A Dynamic Modelling Strategy for Bayesian Computer Model Emulation

Fei Liu[*] and Mike West[†]

Abstract. Computer model evaluation studies build statistical models of deterministic simulation-based predictions of field data to then assess and criticize the computer model and suggest refinements. Computer models are often expensive computationally: statistical models that adequately emulate their key features can be very much more efficient. Gaussian process models are often used as emulators, but the resulting computations lack the ability to scale to higher-dimensional, time-dependent or functional outputs. For some such problems, especially for contexts of time series outputs, building emulators using dynamic linear models provides a computationally attractive alternative as well as a flexible modelling approach capable of emulating a broad range of stochastic structures underlying the input-output simulations. We describe this here, combining Bayesian multivariate dynamic linear models with Gaussian process modelling in an effective manner, and illustrate the approach with data from a hydrological simulation model. The general strategy will be useful for other computer model evaluation studies with time series or functional outputs.

Keywords: Computer model emulation; Dynamic linear model; Forwarding filtering, backward sampling; Gaussian process; Markov chain Monte Carlo; Time-Varying Autoregression.

1 Introduction

1.1 Background and Motivation

Computationally intensive simulation models, or *simulators*, are increasingly used in research on complex systems in science and technology (Kennedy and O'Hagan 2001). Computer model evaluation studies emulate simulator-based input-output relationships using flexible statistical models. Often applied to deterministic simulators, statistical emulators are typically far less computationally expensive than the model simulators of interest. This enables their use as proxies for more extensive study of input-output relationships, as well as to assess, criticize and suggest statistical refinements to the simulator based on analyses of simulator predictions of field data (e.g. Higdon et al. 2004a,b; Goldstein and Rougier 2003; Fuentes et al. 2003; Goldstein and Rougier 2005; Rougier 2008b). Recent applications range from modelling traffic flows (Paulo et al. 2005) to analyzing climate sensitivity (Rougier et al. 2008) with a range of variants of statistical methods (Santner et al. 2003; Bayarri et al. 2005, 2007a,b).

Following Sacks et al. (1989), statistical emulators of computer model simulators

---

[*]Department of Statistics, University of Missouri, Columbia, MO, `liufei@missouri.edu`
[†]Department of Statistical Science, Duke University, Durham, NC, `mw@stat.duke.edu`

are commonly based on flexible Gaussian process (GASP) methodology (e.g. Currin et al. 1991; Welch et al. 1992; Morris et al. 1993; Craig et al. 2001; Kennedy et al. 2002; Rougier 2008a). GASP is justifiably popular in view of its provision of a broad, flexible nonparametric model framework capable of adapting to complex patterns of input-output relationships. However, GASP methodology does not scale easily and is simply computationally infeasible in problems with high-dimensional, time-dependent or functional outputs (Higdon et al. 2008a). This is particularly challenging in contexts with time series outputs, such as in our motivating application contexts and the example in Section 3 of the *logSPM* hydrological model (Kuczera et al. 2006).

With this interest in extending GASP modelling to problems with long time series outputs, we have combined Bayesian multivariate dynamic linear models (DLMs) with Gaussian processes. Specifically, we create multivariate time-varying autoregressive models – a special class of DLMs – in which the stochastic innovations are Gaussian processes over computer model input space. Methodologically, this provides flexibility in representing complex patterns of stochastic structure in outputs, relevant in our application as well as in other areas (e.g. Sanso et al. 2008; Tebaldi and Sanso 2008; Sanso et al. 2007). Computationally, the approach is tractable and scalable as posterior inference via Monte Carlo Markov Chain method efficiently scales based on the forward filtering, backward sampling algorithm (Carter and Kohn 1994; Frühwirth-Schnatter 1994); see also West and Harrison (1997, section 15.2.3).

Section 2 outlines GASP modelling and then describes the novel DLM-GASP synthesis and resulting new class of statistical emulators, with details of MCMC computations for model fitting. Section 3 describes an example of the logSPM hydrological simulator and problem context, and exemplifies the approach in an analysis of the logSPM data. Summary comments appear in Section 4 and supporting technical material in an appendix.

2 DLM-GASP Emulators

2.1 Gaussian Process Models

Begin with a computer model that runs with $d-$dimensional inputs $\boldsymbol{z} = (z_1, \ldots, z_d)'$ and generates univariate outputs $y(\boldsymbol{z})$. GASP uses a Gaussian process prior,

$$y(\cdot) \sim \mathrm{GP}\left(\mu, c(\cdot, \cdot)/\lambda\right), \tag{1}$$

where μ, λ and $c(\cdot, \cdot)$ are the mean, precision and correlation function respectively. The latter often takes the specific, separable form

$$c(\boldsymbol{z}, \boldsymbol{z}^*) = \exp(-\sum_{i=1}^{d} \beta_i \mid z_i - z_i^* \mid^{\alpha_i}) \tag{2}$$

for any two inputs $\boldsymbol{z}, \boldsymbol{z}^*$; the *range* parameters β_i control decay of correlation with respect to distance in individual input dimensions, while the *roughness* parameters α_i control smoothness of the output response.

Write $\boldsymbol{y}_{1:n} = (y(\boldsymbol{z}_1), \ldots, y(\boldsymbol{z}_n))'$ for the vector of outputs of the computer model runs at inputs $\boldsymbol{z}_{1:n} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n\}$. At any query input $\boldsymbol{z}$, due to the Gaussian process assumption, the joint distribution of $(\boldsymbol{y}_{1:n}, y(\boldsymbol{z}))$ is a multivariate normal distribution. Therefore, we can obtain the posterior predictive distribution for $y(\boldsymbol{z})$ as $(y(\boldsymbol{z}) \mid \boldsymbol{z}_{1:n}) \sim \mathrm{N}(\mu_{n+1}(\boldsymbol{z}), v_{n+1}(\boldsymbol{z}))$ where

$$\mu_{n+1}(\boldsymbol{z}) = \mu + \boldsymbol{\gamma}'_{n+1}\boldsymbol{\Gamma}_n^{-1}(\boldsymbol{y}_{1:n} - \mu) \quad \text{and} \quad v_{n+1}(\boldsymbol{z}) = (1 - \boldsymbol{\gamma}'_{n+1}\boldsymbol{\Gamma}_n^{-1}\boldsymbol{\gamma}_{n+1})/\lambda$$

where $\boldsymbol{\gamma}'_{n+1} = (c(\boldsymbol{z}, \boldsymbol{z}_1), \ldots, c(\boldsymbol{z}, \boldsymbol{z}_n))$ and $\boldsymbol{\Gamma}_n = (c(\boldsymbol{z}_i, \boldsymbol{z}_j))_{i,j=1,\ldots,n}$, giving the GASP prediction for the output at any query input.

Dealing with outputs that are time series extends the above to series of $y(\boldsymbol{z})$ values over a time axis, for each input $\boldsymbol{z}$. This could, of course, be done with spatial or other axes too, but we use the time series anchor context here. If outputs are time series, we could simply extend the above to include time t as an input so that the computer model outputs are $y(\boldsymbol{z}, t)$. Viewing t as an extension of the $\boldsymbol{z}$ input vector, this leads to the direct extension of the GASP model above with an additional term for the time dimension. However, this would massively increase the input dimension and computations under the extended GP prior become infeasible with other than trivial length of the time series. Our new DLM-GASP synthesis provides an alternative, tractable approach.

2.2 A DLM-GASP Statistical Model

By way of notation, define the following:

- the training data inputs $\boldsymbol{z}_{1:n} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n\}$, where $\boldsymbol{z}_i$ is a $d-$dimensional input vector, $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{id})'$;

- at each time t, the outputs in the $n-$vector $\boldsymbol{y}_t = (y(\boldsymbol{z}_1, t), \ldots, y(\boldsymbol{z}_n, t))'$;

- at any input $\boldsymbol{z}$, the output time series values up to any chosen time t is the $t-$vector $\boldsymbol{y}_{1:t}(\boldsymbol{z}) = (y(\boldsymbol{z}, 1), \ldots, y(\boldsymbol{z}, t))'$;

- all outputs up to time t are denoted by $\boldsymbol{y}_{1:t}(\boldsymbol{z}_{1:n}) = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t\}$.

Consider the $n-$vector time series $\boldsymbol{y}_t$, $t = 1, \ldots, T$ over a period up to time T. A flexible model for what may be quite intricate inter-dependencies among the elements of $\boldsymbol{y}_t$ over time, including non-stationary, time-varying stochastic structure, is provided by the class of *time-varying autoregressive models* (TVAR), a special subclass of multivariate DLMs (Prado and West 1997; West and Harrison 1997; Aguilar et al. 1999; Prado et al. 2001). To reflect the input space context, we now extend TVAR models to be driven by stochastic innovations that are related over input space via a Gaussian process model. The new DML-GASP model is as follows. For each $\boldsymbol{z}$, we have a TVAR model

$$y(\boldsymbol{z}, t) \quad = \quad \sum_{j=1}^{p} \phi_{t,j} y(\boldsymbol{z}, t-j) + \epsilon_t(\boldsymbol{z}), \tag{3}$$

where the lag p is specified, the autoregressive parameters $\boldsymbol{\phi}_t = (\phi_{t,1}, \ldots, \phi_{t,p})'$ may vary over time, and $\epsilon_t(\boldsymbol{z}) \sim N(0, v_t)$ independently over time. Across series, we link via a GASP model having

$$\mathbb{C}\mathrm{ov}\left(\epsilon_t(\boldsymbol{z}), \epsilon_t(\boldsymbol{z}^*)\right) = v_t c(\boldsymbol{z}, \boldsymbol{z}^*),$$

with $c(\boldsymbol{z}, \boldsymbol{z}^*)$ from equation (2), assuming the same correlation functions $c(\boldsymbol{z}, \boldsymbol{z}^*)$ for all t while allow the variances v_t to change. With $\boldsymbol{\epsilon}_t = (\epsilon_t(\boldsymbol{z}_1), \ldots, \epsilon_t(\boldsymbol{z}_n))'$, we can write these models over all inputs as

$$\boldsymbol{y}_t = \boldsymbol{F}_t' \boldsymbol{\phi}_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathrm{N}(0, v_t \boldsymbol{\Sigma}) \tag{4}$$

where

$$\boldsymbol{F}_t' = \begin{pmatrix} y(\boldsymbol{z}_1, t-1) & \ldots & y(\boldsymbol{z}_1, t-p) \\ \vdots & \ddots & \vdots \\ y(\boldsymbol{z}_n, t-1) & \ldots & y(\boldsymbol{z}_n, t-p) \end{pmatrix}$$

and where $\boldsymbol{\Sigma}$ is the n by n correlation matrix with (i,j) element $c(\boldsymbol{z}_i, \boldsymbol{z}_j)$.

To reflect the fact that the computer model may describe some non-stationary process, we allow the auto-regressive coefficients, $\boldsymbol{\phi}_t = (\phi_{t,1}, \ldots, \phi_{t,p})'$, to vary over time. In particular, we assume the AR coefficients follow a random walk through time,

$$\boldsymbol{\phi}_t = \boldsymbol{\phi}_{t-1} + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \sim \mathrm{N}(0, \boldsymbol{W}_t), \tag{5}$$

where $\boldsymbol{\phi}_0$, v_t and $\boldsymbol{W}_t$ are parameters with subsidiary models and prior distributions to be specified. In our analysis, we fix roughness parameters at $\alpha_i = 2$ throughout, and treat the range parameters $\beta_{1:d}$ as uncertain and also to be estimated.

The model equations (4) and (5) define a multivariate DLM, and we then have access to the full theory of dynamic models (West and Harrison 1997) including the forward filtering analysis and the related backward sampling algorithm that are critical to efficient computation. Some key technical details are given in Appendix 2 here.

2.3   DLM-GASP as an Emulator of GASP

Before proceeding to analysis, we note as an aside that the DLM-GASP model can also be viewed as a flexible, non-stationary approximation to an alternative GASP model that incorporates time into the input variable. In the latter framework, suppose the correlation function to have the separable form $c((\boldsymbol{z}, t), (\boldsymbol{z}^*, t^*)) = c(\boldsymbol{z}, \boldsymbol{z}^*) c_{time}(|t - t^*|)$ where $c_{time}(\cdot)$ is an additional correlation function on the time dimension. Over times $t-1, \ldots, T$, write $\boldsymbol{\Sigma}_{time}$ for the $T \times T$ matrix with (t, s) element $c_{time}(|t - s|)$. Then the GASP model on the extended $(\boldsymbol{z}, t)$ inputs has

$$\mathrm{vec}(\boldsymbol{y}_{1:T}(\boldsymbol{z}_{1:n})) \sim \mathrm{N}_{nT}\left(\mu \times \boldsymbol{1}_{nT}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}_{time}/\lambda\right). \tag{6}$$

Assuming that the correlation $c_{time}(\cdot)$ becomes negligible after p lags, then

$$p(\boldsymbol{y}_t, \boldsymbol{y}_{t-1}, \ldots, \boldsymbol{y}_1 \mid \boldsymbol{\theta}) \approx p\left(\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}, \ldots, \boldsymbol{y}_{t-p+1}, \boldsymbol{\theta}\right)$$

for $t > p$. Using the properties of the Kronecker product listed in Appendix 1, we then obtain the approximations

$$\mathrm{E}\left(\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}, \ldots, \boldsymbol{y}_{t-p}, \boldsymbol{\theta}\right) = \boldsymbol{F}_t' \boldsymbol{\phi} \quad \text{with} \quad \boldsymbol{\phi} = \tilde{\boldsymbol{\Sigma}}_p^{-1} \boldsymbol{\rho}_{t,t-1:t-p}, \tag{7}$$

and

$$\mathrm{V}(\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}, \ldots, \boldsymbol{y}_{t-p}, \boldsymbol{\theta}) = v\boldsymbol{\Sigma} \quad \text{with} \quad v\lambda = 1 - \boldsymbol{\rho}_{t,t-1:t-p}' \tilde{\boldsymbol{\Sigma}}_p^{-1} \boldsymbol{\rho}_{t,t-1:t-p}, \tag{8}$$

where $\boldsymbol{\rho}_{t,r:s} = (c_{time}(|t-r|), \ldots, c_{time}(|t-s|))'$ and $\tilde{\boldsymbol{\Sigma}}_p$ is a $p \times p$ matrix with (i,j) element $c_{time}(|i-j|)$. This is a multivariate normal DLM with constant AR coefficients and innovations variance. Hence we see the TVAR model can be interpreted as approximating this structure but allowing more flexibility through time-varying AR coefficients and innovations variance. Alternatively, we can just view the DLM-GASP as a richer model specification in its own right, allowing that flexibility and generating richer possibilities for correlation structures in the time dimension.

2.4 Model Completion and Prior Specification

Model completion requires specifications for the quantities $\boldsymbol{\phi}_0$, $\{v_t, \boldsymbol{W}_t, t = 1, \ldots, T\}$ and $\beta_{1:d} = (\{\beta_i, i = 1, \ldots, d\}$. This follows standard specifications for all DLM parameters with a conjugate normal/inverse-gamma prior on $(\boldsymbol{\phi}_0, v_0 | D_0)$ and the use of variance discounting to define models for the v_t sequence and specific values for the $\boldsymbol{W}_t$ sequence; See West and Harrison (1997) and the Appendix 2 here.

A key component of the analysis is learning on the range parameters $\beta_{1:d}$ for which we use the half-range correlation idea (Higdon et al. 2008b). Suppose the design inputs have been standardized. Then, for each $i = 1, \ldots, d$, the half-range $\rho_i = \exp(-\beta_i/4)$ is the correlation between $\epsilon_t(\boldsymbol{z})$ and $\epsilon_t(\boldsymbol{z}^*)$ when input $\boldsymbol{z}$ and $\boldsymbol{z}^*$ differ only in the ith component with a difference of 0.5. Following that approach, we use independent $\mathrm{Ga}(1, 0.1)$ priors on ρ_i, placing substantial mass near $\rho_i = 1$ to reflect the expectation that only a subset of the inputs are influential in the computer model. The resulting prior for $\beta_{1:d}$ is then

$$p(\beta_{1:d}) \propto \prod_{i=1}^{d} \left(1 - \exp(-\beta_i/4)\right)^{0.9} \exp(-\beta_i/4). \tag{9}$$

2.5 DLM-GASP Prediction for Simulator Emulation

At any query input $\boldsymbol{z}$, we can predict the output of the computer model as follows based on assumed values of all parameters that will be generated from posterior simulation. Given the realized values of $\epsilon_t(\boldsymbol{z}_i) = y(\boldsymbol{z}_i, t) - \sum_{j=1}^{p} \phi_{t,j} y(\boldsymbol{z}_i, t-j)$ the predictive distribution for $y(\boldsymbol{z}, t)$ conditional on all model parameters, on the output training data $y_{1:T}(\boldsymbol{z}_{1:n})$, and on the "previous" time series values $\boldsymbol{y}_{t-1:t-p}(\boldsymbol{z})$ is normal with mean

$$\boldsymbol{\mu}_t(\boldsymbol{z}) = \sum_{j=1}^{p} \phi_{t,j} y(\boldsymbol{z}, t-j) + v_t^{-1} \boldsymbol{\gamma}_{\boldsymbol{z}}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon}_t$$

and variance $\sigma_t^2(\boldsymbol{z}) = v_t(1 - \boldsymbol{\gamma}_{\boldsymbol{z}}^t \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma}_{\boldsymbol{z}})$, where $\boldsymbol{\gamma}_{\boldsymbol{z}} = (c_1(\boldsymbol{z}, \boldsymbol{z}_1), \ldots, c_1(\boldsymbol{z}, \boldsymbol{z}_n))'$ and $\boldsymbol{\epsilon}_t = (\epsilon_t(\boldsymbol{z}_1), \ldots, \epsilon_t(\boldsymbol{z}_n))'$. These distributions define the emulator that results from our DLM-GASP model. It is apparent that the emulator interpolates the data with $\mu_t(\boldsymbol{z}) = y_t(\boldsymbol{z})$ and $\sigma_t^2(\boldsymbol{z}) = 0$ at the realized points $\boldsymbol{z}_i$; this is a typical feature of the emulators for deterministic computer models that carries over to this new DLM-GASP framework.

2.6   Posterior MCMC Sampling

We use MCMC for simulation of the joint posterior of the unknowns $\boldsymbol{\phi}_{1:T} = \{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_T\}$, $\boldsymbol{v}_{1:T} = (v_1, \ldots v_T)$ and $\beta_{1:d}$. The joint posterior distribution can be written as

$$p(\boldsymbol{\phi}_{1:T}, v_{1:T}, \beta_{1:d} \mid \boldsymbol{y}_{1:T}) \propto p(y_{1:T}|\boldsymbol{\phi}_{1:T}, v_{1:T}, \beta_{1:d})p(\beta_{1:d}) \prod_{t=1}^{T} p(\boldsymbol{\phi}_t, v_t \mid \boldsymbol{y}_{1:t-1}, \boldsymbol{\beta})$$

with conditional likelihood function proportional to

$$p(y_{1:T}|\boldsymbol{\phi}_{1:T}, v_{1:T}, \beta_{1:d}) \propto |\boldsymbol{\Sigma}|^{-(T-p)/2} \exp(-Q/2) \prod_{t=p+1}^{T} v_t^{-n/2}$$

where

$$Q = \sum_{t=p+1}^{T} (\boldsymbol{y}_t - \boldsymbol{F}_t' \boldsymbol{\phi}_t)' \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_t - \boldsymbol{F}_t' \boldsymbol{\phi}_t)/v_t.$$

We use a block Gibbs sampler (Gelfand and Smith 1990) with a Metropolis-within-Gibbs step for $\beta_{1:d}$. The overall MCMC iterates between Gibbs draws from the complete conditional posterior distribution of $(\boldsymbol{\phi}_{1:T}, v_{1:T})$ and Metropolis-Hastings draws whose target distributions are the conditionals for $\beta_{1:d}$. Conditional on $\beta_{1:d}$, the first step here involves running the forward filtering, backward sampling computations to efficiently sample the full sequence of AR parameters and innovation variances; see Appendix 2. A key novelty is the Metropolis-Hastings component for $\beta_{1:d}$ targeting the conditional $p(\beta_{1:d} \mid \boldsymbol{y}_{1:T}, \boldsymbol{v}, \boldsymbol{\phi}_{1:T})$ at each Gibbs iterate. This requires careful choice of the proposal distributions and extensive exploration as discussed in Appendix 3.

Given the draws of the parameters, at any query input $\boldsymbol{z}$ we can then make draws from the emulator sequentially as described in Section 2.5. This generates repeat samples of outcomes $y_{1:T}(\boldsymbol{z})$ to summarize for posterior predictive inferences.

3   Example: LogSPM Hydrological Model

3.1   Background and Problem Context

The hydrological *saturated path models* (logSPMs) of Kuczera et al. (2006) are deterministic simulators of water balance dynamics at catchment scales; these models are

central to environmental models at these scales. The logSPM has 13 inputs: 2 time series inputs of external forces – rainfall and potential evapotranspiration; 3 initial conditions – initial water volume per unit area in the soil, in the ground water and in the river; and 8 coefficients describing water dynamics. The external forces and initial conditions are typically known and here we take them fixed so that we have an input dimension $d = 8$ for water dynamic coefficients z. The logSPM output is a river flow time series. See Reichert et al. (2008) for description of the partial differential equations of this computer model and data source.

We use the first data set in Reichert et al. (2008) as our training data, which consists of $n = 60$ computer model runs with inputs selected according to a 60-point Latin hypercube design (LHD); a 2-dimensional projection of the design after normalization is given in Figure 1. These input points reasonably fill the design space as is important to exploring the global characteristics of the simulator. The corresponding 60 output time series over $T = 1827$ time points outputs are shown in Figure 2.

We analyse the output data on the log scale (right frame in Figure 2), fitting the DLM-GASP model to the $n = 60$ training data inputs and output time series over $t = 1, \ldots, T$. An additional data set of another $n = 60$ runs is also available in Reichert et al. (2008). These runs are based on a 60-point LHD set of inputs that is independent from the designed inputs of the training data, i.e., the design allows different patterns from the training data. We use this data set as our validation data set.

3.2 DLM-GASP Emulation of LogSPM

Based on aspects of the known structure of the LogSPM model (Reichert et al. 2008) we specify $p = 2$ in the TVAR model component, with hyper-parameters defining the initial priors reflecting the known range and scale of outputs while being hedged with substantial uncertainty: $m_0 = 0$, $n_0 = 1$, $d_0 = 1$ and $C_0 = 10I_2$. Discount factors for the analysis exemplified here are $\delta_v = \delta_W = 0.9$ which allow some variation over time in AR parameters and innovations variance; repeat analyses with varying discounts confirm that the global patterns of variation over time in ϕ_t and v_t are evident across models, and the overall conclusions here are not sensitive to choices of discount factors δ_v and δ_W. The MCMC analysis was run for a total of $H = 200,000$ iterations generating draws $\Theta^{(h)}$, $(h = 1, \ldots, H)$, of all uncertain DLM-GASP quantities $\Theta = \{\phi_{1:T}, v_{1:T}, \beta_{1:d}\}$.

Figure 3 shows approximate posteriors for the range parameters $\beta_{1:d}$. That for β_3 is centered around 0.066, which corresponds to a half-range correlation around $\exp(-0.066/4)$, approximately equal to 0.98; in contrast, a β_8 near 2.8 corresponds to a half-range ≈ 0.50. The high correlation in the 3rd dimension and low correlation in the 8th dimension of input space suggest that a more efficient design should put fewer design points in the 3rd than in the 8th dimension. Furthermore, noticing that the smallest correlation in the 3rd direction is still very large, about $\exp(-0.066) \approx 0.94$, one may want to remove the 3rd input from the Gaussian process – that is, posterior inference is suggestive of model refinements and input variable *screening* and selection in a natural manner.

The model identifies time variation in the TVAR model parameters ϕ_{t1}, ϕ_{t2} and v_t via estimated trajectories summarized in Figure 4. Though the TVAR model seems consistent with *local* stationarity as a general rule, with values of $\boldsymbol{\phi}_t$ sitting mainly within the AR(2) stationary region for parameter values (West and Harrison 1997), there are substantial global non-stationarities evident: major stochastic changes in AR parameters and volatility changes in the innovations variance consistent with periods of apparent volatility and spiking behavior in the simulator outputs.

For any of the given validation input points $\boldsymbol{z}$, the DLM-GASP emulator generates draws from the posterior predictive distribution of future outputs from further model runs at $\boldsymbol{z}$. Applying the emulator to all the 60 inputs in the validation data set we obtain posterior predictive draws illustrated in Figure 6; this shows pointwise posterior predictive medians for all 60 inputs in the validation data. These are comparable, in terms of overall nature of patters over time, with the realized validation outputs from the simulator, also displayed. The emulator is defining and predicting global trends, appropriate levels of variability across the set of inputs, and, in particular, is capturing both timing and magnitudes of the peaks and troughs including the marked periods of bursts in volatility and spiking of the output processes.

For each validation run, we can compare the prediction given by the emulator with the true validation data by the mean squared error (MSE), defined as

$$\text{MSE} = \frac{1}{T} \sum_{t=1}^{T} \left(\hat{y}(\boldsymbol{z}, t) - y(\boldsymbol{z}, t) \right)^2 \ ,$$

where $\hat{y}(\boldsymbol{z}, t)$ is the posterior median of the posterior draws at time t, and $y(\boldsymbol{z}, t)$ the true validation data. The MSE can be potentially used as a measure to compare the performance of the present emulator with that of some other emulator, if both are applied to the same data set. For the 60 validation runs in the LogSPM data, the present emulator yields MSEs with mean and standard deviation 1.67 and 1.53, respectively.

We further investigate predictions of several individual runs, making comparisons of simulated predictions from the DLM-GASP emulator with the true outputs for validation sample inputs $\boldsymbol{z}_1$, $\boldsymbol{z}_{20}$, $\boldsymbol{z}_{40}$ and $\boldsymbol{z}_{60}$; see Figure 7. This provides a further, good and subjective view of the ability of the emulator to predict global as well as local characteristics of the simulator-based hydrological process.

4   Further Comments

The DLM-GASP approach is particularly suitable as a candidate emulator for computer models generating long time series outputs, and when there is evidence of dynamic variation in underlying trends as well as in stochastic volatility in the outputs. TVAR models are well-established as flexible, tractable and sometimes interpretable empirical models of what may be quite complex and erratic processes, and we believe this synthesis with GASP will be of interest in computer model evaluation and emulation studies. DLM-GASP provides a computationally feasible approach to simulator emulation even

with very long time series outputs, and will be of use in studies aiming to explore aspects of variation in time and over input space to improve understanding of the underlying model structure as well as in studies where prediction at new design points is of central interest. The flexible model framework will also be relatively easily adaptable to other issues related to the computer experiments, such as design and calibration. We also note the potential for more customized models that build additional structural components into the multivariate DLM, including the opportunity to bring in dynamic regression terms as functions of inputs if desired; the overall methodology will easily extend to accommodate that since dynamic regressions are also special cases of DLMs and can be combined trivially with TVAR model components (West and Harrison 1997).

Supplementary Material: Code and data

Further details and complete information needed to recapitulate the analyses reported are available at www.stat.missouri.edu/~liufei/DlmComputerWeb.zip. This includes the data and the Matlab code to conduct the analysis and a brief readme on use of the code.

References

Aguilar, O., Prado, R., Huerta, G., and West, M. (1999). "Bayesian inference on latent structure in time series (with discussion)." In Bernardo, J., Berger, J., Dawid, A., and Smith, A. (eds.), *Bayesian Statistics 6*, 3–26. Oxford University Press. 395

Bayarri, M., Berger, J., Garcia-Donato, G., Liu, F., Palomo, J., Paulo, R., Sacks, J., Walsh, D., Cafeo, J., and Parthasarathy, R. (2007a). "Computer Model Validation with Functional Outputs." *Annals of Statistics*, 35: 1874–1906. 393

Bayarri, M. J., Berger, J. O., Kennedy, M., Kottas, A., Paulo, R., Sacks, J., Cafeo, J. A., Lin, C. H., and Tu, J. (2005). "Bayesian Validation of a Computer Model for Vehicle Crashworthiness." Technical report, National Institute of Statistical Sciences, RTP, NC, USA. http://www.niss.org/technicalreports.html. 393

Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C.-H., and Tu, J. (2007b). "A Framework for Validation of Computer Models." *Technometrics*, 49(2): 138–154. 393

Carter, C. K. and Kohn, R. (1994). "On Gibbs sampling for state-space models." *Biometrika*, 81: 541–553. 394

Craig, P. S., Goldstein, M., Rougier, J., and Seheult, A. (2001). "Bayesian forecasting for complex systems using computer simulators." *Journal of the American Statistical Association*, 96(454): 717–729. 394

Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1991). "Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments." *Journal of the American Statistical Association*, 86: 953–963. 394

Frühwirth-Schnatter, S. (1994). "Data augmentation and dynamic linear models." *Journal of Time Series Analysis*, 15: 183–202.   394

Fuentes, M., Guttorp, P., and Challenor, P. (2003). "Statistical assessment of numerical models." *International Statistical Review*, 201–221.   393

Gelfand, A. E. and Smith, A. F. M. (1990). "Sampling based approaches to calculating marginal densities." *Journal of the American Statistics Association*, 85: 398–409. 398

Goldstein, M. and Rougier, J. C. (2003). "Calibrated Bayesian forecasting using large computer simulators." Technical report, University of Durham, UK, http://www.maths.dur.ac.uk/stats/physpred/papers/CalibratedBayesian.ps.   393

— (2005). "Probabilistic formulations for transferring inferences from mathematical models to physical systems." *SIAM Journal on Scientific Computing*, 26(2): 467–487.   393

Higdon, D., Gattiker, J., Williams, B., and Rightley, M. (2008a). "Computer Model Calibration Using High-Dimensional Output." *Journal of the American Statistical Association*, 103(482): 570–583.   394

Higdon, D., Kennedy, M., Cavendish, J., Cafeo, J., and Ryne, R. D. (2004a). "Combining field observations and simulations for calibration and prediction." *SIAM Journal of Scientic Computing*, 26: 448–466.   393

Higdon, D., Nakhleh, C., Gattiker, J., and Williams, B. (2008b). "A Bayesian calibration approach to the thermal Problem." *Computer Methods in Applied Mechanics and Engineering (CMAME)*, 197: 2431–2441.   397

Higdon, D., Williams, B., Moore, L., McKay, M., and Keller-McNulty, S. (2004b). "Uncertainty Quantification for Combining Experimental Data and Computer Simulations." Technical Report LA-UR 04-6562, Los Alamos National Laboratories, USA. 393

Kennedy, M. C. and O'Hagan, A. (2001). "Bayesian calibration of computer models (with discussion)." *Journal of the Royal Statistical Society B*, 63: 425–464.   393

Kennedy, M. C., O'Hagan, A., and Higgins, N. (2002). "Bayesian analysis of computer code outputs." In *Quantitative Methods for Current Environmental Issues. C. W. Anderson, V. Barnett, P. C. Chatwin, and A. H. El-Shaarawi (eds.)*, 227–243. Springer-Verlag: London.   394

Kuczera, G., Kavetski, D., Franks, S., and Thyer, M. (2006). "Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters." *Journal of Hydrology*, 331: 161–177.   394, 398

Morris, M. D., Mitchell, T. J., and Ylvisaker, D. (1993). "Bayesian design and analysis of computer experiments: Use of derivatives in surface prediction." *Technometrics*, 35: 243–255.   394

Paulo, R., Lin, J., Rouphail, N., and Sacks, J. (2005). "Calibrating and Validating Deterministic Traffic Models:Application to the HCM Control Delay at Signalized Intersections." *Transportation Research Record: Journal of the Transportation Research Board*, 1920: 95–105. 393

Prado, R., Huerta, G., and West, M. (2001). "Bayesian time-varying autoregressions: Theory, methods and applications." *Resenhas*, 4: 405–422. 395

Prado, R. and West, M. (1997). "Exploratory modelling of multiple non-stationary time series: Latent process structure and decompositions." In Gregoire, T., Brillinger, D., Diggle, P., Russek-Cohen, E., Warren, W., and Wolfinger, R. (eds.), *Modelling Longitudinal and Spatially Correlated Data*, 349–362. New York: Springer-Verlag. 395

Reichert, P., White, G., Bayarri, M., Pitman, E., and Santer, T. (2008). "Mechanism-based Emulation of Dynamic Simulators: Concept and Application in Hydrology." http://www.eawag.ch/kuerze/personen/homepages/reichert/index_EN. 399

Rougier, J. (2008a). "Efficient Emulators for Multivariate Deterministic Functions." *Journal of Computational and Graphical Statistics (In press)*. 394

— (2008b). "Formal Bayes Methods for Model Calibration with Uncertainty." In Beven, K. and Hall, J. (eds.), *Applied Uncertainty Analysis for Flood Risk Management (In press)*. Imperial College Press/World Scientific. Draft version available at http://www.maths.bris.ac.uk/~mazjcr/FRMbox2-4.pdf. 393

Rougier, J., Sexton, D., Murphy, J., and Stainforth, D. (2008). "Analysing the climate sensitivity of the HadSM3 climate model using ensembles from different but related experiments." Technical report, University of Bristol, UK. Available at http://www.maths.bris.ac.uk/~mazjcr/qump1.pdf. 393

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). "Design and analysis of computer experiments (C/R: p423-435)." *Statistical Science*, 4: 409–423. 393

Sanso, B., Forest, C., and Zantedeschi, D. (2007). "Statistical Calibration of Climate System Properties." Technical Report asm2007-06, Dept. of Applied Math & Statistics, University of California, Santa Cruz, http://www.soe.ucsc.edu/research/report?ID=476. 394

— (2008). "Inferring Climate System Properties Using a Computer Model." *Bayesian Analysis*, 3(1): 1–38. 394

Santner, T., Williams, B., and Notz, W. (2003). *The Design and Analysis of Computer Experiments*. New York: Springer-Verlag. 393

Tebaldi, C. and Sanso, B. (2008). "Joint Projections of Temperature and Precipitation Change from Multiple Climate Models: A Hierarchical Bayes Approach." *Journal of the Royal Statistical Society: Series A (In press)*. 394

Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J., and Morris, M. D. (1992). "Screening, predicting, and computer experiments." *Technometrics*, 34: 15–25. 394

West, M. and Harrison, P. (1997). *Bayesian Forecasting and Dynamic Models*. New York: Springer-Verlag, 2 edition. 394, 395, 396, 397, 400, 401, 406

Acknowledgments

1 Kronecker product

The Kronecker product of two matrices $A = (a_{ij})_{i,j}$ and $B = (b_{ij})_{i,j}$ is defined as,

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}.$$

It has the following properties.

1. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ if A and B are both invertible.

2. $|A \otimes B| = |A|^{d_2}|B|^{d_1}$, where d_1 and d_2 are the dimensions of A, B.

3. $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ if the dimensions are matched.
   If we assume $A = U_1 U_1'$, $B = U_2 U_2'$, then $(A \otimes B) = (U_1 \otimes U_2)(U_1 \otimes U_2)'$.

2 DLM and Forward Filtering, Backward Sampling

2.1 DLM Completion

Refer to the DLM representation of the TVAR model of equations (4) and (5) and let $D_t = \{\boldsymbol{y}_t, D_{t-1}\}$ be the information set at time t where D_0 is the initial information set. For this appendix material for notational clarity, we assume all other model parameters not mentioned, including the GASP parameters α_i, β_i, are conditioned upon and absorbed into D_0. The initial prior is $(\boldsymbol{\phi}_0|v_0, D_0) \sim \mathrm{N}\,(\boldsymbol{m}_0, v_0\boldsymbol{C}_0)$ and $(v_0^{-1}|D_0) \sim \mathrm{Ga}(n_0/2, d_0/2)$ with specified initial hyper-parameters. The models for the sequences v_t and $\boldsymbol{W}_t$ are based on standard variance discounting using a variance discount factor δ_v for v_t and one δ_W for $\boldsymbol{W}_t$, each lying in the unit interval and typically taking values very close to 1 to represent relative stability over time in stochastic changes in the $(v_t, \boldsymbol{W}_t)$ sequences. The nature and roles of the two variance discount factors (δ_v, δ_W) is clear in the resulting filtering equations summarized here.

2.2 Forward Filtering Computations

Initializing at $t = 1$, the forward filtering analysis sequences through $t = 1, 2, \ldots$ updating the one-step forecast and posterior distributions as follows, valid for all $t > 0$:

(a) The posterior for $(\boldsymbol{\phi}_{t-1}, v_{t-1}|D_{t-1})$ is the conjugate normal/inverse-gamma form with margins $\boldsymbol{\phi}_{t-1} \sim \mathrm{T}_{n_{t-1}}(\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1})$ and $v_{t-1}^{-1} \sim \mathrm{Ga}(n_{t-1}/2, d_{t-1}/2)$; we note that $s_{t-1} = d_{t-1}/n_{t-1}$ is an implied point estimate of v_{t-1}.

(b) The prior at for $(\boldsymbol{\phi}_t, v_t|D_{t-1})$ has the same form with $\boldsymbol{\phi}_t \sim \mathrm{T}_{\delta_v n_{t-1}}(\boldsymbol{m}_{t-1}, \boldsymbol{R}_t)$ and $v_t^{-1} \sim \mathrm{Ga}(\delta_v n_{t-1}/2, \delta_v d_{t-1}/2)$ where $\boldsymbol{R}_t = \boldsymbol{C}_{t-1} + \boldsymbol{W}_t$. The variance discounting model sets $\boldsymbol{W}_t = \boldsymbol{C}_{t-1}(1 - \delta_W)/\delta_W$ resulting in $\boldsymbol{R}_t = \boldsymbol{C}_{t-1}/\delta_W$. The effect of variance discounting is clear from these equations.

(c) The one-step forecast distribution for $(\boldsymbol{y}_t \mid D_{t-1})$ is $\boldsymbol{y}_t \sim \mathrm{T}_{\delta_v n_{t-1}}(\boldsymbol{f}_t, \boldsymbol{Q}_t)$ with $\boldsymbol{f}_t = \boldsymbol{F}_t' \boldsymbol{m}_{t-1}$ and $\boldsymbol{Q}_t = \boldsymbol{F}_t' \boldsymbol{R}_t \boldsymbol{F}_t + s_{t-1} \boldsymbol{\Sigma}$.

(d) The posterior for $(\boldsymbol{\phi}_t, v_t | D_t)$ is normal/inverse-gamma – as in part (a) following the update to time t. The margins are:

  – $v_t^{-1} \sim \mathrm{Ga}(n_t/2, d_t/2)$ with $n_t = \delta_v n_{t-1} + n$ and $d_t = \delta_v d_{t-1} + s_{t-1} \boldsymbol{e}_t' \boldsymbol{Q}_t^{-1} \boldsymbol{e}_t$ based on realized forecast error vector $\boldsymbol{e}_t = \boldsymbol{y}_t - \boldsymbol{F}_t' \boldsymbol{m}_{t-1}$. The updated point estimate of v_t is $s_t = d_t/n_t$.

  – $\boldsymbol{\phi}_t \sim \mathrm{T}_{n_t}(\boldsymbol{m}_t, \boldsymbol{C}_t)$ where $\boldsymbol{m}_t = \boldsymbol{m}_{t-1} + \boldsymbol{A}_t(\boldsymbol{y}_t - \boldsymbol{F}_t' \boldsymbol{m}_{t-1})$ and $\boldsymbol{C}_t = (\boldsymbol{W}_t - \boldsymbol{A}_t \boldsymbol{Q}_t \boldsymbol{A}_t') s_t/s_{t-1}$ where $\boldsymbol{A}_t = \boldsymbol{R}_t \boldsymbol{F}_t \boldsymbol{Q}_t^{-1}$.

These results follow from the theory of general multivariate DLMs (West and Harrison 1997).

2.3   Backward Sampling Computations

Having completed the forward filtering computations over times $t = 1, \ldots, T$, backward sampling generates a posterior draw for the full sequence of states and variances from the joint posterior $p(\boldsymbol{\phi}_{1:T}, v_{1:T} | D_T)$. This operates in in reverse time, using the general theory of DLMs that generate this computationally efficient algorithm for what can be a very high-dimensional full set of states when T is large. Backward sampling proceeds as follows.

• Initialize at $t = T$: draw v_T^{-1} from $\mathrm{Ga}(n_T/2, d_T/2)$ followed by $\boldsymbol{\phi}_T | v_T$ from $\mathrm{N}(\boldsymbol{m}_T, \boldsymbol{C}_T v_T/s_T)$.

• Set $n_T^* = n_T$, $d_T^* = d_T$, $m_t^* = m_T$ and $C_T^* = C_T$. Then, for $t = T-1, T-2, \ldots, 0$ in sequence, sample:

  – $v_t^{-1} \sim \mathrm{Ga}(n_t^*/2, d_t^*/2)$ where $d_t^* = n_t^* s_t^*$ with $n_t^* = (1 - \delta_v)n_t + \delta_v n_{t+1}^*$ and $1/s_t^* = (1 - \delta_v)/s_t + \delta_v/s_{t+1}^*$;

  – $\boldsymbol{\phi}_t | v_t \sim \mathrm{N}(\boldsymbol{m}_t^*, \boldsymbol{C}_t^* v_t/s_t^*)$ where $\boldsymbol{m}_t^* = (1 - \delta_W)\boldsymbol{m}_t + \delta_W \boldsymbol{m}_{t+1}^*$ and $\boldsymbol{C}_t^* = (1 - \delta_W)\boldsymbol{C}_t/s_t + \delta_W^2 \boldsymbol{C}_{t+1}^*/s_{t+1}^*$

3   MCMC algorithm in DLM-GASP

At each Gibbs iteration h, the overall MCMC steps through the sequence of simulations as follows:

(a) Sample $\beta_{1:d}^{(h)}$ from $p(\beta_{1:d} \mid \boldsymbol{y}_{1:T}, \boldsymbol{\phi}_{1:T}^{(h-1)}, v^{(h-1)})$ by the Metropolis-Hastings algorithm detailed below.

(b) Sample $(\boldsymbol{\phi}_{1:T}, v_{1:T} \mid \boldsymbol{y}_{1:T}, \beta_{1:d}^{(h)})$ using the DLM forward filtering, backward sampling algorithm.

The Metropolis-Hastings algorithm to draw $\beta_{1:d}^{(h)}$ works as follows. We first set $\eta_i = \log(\beta_i)$ and $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_d)$ and propose a new value for η_i as

$$\eta_i^* = \eta_i^{(h-1)} + \mathrm{N}(0, 0.1^2),$$

then setting $\beta_i^* = \eta_i^*$. The choice of the proposal distribution is customized to each application and requires experimentation with the step-width of the random walk proposal here, but the overall strategy is broadly relevant.

Denoting the correlation matrix at the proposal values by $\boldsymbol{\Sigma}_*$, calculate the likelihood function

$$|\boldsymbol{\Sigma}_*|^{-(T-p)/2} \exp(-Q^*/2) \quad \text{where} \quad Q^* = \sum_{t=p+1}^{T} \boldsymbol{e}_t' \boldsymbol{\Sigma}_*^{-1} \boldsymbol{e}_t / v_t.$$

With Jacobian

$$J(\beta_{1:d}) = \prod_{i=1}^{d} \frac{\mathrm{d}\eta_i}{\mathrm{d}\beta_i} = \prod_{i=1}^{d} \beta_i^{-1},$$

this gives the Metropolis-Hastings acceptance ratio

$$r = \max \left\{ \frac{L(\beta_{1:d}^*)\pi(\beta_{1:d}^*)J(\beta_{1:d}^{h-1})}{L(\beta_{1:d}^{(h-1)})\pi(\boldsymbol{\beta}^{(h-1)})J(\beta_{1:d}^*)}, 1 \right\}.$$

Set $\beta_{1:d}^{(h)} = \beta_{1:d}^*$ with probability r; otherwise, set $\beta_{1:d}^{(h)} = \beta_{1:d}^{(h-1)}$. In our example, the proposal distribution for $\beta_{1:d}$ leads to an acceptance rate of about 10%.
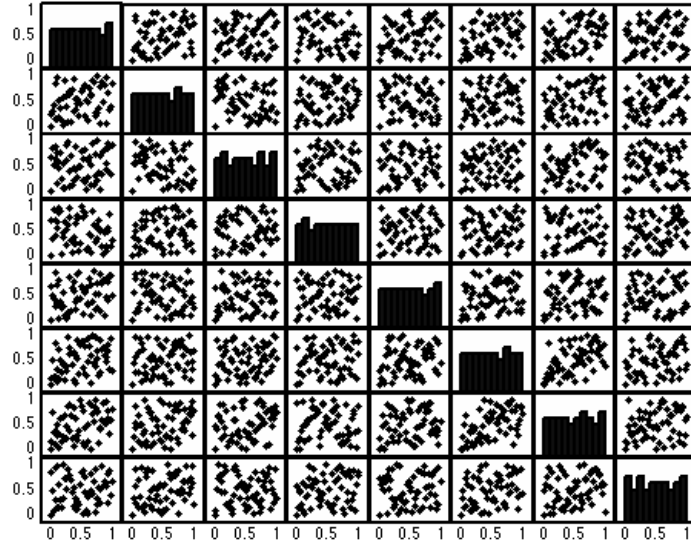
Figure 1: Pairwise scatter plots of normalized values of the $d = 8$ input parameters in the 60 runs of the logSPM model.
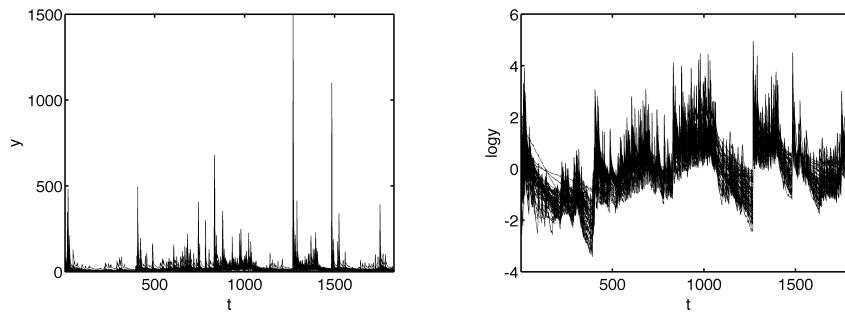

Figure 2: Outputs of the logSPM model at the $n = 60$ selected inputs in the training data set (left), and after log trasnformation and shifting to zero-mean (right).
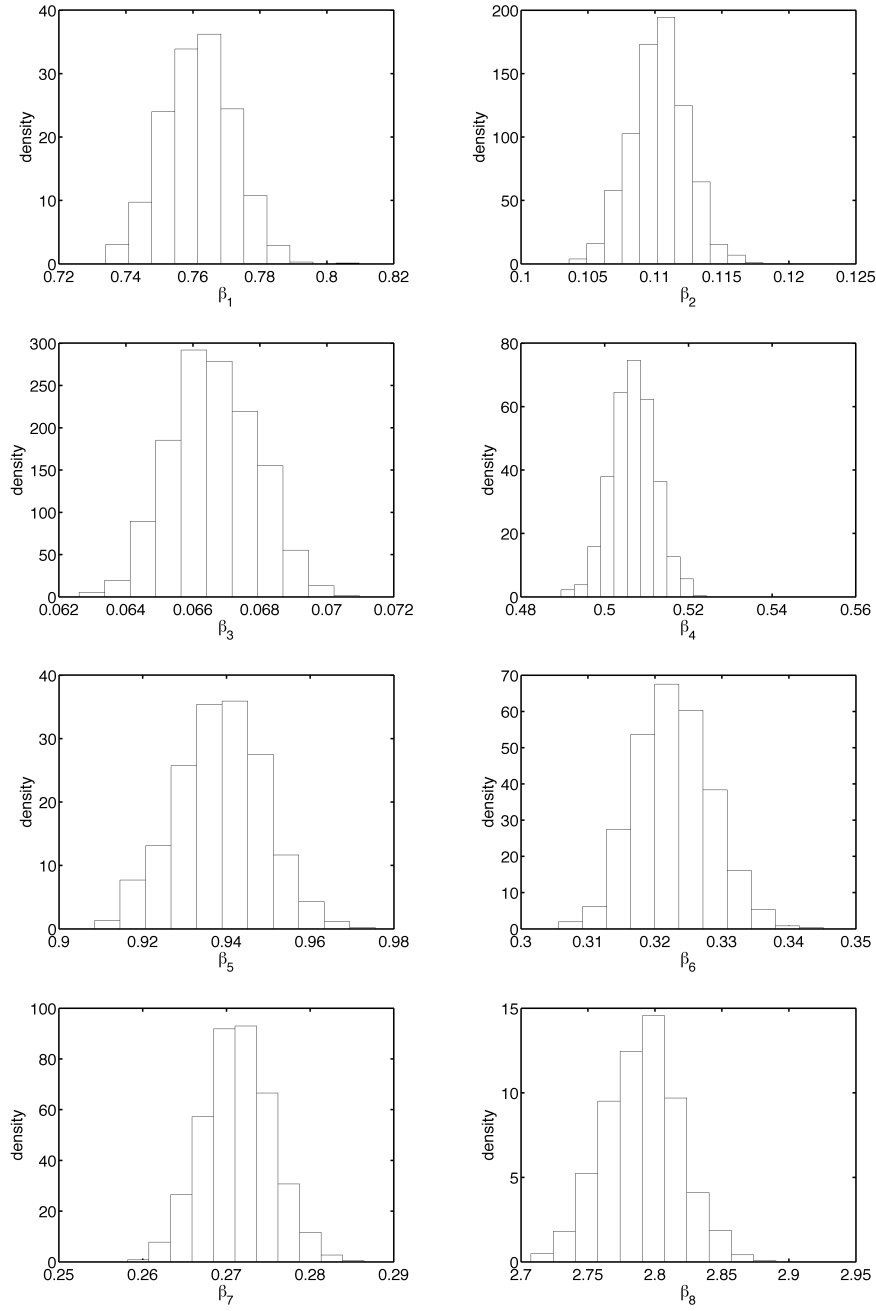
Figure 3: Histograms of posterior samples for the range parameters $\beta_{1:8}$.
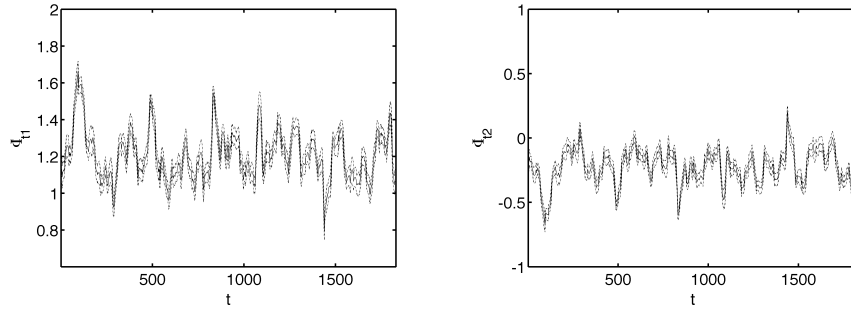
Figure 4: Pointwise posterior medians and 95% credible interval for ϕ_{t1} and ϕ_{t2} defining estimated trajectories over $t = 1, \ldots, T$.
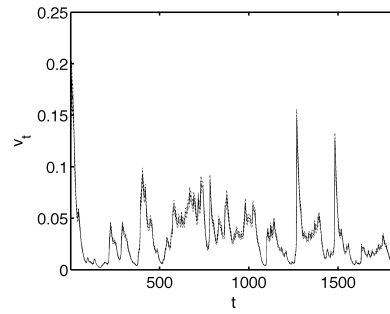

Figure 5: Pointwise posterior medians and 95% credible interval for v_t defining estimated trajectories over $t = 1, \ldots, T$.
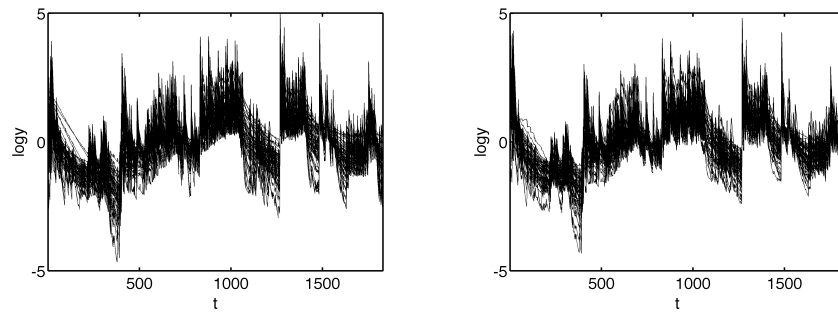

Figure 6:   Validation data output time series (left) and pointwise posterior predictive medians from the DLM-GASP emulator (right).
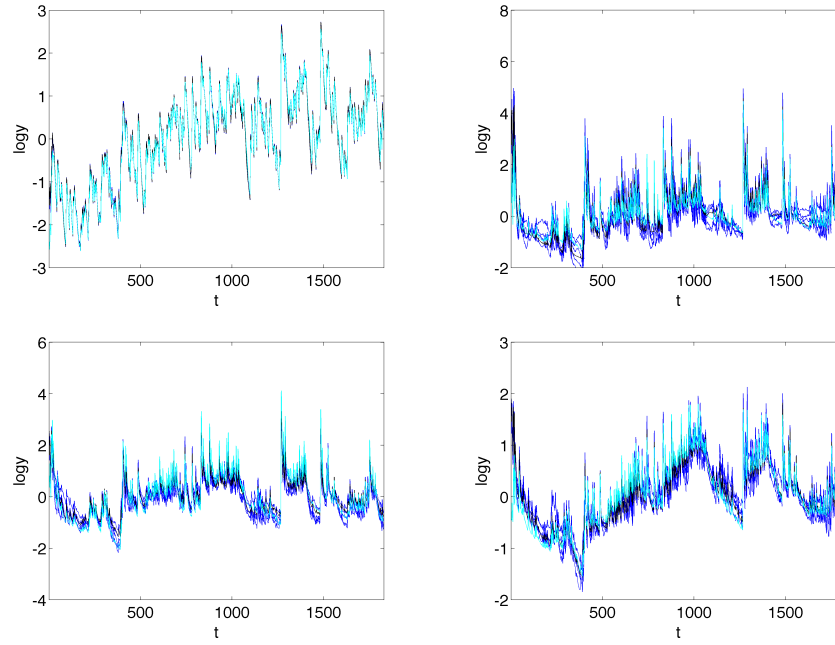
Figure 7: Simulated predictions of outputs at selected validation sample inputs z_1, z_{20}, z_{40} and z_{60}. Light blue curves are the true validation output series and the darker blue curves are 5 randomly selected draws from the posterior predictive distribution of the DLM-GASP emulator. The black curves are the posterior medians.