

POSITION PAPER

Statistical mechanics in climate emulation: Challenges and perspectives

Ivan Sudakow^{1,2,*} , Michael Pokojov³ and Dmitry Lyakhov⁴

¹School of Mathematics and Statistics, The Open University, Milton Keynes, United Kingdom

²Department of Physics, University of Dayton, Dayton, Ohio, USA

³Department of Mathematical Sciences, The University of Texas at El Paso, El Paso, Texas, USA

⁴Visual Computing Center, King Abdullah University of Science and Technology (KAUST), Thuwal, Kingdom of Saudi Arabia

*Corresponding author. E-mail: ivan.sudakow@open.ac.uk

Received: 01 April 2022; **Revised:** 14 August 2022; **Accepted:** 13 September 2022

Keywords: Climate; emulator; machine learning; statistical mechanics; statistical modeling

Abstract

Climate emulators are a powerful instrument for climate modeling, especially in terms of reducing the computational load for simulating spatiotemporal processes associated with climate systems. The most important type of emulators are statistical emulators trained on the output of an ensemble of simulations from various climate models. However, such emulators oftentimes fail to capture the “physics” of a system that can be detrimental for unveiling critical processes that lead to climate tipping points. Historically, statistical mechanics emerged as a tool to resolve the constraints on physics using statistics. We discuss how climate emulators rooted in statistical mechanics and machine learning can give rise to new climate models that are more reliable and require less observational and computational resources. Our goal is to stimulate discussion on how statistical climate emulators can further be improved with the help of statistical mechanics which, in turn, may reignite the interest of statistical community in statistical mechanics of complex systems.

Impact Statement

This perspective paper assesses the potential for improving the performance of climate emulators with the help of techniques from machine learning and statistical mechanics. It is meant to be accessible to a wide readership and to shed light on this emerging field of climate modeling. The paper is jointly written by a physicist, a statistician, and a computational scientist, which guarantees a multifaceted view.

1. Introduction

Incorporating various physical processes into contemporary climate models has greatly improved their predictive power. Nowadays, climate models typically aggregate a broad spectrum of sensor readings and other types of data. A major increase in resolution has resulted in dramatic improvements in weather forecasting quality over the past 30 years (Masson-Delmotte et al., 2021). Unfortunately, the spread between climate model projections has not changed in years, indicating slow progress despite the greatly

  This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

improved physics (Franzke et al., 2022). The climate modeling community has responded to this challenge by identifying a long list of suggested improvements aimed at a comprehensive “multi-physics” climate model. But is “more” necessarily “better?” Adding more physical components comes at a cost as we approach a feasibility limit with respect to model complexity. Computational power keeps pace with demand, but it already takes years to run a set of climate simulations.

The overwhelming complexity of “multi-physics” models and excessive computational resource requirements can be addressed to a certain extent by climate emulators. While no common rigorous definition of climate emulation exists, climate emulators are often grouped into two broad classes: simple or reduced-order deterministic physics models that describe the evolution of some global climate characteristics as output and statistical models serving as “toolboxes” for climate data analysis and statistical projection.

This poses the question if we can develop climate emulators that solely combine only some very basic and most important physics with statistical prediction to produce climate projections. Statistical mechanics offers a potential remedy to this problem. Historically, statistical mechanics applied probability theory and statistics to large ensembles of microscopic entities to explain macroscopic phenomena. These techniques can also drastically reduce the amount of required knowledge about the physics of a particular system. In addition, by employing *machine learning* (ML) (Mehta et al., 2019), we can drastically reduce the computational burden of emulating physical processes.

In classical physical climate emulation, for a given selection of physical models, we calculate their physical properties and describe the associated spatial structures. In contrast, developing a statistical mechanics-based emulator amounts to solving the “inverse problem” of finding a suitable statistical model for a given spatial structure measured from available data and analyzed using ML techniques.

In Section 2, we discuss what a climate emulator is and what is the difference between climate simulation and emulation. In Section 3, we describe how statistical modeling and ML can be used to emulate physical processes, in particular, in connection with climate modeling. In Section 4, we discuss a new type of climate emulators for sea ice modeling. The elements of Earth’s cryosphere, such as the summer Arctic sea ice pack, are declining at unprecedented rates that have far outpaced the projections of physical climate emulators. Understanding key processes, such as the spatial evolution of melt ponds that form atop of the Arctic sea ice and control its optical properties, is crucial for improving climate projections. Finally, Section 5 summarizes the general approach to building statistical mechanics-based climate emulators.

2. What is a Climate Emulator?

In addition to the two traditional—the empirical and the theoretical—paradigms of science, the scientific and technological revolutions of the last century gave rise to the computational (1950s and onward) and the (big) data-driven (2000s and onward) paradigms (Agrawal and Choudhary, 2016; Schleder et al., 2019). The computational paradigm is rooted in and intertwined with the theoretical paradigm and fueled by the rapid development in computational capabilities. It is not only seminal for modern climate research, but owes its recent progress to its paramount importance in climate science (Edwards, 2011). Adopting the computational paradigm, costly or practically infeasible physical experiments, which is often the case in ecological and climate research, are replaced with computer models, also known as simulators (Overstall and Woods, 2016). Being typically based on numerical solutions of large systems of ordinary, stochastic or partial differential equations coupled with algebraic or other types of operator equations and inclusions, a simulator gives rise to a mathematical function (also referred to as forward operator). The latter maps the parameters and input data to the outputs of the system. Due to the high complexity of most simulators, this mapping is typically evaluated only for a small number of points carefully selected from the parameter and the input space as part of a computer experiment (Sacks et al., 1989). These evaluations are then used to put forth a surrogate model, termed (statistical) emulator, designed to mimic the output of the system for any selection of parameter and input values (such as various forcing terms) without running the simulator. The emulator can then be used to replace or supplement the simulator when solving a variety of direct

(inference, prediction, etc.) or indirect (optimization and control, parameter estimation/calibration, model validation, etc.) problems (Overstall and Woods, 2016). We refer to Rougier and Goldstein (2014) for a detailed discussion on climate simulators versus emulators.

The recent IPCC report (Masson-Delmotte et al., 2021) defines climate emulators as a type of simplified or reduced physical deterministic models that form the basis of Earth Systems/Global Climate Models or can be used independently to define projections for critical physical variables describing the most rapid processes in the Earth's system (global temperature change, sea level rise, etc.). However, there exists a solid trend to use an emulator as a tool that is statistically trained on the output of an ensemble of simulations in various climate models (Holden et al., 2015). The use of physics-based emulators requires a quantitative understanding of the framework conditions within which an emulator has acceptable fidelity, either to the full physics model or to the observations of the natural phenomenon. Oftentimes, statistical climate emulators are mere tools to reproduce the climate model output; however, they also have the potential to fill in the gaps in physics-based models when the understanding is not yet sufficient for complete physics-based process modeling. This is a place where statistics meets physics giving rise to statistical mechanics. The central problem of statistical mechanics is computing averages over ensembles of physical quantities, and the principal difficulty is the intractability of those averages for large systems. The standard simulation tool in statistical mechanics is the Monte Carlo method, in particular, the Metropolis algorithm, where a Markov chain starts in some initial state and then “converges” toward an equilibrium state which has to be investigated statistically. A summary of Monte Carlo simulations for climate emulation can be found in Katz (2002) and Baez and Tweed (2013). A particular example of a climate emulator for estimating CO₂ emission based on Monte Carlo techniques is presented in Tsutsui (2021).

In line with the contemporary (big) data-driven paradigm, many researchers are seeking to circumvent or at least minimize the amount of theoretical modeling. Instead of using computer simulations founded upon rational theories (such as fluid dynamics or thermodynamics), the idea they adopt is to directly apply ML techniques to put forth statistical emulators. Simulation-based climate emulators require a lot of computational resources in order to account for a multitude of natural phenomena to provide high-fidelity results. In contrast, data-driven modeling augmented with principal component analysis or nonlinear manifold learning techniques by solely relying on statistical information unwrap a possibility for real-time and low-resource application. Despite numerous attractive sides, purely data-driven approaches bear risks and disadvantages. Such emulators are oftentimes “black boxes” solely developed to optimize some performance metric during the training phase. As such, not only do they typically lack interpretability and transparency, but can be unduly influenced by spurious correlations, may contain and amplify biases, and lead to conclusions that are not backed by the physical system. The recently established field of explainable artificial intelligence or explainable ML (Gagne et al., 2019; Chakraborty et al., 2021; Masrur et al., 2021; Xu et al., 2021) is aimed at improving the explainability of *artificial intelligence* (AI)/machine learning (ML) models by turning them into transparent “glass boxes.” Additionally, the purely data-driven approach may give rise to emulators that violate the laws of physics. This may render predictions unreliable if not outright contradictory and make them unfit for use in decision-making. Accounting for these and other risks and limitations remains a major challenge in data-driven climate research.

3. Statistical Inference and Machine Learning for Emulating Climate Processes

Over the recent years, ML and AI techniques have proved remarkably beneficial in various fields of computer-aided image data analysis, in particular, computer vision and image processing. Despite some recent progress (Kasim et al., 2020; Weber et al., 2020), however, when applied to modeling temporal dynamics of spatially distributed physical systems, these generic approaches suffer from a variety of challenges when trying to mimic the behavior of numerical solutions to the more conventional partial differential equations. To adequately capture the latter dynamics, it is important to preserve intrinsic properties of the underlying differential equation systems (e.g., Lie symmetries, conservation laws, and

symplecticity) in order to get a physically meaningful picture both for short time spans and on long-term horizons. The statistical nature of ML makes it efficacious at discovering some superficial “autoregressive” patterns but fails at unveiling fundamental properties which can otherwise be discovered using mathematical analysis of differential equations. Eventually, it turns out that in order to get meaningful physics (re)produced by neural networks, we have to penalize any violation of constraints which are mathematical consequences of respective physical systems. The traditional multilayer perceptron model augmented with additional N conservation layers allows to enforce N conservation laws by a special form of the loss function which is calculated over the entire output. By so doing, the result will approximately adhere to the constraints manifold (Beucler et al., 2021). The importance of preserving basic conservation quantities for climate emulators is widely discussed in literature (Jensen, 2021). Imbalances in energy or momentum are known to lead to unrealistic long-term behavior even in case of simple mechanical systems. See Kashinath et al. (2021) for detailed discussion.

Weather and climate simulations emerge from a complex interplay of various meteorological phenomena. It is readily possible to derive complex natural phenomena from first principles. Then, instead of directly using information obtained from sensors that is prone to biases and noises or may be confounded with unknown factors, it is often advantageous to simulate synthetic data, for example, by using generative adversarial networks (GANs; Meyer and Nagler, 2021). These synthetic (typically) spatio-temporal data, being unaffected by the aforescribed perils, can then be studied with the aid of statistical methods. Encoding information from such time-series data was the subject of intensive research in ML. Long short-term memory (LSTM) machines, the improved version of the recurrent neural networks, resolved the problem of vanishing or exploding gradients in the back-propagation step performed during the training phase, which allows them to store almost arbitrary long-term dependencies in the input sequences. As recently as in 2018, OpenAI trained a robot using LSTM algorithms to manipulate the human hand with unprecedented accuracy. Nevertheless, this promising architecture had only few applications in complicated tasks of climate prediction with a lot of intrinsic statistical dependencies. Preserving physical laws in time-series learning is one of the perspective research directions in this area.

Modern statistical emulators in climate modeling and research are very versatile. Some of the more prominent approaches include, but are not limited to, Bayesian emulation with Gaussian processes, non- and semiparametric Bayesian emulation, hierarchical models and ensembles, conventional statistical

Table 1. Modern approaches to climate emulation.

Approach	Pros and Cons
Bayesian emulation with Gaussian processes: Drignei and Morris (2006), Ghosh et al. (2014), Overstall and Woods (2016), and Young and Ratto (2011)	Pros: easy estimation and inference; Cons: poor flexibility, poor explainability, not suitable in skewed and heavy-tailed contexts
Non- and semiparametric Bayesian emulation: Antoniano-Villalobos et al. (2020) and Llorente et al. (2021)	Pros: good flexibility; Cons: computationally challenging to calibrate, prone to “curse of dimensionality,” poor explainability
Hierarchical models and ensembles: Castruccio et al. (2014), Schwarber et al. (2019), and Tran et al. (2016)	Pros: improved flexibility (for ensembles), improved explainability and ability to handle multiple scales (for hierarchical models); Cons: harder to train, prone to bias
Conventional statistical learning: Feng et al. (2015) and Nichol et al. (2021)	Pros: good flexibility, easy to train; trade-off between flexibility and explainability usually possible; Cons: prone to “curse of dimensionality”
Deep learning: Guillaumin and Zanna (2021), Kasim et al. (2020), and Weber et al. (2020)	Pros: exceptional flexibility, resistance to “curse of dimensionality;” Cons: computationally challenging to calibrate, reliance on large training datasets, poor explainability
Reservoir computers and echo state networks: Nadiga (2021) and Ouyang and Lu (2018)	Pros: improved flexibility, still fewer parameters/easy to train, analog implementations possible; Cons: no explainability

learning, deep learning, and reservoir computers and echo state networks (ESNs). See Table 1. Often-times, hybrid approaches are employed to improve and customize the trade-off between pros and cons of individual approaches.

While it may also be tempting to try to categorize climate emulators into those based on supervised or unsupervised learning, no “clear-cut” classification appears to be possible for a variety of reasons. First, many climate emulators have both supervised and unsupervised learning aspects to them, for example, may include unsupervised principal component analysis or autoencoders combined with supervised artificial neural network (ANN) models for learning the autoregressive pattern. Second, optimization heuristics based on supervised learning or even reinforcement learning are sometimes used at a “meta” level to calibrate or update the model. Third, other paradigms, such as transfer learning, are sometimes additionally employed. Therefore, instead of forcing climate emulators into the “Procrustean bed” of supervised versus unsupervised learning, we rather decided to structure the following presentation based on the primary type of model(s) employed. Admittedly, even this approach has some degree of subjectivity to it as some emulators can still exhibit a hybrid nature.

3.1. Bayesian emulation with Gaussian processes

While allowing for easy estimation, inference, and uncertainty quantification, Bayesian emulators with Gaussian processes tend to have poor flexibility and explainability and may be not suitable in skewed and heavy-tailed situations. Nonetheless, they remain popular due to their simplicity.

Focusing on dynamical systems described by a system of partial differential equations discretized by a finite difference scheme, Drignei and Morris (2006) proposed an empirical Bayesian approach to computationally efficient surrogates, investigated their approximation quality, and illustrated how this approach can be applied to modeling nonlinear parabolic dynamics of diffusion processes. In the context of design of (computer) experiments, Urban and Fricker (2010) advocated for adopting so-called Latin hypercube designs in lieu of regular designs, especially in high-dimensional parameter spaces. Considering a simple ad hoc multi-physics climate model consisting of several well-known physical models, the “predictive skill” (measured by the root MSE) of both approaches was compared providing strong evidence for the efficacy of the Latin hypercube design at reducing prediction uncertainty. Bayesian emulators for general (finite-dimensional) multivariate models were previously considered by Overstall and Woods (2016), whereas Young and Ratto (2011) specifically concentrated on low-order emulators for linear dynamical systems. Hauser et al. (2012) discussed how ANNs can be used in Bayesian calibration of climate models. Li and Sun (2019) developed a new efficient way of estimating nonstationary mean and/or covariance functions (e.g., in connection with transition between the land and the ocean or between mountains and plains) with application to high-resolution climate model emulation. Using local polynomial approximation of spatially varying parameters in the Matérn covariance function, they proposed a maximum-likelihood estimation procedure and applied it to analyzing precipitation data.

3.2. Non- and semiparametric Bayesian emulation

Non- and semiparametric Bayesian emulation are much more flexible than their Gaussian counterparts but pose major computational challenges at the calibration stage and are prone to the “curse of dimensionality.” Furthermore, they are even more difficult to explain and interpret than Gaussian ones.

Focusing on partially observed Markovian discrete-time processes arising from dynamic state-space models, Ghosh et al. (2014) proposed a Markov chain Monte Carlo (MCMC) scheme for estimating the underlying nonparametric model. Holden et al. (2015) demonstrated how low-rank approximations obtained using singular value decomposition (SVD) can be used in statistical emulation, in particular, with application to modeling the global climate and vegetation fields and how they affected by changes in the Earth’s orbit. To construct a statistical emulator for a forward mapping of interest, Villagran et al. (2016) proposed a nonparametric sampling method to estimate the posterior distribution of respective

parameters by utilizing Voronoi tessellation of the parameter space as an efficient way to generate new points from the posterior distribution without additional evaluations of the forward map.

Adopting the general framework of design and analysis of computer experiments, Antoniano-Villalobos et al. (2020) investigated general multivariate simulators (forward mappings) perturbed by a stochastic “error” term, proposed a statistical emulation procedure based on nonparametric Bayesian density estimation within the one-sample design along with corresponding uncertainty quantification instruments and put forth appropriate probabilistic sensitivity measures. Llorente et al. (2021) presented a new adaptive importance sampling emulation procedure, referred to as regression-based adaptive deep importance sampling framework, based on adaptive regression aimed at minimizing the discrepancy between the proposal and the target density in order to approximate the posterior distribution. Their methodology can be applied to solving both direct (prediction) and inverse (model estimation/calibration) problems, both being of paramount importance in climate modeling. They also developed calibration schemes based on customized MCMC procedure and applied them to several well-known climate models including the community atmospheric model (CAM3.1) provided by the National Center for Atmospheric Research.

Guinness and Hammerling (2018) discussed the impact of data compression on climate models and proposed a statistical compression/decompression algorithm based on summary statistics (akin to features used in ML terminology) and developed a statistical model for the distribution of the full dataset conditioned on summary statistics. Bao et al. (2016) investigated the opposite extreme. They developed an emulator for the CCSM3 climate model conditioned on the past trajectory of atmospheric CO₂ concentrations using just one CO₂ trajectory. To address data scarcity, Meyer et al. (2021) developed a copula-based synthetic data augmentation to facilitate efficacious and reliable training of ML models that are known to heavily rely on voluminous training datasets to assure that a high generalizability level can be attained.

3.3. *Hierarchical models and ensembles*

Hierarchical models and ensembles have a number of attractive properties. While ensembles of emulators allow for improved flexibility, hierarchical models offer improved explainability and ability to handle climatological phenomena at multiple scales. On the downside, they are more complex to train than individual models and can be prone to bias.

Castruccio et al. (2014) developed a computationally efficient statistical emulator for coupled climate models under arbitrary forcing scenarios to predict future temperature and precipitation based on the past trajectories of CO₂ concentrations. Tran et al. (2016) investigated hierarchical climate models constructed using multilevel emulators (e.g., with respect to spatiotemporal or parameter space resolution) coupled through boundary conditions on respective interfaces. Applying this idea to planet simulator (PLASIM) and energy–moisture balance model emulators, the compound emulator was able to explain more than 90% of variation across the validation ensemble. Schwarber et al. (2019) performed extensive impulse testing of some popular simple climate models (SCMs) with respect to three chemical species (CO₂, CH₄, and black carbon) to understand the fundamental gas cycle. While comprehensive SCMs were found to perform better than idealized SCMs, all of them failed to adequately respond to black carbon emission perturbations, suggesting that appropriate modifications to respective emulations procedures need to be made. Holden et al. (2019) provided a description of their new “paleoclimate PLASIM–grid-enabled integrated earth system model (ESM) emulator” PALEO-PGEM and documented how it was applied to obtain high-resolution spatiotemporal climate reconstruction over the past 5 million years as it is related to the evolution of the human species. Gaussian process emulation of the SVD procedure applied to the output from ensembles of intermediate-complexity atmosphere–ocean general circulation model underlies the proposed approach.

Dorheim et al. (2020) described the challenges of calibrating statistical emulators based on their experience with Hector v2.5.0 Simplified Climate Model. In particular, they discovered that the emulator must be constrained with multiple output variables to ensure physicality of the output.

Beusch et al. (2020) employed a modular ESM emulator to produce large “crossbred” multimodel constrained ensembles of regionally optimized land temperature projections. Applied to Coupled Model Intercomparison Project (CMIP6) models, they obtained an ensemble combining the most attractive features of these ESMs at both global and local scales. Tebaldi et al. (2020) discussed how statistical emulation can be used to evaluate climate extreme indices under various future scenarios and proposed an error measure to distinguish between systematic emulation errors and internal variability with application to global temperatures. Miftakhova et al. (2020) proposed a low-dimensional time-series emulator for climate models based on artificially designed uncorrelated CO₂ emissions scenarios. Applied to emulating MAGICC climate model, mean relative out-of-sample forecast errors did not exceed 2%. Yuan et al. (2021) developed an emulator for the mean and variation fields on high-resolution land grids for the global temperature conditioned on green-house gas emissions and showed efficiency under diverse emission scenarios.

3.4. Conventional statistical learning

Climate emulators based on conventional statistical learning techniques, such as nonparametric random forests and support vector machines, are not only quite flexible and easy to train, but offer researchers a trade-off mechanism between flexibility and explainability. At the same time, such emulators can be proved to the “curse of dimensionality” in the presence of noisy or uninformative features.

Using nonparametric random forest regression and computing Gini feature importance measures, Nichol et al. (2021) concluded that the energy exascale earth system model (E3SM) disproportionately relies on some of the climatological quantities when predicting September sea ice averages, which may explain why this model tends to underestimate Arctic sea ice loss. This approach outperformed their earlier contribution (Feng et al., 2015) based on wavelet analysis-support vector machines. Mansfield et al. (2020) proposed an ML approach to uncover relationships between short- and long-term temperature responses to different climate forcing scenarios from a given dataset of climate model simulations. This approach not only accelerates long-term climate emulation, but provides an instrument for early detection of changes and their causes. Watson-Parris (2021) highlighted significant discrepancies between climate and weather emulation from the standpoint of ML methodology involved.

3.5. Deep learning

Deep learning approaches offer exceptional flexibility and resistance to the “curse of dimensionality.” At the same time, they typically rely on large training datasets and can be computationally challenging to calibrate. Furthermore, unless additional techniques such as gradient-weighted class activation mapping (Selvaraju et al., 2017) are employed, deep learning emulators have poor explainability.

In addition to conventional Bayesian framework widely used in climate emulation, alternative ML approaches (both deterministic and frequentist) have recently been proposed. Weber et al. (2020) adopted a convolutional neural network (CNN) approach to emulating precipitation in ESMs and applied it to global 1850–1989 precipitation data. Kasim et al. (2020) developed a deep emulator network search algorithm to search through the space of neural network architectures as to optimally adapt to given dynamics. The proposed approach was applied to a wide variety of situations in physics, climatology, seismology, and other fields and shown to produce computationally efficient and accurate emulation results. Gadat et al. (2021) developed a hybrid downscaling method based on U-Net architecture to learn the relationship between large-scale predictors and a local surface variable of interest over the domain of a regional climate model under consideration. Guillaumin and Zanna (2021) trained a CNN on the outputs of the CM2.6 model to obtain a stochastic Deep Learning parameterization of subgrid momentum forcing within macroscale ocean equations capable of predicting both location and scale parameters of the underlying Gaussian distribution. Xu et al. (2021) discussed a variety of post hoc “explanation methods” based on appropriate feature importance measure in the context of multiple-input-single-output emulators

incorporating a DenseNet encoder. They demonstrated how this approach can be used to visualize features that are important for model prediction.

3.6. Reservoir computers and echo state networks

Reservoir computers and ESNs have improved flexibility over climate emulators based on traditional statistical learning. They typically involve a much smaller number of parameters to tune and, therefore, are significantly easier to train than their deep learning counterparts. Moreover, they allow for analog implementations. Unfortunately, they generally offer no explainability.

In addition to the more prominent ANN models, a very promising new direction is given by reservoir computing (RC). This ML paradigm is capable of capturing the chaotic dynamics typical for highly nonlinear climate systems using reduced-order dynamical models, typically based on a system of nonlinear delay differential equations. In a recent contribution, Nadiga (2021) developed reduced-order climate models using RC and demonstrated that their predictive skill improves upon the linear inverse model and can successfully work even with limited training data. In addition to studying the classical Lorenz dynamics, the proposed approach was applied to emulate the dynamics of the sea surface temperature in the North Atlantic Ocean in the preindustrial control run of the CESM2 climate model. Ouyang and Lu (2018) used ESNs (in addition to multigene genetic programming) for forecasting monthly rainfall and showed that they perform favorably compared to support vector regression.

3.7. Summary

The approaches to climate emulation discussed in this section employ various statistical modeling and ML tools that oftentimes have both supervised and unsupervised aspects to them. Complex physical models derived from first principles (supplemented with appropriate material laws) could give a lot of information about the object under consideration, but typically require major computational resources. For example, Navier–Stokes equations give us full information about incompressible fluid phenomena. Nevertheless, it is hardly possible to use them directly in most practical tasks. Instead, a number of practically feasible “approximate” models were derived. For example, Large Eddy Simulations use averaged velocities and reduce computational complexity for orders of magnitudes. This classical example shows how physical emulators are powerful in complex physical problems. Combined with statistical emulators, we strongly believe that this approach can provide a real possibility for fast and high-fidelity results, which could open a window for predictions in real time.

A survey of statistical models and ML to build climate emulators shows that there is a potential for bridging them via statistical mechanics. In the next section, we introduce a simple didactic example to explain this methodology. We will consider the classical statistical mechanics model that was originally applied to study the spatial structure of physical systems. It has the power of a statistical model but simple computational realizations and clear explainability.

4. Example: Sea Ice Emulator

The emulation of spatial patterns forming due to climate processes can be useful for climate models when the spatial structure defines the physical parametrization of these models. The pattern formation is usually a stochastic process so that respective emulations have probabilistic nature. For example, refer to a stochastic model of multcloud patterns for organized tropical convection (Khouider, 2014) or a probabilistic model emulating forming lakes on permafrost (van Huissteden et al., 2011). In this section, we present an example of climate emulation with application to sea ice modeling. This example is meant to illustrate how statistical physics can serve as a bridge between physics-based climate emulators and statistics-based climate emulators. Having its origins in statistical physics, the famous Ising model employed in our example can also be viewed as a single-layer Boltzmann machine of unsupervised

learning (Welling and Teh, 2003). Additionally, in contrast to “black-box” models produced by general Ising machines or GANs, our emulator is quite easily explainable.

While snow and ice reflect most incident sunlight, melt ponds on the top of the Arctic ice pack and the ocean absorb most of it. The overall reflectance or albedo of sea ice is determined by the evolution of melt pond spatial structure (Perovich et al., 2002, 2009; Polashenski et al., 2012). As melting increases, so does solar absorption, which leads to more melting inducing positive feedback. This ice–albedo feedback has played a significant role in the decline of the summer Arctic ice pack (Perovich et al., 2008) that is melting at precipitous rates that have far outpaced the projections of climate emulators (Serreze et al., 2007). To reproduce observed melt pond spatial configurations, Ma et al. (2019) created a model akin the random field Ising model (RFIM; Krapivsky et al., 2010). The “Ising model” has been widely used in the theory of lattice models of statistical mechanics as a special case of Markov random fields (MRFs) or Markov networks (Izenman, 2021).

The Ising model is the simplest form of a discrete MRFs defined on a discrete lattice Λ of sites where each site takes values from a finite set of states S with probability

$$p(x) = \frac{1}{Z} \exp(-E(x)),$$

where Z is the normalizing constant known as the partition function and $E(x)$ is the energy function. Evaluation of Z requires complex summation that cannot be computed directly, except for trivial cases (see the recent review by Hernandez-Lemus (2021)). For most MRFs, there is no closed form expression for the partition function and, therefore, direct sampling is not feasible. However, we can produce (approximate) samples from these models using MCMC simulations (Izenman, 2021).

In this context, the ponds are modeled using binary variable representing the presence of melt water or ice on the sea ice surface. With the lattice spacing determined by snow topography data as the only measured input into the model, energy minimization drives the system toward realistic pond configurations from an initial random state. The model captures the essential mechanism of pattern formation of Arctic melt ponds, with predictions that agree very closely with observed scaling of pond sizes (Huang et al., 2016). In particular, the energy of the sea ice system is defined as

$$E = \sum_i h_i s_i - \sum_{\langle i, j \rangle} J s_i s_j,$$

where s_i denote binary variables ($s_i = \pm 1$) located at the vertices of a given lattice and J is the coupling (interaction) constant to be taken J sufficiently large. The first sum runs over all bonds (i, j) of the considered lattice, whereas the second runs over all nodes (i) . The random fields h_i are taken according to a given probability distribution $P(h_i)$.

The key factor affecting melt pond configurations is the pre-melt ice topography (Polashenski et al., 2012), in our case, represented by random field h_i . In the spirit of creating order from disorder, these variables are assumed to be independent Gaussian with zero mean and variance σ^2 . The scale (or “bandwidth”) σ in the probability density function $P(h) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{h^2}{2\sigma^2}\right]$ of the underlying distribution is referred to as the “randomness” of the RFIM (Newman and Barkema, 1996). This type of MRFs can be computationally realized through Glauber dynamics (Glauber, 1963), namely, an MCMC method for sampling from a given probability distribution by constructing a Markov chain achieving the desired distribution as its unique stationary distribution (Levin and Peres, 2017).

The pre-melt ice topography is the main parameter for melt pond shaping. Usually, it is generated as independent Gaussian. However, to get a more realistic configuration of melt ponds in emulation processes, appropriate statistical and ML approaches are a more suitable alternative for “substituting” the missing ground truth. For example, statistical mechanics provides the method of “statistical topography,” which has been widely used in various areas ranging from the problems of electronic transport in disordered media to studying patterns of natural coastlines and islands. This method models the shape of random fields, with a special emphasis on contour lines and surfaces of a random potential (Isichenko, 1992). See Adler and Taylor (2007) for a study on statistical topography of Gaussian random

fields. Based on the ideas from statistical topography, Bowen et al. (2018) simulated melt pond topography using random surfaces with level sets representing the water level of melt ponds. They used a finite cosine expansion with a phase given by independent identically distributed (IID) uniform random variables on $[0, 2\pi]$ and amplitude coefficients given by an autoregressive relationship (Kennedy, 2008).

To improve modeling quality of sea ice topography, correlated random Gaussian surfaces of sea ice can be generated using the Fourier filtering method (De Castro et al., 2017). To this end, define a complex function $\eta(\omega_1, \omega_2)$ in the Fourier space

$$\eta(\omega_1, \omega_2) = \sqrt{S(\omega_1, \omega_2)} u(\omega_1, \omega_2) \exp(2\pi i \phi(\omega_1, \omega_2))$$

and take the inverse Fourier transform to recover $h(x) = h(x_1, x_2)$ in the original space with $h(x_1, x_2)$ denoting the height at coordinates (x_1, x_2) . Here, (ω_1, ω_2) is the Fourier frequency, S is a given power spectrum, u 's are independent (not necessarily IID) Gaussian, and ϕ 's are IID uniform random variables on $[0, 2\pi]$ (across (ω_1, ω_2) 's). Applying the inverse discrete Fourier transform to $\eta(\omega_1, \omega_2)$, the surface reads as

$$h(x_1, x_2) = \sum_{\omega_1=0}^{N-1} \sum_{\omega_2=0}^{N-1} \eta(\omega_1, \omega_2) \exp(-2\pi i(\omega_1 x_1 + \omega_2 x_2)).$$

The main challenge of this approach to modeling ice topography is to choose a power spectrum, S , so that it is consistent with the sea ice. Using topography LIDAR data to estimate the covariance may be helpful in this situation (Tilling et al., 2020).

The observational data of melt ponds are very limited, and the best way to enhance the statistical methods of pre-melt ice topography modeling is to use ML algorithms. We propose to employ GANs (Goodfellow et al., 2014; Kashinath et al., 2021). GANs consist of two competing neural networks, namely, a generator $G(z)$ and a discriminator $D(y)$. The generator $G(z)$ maps an input noise vector z to an output “synthetic” sample of the ice topography y . Given sea ice image data consisting of a set of original samples (e.g., realizations $y_1, y_2, \dots, y_n$), the goal of $G(z)$ is to generate data that resemble original data samples. In contrast, the discriminator $D(y)$ is a classifier that takes an input image and attempts to determine whether it is authentic (i.e., comes from the original dataset) or synthetic (i.e., originates from the generator $G(z)$). The output of $D(y)$ is a scalar representing the probability of y to come from the data. After the model has been trained, $G(z)$ can be used to generate new synthetic samples of ice topography, parameterized by the noise vector z . Using real images, we can learn how the vector z can be assigned. Results of GAN ML for sea ice topography are presented in Figure 1.

The ability to efficiently generate realistic pond spatial patterns can be used in global climate models (Pedersen et al., 2009; Flocco et al., 2010). The discussed sea ice emulator provides a framework for

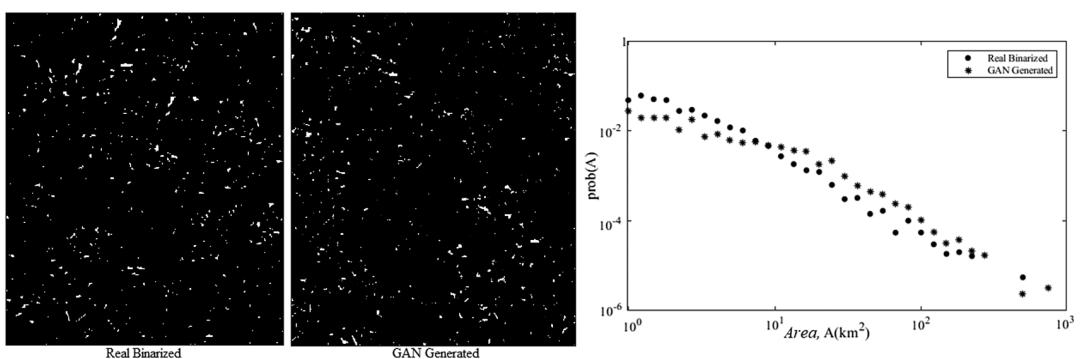


Figure 1. Left: Example of real (but binarized) versus generative adversarial network generated images of melt pond scenes (left panel); Right: Pond size relative frequency for real (dots) versus synthetic ponds (stars).

prescribing a spatial organization of melt ponds based on solely knowing the topography which can be obtained from statistical modeling combined with GAN synthetic emulations.

5. Discussion and Conclusions

The survey of climate emulators performed in this paper suggests that there is no strong consensus on the definition of a climate emulator. Physics-based climate emulators and statistics-based climate emulators are built on different approaches. Physics-based climate emulators are derived from physical principles and serve as a rational instrument for reducing the complexity (or dimensionality) of real-world problems. However, a clear indication exists that statistics-based climate emulators can be even more promising, especially when combined with ML techniques (see Section 3). Having their origin in applied statistics, ML emulators are well suited for performing multiple computational experiments with different sets of parameters or inputs. We argued that the methods of statistical mechanics can make connections between the former two types of emulators. We also provided an example of a climate emulator for sea ice (see Section 4) that offers a way for such connection.

In summary, the general outline of statistical mechanics-based climate emulators typically includes some or all of the following steps (see Figure 2): (a) spatial data acquisition, for example, via remote sensing or sensor networks; emerging data uncertainty (due to many different reasons) at this stage needs to be resolved; (b) real-time data processing using ML to identify system parameters; (c) parallel simulation of (multiple) statistical mechanics models capturing spatial properties of the system triggering critical changes; and (d) Statistical mechanics models implement into ESMs.

In the last step, they can either complement existing modules or can be introduced as new modules. These models offer ensemble outcomes depending on the parameters of emulation. The results should then be studied in terms of the goals considering new uncertainties in the emulation dataset.

The procedure may involve coupling mechanisms since statistical mechanics models “communicate” the identified spatial conditions to the ML module, which, in turn, evaluates the quality of fit based on real-time data and quantifies the risk of critical events in the considered system under consideration.

Adopting this framework, the approach to sea ice emulator presented in Section 4 can be used to improve the representation of landscape change processes in the E3SM climate model, which creates a series of high-speed statistical representations of major atmosphere, land, and ocean processes (Guo et al., 2021). For example, this model relies upon an emulator for fast statistical parameterization of subgrid permafrost lake energy fluxes. Since permafrost lakes can be modeled using statistical mechanics (Sudakov and Vakulenko, 2015), the proposed approach appears promising in this context.

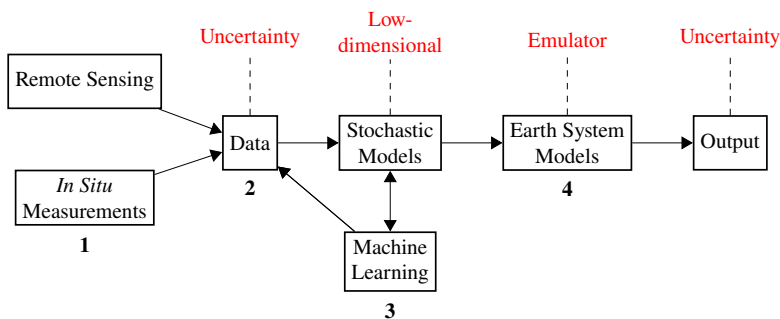


Figure 2. Schematic flowchart for climate emulator development. The numbers correspond to the necessary steps (see in the text). The solid arrows specify one- or two-way relationships between respective logical blocks (in solid rectangles). The dashed lines point to additional properties or clarifications (in red).

The complexity conundrum has adversely affected the applicability of climate models in decision support because policymakers need a rapid response, whereas models take too long to set up and run. High-speed statistical mechanics-based emulators automate the modeling process, thus enabling decision support in various ways that are not readily available now. In particular, climate emulations can be used to decide if a certain set of policies is likely to be efficient or not and truly estimate the uncertainties in climate projections. Coupling the emulators to integrated assessment models will automatically estimate economic impacts of future scenarios and policy options.

As for statistics and ML fields, the advantages of statistical mechanics-based climate emulation are expected to stimulate new theoretical discoveries along with methodological developments and innovative applications (Mecke and Stoyan (2000)). This is clear that interdisciplinary collaborative research involving climate scientists, statistical physicists, as well as data scientists and statisticians will prove crucial in developing a new generation of climate emulators to address contemporary challenges in climate modeling.

Acknowledgments. The authors thank Mr. Andrews T. Anum (Ph.D. candidate in Computational Science at the University of Texas in El Paso, El Paso, Texas, USA) for assistance with preparing BibTeX references. Comments, recommendations, and improvement suggestions from Editor-in-Chief Professor Monteleoni, Editor Professor Rao, and two anonymous referees are greatly appreciated.

Author Contributions. Conceptualization: all authors; Data curation: I.S.; Data visualization: I.S.; Methodology: all authors; Writing—original draft: I.S., M.P. All authors approved the final submitted draft.

Competing Interests. The authors declare no competing interests exist.

Data Availability Statement. The dataset produced as a part of this study has been published and publicly available for download at <https://zenodo.org/record/6602409>.

Ethics Statement. The research meets all ethical guidelines, including adherence to the legal requirements of the United States.

Funding Statement. I.S. gratefully acknowledges support from the Division of Physics at the U.S. National Science Foundation (NSF) through Grant No. PHY-2102906. M.P. was partially supported by the U.S. Department of Education (Award No. P120A180101). D.L. was partially supported by KAUST baseline funding.

References

- Adler RJ and Taylor JE (2007) *Random Fields and Geometry*, Springer Monographs in Mathematics, vol. 80. New York: Springer.
- Agrawal A and Choudhary A (2016) Perspective: Materials informatics and big data: Realization of the “fourth paradigm” of science in materials science. *APL Materials* 4, 053208.
- Antoniano-Villalobos I, Borgonovo E and Lu X (2020) Nonparametric estimation of probabilistic sensitivity measures. *Statistics and Computing* 30(2), 447–467.
- Baez JC and Tweed D (2013) Monte Carlo methods in climate science. *Math Horizons* 21(2), 5–8.
- Bao J, McInerney DJ and Stein ML (2016) A spatial-dependent model for climate emulation. *Environmetrics* 27(7), 396–408.
- Beucler T, Pritchard M, Rasp S, Ott J, Baldi P and Pierre G (2021) Achieving conservation of energy in neural network emulators for climate modeling. *Physical Review Letters* 126(9), 1–7.
- Beusch L, Gudmundsson L and Seneviratne SI (2020) Crossbreeding CMIP6 Earth system models with an emulator for regionally optimized land temperature projections. *Geophysical Research Letters* 47(15), e2019GL086812.
- Bowen B, Strong C and Golden KM (2018) Modeling the fractal geometry of arctic melt ponds using the level sets of random surfaces. *Journal of Fractal Geometry* 5(2), 121–142.
- Castruccio S, McInerney DJ, Stein ML, Crouch FL, Jacob RL and Moyer EJ (2014) Statistical emulation of climate model projections based on precomputed GCM runs. *Journal of Climate* 27(5), 1829–1844.
- Chakraborty D, Başağaoğlu H and Winterle J (2021) Interpretable vs. noninterpretable machine learning models for data-driven hydro-climatological process modeling. *Expert Systems with Applications* 170, 114498.
- De Castro C, Luković M, Andrade RF and Herrmann HJ (2017) The influence of statistical properties of Fourier coefficients on random Gaussian surfaces. *Scientific Reports* 7(1), 1–8.
- Dorheim K, Link R, Hartin C, Kravitz B and Snyder A (2020) Calibrating simple climate models to individual earth system models: Lessons learned from calibrating Hector. *Earth and Space Science* 17(11), e2019EA000980.
- Drignei D and Morris MD (2006) Empirical Bayesian analysis for computer experiments involving finite-difference codes. *Journal of the American Statistical Association* 101(476), 1527–1536.
- Edwards PN (2011) History of climate modeling. *WIREs Climate Change* 2, 128–139.

- Feng Q, Wen X and Li K (2015) Wavelet analysis-support vector machine coupled models for monthly rainfall forecasting in arid regions. *Water Resources Management* 29, 1049–1065.
- Flocco D, Feltham DL and Turner AK (2010) Incorporation of a physically based melt pond scheme into the sea ice component of a climate model. *Journal of Geophysical Research: Oceans* 115(C8).
- Franzke CLE, Blender R, O’Kane TJ and Lembo V (2022) Stochastic methods and complexity science in climate research and modeling. *Frontiers of Physics* 10, 931596.
- Gadat S, Corre L, Doury A, Ribes A and Somot S (2021) Regional climate model emulator based on deep learning: Concept and first evaluation of a novel hybrid downscaling approach. Working Paper No. 21.1233, Toulouse School of Economics.
- Gagne DJ, Haupt SE, Nychka DW and Thompson G (2019) Interpretable deep learning for spatial analysis of severe hailstorms. *Monthly Weather Review* 147, 2827–2845.
- Ghosh A, Mukhopadhyay S, Roy S and Bhattacharya S (2014) Bayesian inference in nonparametric dynamic state-space models. *Statistical Methodology* 21, 35–48.
- Glauber RJ (1963) Time-dependent statistics of the Ising model. *Journal of Mathematical Physics* 4(2), 294–307.
- Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y (2014) Generative adversarial nets. *Advances in Neural Information Processing Systems* 27, 2672–2680.
- Guillaumin AP and Zanna L (2021) Stochastic-deep learning parameterization of ocean momentum forcing. *Journal of Advances in Modeling Earth Systems* 13(9), e2021MS002534.
- Guinness J and Hammerling D (2018) Compression and conditional emulation of climate model output. *Journal of the American Statistical Association* 113(521), 56–67.
- Guo M, Zhuang Q, Yao H, Golub M, Leung LR, Pierson D and Tan Z (2021) Validation and sensitivity analysis of a 1-D lake model across global lakes. *Journal of Geophysical Research: Atmospheres* 126(4), e2020JD033417.
- Hauser T, Keats A and Tarasov L (2012) Artificial neural network assisted Bayesian calibration of climate models. *Climate Dynamics* 39(1–2), 137–154.
- Hernandez-Lemus E (2021) Random fields in physics, biology and data science. *50 Years of Statistical Physics in Mexico: Development, State of the Art and Perspectives*.
- Holden PB, Edwards NR, Garthwaite PH and Wilkinson RD (2015) Emulation and interpretation of high-dimensional climate model outputs. *Journal of Applied Statistics* 42(9), 2038–2055.
- Holden PB, Edwards NR, Rangel TF, Pereira EB, Tran GT and Wilkinson RD (2019) PALEO-PGEM v1.0: A statistical emulator of Pliocene–Pleistocene climate. *Geoscientific Model Development* 12(12), 5137–5155.
- Huang W, Lu P, Lei R, Xie H and Li Z (2016) Melt pond distribution and geometry in high Arctic sea ice derived from aerial investigations. *Annals of Glaciology* 57(73), 105–118.
- Isichenko MB (1992) Percolation, statistical topography, and transport in random media. *Reviews of Modern Physics* 64(4), 961.
- Izenman AJ (2021) Sampling algorithms for discrete Markov random fields and related graphical models. *Journal of the American Statistical Association* 116, 2065–2086.
- Jensen P (2021) Why weather/climate forecasts can be trusted. In *Your Life in Numbers: Modeling Society Through Data*.
- Kashinath K, Mustafa M, Albert A, Wu J, Jiang C, Esmailzadeh S, Azizzadenesheli K, Wang R, Chattopadhyay A, Singh A, Manepalli A, Chirila D, Yu R, Walters R, White B, Xiao H, Tchelepi HA, Marcus P, Anandkumar A, Hassanzadeh P and Prabhat (2021) Physics-informed machine learning: Case studies for weather and climate modelling. *Philosophical Transactions of the Royal Society A* 379(2194), 20200093.
- Kasim M, Watson-Parris D, Deaconu L, Oliver S, Hatfield P, Froula D, Gregori G, Jarvis M, Khatriwala S, Korenaga J, Topp-Mugglestone J, Viezzer E and Vinko SM (2020) Building high accuracy emulators for scientific simulations with deep neural architecture search. *Machine Learning: Science and Technology* 3, 015013.
- Katz RW (2002) Techniques for estimating uncertainty in climate change scenarios and impact studies. *Climate Research* 20(2), 167–185.
- Kennedy ID (2008) The transformation of one-dimensional and two-dimensional autoregressive random fields under coordinate scaling and rotation. Master’s thesis, University of Waterloo.
- Khouider BA (2014) Coarse grained stochastic multi-type particle interacting model for tropical convection: Nearest neighbour interactions. *Communications in Mathematical Sciences* 12, 1379–1407.
- Krapivsky PL, Redner S and Ben-Naim E (2010) *A Kinetic View of Statistical Physics*, 1st Edn. Cambridge, UK: Cambridge University Press.
- Levin DA and Peres Y (2017) *Markov Chains and Mixing Times*, AMS Non-Series Monographs, vol. 107. Providence, RI: American Mathematical Society.
- Li Y and Sun Y (2019) Efficient estimation of nonstationary spatial covariance functions with application to high-resolution climate model emulation. *Statistica Sinica* 29(3), 1209–1231.
- Llorente F, Martino L, Delgado-Gómez D and Camps-Valls G (2021) Deep importance sampling based on regression for model inversion and emulation. *Digital Signal Processing* 116, 103104.
- Ma Y-P, Sudakov I, Strong C and Golden KM (2019) Ising model for melt ponds on Arctic sea ice. *New Journal of Physics* 21(6), 063029.
- Mansfield LA, Nowack PJ, Kasoar M, Everitt RG, Collins WJ and Voulgarakis A (2020) Predicting global patterns of long-term climate change from short-term simulations using machine learning. *npj Climate and Atmospheric Science* 3(1), 1–9.

- Masrur A, Yu M, Mitra P, Peuquet D and Taylor A (2021) Interpretable machine learning for analysing heterogeneous drivers of geographic events in space-time. *International Journal of Geographical Information Science* 36, 692–719.
- Masson-Delmotte V, Zhai P, Pirani A, Connors S, Péan C, Berger S, Caud N, Chen Y, Goldfarb L, Gomis M, Huang M, Leitzell K, Lonnoy E, Matthews J, Maycock T, Waterfield T, Yelekçi O, Yu R and Zhou B (Eds.) (2021) *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press (in press).
- Mecke KR and Stoyan D (2000) *Statistical Physics and Spatial Statistics: The Art of Analyzing and Modeling Spatial Structures and Pattern Formation*, Lecture Notes in Physics, vol. 554. Springer Science & Business Media.
- Mehta P, Bukov M, Wang C-H, Day AG, Richardson C, Fisher CK and Schwab DJ (2019) A high-bias, low-variance introduction to machine learning for physicists. *Physics Reports* 810, 1–124.
- Meyer D and Nagler T (2021) Synthia: Multidimensional synthetic data generation in python. *Journal of Open Source Software* 6 (65), 1–2.
- Meyer D, Nagler T and Hogan RJ (2021) Copula-based synthetic data augmentation for machine-learning emulators. *Geoscientific Model Development* 14(8), 5205–5215.
- Miftakhova A, Judd KL, Lontzek TS and Schmedders K (2020) Statistical approximation of high-dimensional climate models. *Journal of Econometrics* 214(1), 67–80.
- Nadiga BT (2021) Reservoir computing as a tool for climate predictability studies. *Journal of Advances in Modeling Earth Systems* 13, e2020MS002290.
- Newman ME and Barkema GT (1996) Monte Carlo study of the random-field Ising model. *Physical Review E* 53(1), 393.
- Nichol JJ, Peterson MG, Peterson KJ, Fricke GM and Moses ME (2021) Machine learning feature analysis illuminates disparity between E3SM climate models and observed climate change. *Journal of Computational and Applied Mathematics* 395, 113451.
- Ouyang Q and Lu W (2018) Monthly rainfall forecasting using echo state networks coupled with data preprocessing methods. *Water Resources Management* 32, 659–674.
- Overstall AM and Woods DC (2016) Multivariate emulation of computer simulators: Model selection and diagnostics with application to a humanitarian relief model. *Journal of the Royal Statistical Society: Series C* 65(4), 483–505.
- Pedersen CA, Roeckner E, Lüthje M and Winther J-G (2009) A new sea ice albedo scheme including melt ponds for ECHAM5 general circulation model. *Journal of Geophysical Research: Atmospheres* 114(D8).
- Perovich D, Grenfell T, Light B and Hobbs P (2002) Seasonal evolution of the albedo of multiyear Arctic sea ice. *Journal of Geophysical Research: Oceans* 107(C10), SHE 20-1–SHE 20-13.
- Perovich DK, Grenfell TC, Light B, Elder BC, Harbeck J, Polashenski C, Tucker WB and Stelmach C (2009) Transpolar observations of the morphological properties of Arctic sea ice. *Journal of Geophysical Research: Oceans* 114(C1).
- Perovich DK, Richter-Menge JA, Jones KF and Light B (2008) Sunlight, water, and ice: Extreme Arctic sea ice melt during the summer of 2007. *Geophysical Research Letters* 35(11).
- Polashenski C, Perovich D and Courville Z (2012) The mechanisms of sea ice melt pond formation and evolution. *Journal of Geophysical Research: Oceans* 117(C1).
- Rougier J and Goldstein M (2014) Climate simulators and climate projections. *Annual Review of Statistics and Its Application* 1, 103–123.
- Sacks J, Welch W, Mitchell T and Wynn H (1989) Design and analysis of computer experiments. *Statistical Science* 4, 409–435.
- Schleder GR, Padilha ACM, Acosta CM, Costa M and Fazzio A (2019) From DFT to machine learning: Recent approaches to materials science: A review. *Journal of Physics Materials* 2, 032001.
- Schwarber AK, Smith SJ, Hartin CA, Vega-Westhoff BA and Srivier R (2019) Evaluating climate emulation: Fundamental impulse testing of simple climate models. *Earth System Dynamics* 10(4), 729–739.
- Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D and Batra D (2017) Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 618–626.
- Serreze MC, Holland MM and Stroeve J (2007) Perspectives on the Arctic's shrinking sea-ice cover. *Science* 315(5818), 1533–1536.
- Sudakov I and Vakulenko SA (2015) A mathematical model for a positive permafrost carbon–climate feedback. *IMA Journal of Applied Mathematics* 80(3), 811–824.
- Tebaldi C, Armbruster A, Engler H and Link R (2020) Emulating climate extreme indices. *Environmental Research Letters* 15 (7), 074006.
- Tilling R, Kurtz N, Bagnardi M, Petty A and Kwok R (2020) Detection of melt ponds on Arctic summer sea ice from ICESat-2. *Geophysical Research Letters* 47(23), e2020GL090644.
- Tran GT, Oliver KI, Söbester A, Toal DJ, Holden PB, Marsh R, Challenor P and Edwards NR (2016) Building a traceable climate model hierarchy with multi-level emulators. *Advances in Statistical Climatology, Meteorology and Oceanography* 2(1), 17–37.
- Tsutsui J (2021) Minimal CMIP emulator (MCE v1. 2): A new simplified method for probabilistic climate projections. *Geoscientific Model Development Discussions*, pp. 1–29.
- Urban NM and Fricker TE (2010) A comparison of Latin hypercube and grid ensemble designs for the multivariate emulation of an Earth system model. *Computers & Geosciences* 36(6), 746–755.

- van Huissteden J, Berrittella C, Parmentier FJW, Mi Y, Maximov TC and Dolman AJ** (2011) Methane emissions from permafrost thaw lakes limited by lake drainage. *Nature Climate Change* 1, 119–123.
- Villagran A, Huerta G, Vannucci M, Jackson CS and Nosedal A** (2016) Non-parametric sampling approximation via Voronoi tessellations. *Communications in Statistics—Simulation and Computation* 45(2), 717–736.
- Watson-Parris D** (2021) Machine learning for weather and climate are worlds apart. *Philosophical Transactions of the Royal Society A* 379(2194), 20200098.
- Weber T, Corotan A, Hutchinson B, Kravitz B and Link R** (2020) Deep learning for creating surrogate models of precipitation in Earth system models. *Atmospheric Chemistry and Physics* 20(4), 2303–2317.
- Welling M and Teh YW** (2003) Approximate inference in Boltzmann machines. *Artificial Intelligence* 143(1), 19–50.
- Xu W, Luo X, Ren Y, Park JH, Yoo S and Nadiga BT** (2021). Feature importance in a deep learning climate emulator. In *AIMOCC—AI: Modeling Oceans and Climate Change*, pp. 1–9.
- Young PC and Ratto M** (2011) Statistical emulation of large linear dynamic models. *Technometrics* 53(1), 29–43.
- Yuan X-C, Zhang N, Wang W-Z and Wei Y-M** (2021) Large-scale emulation of spatio-temporal variation in temperature under climate change. *Environmental Research Letters* 16(1), 014041.