Procedia Computer Science

Volume 29, 2014, Pages 1146–1155

ICCS 2014. 14th International Conference on Computational Science

Stochastic Parameterization to Represent Variability and Extremes in Climate Modeling

R. Langan[1,3]*, R. Archibald[1,3], M. Plumlee[5], S. Mahajan[1,2], D. Ricciuto[1,6], C. Yang[4], R. Mei[1,2], J. Mao[1,6], X. Shi[1,6], and J. S. Fu[4].

[1] Climate Change Science Institute, Oak Ridge National Laboratory, Oak Ridge, TN
[2] Computational Earth Sciences Group, Oak Ridge National Laboratory, Oak Ridge, TN
[3] Computational and Applied Mathematics Group, Oak Ridge National Laboratory, Oak Ridge, TN
[4] Department of Civil & Environmental Engineering University of Tennessee Knoxville, TN
[5] Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA
[6] Environmental Science Division, Oak Ridge National Laboratory, Oak Ridge, TN

Abstract

Unresolved sub-grid processes, those which are too small or dissipate too quickly to be captured within a model's spatial resolution, are not adequately parameterized by conventional numerical climate models. Sub-grid heterogeneity is lost in parameterizations that quantify only the 'bulk effect' of sub-grid dynamics on the resolved scales. A unique solution, one unreliant on increased grid resolution, is the employment of stochastic parameterization of the sub-grid to reintroduce variability. We administer this approach in a coupled land-atmosphere model, one that combines the single-column Community Atmosphere Model (CAM-SC) and the single-point Community Land Model (CLM-SP), by incorporating a stochastic representation of sub-grid latent heat flux to force the distribution of precipitation. Sub-grid differences in surface latent heat flux arise from the mosaic of Plant Functional Types (PFT) that describe terrestrial land cover. With the introduction of a stochastic parameterization framework to affect the distribution of sub-grid PFT's, we alter the distribution of convective precipitation over regions with high PFT variability. The stochastically forced precipitation probability density functions (pdf) show lengthened tails, demonstrating the retrieval of rare events. Through model data analysis we show that the stochastic model increases both the frequency and intensity of rare events in comparison to conventional deterministic parameterization.

Keywords:

1 Introduction

The deterministic numerical methods used in climate modeling, which rely on discretization, lead to scale separation and prevent the explicit resolution of small-scale dynamics. Conven-

*Corresponding Author: `langanrt@ornl.gov`

 Selection and peer-review under responsibility of the Scientific Programme Committee of ICCS 2014

doi:10.1016/j.procs.2014.05.103

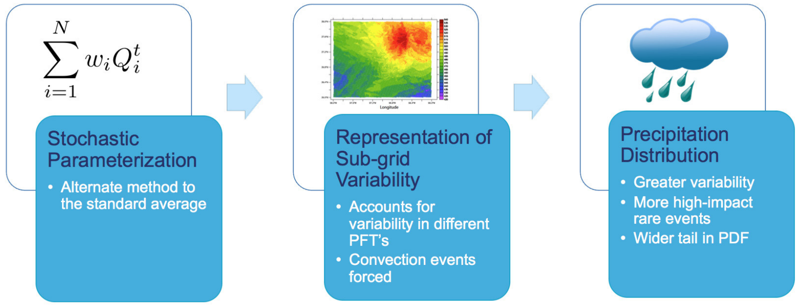


Figure 1: The stochastic model is focused on the weighted average calculation of the latent heat flux, where sub-grid variability better represents and the pdf tails of atmospheric processes like precipitation.

tional closure to this problem involves incorporating the average of sub-grid flows as initial conditions. However, the chaotic behavior of the nonlinear differential equations that describe the climate imply that errors in initial conditions can impact resolved scales. For instance, the use of averaging based parameterizations is known to cause model dampening of variation in climate processes. Statistical methods are widely accepted as an approach to quantifying and remediating model errors symptomatic of the closure problem. Stochastic ensemble modeling, for example, is used to forecast the overall predictability of the climate system through a sampling of initial observation error. In the small-scales, stochastic physics can enhance parameterizations of the sub-grid and simulate the dynamical variability that is characteristic of the climate system [5]. Stochastic forcing in climate has the ability to bridge multi-scale interactions, because of this, it is clear that stochastic forcing will play a primary role in our understanding of non-gaussian statistics and the dynamics of extreme events [14]. With regard to land models, stochastic methods can broaden the influence of surface heterogeneity by channeling variance through sub-grid land cover distributions. In effect, stochastically perturbed land distributions give rise to a wider array of dynamical event outcomes, including extreme or rare events. This paper presents a model that introduces stochastic error to the distribution of sub-grid PFT's to improve latent heat flux variability and extremes on the land boundary layer.

One challenge associated with the closure problem is model underestimation of both precipitation variability and extreme precipitation events [4]. This research offers the statistical representation of land-atmosphere latent heat fluxes as a tool for enhancing extreme precipitation modeling. Figure 1 depicts a stochastic approach to the closure problem for latent heat flux. We target the representation of latent heat flux on the land boundary layer because of its dominant role as a driver of convective precipitation. Latent heat flux rates over land are derived from the phenological characteristics associated with particular PFT's. The diversity of PFT's that constitute a heterogeneous land cover instill variability in the heat fluxes of the land-atmosphere interface. To reproduce the variability of latent heat flux, we develop a stochastic framework that perturbs the sub-grid PFT distribution. We then examine how the incorporation of latent heat flux variability benefits precipitation simulation in the coupled CAM-SC and CLM-SP model.

The focus of our alternate PFT parameterization, is to statistically stimulate variance in the latent heat exchanges of the land-atmosphere interface. Our stochastic parameterization uses an adaptive methodology which allows for the control and optimization of incorporated stochastic

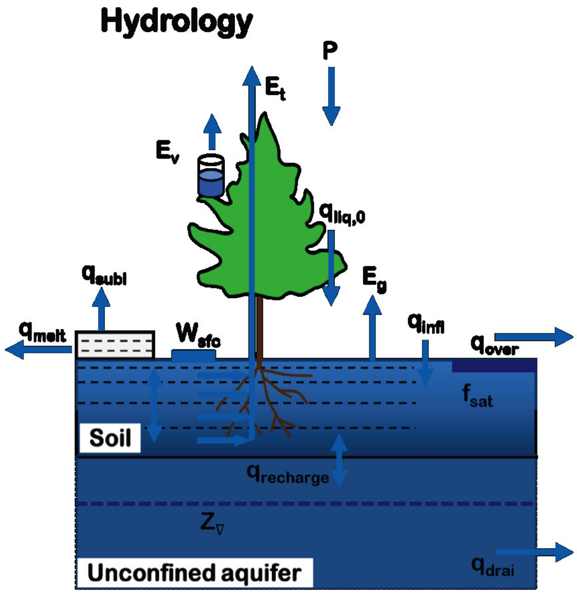


Figure 2: The CLM incorporates the terrestrial ecological impacts on hydrology. (source [10])

intensity. We use a control parameter to facilitate the study of weighted PFT distributions; through the analysis of different noise intensity models we determine the optimal noise model for achieving latent heat flux variability by comparison to observational data. The model is distinct in its use of parameter, β, as well as a Dirichlet distribution, a multivariate distribution of non-negative variables summing to unity with single parametric control of the variance [6]. The Dirichlet distribution, well-suited to statistically model weighted average sums, has been used in a broad number of fields, including evolutionary theory [12], Bayesian statistics [11], geology [3], forensics [8], econometrics [7], transport [1], and population biology [13] to name a few prominent studies. This method of stochastic modeling has yet to be utilized in climate studies.

1.1 CESM

Both the CLM and CAM are components of the Community Earth System Model (CESM), a climate model designed through collaboration with universities, national labs, and the National Center for Atmospheric Research (NCAR). The CESM uses a coupled-climate-system approach combining atmosphere, land surface, ocean and sea ice models to produce a comprehensive model of climate dynamics. This interdisciplinary model was developed to study changes in climate as well as the complex interactions of component processes such as cloud physics, radiative transfer, and boundary layer processes as well as their overall influence on large scale circulation.

The land component of the CESM, the CLM, is well-suited to studies examining the impact of terrestrial ecosystems on the cycling of energy, gases, and water. The biogeophysics component simulates the radiative transfer between the land-atmosphere boundary and incorporates variables such as: reflected solar radiation, surface stresses, latent heat flux, and sensible heat flux. Initial conditions for the CLM can incorporate vegetative and terrestrial influence on climate processes and enable a more precise modeling of biogeophysical fluxes. A few of these significant vegetative characteristics include: stomal structure, albedo, roughness length, and canopy height [2]. When coupled to the CAM, the fluxes associated with the terrestrial ecology are passed to the atmospheric model where they influence larger circulatory processes. Figure 2 depicts the fluxes on the land boundary layer that drive the hydrologic cycle. In the CLM, convective precipitation is parameterized to be dependent on the convective available potential energy (CAPE), a process strongly influenced by these surface heat fluxes [16].

2 Stochastic Model

We divide the description of the stochastic model in two parts; parameterization and stochastic distribution. First we explain the sub-grid latent heat flux parameterization used in the CLM and then present our stochastic distribution that captures the variability of the sub-grid latent heat flux.

2.1 Parameterization

Typical climate models have a horizontal spatial resolution of about $150km$, however, within this sub-grid scale a wide variety of surface features interact with and influence the climate system. Large-scale surface heterogeneity in the CLM is represented by five primary land cover types, namely, glacier, lake, wetland, urban and vegetated as seen in figure 3. PFT's are used to further classify heterogeneities within the vegetated land-unit where they quantify ecological impact on atmospheric circulations. In this way, the CLM represents the entire domain of plant species with fifteen PFT's that are categorized based on their flux-influencing characteristics. Datasets describing the fractional abundance of sub-grid PFT's are then used to generate initial conditions and thus relay the sub-grid impact of the land component. The CLM's Present Day PFT dataset [15] uses satellite products to obtain the distribution and areal extent of PFT's as well as their leaf area index values. These PFT blueprints enable sub-grid heterogeneity parameterization since they reproduce the physical properties of land cover while sustaining the multiple PFT representation.

2.2 Stochastic Distribution

Conventional models incorporate small-scale dynamics by passing up the average influence of sub-grid features on the climate system. The consequence is a loss of information, including the variance of climate-influencing surface characteristics like land composition. We begin with a description of the conventional parameterization where sub-grid latent heat flux is computed as the average of the column at each time-step:

$$\tilde{Q}_i^t = \sum_{i=1}^{N} \alpha_i Q_i^t, \quad \text{where} \quad \sum_{i=1}^{N} \alpha_i = 1. \tag{1}$$

In equation (1), N represents the number of PFTs, α_i is the proportion of the i^{th} functional type and Q_i is the latent heat flux for a specific functional type. The values of α_i are quantities obtained from datasets like the Present Day PFT dataset. We modify the conventional representation (1) and generate stochastic sub-grid representation the following Dirichlet distribution, parameterized:

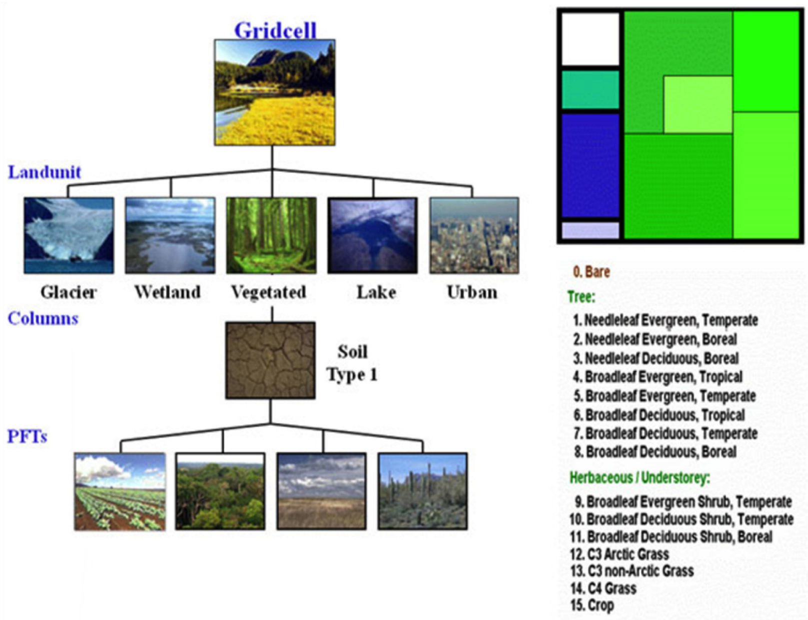


Figure 3: source [9]

$$D(\alpha,\beta), \quad \text{where} \quad \alpha \in \mathbb{R}^N_{>0}, \quad \sum_{i=1}^{N} \alpha_i = 1, \quad \text{and} \quad \beta \in \mathbb{R}_{>0}. \tag{2}$$

The value of α controls the expected value of the output, i.e.,

$$\text{if} \quad \mathbf{W} \sim D(\alpha,\beta) \quad \text{then} \quad \mathbb{E}W_i = \alpha_i. \tag{3}$$

Finally, the value of β controls the variance,

$$\text{var}(W_i) = \alpha_i(1-\alpha_i)/(\beta+1), \tag{4}$$

framing a process in which a large value of β will have little variance and vice versa. The most efficient method for producing draws from the Dirichlet distribution is the following algorithm, based on the work [6]:

Data: Given the PFT Dataset α and variance parameter β
Result: Dirichlet distributed PFT Dataset, $\mathbf{W} \sim D(\alpha,\beta)$
for *i=1 to N* **do**
  $w_i^* \sim \text{Gamma}(\beta, \alpha_i)$
end
for *i=1 to N* **do**
  $w_i = \frac{w_i^*}{\sum_{i=1}^{N} w_i^*}$
end

Algorithm 1: Generating draws from the Dirichlet distribution from PFT dataset

The pdf of this particular parametrization is not commonly used:

$$f(\alpha,\beta) = \frac{\Gamma(\beta)}{\prod_{i=1}^{N}\Gamma(\alpha_i)} \prod_{i=1}^{N} x_i^{\beta\alpha_i - 1} \tag{5}$$

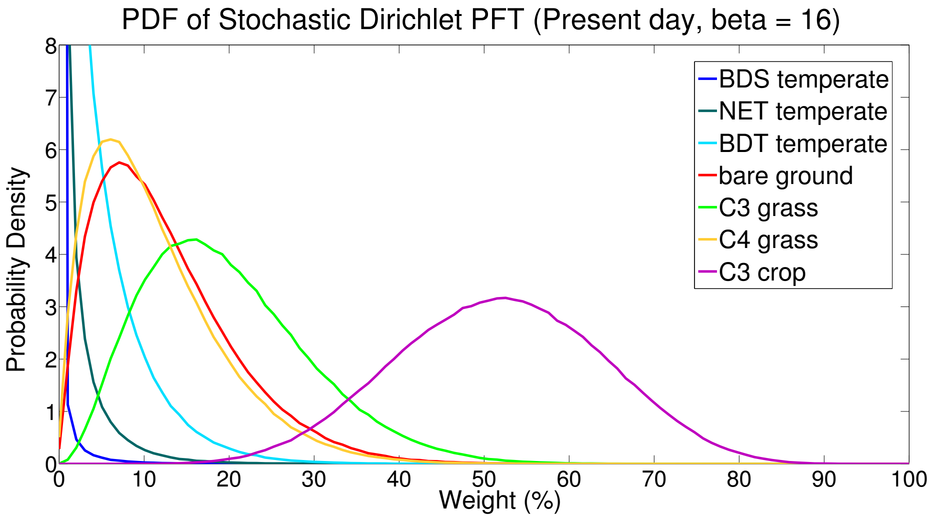


Figure 4: Dirichlet distribution for the PFT dataset at the Lamont, Oklahoma–present day timeframe, T42 spectral resolution, and $\beta = 16$.

The result will depend on the parameter β, and a larger value of β signifies greater confidence in the PFT dataset. As $\beta \to \infty$ the samples will converge to α_i in probability; this convergence result can be derived via Chebyshev's inequality based on the expression for the variance.

3 Methods

We use the single-column feature of the CESM to simulate the land and atmosphere at T42 spectral resolution–approximately 2.8 degrees. The Atmospheric Radiation Measurement (ARM) Southern Great Plains (SGP) facility, funded by the Department of Energy (DOE), provides detailed atmospheric measurements and these data products are used to form our observational datasets [15]. To enable the comparison of model output against real-world data, single-column simulations are run over the Lamont, Oklahoma ARM site. Our control simulation uses the original weighted average method given in (1) to calculate the latent heat flux at each time-step for the PFT dataset. Stochastic simulations incorporate the Dirichlet distribution and β parameter given in (2), modeling the latent heat flux as:

$$\tilde{Q}_i^t = \sum_{i=1}^{N} W_i Q_i^t, \quad \text{where} \quad \sum_{i=1}^{N} W_i = 1, \quad \text{and} \quad \mathbf{W} \sim D(\alpha, \beta). \tag{6}$$

We simulate the stochastic model 50 times each for variation parameter with values $\beta = \{16, 8, 4, 2\}$ to survey the impact of incorporated variance. ARM SGP observational data is then used to tune the model variation parameter, β. Figure 4 is an example Dirichlet distribution generated by our stochastic process that uses the Present Day Dataset for PFT representation, the ARM Lamont site measurement data from time period: 1994-2013, T42 spectral resolution, and $\beta = 16$.

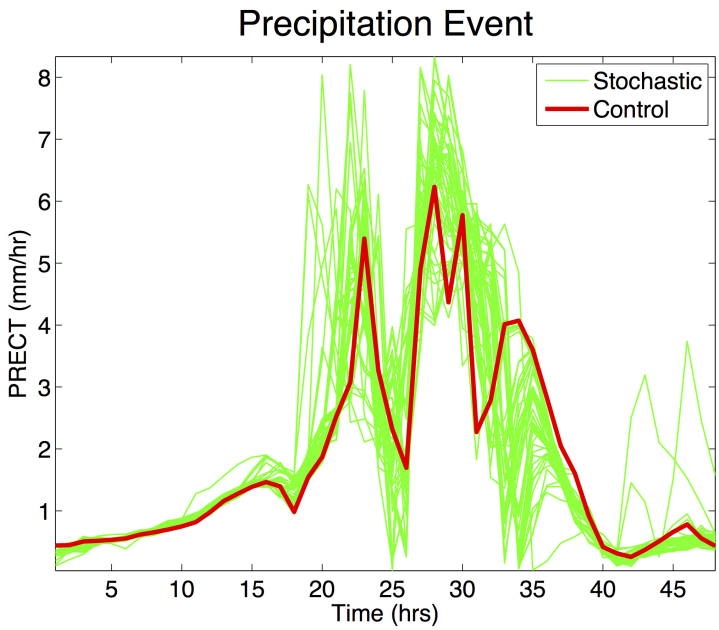


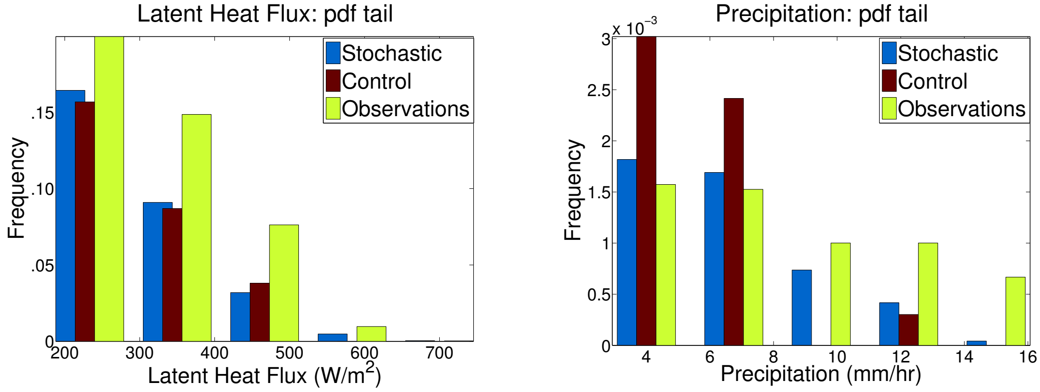
Figure 5: Model simulations of a heavy rainfall event, during late spring, where the precipitation timeline of the control model is in red and the fifty ensemble members of the stochastic model are in green.

4 Results

The single-column model for CESM is a strongly constrained climate simulation, where each stochastic run has the same boundary conditions along the entire column. Each simulation is run over a one year time frame with a time-step of 20 minutes. Part of our analysis considers extreme precipitation events or events occurring less than 10% of the time. We also consider the duration in which rainfall intensity remains at the 'extreme' threshold, a factor used in determining when an extreme event becomes a natural disaster.

Figure 5 depicts a rare, heavy, rainfall event taking place in the control model, visible in red, over the course of two days in the late spring over the Lamont, Oklahoma location. The total average precipitation projected by the control model is $86.6mm$. The fifty stochastic ensemble runs, each utilizing the same single-column boundary conditions, are depicted in green and are forced with variation parameter $\beta = 4$. The boundary conditions strongly constrain the single-column model and all stochastic runs follow the same rain events. Despite this, it can been seen that some stochastic ensemble members are capable of simulating intensifying precipitation for a longer duration. Stochastic ensemble members diversify total rainfall spread in this 48hr period: ensemble members have a mean and standard deviation of $88.9 \pm 3mm$. Collectively, the stochastic ensembles widen the spread of precipitation intensity; individually, certain stochastic ensembles signal a more extreme event than projected by the control. In this way, the ensembles of our stochastic model are more likely to project potential natural disasters.

The effect of the stochastic simulations for the single-column model have their most dramatic effect on the tails of the pdfs. Figure 6 depicts the tails of the latent heat flux and precipitation for the control simulation, the stochastic simulation: $\beta = 4$, and the observational dataset



(a) The pdf tail of the latent heat flux for the stochastic model, the control and the ARM data from Lamont, OK.

(b) The pdf tail of precipitation for the stochastic model, the control and the ARM data from Lamont, OK.

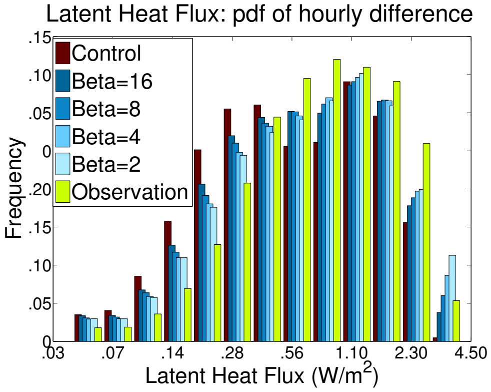
Figure 6

provided by ARM[1]. It can seen in Figure 6 that the observational data display long tail behavior that is not simulated by the control. This demonstrates that a tailored stochastic forcing of variability to the single-column model improves simulation of real-world extreme precipitation. Since the stochastic model acts on the latent heat flux at each time-step, an examination of the hourly difference in latent heat flux provides a sensitivity measure of stochastic impact to single-column simulations. Figure 7a portrays the pdf fit of the hourly difference in latent heat flux of the following data: observational measurements of latent heat flux, the control model simulation of the variable, and fifty ensemble members of the stochastic simulation with variation parameters $\beta = \{16, 8, 4, 2\}$. Figure 7b shows the relative error in simulated latent heat flux for both the control model and and the stochastic model as compared to the measured data. The stochastic model has a lower relative error for each of the variational parameter values, and $\beta = 4$ has the lowest relative error. Thus, the stochastic parameterization with noise value given by $\beta = 4$, provides the best model for latent heat flux variability.

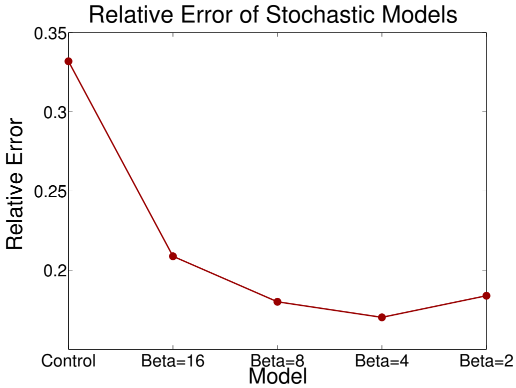
5 Conclusion

For the foreseeable future–as long as grid-size is greatly limited by computational complexity–climate models will require innovative parameterizations to capture features of the climate system. Parameterizations based on averaging techniques dampen variability and as a result, models underpredict extreme events. This research addresses model deficiency in the case of extreme precipitation with a novel stochastic approach to sub-grid parameterization. We show that our method, which statistically represents PFT's, improves the representation of latent heat flux variability and this boosts the frequency of simulated rare precipitation. Consequentially, our model narrows the gap between forecasted rare precipitation events and their frequency in nature. Also, we demonstrate how the unique statistical approach, with tunable beta parameter, provides a chassis for noise model optimization and validation with observational data. The parametrization advances implemented in this study have the potential to improve warning systems for disasters like flooding and drought.

[1] Data available at `http://www.arm.gov`



(a) Frequency estimation of the pdf showing the hourly difference in latent heat flux for the stochastic model with different beta values, control simulation, and ARM observational data.



(b) Relative error of the frequency estimation of the pdf for the hourly difference in latent heat flux for the stochastic model for each beta value and the control, as compared to the observational data.

Figure 7

5.1 Acknowledgments

The submitted manuscript has been authored in part by contractors [UT-Battelle LLC, manager of Oak Ridge National Laboratory (ORNL)] of the U.S. Government under Contract No. DE-AC05-00OR22725. Accordingly, the U.S. Government retains a non-exclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes. Special thanks to the Research Alliance in Math and Science (RAMS) program and the Climate Change Science Institute (CCSI) at ORNL.

References

[1] J. Bakosi and J. R. Ristorcelli. A stochastic diffusion process for lochner's generalized dirichlet distribution. *Journal of Mathematical Physics*, 54(10), 2013.

[2] G. B. Bonan, S. Levis, L. Kergoat, and K. W. Oleson. Landscapes as patches of plant functional types: An integrating concept for climate and ecosystem models. *Global Biogeochemical Cycles*, 16(2), 2002.

[3] F. Chayes. Numerical correlation and petrographic variation. *Journal of Geology*, 70, 1962.

[4] A. M. DeAngelis, A. J. Broccoli, and S. G. Decker. A comparison of cmip3 simulations of precipitation over north america with observations: Daily statistics and circulation features accompanying extreme events. *J. Climate*, 26, 2013.

[5] R. H. Gardner, B. Rojder, and U. Bergstrom. Prism: A systematic method for determining the effect of parameter uncertainties on model predictions. Technical report, Studsvik Energiteknik AB report/NW-83/555, Nyköping, Sweden, 1983.

[6] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian data analysis*. CRC press, 2003.

[7] C. Gourieroux and J. Jasiak. Multivariate jacobi process with application to smooth transitions. *Journal of Econometrics*, 131, 2006.

[8] K. Lange. Applications of the dirichlet distribution to forensic match probabilities. *Genetica*, 96(1-2), 1995.

[9] NCAR. Community land model: Surface heterogeneity, 2013.

[10] K. W. Oleson, D. M. Lawrence, G. B. Bonan, B. Drewniak, M. Huang, C. D. Koven, S. Levis, F. Li, W. J. Riley, Z. M. Subin, S. C. Swenson, and P. E. Thornton. Technical description of version 4.5 of the community land model (clm). Technical Report NCAR/TN-503+STR, Climate and Global Dynamics Division, NCAR Earth System Laboratory, 2013.

[11] C. D. M. Paulino and C. A. De Bragança Pereira. Bayesian methods for categorical data under informative general censoring. *Biometrika*, 82(2), 1995.

[12] K. Pearson. Mathematical contributions to the theory of evolution. on a form of spurious correlation which may arise when indices are used in the measurement of organs. *Royal Society of London Proceedings Series I*, 60, 1896.

[13] M. Steinrucken, Y. X. R. Wang, and Y. S. Song. An explicit transition density expansion for a multi-allelic wright–fisher diffusion with general diploid selection. *Theoretical Population Biology*, 83, 2013.

[14] P. Sura. Stochastic models of climate extremes: Theory and observations. *Extremes in a Changing Climate–Detection, Analysis and Uncertainty. Water Science and Technology Library*, 65, 2013.

[15] S. Xie, R. B. McCoy, S.A. Klein, R. T. Cederwall, W. J. Wiscombe, E.E. Clothiaux, K. L. Gaustad, J. C. Golaz, S. D. Hall, M. P. Jensen, K. L. Johnson, Y. Lin, C. N. Long, J. H. Mather, R. A. McCord, S. A. McFarlane, G. Palanisamy, Y. Shi, and D.D. Turner. Clouds and more: Arm climate modeling best estimate data. *Bull. Amer. Meteor. Soc.*, 91, 2013.

[16] G. J. Zhang and N. A McFarlane. Sensitivity of climate simulations to the parameterization of cumulus convection in the canadian climate centre general circulation model. *Atmosphere-Ocean*, 33(3), 1995.