SANDIA REPORT

SAND2011-6496
Unlimited Release
Printed September, 2011

Statistical surrogate models for prediction of high-consequence climate change

Richard V. Field, Jr.
Paul Constantine
Mark Boslough

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online ordering: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2011-6496
Unlimited Release
Printed September, 2011

Statistical surrogate models for prediction of high-consequence climate change

Richard V. Field, Jr.
Component Science & Mechanics
Sandia National Laboratories
Albuquerque, NM 87185-0346
rvfield@sandia.gov

Paul Constantine
Optimization and Uncertainty Quantification
Sandia National Laboratories
Albuquerque, NM 87185-1318
pconsta@sandia.gov

Mark Boslough
Discrete Math & Complex Systems
Sandia National Laboratories
Albuquerque, NM 87185-0379
mbboslo@sandia.gov

Abstract

In safety engineering, performance metrics are defined using probabilistic risk assessments focused on the low-probability, high-consequence tail of the distribution of

possible events, as opposed to best estimates based on central tendencies. We frame the climate change problem and its associated risks in a similar manner. To properly explore the tails of the distribution requires extensive sampling, which is not possible with existing coupled atmospheric models due to the high computational cost of each simulation. We therefore propose the use of specialized statistical surrogate models (SSMs) for the purpose of exploring the probability law of various climate variables of interest. A SSM is different than a deterministic surrogate model in that it represents each climate variable of interest as a space/time random field. The SSM can be calibrated to available spatial and temporal data from existing climate databases, e.g., the Program for Climate Model Diagnosis and Intercomparison (PCMDI), or to a collection of outputs from a General Circulation Model (GCM), e.g., the Community Earth System Model (CESM) and its predecessors. Because of its reduced size and complexity, the realization of a large number of independent model outputs from a SSM becomes computationally straightforward, so that quantifying the risk associated with low-probability, high-consequence climate events becomes feasible. A Bayesian framework is developed to provide quantitative measures of confidence, via Bayesian credible intervals, in the use of the proposed approach to assess these risks.

Acknowledgments

The authors would like to thank our colleagues Michael Levy of Org. 1442 for providing the CCSM data on precipitation rate used in Section 3.2, and Timothy Trucano of Org. 1441 for his many useful comments on this paper.

Contents

1 Introduction and motivation . . . 9
2 Statistical surrogate model for climate variables . . . 11
2.1 Model definition . . . 11
2.2 Model calibration . . . 12
2.3 Model prediction and credibility . . . 13
2.3.1 Point estimates . . . 14
2.3.2 Bayesian approach . . . 14
3 Application to CCSM data . . . 17
3.1 Surface temperature . . . 17
3.2 Precipitation rate . . . 20
4 Conclusions . . . 27

Appendix

A Calibration of the K-L representation for a Gaussian random field . . . 31
B A generalized normal distribution . . . 33

Figures

1 First CCSM data set: average surface air temperature during December. 18
2 Sample coefficient of skewness of all CCSM surface air temperature data. 18
3 Generalized normal distribution with positive, negative, and zero coefficients of skewness. 19
4 Estimate of exceedance probability $1 - p_1 = \Pr\left(\bar{A} > 2\right)$ using the statistical surrogate model calibrated for data on air surface temperature. 20
5 Estimate of exceedance probability defined by Eq. (15) using the statistical surrogate model calibrated for air surface temperature. 21
6 First CCSM data set: precipitation rate during June, July, August. 22
7 Estimate of exceedance probability $p_3 = \Pr\left(\bar{A} \leq -1\right)$ using the statistical surrogate model calibrated for data on precipitation rate. 23
8 Bayesian posterior analysis for precipitation rate data. 25
9 Estimates of p_4 with 90% Bayesian credible intervals. 26

1 Introduction and motivation

The US Climate Change Research Program's (CCRP) long-term performance measure is to "deliver improved scientific data and models about the potential response of the Earth's climate and terrestrial biosphere to increased greenhouse gas levels for policy makers to determine safe levels of greenhouse gases in the atmosphere." One interpretation of this goal is that: (1) there is a well-defined threshold above which levels of greenhouse gases are "unsafe"; (2) this threshold is possible to determine; and (3) the Earth's climate has not yet crossed the threshold. The CCRP meaning of "safe" is presumed to be the level "that would prevent dangerous anthropogenic interference with the climate system" referenced by the 1992 United Nations Framework Convention on Climate Change (UNFCCC) [23]. Schneider and Lane [22] proposed metrics for dangerous climate change, which spanned the sustainability measures of water, energy, health, agriculture, and biodiversity, and included risks associated with extreme weather events and irreversible cascading chains of events beyond "tipping points". As a consequence, we are motivated by metrics associated with high-consequence climate changes, which we define as the changes that would be experienced if the true climate sensitivity is in the upper tail of its underlying probability distribution. Climate sensitivity is commonly defined as the change in the global mean surface temperature after the climate system has reached a new equilibrium in response to a doubling of the CO_2 concentration in the atmosphere.

The large and growing body of literature on global climate change is mostly written from a scientific perspective that focuses on the most probable future. A scientific approach is the most appropriate method for gaining understanding of natural systems by applying physically sound theory, empirical observations, and validated models. Scientifically conservative estimates are the ones that deviate the least from prior expectations. Scientific conservatism, when applied to climate change, tends to downplay the degree of change, and virtually all the climate change literature uses the term "conservative" in the opposite sense from that of safety engineers. The common approach is to generate probability density functions (PDFs) that encapsulate the best estimate of the future, plus some bounds on its uncertainty. The lower bound on expected climate change is the scientifically conservative estimate.

The Intergovernmental Panel on Climate Change (IPCC) reports present climate forecasts as assessments of the most probable future, with the tendency to err on the side of scientific conservatism. For example, the Fourth Assessment Report (AR4) of the IPCC provides a graph of "warming by 2090-2099 relative to 1980-1999 for non-mitigation scenarios" in terms of "best estimate and likely ranges of warming." "Likely" is defined by the AR4 as an outcome that occurs with a probability of more than 66%. Thus, the ranges provided by the IPCC for various scenarios tend to be of the most interest to decision makers because they are the most probable. Unfortunately, they are often treated as accurate forecasts to be used as the basis for informing policy decisions, as opposed to using high-consequence

forecasts as suggested by Palmer [18].

The development of methods to quantify the uncertainty in climate sensitivity (and other climate system response characteristics) is a topic of ongoing research. Some assessments result in the generation of PDFs rather than the simple "likelihood bounds" as provided by the IPCC. These studies consistently show that the high-end sensitivities have a significant probability. For example, Forest, et al., [6], give a 5-to-95% confidence interval of 1.4 to 7.7 °C climate sensitivity with a distribution that is strongly skewed with a sharp cutoff at the low end and a fat tail at the high end. The sharp low-end cutoff is expected, because the best understood feedbacks are strongly positive (e.g. water vapor) and any perfectly-canceling negative feedback would need to cancel not only the forcing but also the large positive feedbacks. Andronova & Schlesinger [1] calculated a distribution with a 10% probability of climate sensitivity greater than 6.8 °C. Skewed distributions with a high-sensitivity tail are characteristic of climate sensitivity PDFs, and Roe and Baker [20] argue that such skewness is an inevitable consequence of the nature of the climate system and the inherent uncertainty in the feedbacks. More recent research suggests that the upper bound of the confidence interval in climate sensitivity may also have been underestimated [14] due to compensating feedbacks assumed in previous studies. This interpretation further suggests the need for a safety engineering approach to characterize the mechanisms that can lead to high sensitivity.

Murphy, et al. [17] used the ensemble method in a "perturbed physics" approach, which systematically varied 29 model parameters to determine a probability distribution function that has a 5% to 95% range of 2.4°C to 5.4 °C, with a median of 3.5°C and a most probable value of 3.2°C. This sophisticated analysis makes use of more advanced models and is consistent with the transient effects of climate change and forcing. Processes that determine climate sensitivity are varied systematically and uncertainties are weighted according to an objective index. However, the perturbed physics approach is prohibitively expensive in many cases due to the cost of the coupled climate model. In practice, one must construct a cheaper surrogate of a quantity of interest – say surface temperature or precipitation – whose samples are used to approximate statistics. Constructing a surrogate model from scattered data poses challenges for many existing techniques. In what follows, we propose a translation random field surrogate for a quantity of interest tuned to a set of output data from a coupled climate model. With this surrogate, we are able to explore tail probabilities of the quantity of interest that are essential to formal study of the consequences of the changing climate system.

The outline of our paper is as follows. The statistical surrogate model (SSM) is developed in Section 2, with emphasis on calibration of the SSM to available data, the use of the SSM to make predictions, and the development of Bayesian methods to assess model credibility. The general framework is then applied to two different climate variables of interest in Section 3: global mean surface temperature and precipitation rate.

2 Statistical surrogate model for climate variables

Let $A(\lambda, \phi, z)$ denote a climate variable of interest, *e.g.*, temperature, precipitation rate, total cloud fraction, *etc.*, where $0 \leq \lambda < 2\pi$, $-\pi/2 \leq \phi < \pi/2$, and $z \geq 0$ are spatial coordinates denoting longitude, latitude, and geodesic altitude, respectively. To simplify notation, we use $\mathbf{u} = (\lambda, \phi, z)^T \in D = [0, 2\pi) \times [-\pi/2, \pi/2) \times [0, \infty)$ to collectively represent all spatial coordinates and simply write $A(\mathbf{u})$, $\mathbf{u} \in D$.

To represent variability and uncertainty in climate variables, we model $A(\mathbf{u})$ as a random field, that is, $A(\mathbf{u})$ is a random variable for every fixed $\mathbf{u} \in D$. Hence, $A = A(\mathbf{u}, \omega)$ also depends on additional argument $\omega \in \Omega$, where Ω represents the appropriate sample space of A. It is common practice to omit the explicit depiction of the functional dependence of A on ω and we shall do so here. Further, our convention is to use a capital letter or symbol to denote any random quantity; lowercase letters and symbols are reserved for deterministic quantities.

Our objective is to develop a class of surrogate models for A that can be used to make defensible predictions about the probability of various climate change scenarios of interest. Special emphasis is placed on the tails of the probability distribution, the calibration of the surrogate model to available data from current and future runs of complex climate models, and to providing a quantitative measure of confidence in any model predictions. Throughout the discussion we assume A is a scalar quantity; the approach can be extended to represent vector-valued climate variables.

2.1 Model definition

Let $G(\mathbf{u})$ be a real-valued Gaussian random field with zero mean, unit variance, and covariance function $c(\mathbf{u}, \mathbf{v}) = \mathrm{E}[G(\mathbf{u})\, G(\mathbf{v})]$, where $\mathrm{E}[X]$ denotes the expected value of random variable X. We model A by a monotonic transformation of G of the following form

$$A(\mathbf{u}) = F_A^{-1} \circ \Phi\left(G(\mathbf{u})\right), \tag{1}$$

where $F_A = F_A(a|\boldsymbol{\theta})$ is an arbitrary cumulative distribution function (CDF) that depends on parameter vector $\boldsymbol{\theta} \in \mathbb{R}^d$, and Φ is the CDF of a $N(0, 1)$ random variable, that is, a Gaussian or normal random variable with zero mean and unit variance. Herein, we assume F_A is absolutely continuous, that is, we assume there exists an integrable function f_A such that $f_A(a|\boldsymbol{\theta}) = \mathrm{d}F_A(a|\boldsymbol{\theta})/\mathrm{d}a$ is a probability density function (PDF). It can be shown that random field A has marginal CDF F_A and marginal PDF f_A. This class of model is a special type of non-Gaussian random field referred to as a translation random field (see [10, Section 3.1.1] and [11]); it is common to refer to G as the Gaussian image of A.

By careful selection of the properties of G, as well as the functional form of F_A and its parameters $\boldsymbol{\theta}$, it is possible to calibrate A to match statistical estimates of the mean, covariance, and marginal probability distribution functions of a climate variable of interest. The translation random field model defined by Eq. (1) is a very flexible model and has proved effective for a variety of applications in stochastic mechanics (see, for example, [4, 5]) and the modeling of various environmental phenomena [8, 9].

2.2 Model calibration

Let $z_k(\mathbf{u})$, $k = 1, \ldots, n$, denote the available data on A, that is, results from a sequence of climate model calculations. We use $\mathbf{z} = (z_1(\mathbf{u}), \ldots, z_n(\mathbf{u}))^T$ to represent the collection of available data to simplify notation. Calibration of the random field model defined by Eq. (1) to $\mathbf{z}$ requires three steps:

1. Choose the functional form for F_A, the marginal CDF of A defined by Eq. (1).
2. Calibrate $\boldsymbol{\theta}$, the associated parameters of F_A.
3. Construct the covariance function, $c(\mathbf{u}, \mathbf{v})$, of G, the Gaussian image of A.

The objective of step 1 is to select a marginal distribution function F_A that is sufficiently flexible to capture any desired behavior observed in the data $\mathbf{z}$, and is consistent with the known physics. For example, if A models precipitation rate, the distribution function must have support on the positive real line with positive skewness; the lognormal distribution satisfies these constraints and is often used to model precipitation rate [21].

Calibration of parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)^T$ to be consistent with available data $\mathbf{z}$, that is, step two of the calibration procedure, can be completed by one of two methods: the method of moments, or the method of maximum likelihood [15, Chapter 6]. For the former, we choose $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ such that $f_A(a|\hat{\boldsymbol{\theta}})$ satisfies

$$\int_{\mathbb{R}} a^q \, f_A(a|\hat{\boldsymbol{\theta}}) \, \mathrm{d}a = \frac{1}{n} \sum_{k=1}^{n} (\mathbf{z})^q \,, \;\; q = 1, \ldots, d. \tag{2}$$

For the latter, we choose $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ that maximizes

$$l(\mathbf{z}|\boldsymbol{\theta}) = \prod_{k=1}^{n} f_A(z_k|\boldsymbol{\theta}), \tag{3}$$

where $l(\mathbf{z}|\boldsymbol{\theta})$, viewed as a function of $\boldsymbol{\theta}$, represents the likelihood that the data has PDF $f_A(a|\boldsymbol{\theta})$. We note that, by either method, the calibrated value for $\boldsymbol{\theta}$ may depend on $\mathbf{u} \in D$.

Upon completion of steps 1 and 2 in the calibration procedure, the available data on A can be mapped to available data on G, *i.e.* $g_k(\mathbf{u}) = \Phi^{-1} \circ F_A(z_k(\mathbf{u})|\hat{\boldsymbol{\theta}})$, $k = 1, \ldots, n$. The objective for step three of the calibration procedure is to construct covariance function $c(\mathbf{u}, \mathbf{v}) = \mathrm{E}[G(\mathbf{u})\, G(\mathbf{v})]$ of G, a Gaussian random field with zero mean and unit variance, that is consistent with the data $\{g_k(\mathbf{u}),\ k = 1, \ldots, n\}$. One approach is to utilize the Karhunen-Loéve (K-L) representation for a Gaussian random field, *i.e.*

$$G(\mathbf{u}) = \sum_{k \geq 1} \sqrt{\zeta_k}\, \psi_k(\mathbf{u})\, W_k,\ \ \mathbf{u} \in D, \tag{4}$$

where $\{\zeta_k, \psi_k(\mathbf{u}),\ k \geq 1\}$ are the eigenvalues and eigenfunctions, respectively, of $c(\mathbf{u}, \mathbf{v}) = \mathrm{E}[G(\mathbf{u})\, G(\mathbf{v})]$, and satisfy the integral equation

$$\int_D c(\mathbf{u}, \mathbf{v})\, \psi_k(\mathbf{v})\, \mathrm{d}\mathbf{v} = \zeta_k\, \psi_k(\mathbf{u}), \tag{5}$$

and $\{W_k, k \geq 1\}$ is a collection of independent and identically distributed (iid) $N(0, 1)$ random variables. In practice, the infinite sum defined by Eq. (4) is truncated at $r \geq 1$ terms; we refer to the truncated sum as the K-L approximation for G. Calibration of the K-L approximation model requires a method to choose: (i) r, the number of terms retained in the sum, and (ii) the associated eigenvalues and eigenvectors, $(\zeta_k, \psi_k(\mathbf{u}))$, $k = 1, \ldots, r$. An efficient method to do this based on the the singular value decomposition of a matrix containing the available data $\{g_k(\mathbf{u}),\ k = 1, \ldots, n\}$ is presented in A.

2.3 Model prediction and credibility

Let $h(A(\mathbf{u}))$ be a particular property of A that is of interest, e.g., a functional of climate field data such as the extreme or average value. Our objective is to predict values for

$$p(\boldsymbol{\theta}) = \Pr\big(h(A(\mathbf{u})) \in S_h \mid \boldsymbol{\theta}\big) = \int_{S_h} h(a)\, f_A(a|\boldsymbol{\theta})\, \mathrm{d}a, \tag{6}$$

where S_h denotes an appropriate "safe set," that is, a set such that if the quantity of interest $h(A(\mathbf{u}))$ remains within S_h, there is no cause for alarm. We note that it is oftentimes more useful to report values for $1 - p$ which corresponds to the probability that the quantity of interest departs from the safe set. The concept of a "tipping point" described in climate literature [16] is an example of the boundary of a safe set for this study.

Exact solutions to the integral defined by Eq. (6) do not exist in general. However, because generating a large number of samples from the surrogate model is computationally inexpensive, we can make use of straightforward Monte Carlo simulation to estimate the value for the integral in Eq. (6) with high accuracy. Further, as indicated by Eq. (6), our prediction p depends on the parameter vector $\boldsymbol{\theta}$ which, due to limited data, is also uncertain. We can handle this issue in one of two ways, as described in Sections 2.3.1 and 2.3.2.

2.3.1 Point estimates

The calibrated model for A can be used directly to provide point estimates for p defined by Eq. (6). Let $\hat{p}$ denote an estimate for p. This estimate can be achieved by the following steps:

1. Generate $m \gg 1$ samples of the calibrated K-L approximation for G using a truncated version of Eq. (4).
2. Translate each sample of G using Eq. (1) with $\boldsymbol{\theta}$ replaced by $\hat{\boldsymbol{\theta}}$ defined in Section 2.2.
3. Evaluate the quantity of interest for each sample of A, herein denoted by h_k, $k = 1, \ldots, m$.

The point estimate for p defined by Eq. (6) is then given by

$$\hat{p} = \frac{1}{m} \sum_{k=1}^{m} \mathbf{1}\left(h_k \in S\right), \tag{7}$$

where $\mathbf{1}(\,\cdot\,)$ denotes the indicator function, *i.e.* $\mathbf{1}(B) = 1$ if event B is true and is zero otherwise. $\hat{p}$ defined by Eq. (7) is an unbiased estimator for p with variance proportional to $1/m$. Hence, the point estimate can be made very accurate because generating many samples of the surrogate model (large m) is inexpensive.

2.3.2 Bayesian approach

The approach of the previous section provides only point estimates for p defined by Eq. (6). We have limited confidence in these estimates due to the fact that we have limited data on A, and we would like to quantify our level of confidence in some manner. One approach is to instead treat $\boldsymbol{\theta} \in \mathbb{R}^d$, the parameters of the marginal CDF of A, as a random vector and apply a Bayesian approach. Bayesian credible sets can then be used to quantify prediction confidence.

Let $\boldsymbol{\Theta}$ be a random vector with d coordinates and support $D_{\boldsymbol{\Theta}} \subset \mathbb{R}^d$ representing the uncertainty in the parameters of F_A defined by Eq. (1). Given a prior PDF, denoted by $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$, that describes our knowledge on the uncertainty of $\boldsymbol{\Theta}$ prior to studying the available data, the posterior PDF for $\boldsymbol{\Theta}$ given the data is [24, Section 2.1]

$$f_{\boldsymbol{\Theta}|\mathbf{z}}(\boldsymbol{\theta}|\mathbf{z}) \propto l(\mathbf{z}|\boldsymbol{\theta})\, f_{\boldsymbol{\Theta}}(\boldsymbol{\theta}), \tag{8}$$

where $\mathbf{z} = (z_1(\mathbf{u}), \ldots, z_n(\mathbf{u}))^T$ denotes the available data on the climate variable of interest, l denotes the likelihood function defined by Eq. (3), and the $\propto$ symbol is used to denote that the left and right side of Eq. (8) are equal to within a normalizing constant.

With $\boldsymbol{\theta}$ replaced by $\boldsymbol{\Theta}$, the conditional probability $p(\boldsymbol{\Theta})$ defined by Eq. (6) is a random variable taking values in $[0, 1]$. It is possible to find the probability that $p(\boldsymbol{\Theta})$ belongs to any subset C of $[0, 1]$; this probability is equal to $\int_{C_{\boldsymbol{\Theta}}} p(\boldsymbol{\theta})\, f_{\boldsymbol{\Theta}|\mathbf{z}}(\boldsymbol{\theta}|\mathbf{z})\, \mathrm{d}\boldsymbol{\theta}$, where $C_{\boldsymbol{\Theta}} = \{\boldsymbol{\theta} \in D_{\boldsymbol{\Theta}} : p(\boldsymbol{\theta}) \in C\}$. In particular, the $100(1-\alpha)\%$ credible set for $p(\boldsymbol{\Theta})$ is the set C such that [2, Section 4.3.2]

$$1 - \alpha \leq \int_{\{\boldsymbol{\theta} \in D_{\boldsymbol{\Theta}} \,:\, p(\boldsymbol{\theta}) \in C\}} p(\boldsymbol{\theta})\, f_{\boldsymbol{\Theta}|\mathbf{z}}(\boldsymbol{\theta}|\mathbf{z})\, \mathrm{d}\boldsymbol{\theta}. \tag{9}$$

For the special case when random variable $p(\boldsymbol{\Theta})$ has uni-modal density, the credible set can be expressed as a credible interval, *i.e.* $C = [a, b]$.

The implementation of the Bayesian approach described in this section is as follows:

1. Postulate $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$, the prior PDF for model parameters $\boldsymbol{\Theta}$; non-informative priors [24, Section 2.3] may be used for the case when no information on $f_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ is available.

2. Compute the likelihood function and posterior PDF for $\boldsymbol{\Theta}$ using Eq. (8).

3. Draw $m_1 \gg 1$ independent samples of random vector $\boldsymbol{\Theta}$ from $f_{\boldsymbol{\Theta}|\mathbf{z}}(\boldsymbol{\theta}|\mathbf{z})$; Markov Chain Monte Carlo methods can be used for this step [7, 12].

4. For each sample $\boldsymbol{\theta}_k$ of $\boldsymbol{\Theta}$, $k = 1, \ldots, m_1$, draw m_2 independent samples from $f_A(a|\boldsymbol{\theta}_k, \mathbf{z})$.

5. Utilize the procedure in the previous section to estimate

$$\hat{p}_k = \frac{1}{m_2} \sum_{j=1}^{m_2} \mathbf{1}(h_{k,j} \in S), \tag{10}$$

   where $h_{k,j}$ denotes the jth random sample drawn from $f_A(a|\boldsymbol{\theta}_k, \mathbf{z})$.

6. Estimate the endpoints of the $100(1-\alpha)\%$ credible interval for $p(\boldsymbol{\Theta})$, denoted by $\hat{a} \leq \hat{b}$, where $\hat{a}$ and $\hat{b}$ satisfy

$$\begin{aligned} \frac{1}{m_1} \sum_{k=1}^{m_1} 1(\hat{p}_k \leq \hat{a}) &= \frac{\alpha}{2} \\ \frac{1}{m_1} \sum_{k=1}^{m_1} 1(\hat{p}_k > \hat{b}) &= 1 - \frac{\alpha}{2} \end{aligned} \tag{11}$$

We conclude this section by noting that we have ignored any uncertainty in the parametric description of the Gaussian image of A, that is, in the eigenvalues and eigenvectors of the K-L approximation for G defined by Eq. (4). This information can also be included in the Bayesian approach outlined above if necessary.

3 Application to CCSM data

To illustrate the use of the statistical surrogate model for climate variables developed in Section 2, we consider two collections of output from the Community Climate System Model (CCSM) v3.0, a fully-coupled global climate model sponsored by the National Center for Atmospheric Research (NCAR). The first collection, presented in Section 3.1, corresponds to average December surface temperature for years 1990–1999 based on a collection of 8 different model runs obtained from the Program for Climate Model Diagnosis and Intercomparison [19]. We calibrate the surrogate model to the available model data and make various point predictions that may be of interest. In Section 3.2, we study average precipitation rate in June, July, and August over a 54-year period assuming a "cyclic Y2K ocean model." We apply the calibrated surrogate model to study the probability that the precipitation rate falls below certain thresholds, then utilize the Bayesian approach described in Section 2.3.2 to quantify our confidence in these predictions.

3.1 Surface temperature

Let $z_k(\mathbf{u})$, $k = 1, \ldots, n = 80$, denote the available surface temperature data; Fig. 1 illustrates the first data set, $z_1(\mathbf{u})$, as contours of average surface air temperature as a function of longitude, λ, and latitude ϕ. Upon inspection, we find that the data exhibits considerable skewness, *i.e.* the probability distribution of surface temperature is not symmetric about its mean value, and the degree of skewness changes with spatial location. An estimate of the coefficient of skewness is given by

$$\hat{\gamma}_3(\mathbf{u}) = \frac{n\sqrt{n-1}}{n-2} \frac{\sum_{k=1}^{n} \left(z_k(\mathbf{u}) - \hat{\mu}(\mathbf{u})\right)^3}{\left(\sum_{k=1}^{n} \left(z_k(\mathbf{u}) - \hat{\mu}(\mathbf{u})\right)^2\right)^{3/2}} \tag{12}$$

where $\hat{\mu}(\mathbf{u}) = (1/n) \sum_{k=1}^{n} z_k(\mathbf{u})$ is an unbiased estimator for the mean surface temperature. The coefficient of skewness of the data is illustrated by Fig. 2. Regions of large positive skewness, *e.g.* parts of South America, correspond to locales where infrequent but large increases in temperature can be expected. It therefore seems appropriate to choose a non-Gaussian distribution that is able to match the large range of observed skewness exemplified by Fig. 2. In regions with near zero skewness, a Gaussian distribution may be adequate.

Because of these observations, we choose F_A to be the CDF of a generalized version of the Gaussian distribution [13], that is, $F_A(a|\boldsymbol{\theta}) = \int_{-\infty}^{a} f_A(\xi|\boldsymbol{\theta})\, \mathrm{d}\xi$, where

$$f_A(a|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\,(\theta_2 - \theta_3(a - \theta_1))}\, e^{-y^2/2}, \tag{13}$$

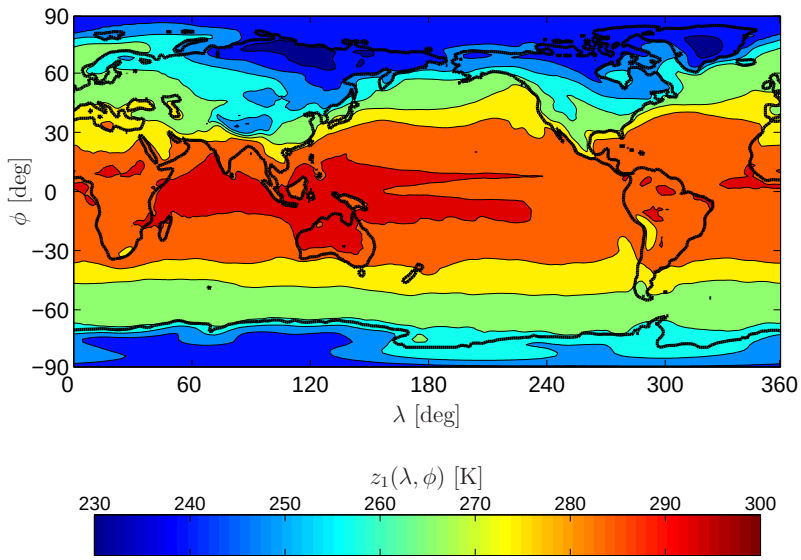

Figure 1. First CCSM data set: average surface air temperature during December (from [19]).

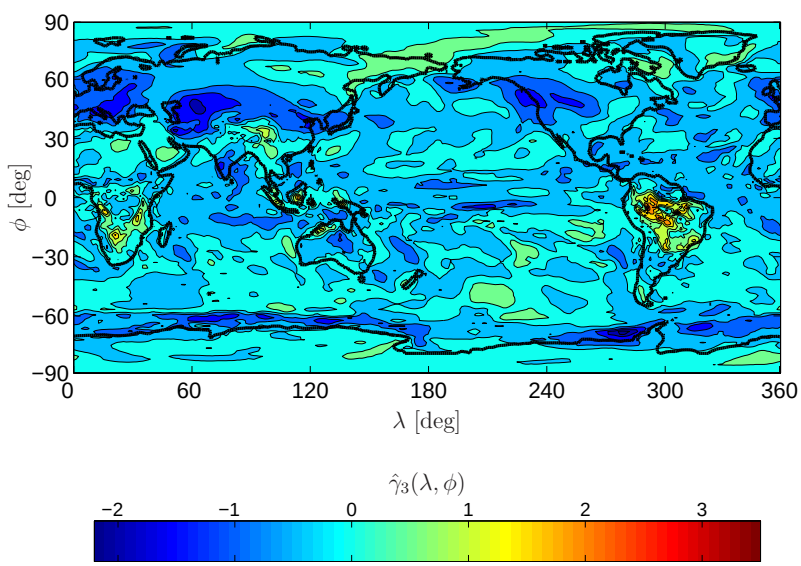

Figure 2. Sample coefficient of skewness of all CCSM surface air temperature data.

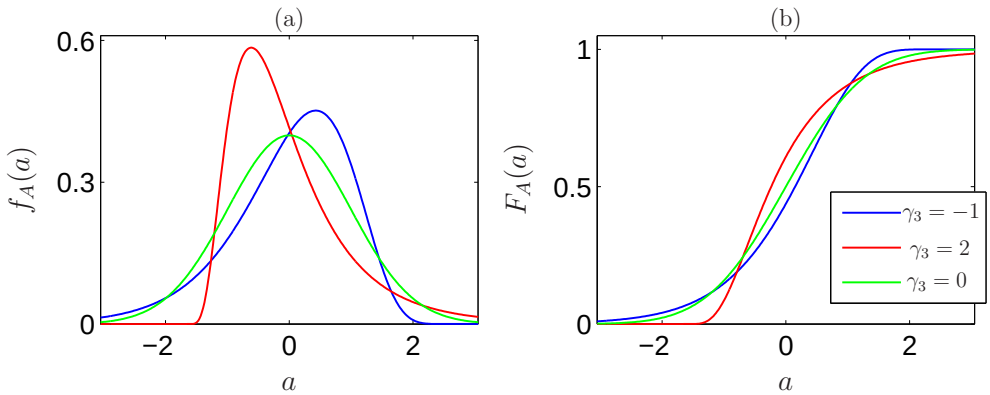

Figure 3. Generalized normal distribution with positive ($\gamma_3 = 2$), negative ($\gamma_3 = -1$), and zero ($\gamma_3 = 0$) coefficients of skewness: (a) the PDFs, and (b) the CDFs. Each has zero mean and unit variance. For $\gamma_3 = 0$, F_A is a Gaussian distribution.

is the corresponding PDF, $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^T$ is a vector of model parameters with support $\theta_1, \theta_3 \in \mathbb{R}$, $\theta_2 > 0$ and, assuming $\theta_3 \neq 0$,

$$y = -\frac{1}{\theta_3} \ln \left(1 - \frac{\theta_3 (a - \theta_1)}{\theta_2} \right).$$

The CDF and PDF defined by Eq. (13) are illustrated by Figs. 3(a) and (b), respectively, for various values of parameters θ_1, θ_2, and θ_3; more details on this distribution are provided in B.

For illustrative purposes, we make point predictions for two potential climate change scenarios of interest. First, we develop an estimate of

$$p_1 = \Pr\left(h_1(A(\mathbf{u})) \in S_1\right) = \Pr\left(\bar{A}(\mathbf{u}) \leq 2\right), \tag{14}$$

where $\bar{A}(\mathbf{u}) = (A(\mathbf{u}) - \mathrm{E}[A(\mathbf{u})])/\sqrt{\mathrm{Var}[A(\mathbf{u})]}$ is a random field with zero mean and unit variance. By Eq. (14), $1 - p_1$ corresponds to the probability that A will exceed the so-called "2-σ level" and is illustrated by Fig. 4 as a function of latitude and longitude. This result illustrates a point estimate of the probability of higher than average air temperature during December based on the available CCSM data.

Next let $D' \subset D$ be a region of interest whose boundary is defined by lines of longitude 170 deg and 240 deg and lines of latitude −30 deg and 30 deg (this corresponds roughly

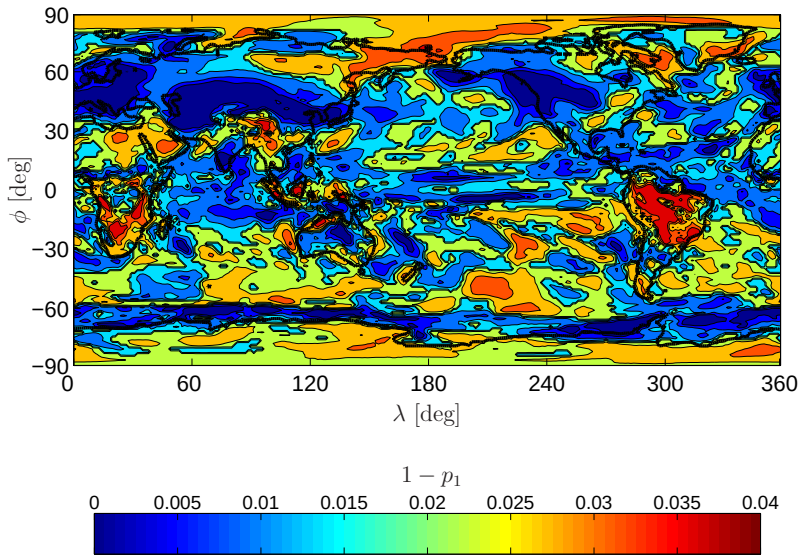

Figure 4. Estimate of exceedance probability $1 - p_1 = \Pr\left(\bar{A} > 2\right)$ using the statistical surrogate model calibrated for data on air surface temperature.

to the Pacific ocean). Suppose the maximum average surface temperature within D' is a quantity of interest so that

$$p_2 = \Pr\left(h_2(A(\mathbf{u})) \in S_2,\ \forall \mathbf{u} \in D'\right) = \Pr\left(\max_{\mathbf{u} \in D'} A(\mathbf{u}) \leq t\right). \tag{15}$$

Figure 5 illustrates a point estimate of $1 - p_2$ as a function of temperature t; the estimate is based on 20,000 independent samples of A, the statistical surrogate model calibrated for average surface temperature.

3.2 Precipitation rate

The proposed statistical surrogate model has also been calibrated to CCSM model predictions of precipitation rate. This data set consists of monthly average precipitation rates for June, July, and August, over a 54-year period. Let $z_k(\mathbf{u})$, $k = 1, \ldots, n = 162$, denote the available data on precipitation rate; Fig, 6 illustrates $z_1(\mathbf{u})$ in units of mm/hr as a function of longitude, λ, and latitude ϕ. As mentioned, the lognormal distribution is often used to

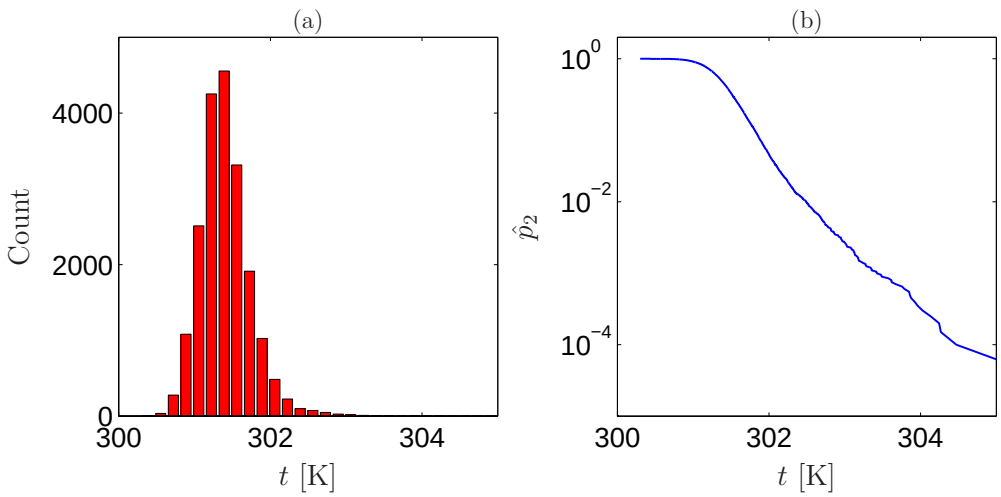

Figure 5. Estimate of exceedance probability defined by Eq. (15) using the statistical surrogate model calibrated for air surface temperature: (a) histogram of 20,000 Monte Carlo samples of $\max_{\mathbf{u}\in D'} A(\mathbf{u})$, and (b) corresponding estimate for p_2.

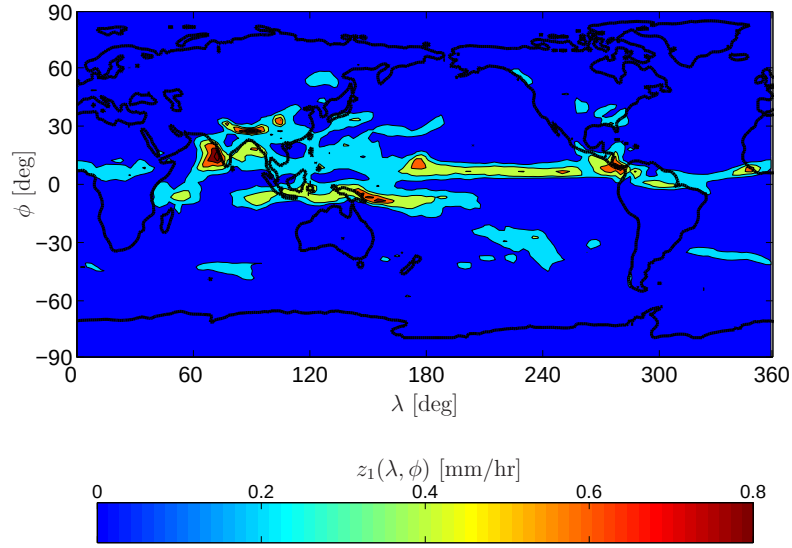

Figure 6. First CCSM data set: precipitation rate during June, July, August.

model precipitation rate [21]; the corresponding PDF is given by

$$f_A(a|\theta_1, \theta_2) = \frac{1}{\sqrt{2\pi}\,\theta_2\,a}\, e^{-(\ln a - \theta_1)^2/(2\,\theta_2^2)},\ \ a > 0, \tag{16}$$

where $\theta_1 \in \mathbb{R}$ and $\theta_2 > 0$ are model parameters.

We first consider point estimates of

$$p_3 = \Pr\left(\bar{A}(\mathbf{u}) \leq -1\right), \tag{17}$$

where $\bar{A}$ is defined by Eq. (14); p_3 corresponds to the probability that the precipitation rate will be equal to or less than the "1-σ level". Figure 7 illustrates an estimate of p_3 using the surrogate model calibrated to the available data; this result illustrates a point estimate of the probability of drier than average conditions. Estimates of the model parameters, $\hat{\theta}_1 = n^{-1}\sum_{k=1}^{n} \ln z_k$ and $\hat{\theta}_2^2 = n^{-1}\sum_{k=1}^{n}(\ln z_k - \hat{\theta}_1)^2$, are obtained using the method of maximum likelihood, Eq. (3).

A more useful result for the assessment of risk is to apply the Bayesian analysis described in Section 2.3.2. We predict

$$p_4(\rho) = \Pr\left(A(\mathbf{u}) \leq \rho\right), \tag{18}$$

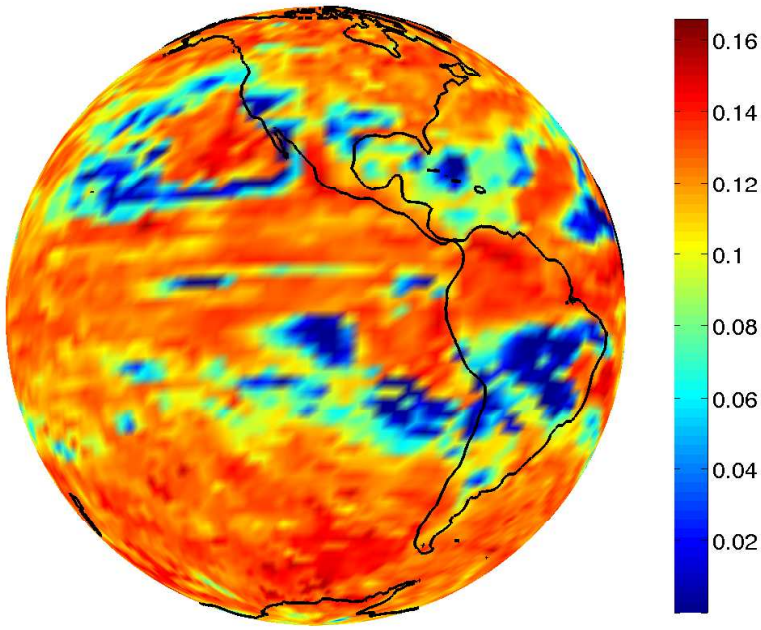

Figure 7. Estimate of exceedance probability $p_3 = \Pr\left(\bar{A} \leq -1\right)$ using the statistical surrogate model calibrated for data on precipitation rate.

the probability that the precipitation rate is less than or equal to threshold ρ, and provide a measure of confidence on our predictions of p_4 as a function of ρ. To simplify the discussion, we present such an analysis for a fixed spatial location, *i.e.* a fixed $\mathbf{u} \in D$, that is roughly located at the city of Albuquerque, New Mexico. The analysis can be generalized to consider the entire domain D, but we focus here on a single grid point to clarify the discussion.

Following the procedure outlined in Section 2.3.2, we first model $\boldsymbol{\theta} = (\theta_1, \theta_2)^T$, the parameters of the lognormal PDF defined by Eq. (16), as $\boldsymbol{\Theta} = (\Theta_1, \Theta_2)^T$ a random vector with independent coordinates and non-informative prior PDF

$$f_{\boldsymbol{\Theta}}(\theta_1, \theta_2) \propto \text{ const.}, \ \theta_1 \in \mathbb{R}, \theta_2 > 0. \tag{19}$$

Given the available CCSM data on precipitation rate, $\mathbf{z}$, application of Eqs. (8) and (16) yields the posterior PDF for the model parameters, *i.e.*

$$f_{\boldsymbol{\Theta}|\mathbf{z}}(\theta_1, \theta_2|\mathbf{z}) \propto \theta_2^{-n} \exp\left(-\frac{n}{2\,\theta_2^2}\left(\hat{s}^2 - 2\hat{\mu}\theta_1 + \theta_1^2\right)\right), \ \theta_1 \in \mathbb{R}, \theta_2 > 0, \tag{20}$$

where $\hat{\mu} = n^{-1}\sum_{k=1}^{n} \ln z_k$ and $\hat{s}^2 = n^{-1}\sum_{k=1}^{n}(\ln z_k)^2$ denote the sample mean and sample mean square of the natural log of the data. The posterior PDF for $\boldsymbol{\Theta}$ is illustrated by Fig. 8(a).

Following step 3, the Markov Chain Monte Carlo (MCMC) method [12], was used to draw 10,000 independent samples of random vector $\boldsymbol{\Theta}$ from $f_{\boldsymbol{\Theta}|\mathbf{z}}(\theta_1, \theta_2|\mathbf{z})$ defined by Eq. (20). These samples are illustrated by Fig. 8(b); level contours of the posterior PDF are also plotted to illustrate the performance of the MCMC method.

We next apply steps 4–6 from Section 2.3.2 to calculate the $100(1-\alpha)\%$ Bayesian credible interval for p_4 defined by Eq. (18); for calculations, we assume $\alpha = 0.1$. Two cases are studied, as illustrated by Figs. 9(a) and (b), respectively. In each case, the estimate of $p_4(\rho)$ is denoted by a blue solid line, while the corresponding 90% Bayesian credible intervals are illustrated by red dashed lines. In panel (a), we restrict the analysis to half of the available data points. The resulting credible intervals are wider, indicating a lower degree of model credibility, when compared to panel (b), which corresponds to the same analysis based on the full data set. In addition, we note that the width of the credible intervals changes with ρ.

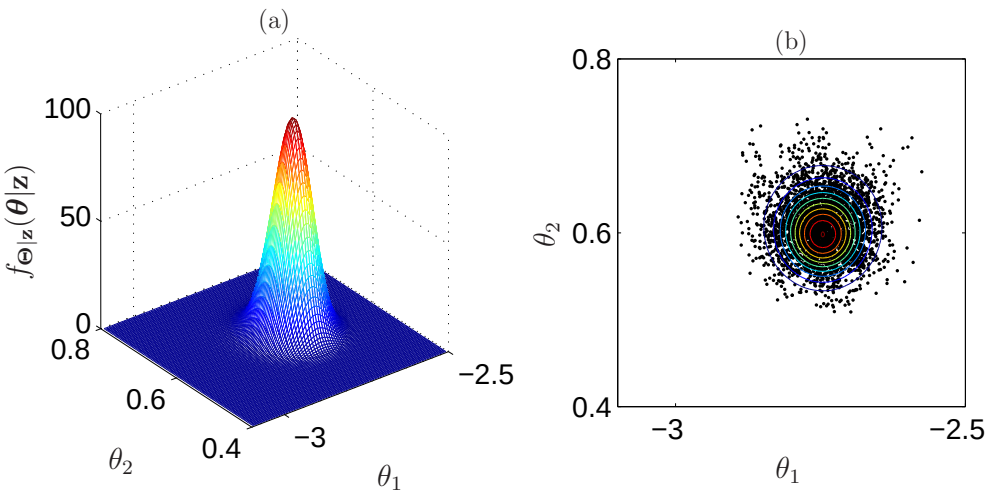

Figure 8. Bayesian posterior analysis for precipitation rate data: (a) posterior PDF $f_{\mathbf{\Theta}|\mathbf{z}}(\boldsymbol{\theta}|\mathbf{z})$, and (b) 10,000 independent samples of $\mathbf{\Theta}$ using MCMC. Level contours of $f_{\mathbf{\Theta}|\mathbf{z}}(\boldsymbol{\theta}|\mathbf{z})$ are also shown in panel (b).

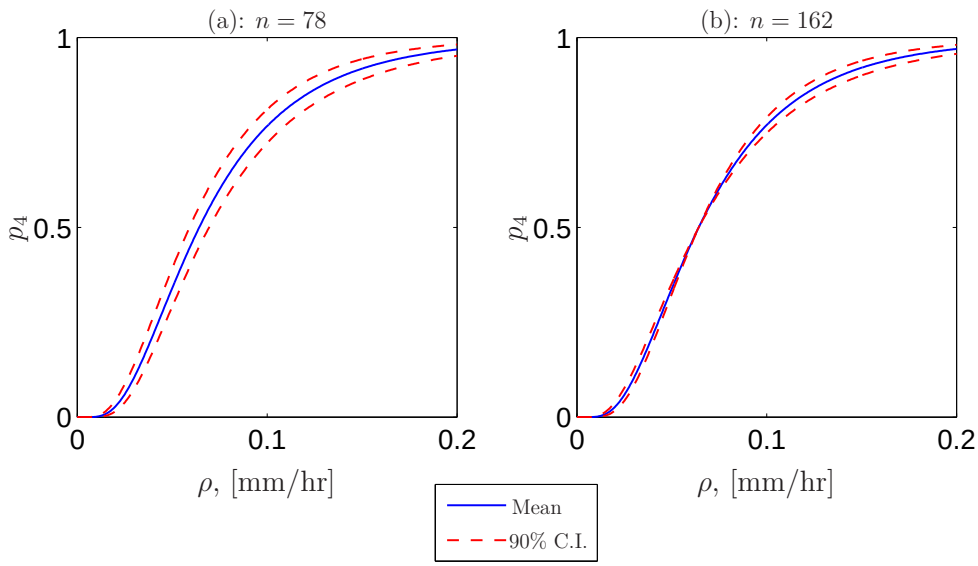

Figure 9. Estimates of p_4 with 90% Bayesian credible intervals for the case of: (a) $n = 78$ data points, and (b) $n = 162$ data points.

4 Conclusions

We have formulated risk assessment in climate change impact studies in a framework similar to that used in safety engineering, by acknowledging that probabilistic risk assessments focused on low-probability, high-consequence climate events are perhaps more appropriate than studies focused simply on best estimates. To aid in this study, we have developed specialized statistical surrogate models (SSMs) that can be used to make predictions about the tails of the associated probability distributions. The SSM represents each climate variable output of interest as a space/time random field, and can be calibrated to available spatial and temporal data from existing climate databases, or from a collection of outputs from Global Circulation Models. Due to its reduced size and complexity, the realization of a large number of independent model outputs from a SSM becomes computationally straightforward, so that estimates of low-probability, high-consequence climate events becomes feasible. A Bayesian framework was also developed to provide quantitative measures of confidence, via Bayesian credible intervals, in the use of the proposed SSM as a statistical replacement for the associated GCM.

References

[1] N. G. Andronova and M. E. Schlesinger. Objective estimation of the probability density function for climate sensitivity. *Journal of Geophysical Research*, 106(D19):22,605–22,611, 2001.

[2] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*, Springer, 1985.

[3] N. A. C. Cressie. *Statistics for Spatial Data*, Wiley, 1993.

[4] R. V. Field, Jr., T. S. Edwards, and J. W. Rouse. Modeling of atmospheric temperature fluctuations by translations of oscillatory random processes with application to spacecraft atmospheric re-entry. *Probabilistic Engineering Mechanics*, 26(2):231–239, 2001.

[5] R. V. Field, Jr. and M. Grigoriu. Optimal stochastic models for spacecraft atmospheric re-entry. *Journal of Sound and Vibration*, 290(3–5):991–1014, 2006.

[6] C. E. Forest, P. H. Stone, A. P. Sokolov, M. R. Allen, and M. D. Webster. Quantifying uncertainties in climate system properties with the use of recent climate observations. *Science*, 295(5522):113–117, 2002.

[7] D. Gamerman and H. F. Lopes. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Chapman & Hall, 2006.

[8] M. Gioffrè, V. Gusella, and M. Grigoriu. Simulation of non-Gaussian field applied to wind pressure fluctuations *Probabilistic Engineering Mechanics*, 15(4):339–345, 2000.

[9] M. Grigoriu. Extremes of wave forces. *Journal of Engineering Mechanics*, 110(12):1485–1494, December 1984.

[10] M. Grigoriu *Applied Non-Gaussian Processes*, P T R Prentice-Hall, Englewood Cliffs, NJ, 1995.

[11] M. Grigoriu. Existence and construction of translation models for stationary non-Gaussian processes. *Probabilistic Engineering Mechanics*, 24(4):545–551, October 2009.

[12] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1): 97–109, 1970.

[13] J. R. M. Hosking. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B*, 52(1): 105–124, 1990.

[14] P. Huybers. Compensation between model feedbacks and curtailment of climate sensitivity. *Journal of Climate*, 23: 3009–3018, 2010.

[15] E. S. Keeping. *Introduction to Statistical Inference*, Dover, 1995.

[16] T. M. Lenton, H. Held, E. Kriegler, J. W. Hall, W. Lucht, S. Rahmstorf, and H. J. Schellnhuber. Tipping elements in the Earth's climate system. *Proceedings of the National Academy of Sciences*, 105(6): 1786–1793, 2008.

[17] J. Murphy, D. Sexton, D. Barnett, G. Jones, M. Webb, M. Collins, and D. Stainforth. Quantification of modeling uncertainties in a large ensemble of climate change simulations. *Nature*, 430(7001):768–772, 2004.

[18] T. N. Palmer. The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Quarterly Journal of the Royal Meteorological Society*, 128(581):747–774, 2002.

[19] Program for Climate Model Diagnosis and Intercomparison. Lawrence Livermore National Laboratory, *http://www-pcmdi.llnl.gov*.

[20] G. H. Roe and M. B. Baker. Why is climate sensitivity so unpredictable? *Science*, 318(5850):629–632, 2007.

[21] H. Sauvageot. The probability density function of rain rate and the estimation of rainfall by integral areas. *Journal of Applied Meteorology*, 33(11): 1255–1262, 1994.

[22] S. H. Schneider and J. Lane. An overview of "dangerous" climate change. *Avoiding Dangerous Climate Change*, Cambridge University Press, 2006.

[23] UNFCC. United Nations Framework Convention on Climate Change. *United Nations*, 1992, http://unfccc.int/resource/docs/convkp/conveng.pdf.

[24] A. Zellner. *An Introduction to Bayesian Inference in Econometrics*. John Wiley & Sons, Inc., 1971.

A Calibration of the K-L representation for a Gaussian random field

We present a method to calibrate a covariance model of a Gaussian random field based on the singular value decomposition (SVD) of the data matrix of measurements described in Section 2.2. In practice, the data exist as point measurements over the spatial domain D; let $\mathbf{u}_j \in D$, $j = 1 \ldots m$ be the locations of measurements. We assume samples of the surrogate random field model occur at the $\mathbf{u}_j$. If we wish to sample the model at other points in D, we can use standard spatial interpolation techniques [3].

The kth measurement at the point $\mathbf{u}_j$ is written $g_k(\mathbf{u}_j)$, which we assume comes from a zero mean random field. Define the $m \times n$ data matrix $\mathbf{x}$ with entries

$$\mathbf{x}_{jk} = g_k(\mathbf{u}_j). \tag{21}$$

To calibrate an empirical covariance matrix, we compute the SVD of the zero-mean samples

$$\mathbf{x} = \mathbf{y}\,\boldsymbol{\sigma}\,\mathbf{v}^T. \tag{22}$$

We then pose the following surrogate model for the Gaussian random field $G(\mathbf{u}_j)$:

$$\begin{bmatrix} G(\mathbf{u}_1) \\ \vdots \\ G(\mathbf{u}_m) \end{bmatrix} \approx \frac{1}{\sqrt{n-1}} \sum_{i=1}^{n-1} \sigma_i \mathbf{y}_i W_i, \tag{23}$$

where $\mathbf{y}_i$ are the left singular vectors and σ_i are the singular values. This is similar to the truncated Karhunen-Loéve expansion defined by Eq. (4), since $\mathbf{y}_i$ are the eigenvectors of the empirical covariance matrix $\frac{1}{n-1}\mathbf{x}\,\mathbf{x}^T$, and $\sigma_i^2/(n-1)$ are its eigenvalues. Notice that the maximum possible truncation level is set by the number of available data n. However, we may be able to further truncate the expansion to less than n terms depending on the decay of the singular values. To generate samples of the calibrated field $G(\mathbf{u}_j)$, we draw samples from the iid Gaussian W_i and apply the transformation given by $\mathbf{y}\,\boldsymbol{\sigma}$.

B A generalized normal distribution

Let A be a random variable with PDF $f_A(a|\theta_1, \theta_2, \theta_3)$ defined by Eq. (13). We can show that the support of the distribution, denoted by D_A, is given by

$$D_A = \begin{cases} (-\infty, \theta_1 + \theta_2/\theta_3), & \theta_3 > 0; \\ (\theta_1 + \theta_2/\theta_3, \infty), & \theta_3 < 0; \\ (-\infty, \infty), & \theta_3 = 0, \end{cases} \tag{24}$$

meaning that the PDF defined by Eq. (13) is zero for all $a \notin D_A$. It follows that the mean, variance, and coefficient of skewness of A are

$$\begin{aligned} \mathrm{E}[A] &= \int_{D_A} a\, f_A(a)\, \mathrm{d}a = \theta_1 - \frac{\theta_2}{\theta_3}\left(e^{\theta_3^2/2} - 1\right), \\ \mathrm{var}[A] &= \int_{D_A} (a - \mathrm{E}[A])^2\, f_A(a)\, \mathrm{d}a = \left(\frac{\theta_2}{\theta_3}\right)^2 e^{\theta_3^2}\left(e^{\theta_3^2} - 1\right), \\ \mathrm{skew}[A] &= (\mathrm{var}[A])^{-3/2} \int_{D_A} (a - \mathrm{E}[A])^3\, f_A(a)\, \mathrm{d}a = \frac{3e^{\theta_3^2} - e^{3\theta_3^2} - 2}{\left(e^{\theta_3^2} - 1\right)^{3/2}}\, \mathrm{sign}(\theta_3). \end{aligned} \tag{25}$$

Further, A is a Gaussian random variable with zero skewness if, and only if, $\theta_3 = 0$; A is a non-Gaussian random variable with positive (negative) skewness when $\theta_3 < 0$ ($\theta_3 > 0$).

Given n independent samples of A, denoted by $\mathbf{z} = (z_1, \ldots, z_n)^T$, the likelihood that this collection of samples was drawn from the generalized normal distribution is given by Eq. (3). This is often conveniently expressed as the negative of the log of the likelihood function, that is, $-\ln l(\mathbf{z}|\boldsymbol{\theta})$ where, by Eqs. (3) and (13)

$$\begin{aligned} -\ln l(\mathbf{z}|\boldsymbol{\theta}) = \frac{n}{2}\, \ln 2\pi &+ \sum_{k=1}^{n} \frac{1}{2\,\theta_3^2}\left[\ln\left(1 - \frac{\theta_3(z_k - \theta_1)}{\theta_2}\right)\right]^2 \\ &+ \ln\left(\theta_2 - \theta_3(z_k - \theta_1)\right). \end{aligned} \tag{26}$$

It follows that the values for $\boldsymbol{\theta}$ that maximize the likelihood function also minimize Eq. 26 and are the solution $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)^T$ to the following set of equations

$$\begin{aligned} \sum_{k=1}^{n} \frac{\hat{\theta}_3}{\hat{\theta}_2 - \hat{\theta}_3(z_k - \hat{\theta}_1)} + \frac{\ln \xi_k}{\hat{\theta}_2\, \hat{\theta}_3\, \xi_k} &= 0 \\ \sum_{k=1}^{n} \frac{1}{\hat{\theta}_2 - \hat{\theta}_3(z_k - \hat{\theta}_1)} + \frac{\ln \xi_k}{\hat{\theta}_2^2\, \hat{\theta}_3\, \xi_k}\,(z_k - \hat{\theta}_1) &= 0 \\ \sum_{k=1}^{n} \frac{\hat{\theta}_1 - z_k}{\hat{\theta}_2 - \hat{\theta}_3(z_k - \hat{\theta}_1)} - \frac{(\ln \xi_k)^2}{\hat{\theta}_3^3} - \frac{\ln \xi_k}{\hat{\theta}_2\, \hat{\theta}_3^2\, \xi_k}\,(z_k - \hat{\theta}_1) &= 0 \end{aligned} \tag{27}$$

where $\xi_k = 1 - (\hat{\theta}_3/\hat{\theta}_2)(z_k - \hat{\theta}_1)$.

DISTRIBUTION:

1	MS 0346	R. V. Field, Jr., 1526
1	MS 0370	G. Backus, 1465
1	MS 0370	M. Boslough, 1465
1	MS 0370	M. Taylor, 1442
1	MS 0370	T. Trucano, 1441
1	MS 0847	P. Wilson, 1520
1	MS 1070	C. Wong, 1526
1	MS 1318	P. Constantine, 1441
1	MS 1318	K. Dalbey, 1441
1	MS 1318	M. Eldred, 1441
1	MS 1318	J. Stewart, 1441
1	MS 1318	L. Swiler, 1441
1	MS 9051	B. J. Debusschere, 8351
1	MS 9051	H. Najm, 8351
1	MS 9051	C. Safta, 8954
1	MS 9051	K. Sargsyan, 8351
1	MS 0899	Technical Library, 9536 (electronic copy)

Sandia National Laboratories