. STL decomposition of $\text{Log}(EV)$.

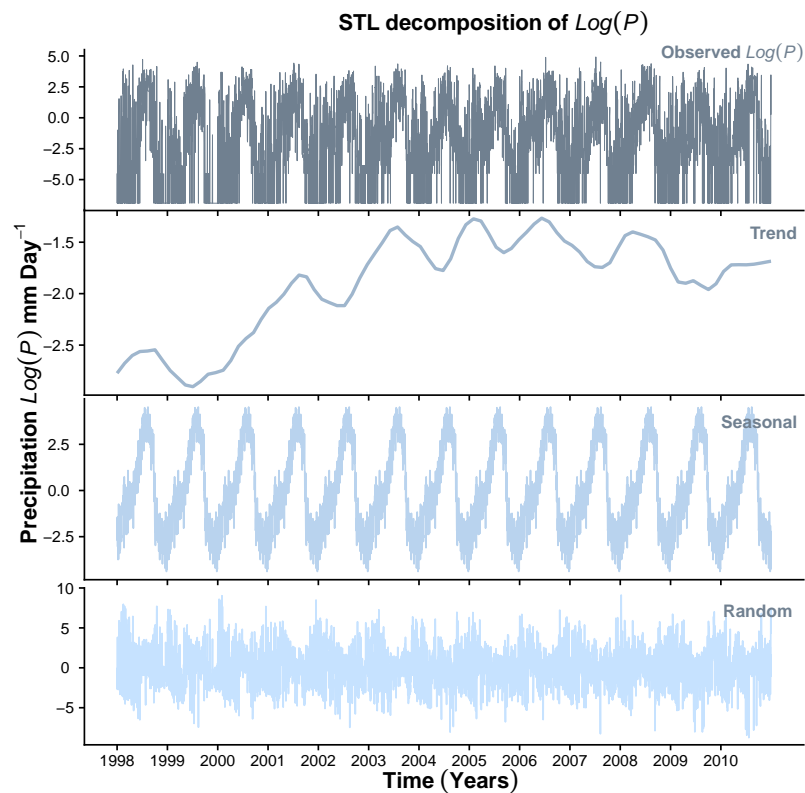


Figure 4. STL decomposition of $\text{Log}(P)$.

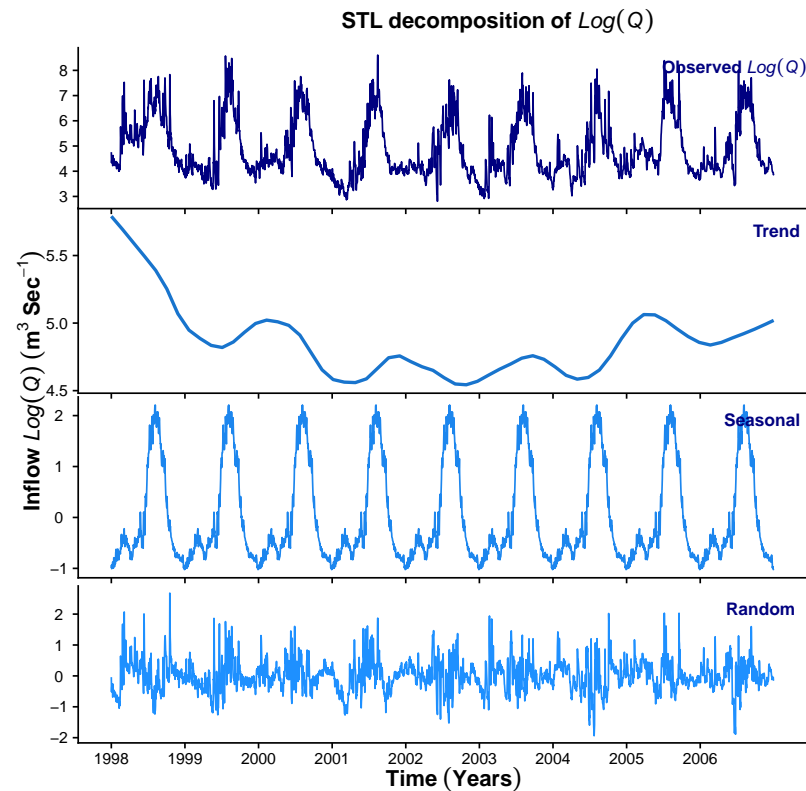


Figure 5. STL decomposition of $\text{Log}(Q)$.

To fit extreme values adequately in the synthetic series, we found that fitting the GEV distribution to extreme EV data values above the 95th percentile provided the best results. For fitting the GP distribution to the precipitation data, we also found that fitting to data above the 95th percentile was optimal, whereas for the Q data we found that fitting to the 99th percentile was the best option. We suspect that fitting our model to a time series with a lower degree of auto-correlation is more problematic due to the greater likelihood of large differentials between sequential data points. In our case study, this is due to the nature of P data that can change from high to low daily even during monsoon season. There is ample room for further investigation in this area.

Figures A1, A2 and A3 (provided in appendix section) show a further four sample synthetic series generated using HMM_GP for EV , P and Q respectively, each exhibiting similar statistical properties to those in Figures 6, 7 and 8. We conclude that the synthetic series we have produced for EV , P and Q has a similar statistical characteristics to their original series. Thus, the HMM_GP framework is shown to effectively simulate dynamics of climatic variable along with river runoff. In fact, in previous work by authors, the proposed modelling schematics is shown to generate statistical synthetics time series for a range of applications including energy demand series at different resolution in ranging (5 minutes to 30 minutes) [58] and for Scottish rivers inflow sequences (15 minutes inflow) [17]. In the present work model is first time applied for simulating statistical dynamics of climatic variables and a novel climate module is developed.

5.3. Calibration of 'climatic module'

Climate module is novel feature of HMM_GP model and has been intended to establish a mathematical association between the trends components of climatic variables (EV and P) with the trends of (Q) using a simple multiple regression-based model. Key underpinning idea is that such a relationship can be used to project changes in trend of future Q as a response to change in EV and P . Seasonal component can be explored in within the same conceptual framework, though is not focus of present work. The output from R [65] for the

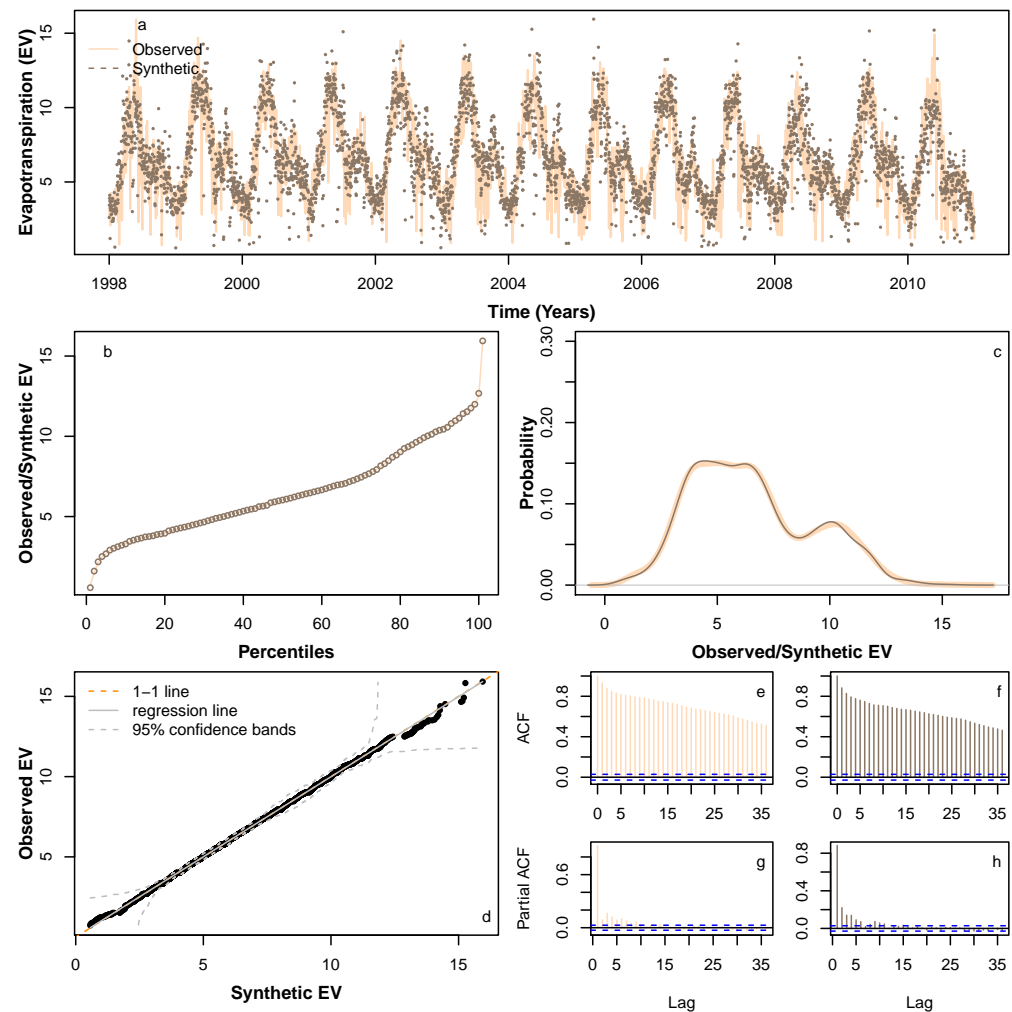


Figure 6. Comparison of various statistical characteristics of a sample synthetic *EV* time series for 1998-2010 (dotted brown lines), based on learning *EV* pattern from 1998-2010, with observed *EV* 1998-2010 time series (solid peach lines). Comparing (a) Realisation; (b) Percentiles; (c) Probability density distribution; (d) QQ-plot; (e) ACF for observed; (f) ACF for a sample synthetic series; (g) PACF for observed; (h) PACF for a sample synthetic series.

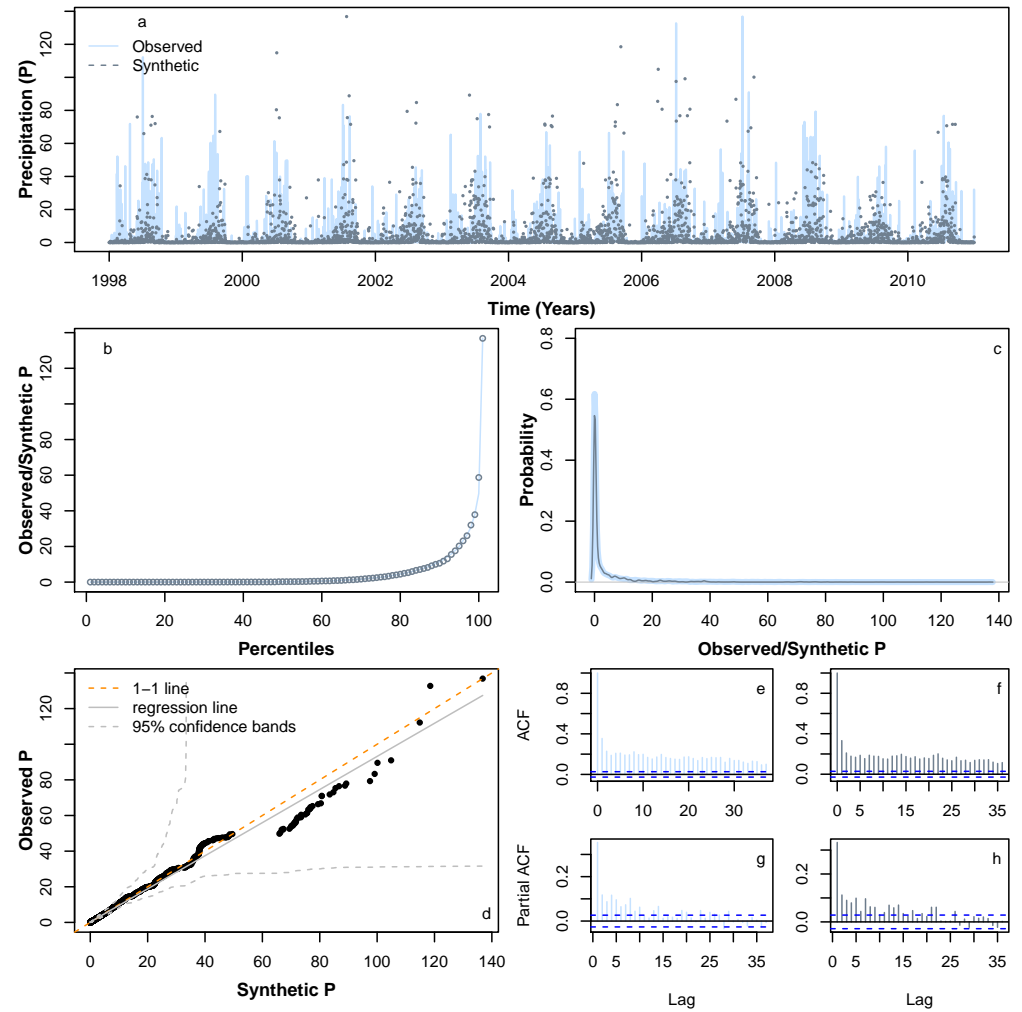


Figure 7. Comparison of various statistical characteristics of a sample synthetic P time series for 1998-2010 (dotted lines in slate grey), based on learning P pattern from 1998-2010, with observed P 1998-2010 time series (solid lines in light slate grey). Comparing (a) Realisation; (b) Percentiles; (c) Probability density distribution; (d) QQ-plot; (e) ACF for observed; (f) ACF for a sample synthetic series; (g) PACF for observed; (h) PACF for a sample synthetic series.

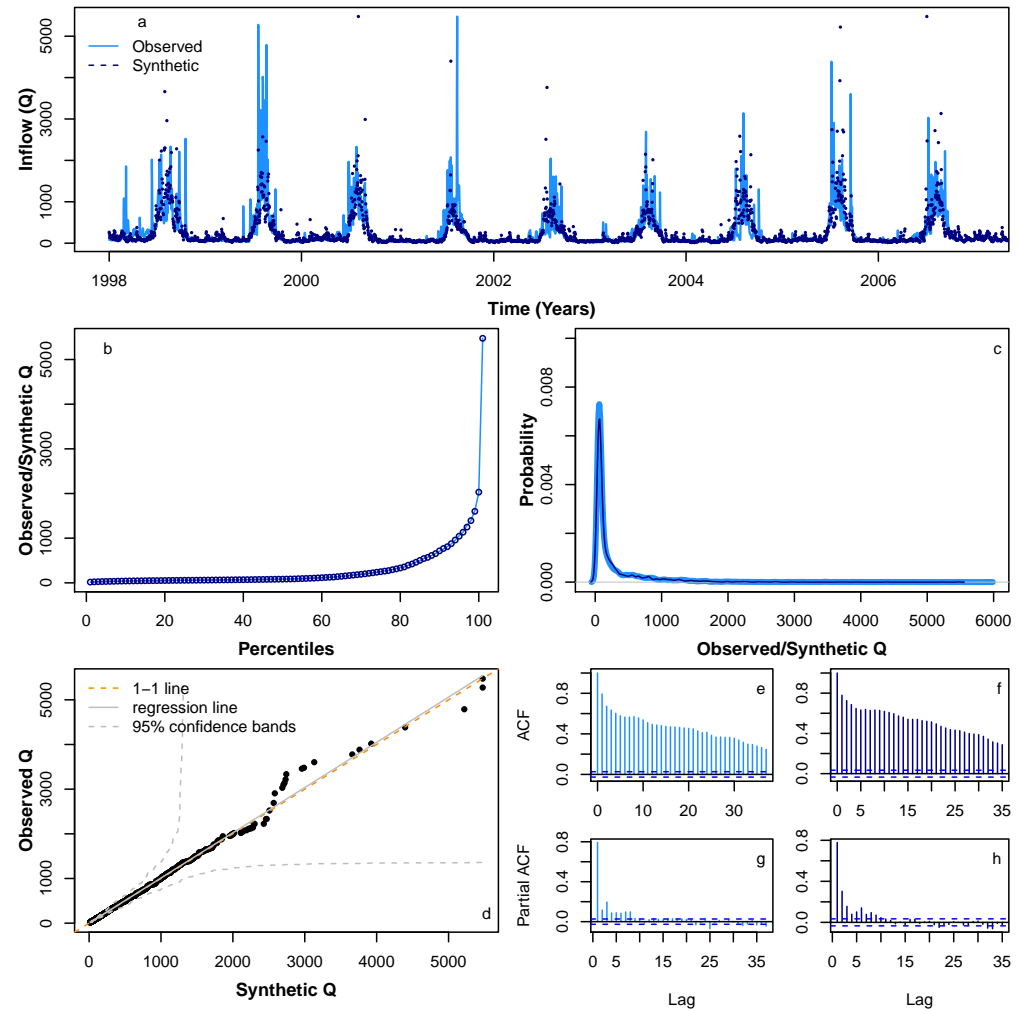


Figure 8. Comparison of various statistical characteristics of a sample synthetic Q time series for 1998-2006 (dotted navy lines), based on learning Q pattern from 1998-2006, with observed Q 1998-2006 time series (solid blue lines). Comparing (a) realisation; (b) Percentiles; (c) Probability density distribution; (d) QQ-plot; (e) ACF for observed; (f) ACF for a sample synthetic series; (g) PACF for observed; (h) PACF for a sample synthetic series.

linear model linking the trend of $\log(Q)$ to the trend of $\log(P)$ and trend of $\log(EV)$ gives the following model summary:

```
Call:
lm(formula = Trend_TS_Inflow_Log ~
Trend_TS_PrecL_Log + Trend_TS_EvoPL_Log)

Residuals:
Min      1Q  Median      3Q      Max
-0.22324 -0.06903 -0.00638  0.06717  0.16561

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    12.235107   0.049441    247.5  <2e-16 ***
Trend_TS_PrecL_Log -0.288592   0.002832   -101.9  <2e-16 ***
Trend_TS_EvoPL_Log -4.416697   0.028030   -157.6  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.08359 on 3282 degrees of freedom
Multiple R-squared:  0.8992, Adjusted R-squared:  0.8992
F-statistic: 1.465e+04 on 2 and 3282 DF,  p-value: < 2.2e-16
```

where we can see that the linear model, we have proposed is a reasonable fit as the residual standard error is low and the p-values for both P and EV are extremely small (less than 0.05) and suggest that both P and EV in the model are statistically significant and we should retain them. R-square values is $0.89 \sim 0.9$, which is considerably high, suggesting a ‘strong’ model fit has been achieved. The overall p-value of the regression model is also small (less than 0.05 as indicated in the last row), suggesting modelling is statistically significant. To assess the capabilities of ?climate module? in simulating trend components of Q , Figure 9 Upper panel compares the results for the modelled Q trend (dashed lines in light blue) superimposed on the observed Q trend (solid line in navy), showing a reasonable model fit. We note that the fit is not perfect, in particular, that the peaks and troughs in the trend are not always concurrent, however, the overall general trend is followed with good agreement. To further analyse the discrepancies in the match and the magnitude, in 9 Lower panel, we plotted histogram of residual errors (defined as difference between ‘observed’ and ‘predicted’ trend of Q). It is interesting to notice that most of the residual error (approximately 90%) is distributed within the range $[-0.11, 0.13]$ with some extreme values reaching the range $[-0.2, 0.2]$. Table 3 provides a detailed percentile distribution of residual error.

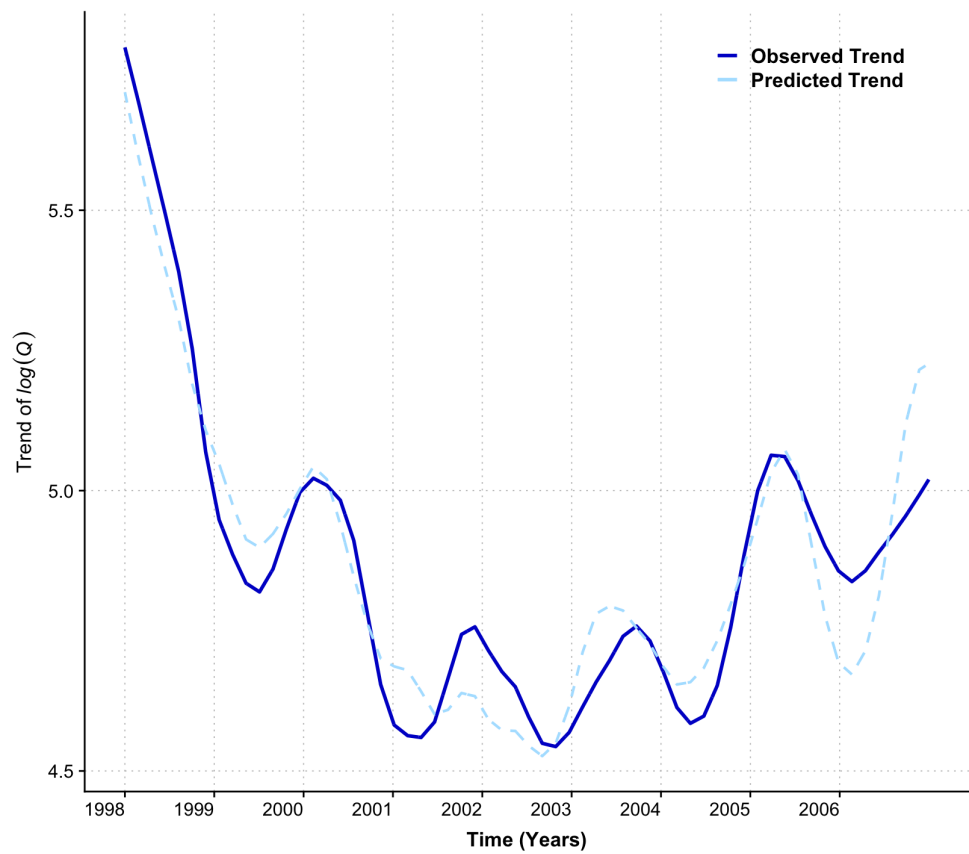
Table 3: Percentiles for residual error

Percentiles	0 th	5 th	10 th	25 th	50 th	75 th	90 th	95 th	100 th
residual error	-0.22	-0.11	-0.10	-0.70	-0.01	0.07	0.11	0.13	0.17

In future versions of our work, we intend to improve this model and consider this further in the final discussion, in particular considering possible lags in the data.

5.4. Model prediction

Figure 10 shows the predicted Q for 2007 to 2010 based on learning the behaviour of the Q time series for 1998-2006 based on the STL seasonal decomposition, using an HMM to learn the random component for 1998-2006 and the linear model for the Q trend whose coefficients are found using the P and EV trend data for 1998-2006. The resultant predicted Q for 2007-2010 closely follows the timing of peaks and troughs of the original



Histogram of residual error (Observed – Predicted Trend Q)

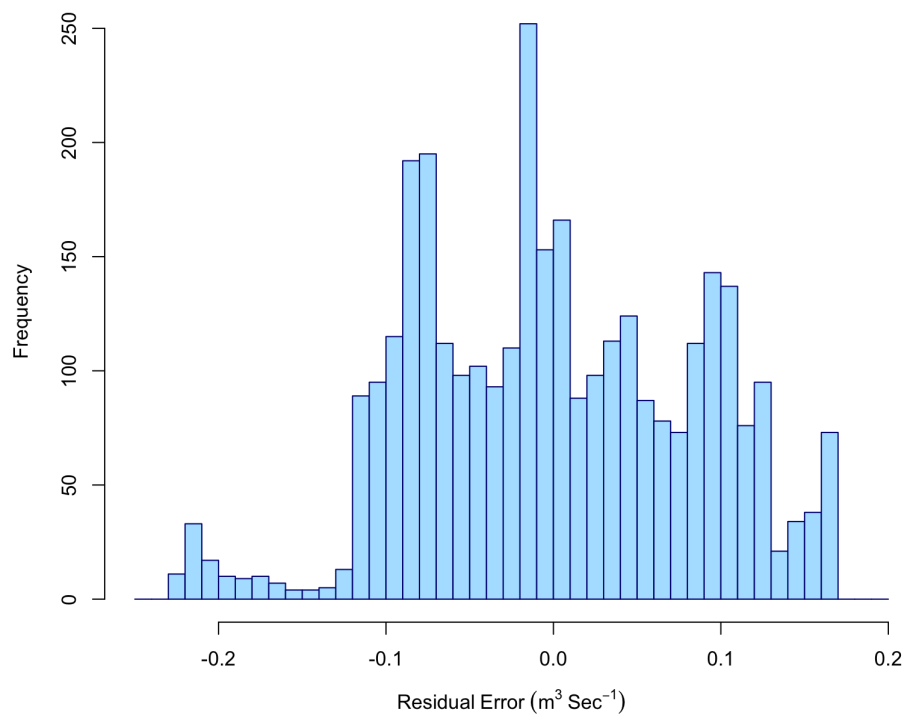


Figure 9. Upper Panel: Compares observed trend component of $\log(Q)$ (solid lines in Navy) with predicted trends of $\log(Q)$ for years 1998-2006 simulated by 'climate module' (dashed lines in light blue) in response to trend of $\log(EV)$ and $\log(P)$ as independent variables. Lower Panel: Histogram of residual error distribution

data (Figure 10a). Due to our modelling being on a probabilistic basis the most extreme values in the synthetic series do not always match the years in the original Q time series when the most extreme events occur. However, the percentiles (Figure 10b), probability density (Figure 10c), ACF and PACF profiles (Figures 10e,f,g&h) have a similar statistical profile to the original series. Figure 11 shows a further four sample synthetic series for Q , each exhibiting similar statistical properties to those in Figure 10. **Therefore, we have shown that for several years (less than the initial learning period), we can predict the vast majority of the behaviour exhibited by the original series, barring identifying the correct years when the most extreme events occur.**

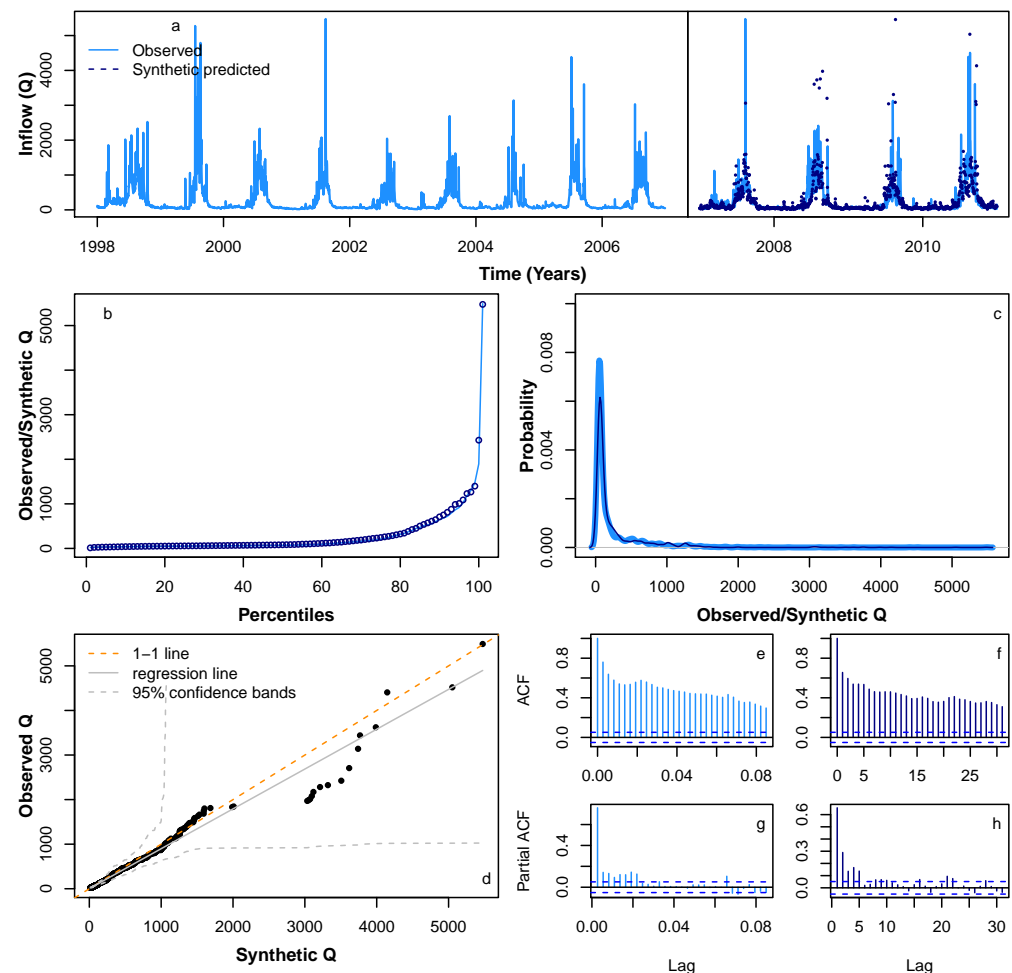


Figure 10. Comparison of various statistical characteristics of a sample predicted Q for 2007-2010 (dotted navy lines), based on learning inflow pattern from 1998-2006, with observed Q 2007-2010 time series (solid blue lines). Based on learning Q for 1998-2006, comparing (a) realisation of observed Q time series for 1998-2010 (solid blue lines) and a sample predicted time series for 2007-2010 (dotted navy lines); (b) Percentiles for observed Q time series (solid blue lines) and a sample predicted time series for 2007-2010 (dotted navy lines); (c) Probability density distribution for observed Q time series (solid blue lines) and a sample predicted time series for 2007-2010 (dotted navy lines); (d) QQ-plot for observed Q time series (solid blue lines) and a sample predicted time series for 2007-2010 (dotted navy lines); (e) ACF for observed Q 2007-2010 time series; (f) ACF for a sample predicted Q 2007-2010 time series; (g) PACF for observed Q 2007-2010 time series; (h) PACF for a sample predicted Q 2007-2010 time series.

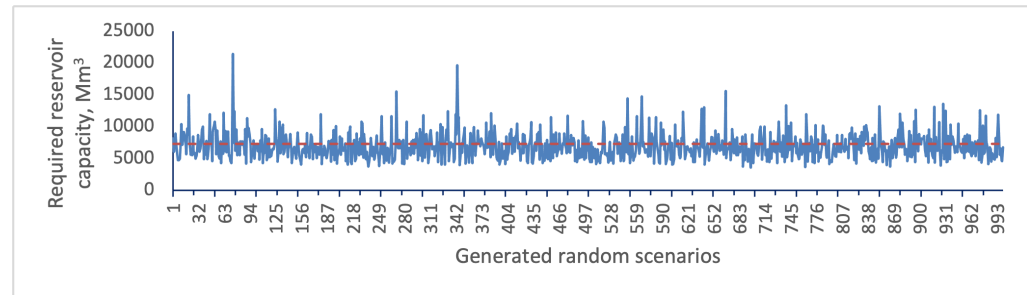


Figure 12. Reservoir capacity uncertainty using the generated random inflow scenarios.

5.5. Model Application: Uncertainty in reservoir capacity estimates

The required capacity to meet the existing demands at the Pong without failure when fed with the different Q scenarios. A simple technique for obtaining the failure-free capacity estimate is the sequent peak algorithm (SPA) proposed by [63].

$$K_{t+1} = \max(0, K_t + D_t - Q_t); t \in N \quad (2)$$

$$K_a = \max(K_{t+1}) \quad (3)$$

where, K_a is reservoir capacity, K_{t+1} and K_t are respectively the sequential deficits at the end and start of time period t , D_t is the demand during t , Q_t is the inflow during t and N is the number of months in the data record. The SPA is a critical period reservoir sizing technique and like all such techniques assumes that the reservoir is full at start and end of the cycle, i.e., $K_0 = K_N = 0$. If, however, this is untrue, i.e. $K_N \neq 0$, the SPA cycle is repeated by setting the initial deficit to K_N , i.e. $K_0 = K_N$. This second iteration should end with K_N unless the demand is unrealistic, e.g. such as attempting to take a demand higher than the mean annual runoff from the reservoir. In this sense, the assumption of an initially full reservoir is not crucial for the SPA because if this assumption is not valid, it will become evident at the end of the first cycle and corrected for during the second cycle.

To estimate uncertainty in reservoir capacity, HMM_GP model along with ?climate module? is applied to generate 1000 synthetic Q sequences. Population of reservoir capacity based on existing monthly irrigation releases at the Pong are summarised in the Figure 12. The horizontal dashed line represents the existing (or historic) capacity of 7290 Mm^3 . As Figure 12 clearly shows, there is wide variability in the required reservoir capacity for each runoff scenario. Although the existing capacity of the Pong is 7290 Mm^3 , the required capacity estimates based on the simulated current runoff series could be as low as 3545 Mm^3 or as high as 21452 Mm^3 . These, respectively, represent under-design and over-design situations relative to the existing capacity at the Pong reservoir. The implication of under design is that the reservoir will fail frequently to meet the demand.

The effect of ENSO on the capacity estimates broadly follows the effect on runoff. Thus, as the rainfall and hence runoff decreases, the capacity required for meeting the demand increases. The large arrays of possibilities in the impact of ENSO are bound to complicate decision-making regarding adaptation and mitigation.

6. Discussion

From the initial data that we have obtained for the Beas river basin (EV , P and Q) we have demonstrated that we can successfully produce a suite of synthetic time series exhibiting similar statistical characteristics to the original historic data, following seasonal highs and lows well. Moreover, we have linked the trend of the Q to the Pong dam to local P and EV . Using this link to the trend in climatic conditions we have also demonstrated that predictions for inflow based on seasonality and random flow learnt from historical river Q data and climatic trend are statistically accurate. With the ability to adjust the Q trend based on the amount of P and other climatic conditions we envisage that such synthetic time series can be used as input to crop growth models to help understand the

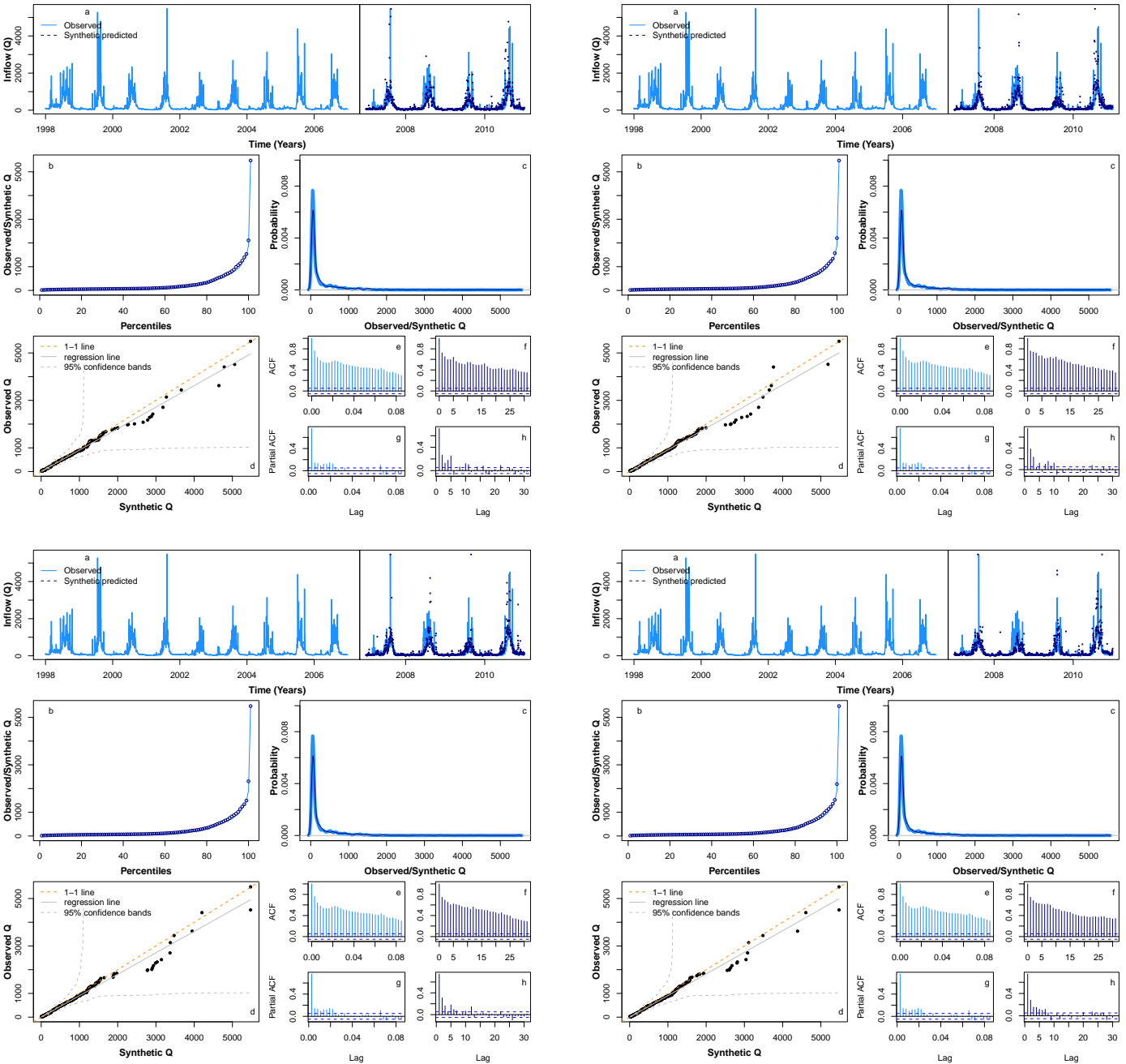


Figure 11. Comparison of four sample predicted inflow time series for 2007-2010. See Fig 10 for individual figure details.

39 effects of changes in climatic factors on crop selection and output. We have based this work
40 on data available for the river Beas in the north of India and this approach can easily be
41 translated to other rivers in India influenced by monsoon conditions and extreme weather
42 events. Of course, this depends on good quality historical data on which accurate machine
43 learning can be built.

44 The accuracy of our model depends on a suitable length of historical data to train data
45 to capture long-term changes in patterns. In the case we have predicted 4 years of Q based
46 on 9 years of learning. Given that ENSO type events can occur every 2-7 years and have a
47 variable duration, we would expect more accurate results with historical data that includes
48 multiple numbers of ENSO type occurrences. We envisage that if the learning phase of our
49 model has a longer duration, then our model will have a greater likelihood of simulating
50 extreme value occurrences.

51 As well as the length of data, extending the breadth of data used for historical learning
52 may also improve prediction results. In this study, we have limited linking the trend of the
53 Q to the EV and P in the Beas river basin (up to the Pong dam), and it is conceivable that
54 other explanatory variables may improve the linear model for Q trend (Figure 9). We know
55 that P in the uppermost reaches of the river Beas falls as snow during the winter months,
56 and that snowmelt contributes significantly to Q [66]. If sufficient historical data for EV
57 and P in the upper reaches of the river Beas is available, further investigation to fitting a
58 multiple linear regression model to EV and P from a wider catchment could be beneficial.
59 Note that this would also require an understanding of any time lag that occurs between
60 effects of EV and P in the upper catchment area and the river Q at the Pong dam.

61 Although the data used for our studies have been decomposed using the statistical
62 STL approach, there perhaps could be more deterministic ways of extracting the various
63 components from the data. For instance, if we consider the seasonal patterns in the data
64 that occur over fixed periods, one could use tools such as Fourier and Wavelet transforms
65 for analysis. The data could also be pre-conditioned with such transforms to provide a
66 way to filter noise, or even compress the data. Data decomposed with Fourier/Wavelet
67 transforms could provide a robust method not only to study the regularities present but also
68 to produce synthetic data that could be used for predictions. The challenge in this aspect of
69 research would be to maintain the statistical properties of the data before and after such a
70 transform. Moreover, Fourier analysis in particular is restricted to finite energy functions,
71 which may interfere with the long-term progression trend of the data. Nonetheless, we
72 could devise techniques to tackle such issues and produce hybrid methodologies to better
73 analyse and model the data [67].

74 Linked to exploring further avenues for decomposing the data we note that our
75 modelling approach has been to restrict the STL decomposed seasonal component to be
76 unchanged over each year. It is conceivable that this is too restrictive, and there could be
77 a benefit in allowing the seasonal component to evolve over a period of time. However,
78 such a change would also require changes in the assembling of our inflow predictions. The
79 suggestions we have discussed here provide a number of future possibilities to extend
80 our statistical modelling approach. As the motivation for modelling runoff is to provide
81 suitable input to ascertain probabilities of crop success in the face of ENSO type events, the
82 biggest determinant of the most suitable improvements to investigate in the future will be
83 the utility of the synthetic and predicted series in crop development models.

84 **Author Contributions:** For research articles with several authors, a short paragraph specifying their
85 individual contributions must be provided. The following statements should be used “Conceptualiza-
86 tion, X.X. and Y.Y.; methodology, X.X.; software, X.X.; validation, X.X., Y.Y. and Z.Z.; formal analysis,
87 X.X.; investigation, X.X.; resources, X.X.; data curation, X.X.; writing—original draft preparation,
88 X.X.; writing—review and editing, X.X.; visualization, X.X.; supervision, X.X.; project administration,
89 X.X.; funding acquisition, Y.Y. All authors have read and agreed to the published version of the
90 manuscript.”, please turn to the [CRediT taxonomy](#) for the term explanation. Authorship must be
91 limited to those who have contributed substantially to the work reported.

Funding: Please add: "This research received no external funding" or "This research was funded by NAME OF FUNDER grant number XXX." and and "The APC was funded by XXX". Check carefully that the details given are accurate and use the standard spelling of funding agency names at <https://search.crossref.org/funding>, any errors may affect your future funding.

Acknowledgments: In this section you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

References

- Gleick, P. H. Streamflow and the Water Cycle, In *Encyclopedia of Climate and Weather*; Oxford University Press: New York, 1996; pp. 817–823.
- Johnson, A. C.; Acreman, M. C.; Dunbar, M. J.; Feist, S. W.; Giacomello, A. M.; Gozlan, R. E.; Hinsley, S. A.; Ibbotson, A. T.; Jarvie, H. P.; Jones, J. I.; Longshaw, M.; Maberly, S. C.; Marsh, T. J.; Neal, C.; Newman, J. R.; Nunn, M. A.; Pickup, R. W.; Reynard, N. S.; Sullivan, C. A.; Sumpter, J. P.; Williams, R. J. The British river of the future: How climate change and human activity might affect two contrasting river ecosystems in England. *Science of The Total Environment*, **2009**, 407(17), 4787-4798.
- Robins, P. E.; Skov, M. W.; Lewis, M. J.; Giménez, L.; Davies, A. G.; Malham, S. K.; Neill, S. P.; McDonald, J. E.; Whitton, T. A.; Jackson S. E. and Jago, C. F. Impact of climate change on UK estuaries: A review of past trends and potential projections. *Estuarine, Coastal and Shelf Science*, **2016**, 169, 119-135.
- Visser-Quinn, A.; Beevers, L. and Patidar, S. A coupled modelling framework to assess the hydroecological impact of climate change. *Environmental Modelling and Software*, **2019**, 114, 12-28.
- Walther, G. R.; Post, E.; Convey, P.; Menzel, A.; Parmesan, C.; Beebee, T. J.; Fromentin, J. M.; Hoegh-Guldberg, O. and Bairlein, F. *Ecological responses to recent climate change*. *Nature*, **2002**, 416, 386-395.
- Whitehead, P. G.; Wade, A. J. and Butterfield, D. Potential impacts of climate change on water quality and ecology in six UK rivers. *Hydrology Research*, **2009**, 40(2-3), 113-122.
- Visser-Quinn, A.; Beevers, L.; Collet, L.; Formetta, G.; Smith, K.; Wanders, N.; Thober, S.; Pan, M. and Kumar, R. Spatio-temporal analysis of compound hydro-hazard extreme across the UK. *Advances in Water Resources*, **2019**, 130, 77-90.
- Seneviratne, S.; Nicholls, N.; Easterling, D.; Goodess, C.; Kanae, S.; Kossin, J.; Luo, Y.; Marengo, J.; Mc Innes, K.; Rahimi, M.; Reichstein, M.; Sorteberg, A.; Vera, C.; Zhang, X.; Rusticucci, M.; Semenov, V.; Alexander, L.; Allen, S. and Benito, G. Changes in climate extremes and their impacts on the natural physical environment, In *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC), Cambridge University Press: Cambridge, UK, and New York, NY, USA, 2012, pp. 109-230.
- Walker, G.T. Correlation in seasonal variation of weather. *Q. J. R. Meteorol. Soc.*, **1918**, 44, 223–224.
- Webster, P. J.; Yang, S. Monsoon and ENSO: Selectively interactive systems. *Q. J. R. Meteorol. Soc.*, **1992**, 118, 877-926.
- Webster, P. J.; Magaña, V. O.; Palmer, T. N.; Shukla, J.; Tomas, R. A.; Yanai, M.; Yasunari, T. Monsoons: Processes, predictability, and the prospects for prediction. *J. Geophys. Res.*, **1998**, 103, 14451-14510.
- McPhaden, Michael J.; Zebiak, Stephen E.; Glantz, Michael H. ENSO as an Integrating Concept in Earth Science. *Science*, **2006**, 314 (5806), 1740-1745.
- Collins, M.; An, Soon-II; Cai, W.; Ganachaud, A.; Guilyardi, E.; Jin, F. F.; Jochum, M.; Lengaigne, M.; Power, S.; Timmermann, A.; Vecchi, G.; Wittenberg, A. The impact of global warming on the tropical Pacific Ocean and El Niño. *Nature Geoscience*, **2010**, 3, 391-397.
- Cai, W.; McPhaden M. J.; Grimm, A. M.; Rodrigues, R. R.; Taschetto, A. S.; Garreaud, R. D.; Dewitte, B.; Poveda, G.; Ham, Y. G.; Santoso, A.; Ng, B.; Anderson, W.; Wang, G.; Geng, T.; Jo, H. S.; Marengo, J. A.; Alves, L. M.; Osman, M.; Li, S.; Wu, L.; Karamperidou, C.; Takahashi, K.; Vera, C. Climate impacts of the El Niño Southern Oscillation on South America. *Nature Reviews Earth & Environment*, **2020**, 1, 215-231.
- Fan, F.; Dong, X.; Fang, X.; Xue, F.; Zheng, F.; Zhu, J. Revisiting the relationship between the South Asian summer monsoon drought and El Niño warming pattern. *Atmospheric Science Letters*, **2017**, 18 (4), 175-182.
- Early Warning Early Action report on food security and agriculture (April-June 2019). *Licence: CC BY-NC-SA 3.0 IGO*, **2019**, Rome.
- Patidar, S.; Allen, D.; Haynes, R.; Haynes, H. Stochastic modelling of flow sequences for improved prediction of fluvial flood hazards. *Geological Society, London, Special Publications (Geological Society of London)*, **2018**, 488, SP488–4.
- Gergis, J. L. and Fowler, A. M. A history of ENSO events since A.D. 1525: implications for future climate change. *Climatic Change*, **2009**, 92, 343-387.
- Tsonis, A. A.; Hunt, A. G. and Elsner, J. B. On the relation between ENSO and global climate change. *Meteorology and Atmospheric Physics*, **2003**, 84, 229-242.
- Trenberth, K. E. and Hoar, T. J. El Niño and climate change. *Geophysical Research Letters*, **1997**, 24(23), 3057-3060.
- Wang, B.; Luo, X.; Yang, Y. M.; Sun, W.; Cane, M. A.; Cai, W.; Yeh, S. W. and Liu, J. Historical change of El Niño properties sheds light on future changes of extreme El Niño. *Proceedings of the National Academy of Sciences of the United States of America*, **2019**, 116(45), 22512-22517.
- Stevenson, S. L. Significant changes to ENSO strength and impacts in the twenty-first century: Results from CMIP5. *Geophysical Research Letters*, **2012**, 39(17), L17703.

23. Bhandari, S.; A. Kalra,; Tamaddun, K. and Ahmad, S. Relationship between Ocean-Atmospheric Climate Variables and Regional Streamflow of the Conterminous United States. *Hydrology*, **2018**, 5(2), 30.
24. Bradley, R. S.; Diaz, H. F.; Kiladis, G. N. and Eischeid, J. K. ENSO signal in continental temperature and precipitation records. *Nature*, **1987**, 327, 497-501.
25. Halpert, M. S. and Ropelewski, C. F. Surface temperature patterns associated with the Southern Oscillation. *Journal of Climate*, **1992**, 5(6), 577-593.
26. Ropelewski, C. F. and Halpert, M. S. Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Monthly Weather Review*, **1987**, 115(8), 1606-1626.
27. Ropelewski, C. F. and Halpert, M. S. Precipitation patterns associated with the high index phase of the Southern Oscillation. *Journal of Climate*, **1989**, 2(3), 268-284.
28. Ropelewski, C. F. and Halpert, M. S. Quantifying Southern Oscillation?precipitation relationships. *Journal of Climate*, **1989**, 9(5), 1043-1059.
29. Chen, D.; Cane, M. A. El Niño prediction and predictability. *Journal of Computational Physics*, **2008**, 227 (7), 3625-3640.
30. Clarke, A. J. El Niño physics and El Niño predictability. *Annual Review of Marine Science*, **2014**, 6 (1), 79-99.
31. Dijkstra, H. A.; Petersik, P.; Hernández-Garcia, E.; López, C. The application of machine learning techniques to improve El Niño prediction skill. *Frontiers of Physics*, **2019**, 7, 1-13.
32. Ham, Y. G.; Kim, J. H.; Luo, J. J. Deep Learning for Multiyear ENSO Forecasts. *Nature*, **2019**, 573 (7775), 568-572.
33. Tamaddun, K. A.; Kalra, A.; Bernardez, M. and Ahmad, S. Effects of ENSO on Temperature, Precipitation, and Potential Evapotranspiration of North India's Monsoon: An Analysis of Trend and Entropy. *Water*, **2019**, 11(2), 189.
34. Islam, M. N.; Sivakumar, B. Characterization and prediction of runoff dynamics: a nonlinear dynamical view. *Adv. Water Resour.*, **2002**, 25, 179-190.
35. Jayawardena, A. W.; Lai, F. Analysis and prediction of chaos in rainfall and stream flow time series. *J. Hydrol.*, **1994**, 153, 23-52.
36. Porporato, A.; Ridolfi, L. Nonlinear analysis of river flow time sequences. *Water Resour. Res.*, **1997**, 33(6), 1353-1367.
37. Liu, Q.; Islam, S.; Rodriguez-Iturbe, I.; Le, Y. Phase-space analysis of daily streamflow: characterization and prediction. *Adv. Water Resour.*, **1998**, 21, 463-475.
38. Dhanya, C. T.; Kumar, Nagesh D. Multivariate nonlinear ensemble prediction of daily chaotic rainfall with climate inputs. *Journal of Hydrology*, **2010**, 403(3-4), 292-306.
39. Lu, Z. Q.; Berliner, L. M. Markov switching time series models with application to a daily runoff series. *Water Resour. Res.*, **1999**, 35(2), 523-534.
40. Lampinen, J.; Vehtari, A. Bayesian approach for neural networks?review and case studies. *Neural Networks*, **2001**, 14, 257-274.
41. Shinohara, Y.; Kumagai, T.; Otsuki, K.; Kume, A.; Wada, N. Impact of climate change on runoff from a mid-latitude mountainous catchment in central Japan. *Hydrological Processes*, **2009**, 23(10), 1418-1429.
42. Halff, A. H.; Halff, H. M.; Azmoodeh, M. Predicting runoff from rainfall using neural networks. *Proceedings of Engineering Hydrology, Am. Soc. of Civ. Eng.*, **1993**, 23(10), 760-765.
43. Kabir, S.; Patidar, S.; Pender, G. Investigating capabilities of machine learning techniques in forecasting stream flow. *Proceedings of the Institution of Civil Engineers-Water Management*, **2020**, 173(2), 69-86.
44. Frigessi, A.; Haug, O.; Rue, H. A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes*, **2002**, 5, 219-235.
45. Embrechts, P.; Kluppelberg, C.; Mikosch, T. *Modelling Extremal Events*; Springer: Berlin, Heidelberg, 1997.
46. Carreau, J.; Naveau, P.; Sauquet, E. A statistical rainfall?runoff mixture model with heavy?tailed components. *Water Resources Research*, **2009**, 45(10).
47. Chavez-Demoulin, V.; Davison, A. C. Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society, Applied Statistics, Series C*, **2005**, 54(1), 207-222.
48. Coulibaly, P.; Bobée, B.; Anctil, F. Improving extreme hydrologic events forecasting using a new criterion for artificial neural network selection. *Hydrol. Processes*, **2001**, 15, 1533-1536.
49. Comte, L.; Buisson, L.; Daufrèsne, M. and Grenouillet, G. Climate?induced changes in the distribution of freshwater fish: observed and predicted trends. *Freshwater ecology*, **2013**, 58(4), 625-639.
50. Hulme, M. Attributing weather extremes to ?climate change?: A review. *Progress in Physical Geography: Earth and Environment*, **2014**, 38(4), 499-511.
51. Thompson, R. M.; Beardall, J.; Beringer, J.; Grace, M. and Sardina, P. Means and extremes: building variability into community level climate change experiments. *Ecology Letters*, **2013**, 16(6), 799-806.
52. Woodward, G.; Bonada, N.; Brown, L. E.; Death, R. G.; Durance, I.; Gray, C.; Hladysz, S.; Ledger, M. E.; Milner, A. M.; Ormerod, S. J.; Thompson, R. M. and Pawar, S. The effects of climatic fluctuations and extreme events on running water ecosystems. *Philosophical Transactions of the Royal Society B (Biological Sciences)*, **2016**, 371(1694).
53. Abrahams, C.; Brown, L.; Dale, K.; Edwards, F.; Jeffries, M. J.; Klaar, M.; Ledger, M. E.; May, L.; Milner, A. M.; Murphy, J. A. and Robertson, W. G. The impact of extreme events on freshwater ecosystems. In *Ecological issues special publication*; British Ecological Society: London, 2013.

54. Adebayo, A. J.; Soundharajan, B. S.; Ojha, C.S.P.; Remesan, R. Effect of hedging-integrated rule curves on the performance of the Pong reservoir (India) during scenario-neutral climate change perturbations. *Water resources management, Springer*, **2016**, 30(2), 445–470.
55. Soundharajan, B. S.; Adebayo, Adeloye; Remesan, Renji; Ojha, Chandra S.P. Simulating the performance of the Pong Reservoir in India under climate change perturbations. *Dooge-Nash International Symposium* **2014**.
56. Ncube, S.; Beevers, L.; Adeloye, A.; Visset, A. Assessment of freshwater ecosystem services in the Beas River Basin, Himalayas region, India. In *Proceedings of the International Association of Hydrological Sciences; 8th International Symposium on Integrated Water Resources Management 2018*, Beijing, China, 13-15 Jun 2018.
57. Prasad, V. H. and Roy, P. S. Estimation of Snowmelt Runoff in Beas Basin, India. *Geocarto International*, **2005**, 20(2), 41–47.
58. Patidar, S.; Jenkins, D. P.; Peacock, A. and McCallum, P. A hybrid system of data-driven approaches for simulating residential energy demand profiles. *Journal of Building Performance Simulation*, **2021**, 14(3), 277-302.
59. Cleveland, R. B.; Cleveland, W. S.; McRae, J. E.; Terpenning, I. STL: a seasonal-trend decomposition. *Journal of official statistics* **1990**, 6(1), 3–73.
60. Huang, Y.; Schmit, F. G.; Lu, Z. and Liu, Y. Analysis of daily river flow fluctuations using Empirical Mode Decomposition and arbitrary order Hilbert spectral analysis. *Journal of Hydrology*, **2009**, 373(1–2), 103–111.
61. Agana, N. A. and Homaifar, A. EMD-Based Predictive Deep Belief Network for Time Series Prediction: An Application to Drought Forecasting. *Journal of Hydrology*, **2018**, 5(1), 18.
62. Pender, D.; Patidar, S.; Pender, G.; Haynes, H. Stochastic simulation of daily streamflow sequences using a hidden Markov model. *Hydrology Research* **2015**, 47(1), 75-88.
63. Metcalfe, A. V.; Cowpertwait, P. S. In *Introductory time series with R*; Springer: New York, NY, 2009.
64. United States Climate Prediction Center, Historical El Niño/ La Niña episodes (1950 - present), Maryland , USA, 2019. Available online: https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php (accessed on 12 May 2021).
65. The R Project for Statistical Computing. Available online: <https://www.r-project.org/> (accessed on 03 Feb 2021).
66. Jain, S. K.; Ajanta, G.; Saraf, A. K. Assessment of snowmelt runoff using remote sensing and effect of climate change on runoff. *Water resources management, Springer*, **2010**, 24(9), 1763–1777.
67. Krishna, B.; Satyaji Rao, Y. R.; Nayak, P. C. Time Series Modeling of River Flow Using Wavelet Neural Networks. *Journal of Water Resource and Protection*, **2011**, 3(1), Article ID: 3778.

99 7. Appendix

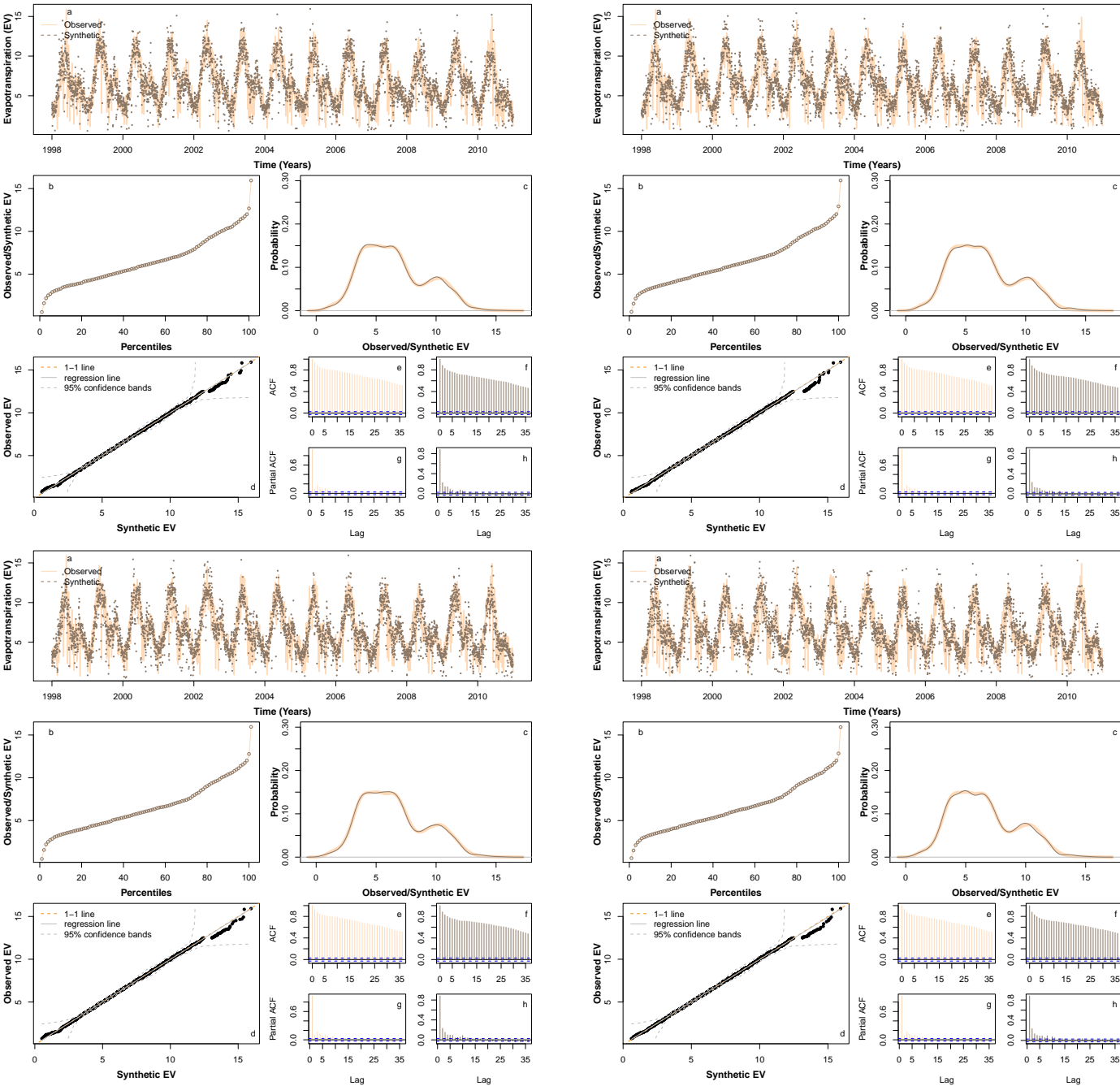


Figure A1. Comparison of four sample synthetic EV time series for 1998-2010. See Fig 6 for individual figure details.

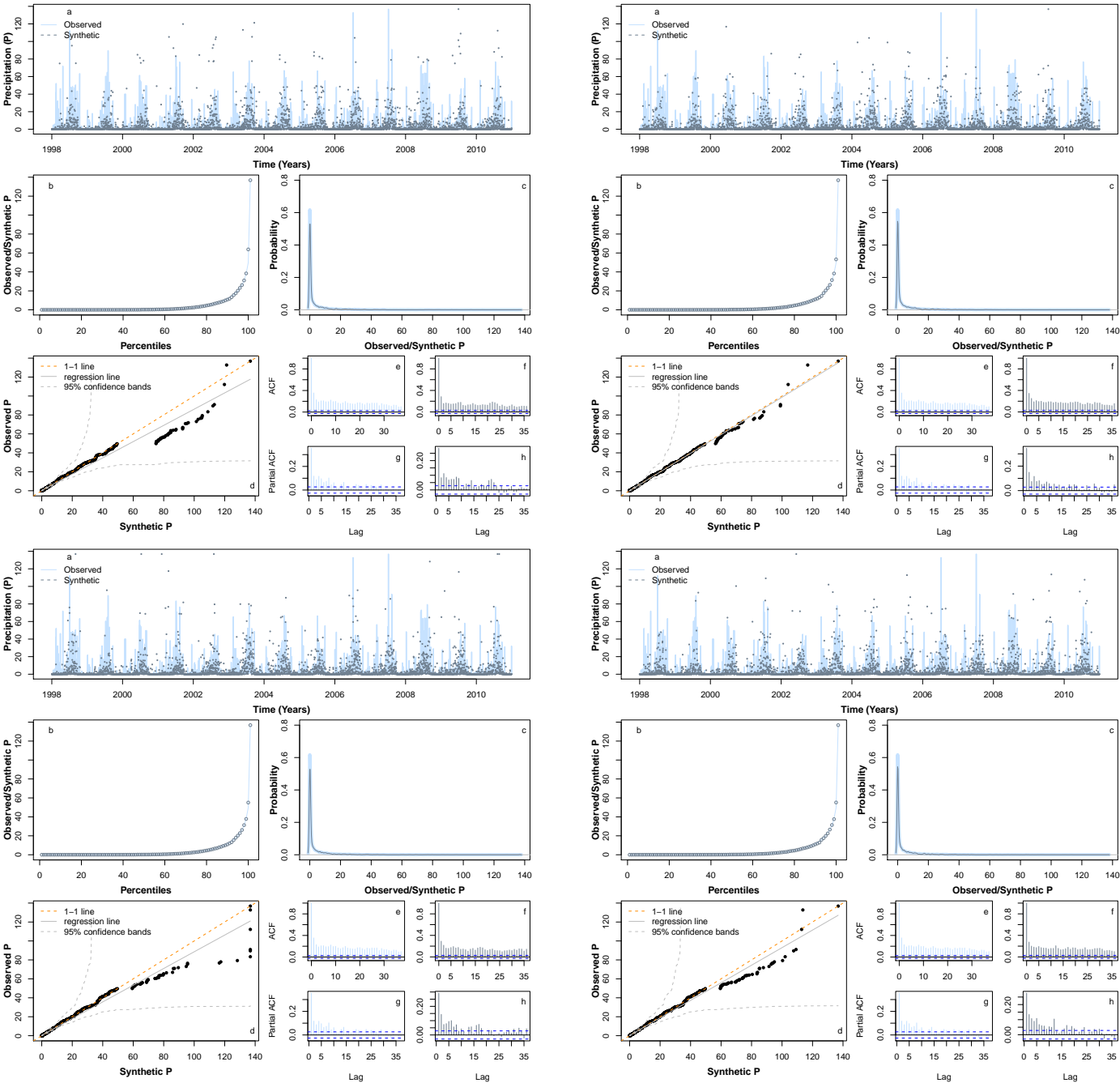


Figure A2. Comparison of four sample synthetic P time series for 1998-2010. See Fig 7 for individual figure details.

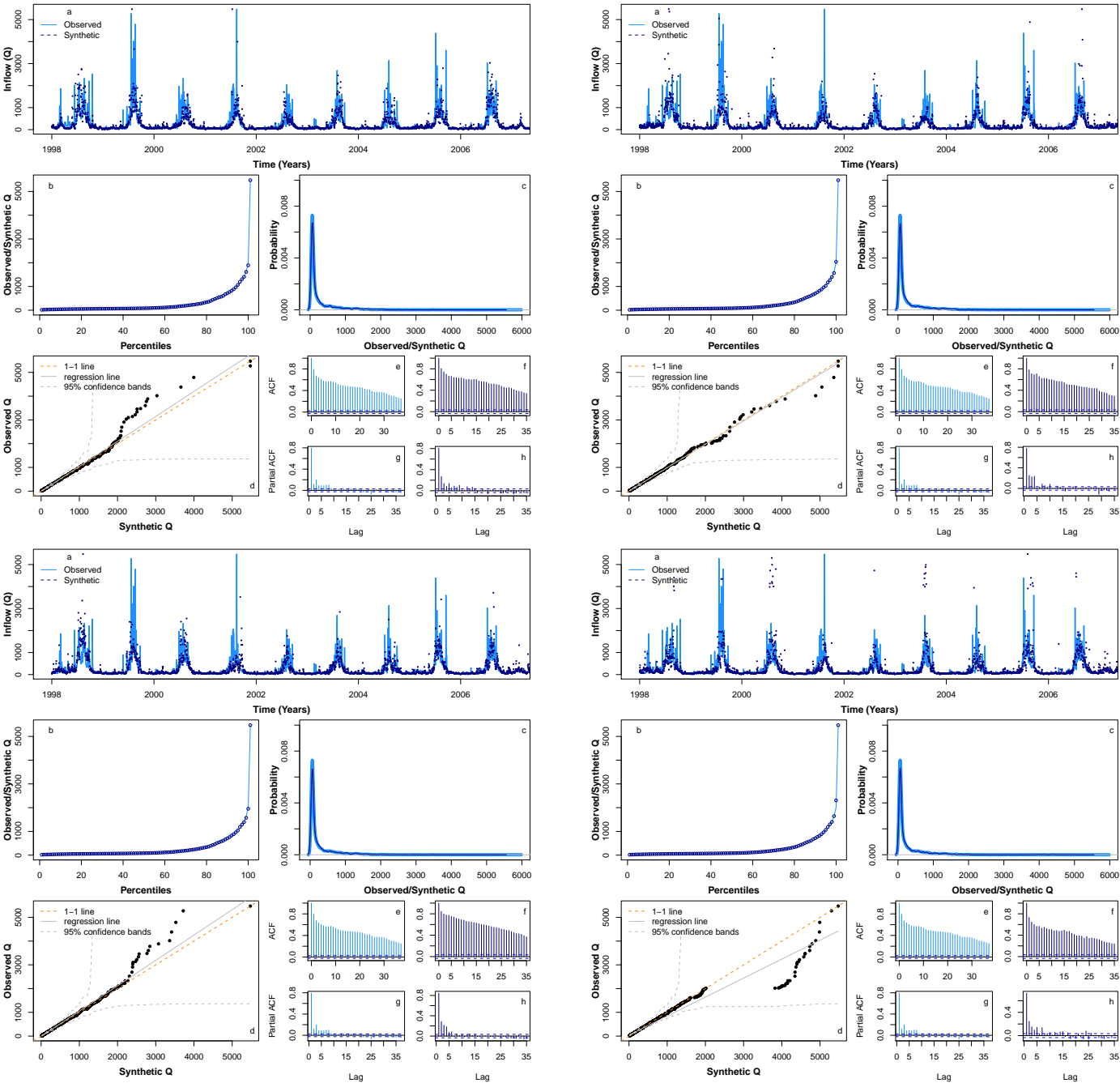


Figure A3. Comparison of four sample synthetic Q time series for 1998-2006. See Fig 8 for individual figure details.