Statistical Science
2020, Vol. 35, No. 1, 31–41
https://doi.org/10.1214/19-STS730

Some Statistical Issues in Climate Science

Michael L. Stein

Abstract. Climate science is a field that is arguably both data-rich and data-poor. Data rich in that huge and quickly increasing amounts of data about the state of the climate are collected every day. Data poor in that important aspects of the climate are still undersampled, such as the deep oceans and some characteristics of the upper atmosphere. Data rich in that modern climate models can produce climatological quantities over long time periods with global coverage, including quantities that are difficult to measure and under conditions for which there is no data presently. Data poor in that the correspondence between climate model output to the actual climate, especially for future climate change due to human activities, is difficult to assess. The scope for fruitful interactions between climate scientists and statisticians is great, but requires serious commitments from researchers in both disciplines to understand the scientific and statistical nuances arising from the complex relationships between the data and the real-world problems. This paper describes a small fraction of some of the intellectual challenges that occur at the interface between climate science and statistics, including inferences for extremes for processes with seasonality and long-term trends, the use of climate model ensembles for studying extremes, the scope for using new data sources for studying space-time characteristics of environmental processes and a discussion of non-Gaussian space-time process models for climate variables. The paper concludes with a call to the statistical community to become more engaged in one of the great scientific and policy issues of our time, anthropogenic climate change and its impacts.

Key words and phrases: Statistical climatology, climate extremes, Argo network, non-Gaussian processes.

1. INTRODUCTION

Climatology, like all areas of science, requires a combination of theory and observation to advance. And, without weighing in on the controversy of whether computation should be viewed as a "third leg" of science (Vardi, 2010), there can be no doubt that modern computation has had a huge impact on climate science. Of course, much was understood about the Earth's climate before the advent of modern computers. Indeed, the idea that gases in the atmosphere could warm surface temperature was first proposed by Joseph Fourier in the 1820s. In 1896, Svante Arrhenius estimated the impact of CO_2 levels on mean surface temperature based on a theoretical calculation using, among many other factors, estimates of the absorption characteristics of CO_2 and water vapor (Arrhenius, 1896). In particular, under the plausible approximation that relative humidities stay constant as temperatures change, he estimated the effect of a doubling of CO_2 concentration on the mean equilibrium surface temperature of the Earth by 10° latitude bands between 60°S and 70°N, finding increases between 4.95°C and 6.05°C depending on latitude band. Based on better estimates of the absorptive properties of CO_2 and water vapor, Arrhenius later updated his estimate of the effect of doubling CO_2 on equilibrium global mean surface temperature to 4°C (Arrhenius, 1908), placing his estimate within the current range of likely values according to the most recent Intergovernmental Panel on Climate Change report of 1.5°–4.5°C (IPCC, 2013). Arrhenius (1896) is surprisingly readable and I recommend it to anyone who might think that only direct observations of surface temperature provide useful information about climate change.

Of course, observations do play a critical role in our understanding of climatology. As in astrophysics, the data in climatology is largely observational, although humanity is presently undertaking an uncontrolled experiment on the Earth's climate (Ramanathan, 1988). Furthermore, while climatologists study the climate of planets other than the

Michael L. Stein is Distinguished Professor of Statistics, Department of Statistics, Rutgers University, Piscataway, New Jersey 08854, USA (e-mail: stein@galton.uchicago.edu).

Earth (Pierrehumbert, 2010), the amount of data on the climate of other planets, especially planets with an atmosphere similar to Earth's, is severely limited. Thus, global climate models, with which one can do controlled experiments, necessarily play a large role in making qualitative and quantitative inferences about the Earth's climate and how it will change due to anthropogenic effects (IPCC, 2013).

In thinking through the role of statistics in learning about the Earth's climate, it is worthwhile to provide some background on the nature of both modern observations and climate models. First, it is important to recognize that even with the massive amount of data now collected about the atmosphere and oceans, we still only observe a tiny fraction of the information that would be needed to determine the present state of the Earth's climate system. Thus, it is quite common to use what are called data products, which are some combination of observations and physical and/or statistical models, rather than direct *in situ* measurements, in analyses of the climate system. Current reanalysis data products (Balsamo et al., 2015, Saha et al., 2010) make use of *in situ* and satellite observations combined with a high-resolution climate model to obtain estimates of the past states of the atmosphere-ocean-land surface-sea ice system at a sub-daily time scale. While reanalysis products are constantly improving, they still have known problems, especially with precipitation (Peña-Arancibia et al., 2013). Satellite measurements of the climate system can also reasonably be called data products, as the raw measurements, photon fluxes within various frequency bands, have to be processed to produce climatological variables of interest such as surface temperature (Hollmann et al., 2013). Even simple rain gauges can have substantial biases in their measurements of precipitation (Rodda and Dixon, 2012), especially in high winds, which can have important implications for inferences about extreme precipitation. Paleoclimate records, which play an important role in understanding and constraining the impacts of various forcings on climate, have myriad scientific and statistical challenges in converting observations into estimates of past climatology and quantifying the uncertainties in them (IPCC, 2013, Chapter 5). It is incumbent on statisticians who work with climatological data to make themselves aware of potential problems with any observational data or data product they might use.

Observational data and data products tell us about the past and present, but we need models to tell us about the future. Quantitative climate models can range from simple one-dimensional energy balance models, which is essentially what Arrhenius used, to modern Earth system models, which simulate physical, chemical and biological aspects of the atmosphere-ocean-land surface-sea ice system on a global scale (Edwards, 2011). These models are generally deterministic, although there is substantial research showing that including stochastic elements in climate models can reduce biases (Berner et al., 2017). My personal view is that stochastic climate models will be essential to obtain output with realistic space-time variability and to accurately represent the uncertainties in these models. I see great opportunities for statisticians and applied probabilists who could establish effective collaborations with the groups working in this area.

Global climate models take account of much that is known about the climate system, but memory and computational limitations make it impossible for global models to take direct account of processes that occur at fine spatial scales such as atmospheric turbulence and cloud formation. Thus, global models are forced to either leave out such processes or to "parameterize" them in some form, which means replacing a description based on physical laws with an empirical representation that is meant to mimic the essential properties of the process. These empirical representations generally include tunable quantities (parameters) that determine the detailed mechanism of the parameterization, such as the atmospheric conditions under which a certain type of cloud forms. There is considerable research on selecting these parameters and their effect on climate projections (Neelin et al., 2010, Sexton et al., 2012), but it is not necessarily clear when parameterizations that work well for the present climate will work well for assessing future changes in climate.

In addition to selecting a model, which includes setting parameter values and the resolution of the model, it is also necessary to set initial conditions and any forcings that are not calculated as part of the model. These external forcings include orbital variations, solar output, volcanoes and, critically, human influences such as greenhouse gas emissions and land use changes. Because of the sensitivity of complex climate models to initial conditions, output even from deterministic models may be treated as effectively stochastic. Indeed, one of the most important papers on chaotic dynamical systems, Lorenz (1963), appeared in *Journal of the Atmospheric Sciences* and studied a model for hydrodynamic flow made up of a system of three nonlinear ordinary differential equations that exhibits the extreme sensitivity to initial conditions characteristic of chaotic dynamical systems. Kay et al. (2015) make use of this extreme sensitivity to initial conditions in climate models to produce an ensemble of 40 runs of the Earth system model CESM that differ only in the initial temperature field perturbed at just above machine precision. After the first several weeks of the runs, at least the atmospheric conditions in the different runs appear to be largely independent. In addition to initial condition ensembles as in Kay et al. (2015), there are also ensembles in which other components of a model run, such as parameters, forcings or the climate model itself, are varied

(Tebaldi and Knutti, 2007, Jun, Knutti and Nychka, 2008, Yokohata et al., 2010, Collins et al., 2011). Thus, there is a large literature that focuses on statistical analysis of climate model output, as well as design of ensemble experiments (Taylor, Stouffer and Meehl, 2012, Mearns et al., 2013). Such analyses provide critical information about the properties of climate models, how climate models differ (Jun, Knutti and Nychka, 2008) and how they compare to observed climatology (Rahmstorf et al., 2007). Nevertheless, having more and more climate models and model output is not the same as having more and more observational data, as there may be biases common to all climate models in terms of how the climate will change in the future.

Because of the important but different limitations of model output and observational records, intelligent inferences with reasonable uncertainties about future climate demand a combination of information from climate models and observations. A Bayesian approach provides a natural way to combine such information in a way that accounts for multiple sources of uncertainty (Berliner, Levine and Shea, 2000), but developing a full Bayesian model for these disparate sources of information may be problematic when trying to reach a broad consensus. Ribes et al. (2017) review methods for estimating the impacts of natural and anthropogenic changes in forcings on climate and describe a new method that attempts to avoid making strong assumptions about climate model uncertainty, although it does assume that, for the climate characteristic under study, the difference between the true climate and a climate model is statistically indistinguishable from the difference between two climate models. This study, like many others in the literature, concludes that most of the observed warming in recent decades can be attributed to anthropogenic forcings.

This paper is a highly personal view on what I view as some of the interesting challenges at the intersection of climate science and statistics. It is not a review paper on statistical climatology and, despite the large number of references, it does not provide a meaningful review of the literature in the field. Section 2, the longest, considers climate extremes, mainly because it is a current interest of mine, but also because it is an area in which sophisticated statistical methods play a central role. Section 3 discusses the role of new data sources in climate science with a focus on a few examples that might particularly interest statisticians. Section 4 discusses the role of and need for non-Gaussian stochastic process models in climate science and Section 5 gives some general comments on the existing and possible future roles for statisticians in climate science.

2. EXTREMES

Many of the most dramatic impacts of climate on society are through extreme events, such as heat waves, high winds, flooding or droughts (Field et al., 2012). Since observational records are often of limited duration, it is important to use the best statistical methods when making inferences about future extremes. The problem is only made more difficult by the rapidly changing climate due to human influences. In many applications, the process of interest is a time series observed at regularly spaced intervals, such as daily or hourly, so we will focus on this setting for now.

There are two main approaches to statistical modeling of extremes of a time series. The first is based on considering maxima of the series over blocks of observations of some fixed length. The second is based on exceedances of the process over some high threshold, often fixed, but perhaps changing over time. When individual observations are independent and identically distributed, under a regularity condition on the upper tail of their common distribution, properly normalized versions of the block maxima or exceedances over a high threshold converge in distribution to a nontrivial limit. For example, under appropriate regularity, the conditional distribution of an exceedance over a high threshold μ approximately follows a generalized Pareto distribution with location parameter μ, scale parameter σ and shape parameter ξ, with survival function given by

$$S(x) = \left(1 + \xi \frac{x-\mu}{\sigma}\right)^{-1/\xi}$$

for $x > \mu$ when $\xi > 0$ and for $\mu < x < \mu - \sigma/\xi$ when $\xi < 0$ (de Haan and Ferreira, 2006). The distribution is defined by continuity when $\xi = 0$, in which case $S(x) = e^{-(x-\mu)/\sigma}$ for $x > \mu$. There is a third approach to extremes based on point process approximations, but it is less frequently used and is closely related to the exceedance approach.

All of these approaches are motivated by a common underlying theory that relates the behavior of block maxima or exceedances to the behavior of the upper tails of the distributions of the individual observations in the time series (de Haan and Ferreira, 2006). If the time series is stationary, then this distribution is unambiguous because it is the same for all time points. However, most climate variables observed at a sub-annual time scale have substantial seasonal patterns that make an assumption of stationarity untenable. Moreover, there is strong evidence for trends in location and scale parameters of climate extremes due to climate change (Westra, Alexander and Zwiers, 2013, Zwiers et al., 2013, Roth, Jongbloed and Buishand, 2018). There is a considerable literature showing that extreme value results can be extended to cover many, but not all, nonstationary processes (Cheng et al., 2014, Einmahl, de Haan and Zhou, 2016, Stein, 2017).

2.1 Extremes of Time Series

For daily climate measurements, such as daily maximum temperature or daily precipitation, it is convenient to study annual maxima and assume these follow a generalized extreme value distribution. In one fell swoop, this simplification largely eliminates the problems of seasonality and temporal dependence; seasonality trivially goes away when considering annual maxima and dependence largely disappears because annual maxima generally show weak temporal dependence.

Despite these conveniences, there are good reasons for preferring to model and understand the seasonality and the temporal dependence of extremes. When extreme events occur is often important; for example, late spring or early fall frosts can have major impacts on agriculture. Clustering of extreme temperature events is also of practical relevance; for example, multiple extremely hot days in a short time span may have larger health (Anderson and Bell, 2009, Gasparrini and Armstrong, 2011) or agricultural impacts (Troy, Kipgen and Pal, 2015) than isolated hot days. Thus, in order to develop a full understanding of temperature extremes, it would be desirable to have a single model that takes account of both seasonality and dependence in extremes. Dealing with only the seasonality can be addressed by, for example, allowing the parameters of a generalized Pareto distribution to depend on time of year. To take account of climate change, it is important to include not just a long-term trend but also changes in seasonal patterns. For example, output from a range of climate models (Rummukainen, 2012, Huang et al., 2016, Haugen et al., 2018) indicates that the lowest quantiles of temperature will warm more in Winter than in Summer in many locations.

How to deal with the clustering is less clear. One common approach to handling clusters of extremes is declustering: remove all but the most extreme event in a period with temporally nearby multiple extremes before fitting a generalized Pareto distribution (Ferro and Segers, 2003). Defining clusters is problematic and, in any case, when one is specifically interested in the clusters themselves, declustering can be at best only part of the solution. Furthermore, Fawcett and Walshaw (2007) find that declustering methods can lead to substantial biases when estimating extremes. Another statistical approach to dependence in extremes is to estimate some index of dependence (Davison and Huser, 2015). However, such indices do not fully characterize the dependence and thus may not be adequate for some impacts. As with many problems relating to rare events, a marked point process approach should in principle be appropriate. For example, following the usual point process approach to extremes, the first day over some time period with an exceedance over some high threshold T and the observed value of this exceedance can be viewed as an event of the point process in $(-\infty, \infty) \times (T, \infty)$ and then the mark can be the times and values of further exceedances that are deemed to be part of a cluster that begins with this first exceedance. The great flexibility of what marks can represent in the marked point process formulation is both an attraction and a liability; in the present setting, one would need to have a model for marks that included a random number of exceedances over the threshold at random times and by random amounts.

2.2 Extremes of Space-Time Processes

An open area of statistics with important implications for climate science is the characterization and estimation of climatological extremes that takes proper account of extreme events as coherent meteorological phenomena in space and time. It is typical to study both temperature and precipitation extremes by considering daily values at individual locations. This choice may make sense for temperature because daily highs (or lows) of temperature are a reasonable summary of the potential impact of temperature on human health or agriculture. In contrast, the daily division is not as meaningful for precipitation, since 20 cm of rain in 24 hours will have a similar impact whether it falls between midnight and midnight or between noon and noon. Perhaps more importantly, the entire spatial or space-time pattern of extreme events may matter for some impacts. For example, for temperature, the spatial extent of a heat wave affects peak electricity consumption, and for precipitation, the entire space-time pattern of precipitation within a watershed is critical for flooding. Touma et al. (2018) recognize this problem and propose using spatial variograms of indicators of exceedances over a high threshold as a tool for describing the spatial extent of extremes, although these variograms do not fully characterize the relevant space-time relationships.

Statistical extreme value methods are well accepted in the environmental sciences and have a particularly long history in the study of floods (Institute of Hydrology, 1975, Hosking, Wallis and Wood, 1985). For many observational records, it is difficult to accurately estimate the parameters, especially the shape parameter, of an extreme value model based on observations at a single site. Thus, for example, it is common in hydrology to assume that the shape parameter of extreme value distributions does not vary within some region and there is considerable empirical evidence supporting this practice (Katz, Parlange and Naveau, 2002). More recent work uses Bayesian hierarchical models to allow parameters of extreme value distributions to vary smoothly in space (Cooley and Sain, 2010). These methods often assume that, conditional on the spatially varying parameters of the extreme value distributions, the extreme observations are conditionally independent (Cooley et al., 2012, Jalbert et al., 2017), although Opitz et al. (2018) allow spatial dependence in both the parameters of the model and the realized climate variable given these parameters. In another important research thread, Davison, Padoan and Ribatet (2012) review recent methods for accounting for spatial dependence in block maxima, where, for each time block, the maximum is taken separately at each spatial location. While all of these methods take account of spatial information, they do not seek to model extreme events as phenomena in space and time and, thus, do not address the need identified in the previous paragraph.

2.3 Climate Model Output as a Tool for Studying Extremes

It can be difficult to assess when complex methods, such as those based on borrowing strength spatially to estimate extremes, work well just using observational data. Even in cases with long observational records, evaluating the performance of a method in a particular application can be difficult due to possible nonstationarities caused by local effects such as urban heat islands, inconsistency in measurement instruments and climate change. Furthermore, however long the observational record, there is always a temptation to make inferences about extremes, such as the 500-year flood, that go beyond the direct observational evidence available in even the longest records. Thus, standard approaches for evaluating statistical methods in applications such as cross-validation are of limited utility when studying extremes.

Output from global climate models provides a promising alternative for evaluating the sampling properties of statistical approaches for estimating climate extremes. Although extreme value methods have been regularly applied to climate model output (Katz and Brown, 1992, Kharin and Zwiers, 2005, Kharin et al., 2013), these studies have focused more on studying the properties of the model output than on evaluating statistical methods. An important advantage of climate model output compared to observational records are the long time series with no missing values (Sterl et al., 2012). In recent years, a number of groups have produced climate model runs totaling thousands of years. All of the output can be used to get accurate estimates of the extremes, which can then be used to evaluate estimates based on a subset of the output that is comparable to a typical observational record. We do not need to assume that the models accurately capture climate extremes and how they might change in the future to be useful as a testbed for statistical methods, only that they share some of the space-time complexity of the actual climate system.

Extended model output comes in two forms: single runs of a model over a long period (Paynter et al., 2018) and initial condition ensembles (Kay et al., 2015). For example, Huang et al. (2016) used the last 1000 years of three multimillennial climate model runs with fixed forcings that differ only in their concentrations of CO_2 in the atmosphere to compare distributions of annual extremes of daily temperature for the three concentrations. The length of these runs made it possible to investigate whether a year is sufficiently long block length for block maxima and minima to be well approximated by generalized extreme value distributions. Specifically, Huang et al. (2016) fit generalized extreme value distributions using blocks of size 1, 2, 5 and 10 years and found that estimated shape parameters did not systematically vary for hot extremes but did show some distinct geographic patterns in how they varied for cold extremes, suggesting that blocks of one year may not be adequately long. The long runs also made it possible to evaluate resampling approaches to uncertainty assessment. Specifically, uncertainties of estimates of long return periods were obtained using block bootstrapping for blocks of either one year or a decade. Decadal blocks would generally be too long to apply to observational records but are of a reasonable length for 1000 years of model output. Huang et al. (2016) found that block bootstrap standard errors for return periods showed no systematic changes when blocks were increased from one year to ten. This result is in contrast to what would happen for annual averages of temperature, for which climate model output shows significant dependence (Castruccio et al., 2014), but annual maxima and minima are more singular events than long-term averages and, at least for this model, show no substantial dependence.

Model runs at fixed forcings avoid the problems of transient climates, which simplifies the statistical issues, but reduces their relevance to understanding the Earth's changing rapidly climate in the coming decades. For this purpose, an initial condition ensemble under one or more plausible forcing scenarios is more useful. The importance of large initial condition ensembles to the study of climate variability is reflected in the recently held Large Ensembles Workshop in Boulder, Colorado. The value of such ensembles as a testbed for how well statistical methods actually work when applied to complex, nonstationary space-time processes may not be as well appreciated.

Due to the sensitivity of these models to initial conditions, it is commonly assumed that the different runs within an ensemble can be treated as independent and identically distributed realizations of the same multivariate space-time process, although Corti et al. (2015) note that initial conditions can have a substantial impact on certain ocean characteristics for five years or more. The LENS ensemble (Kay et al., 2015) is a well-known example of such an ensemble and includes 40 runs of the same climate model, CESM, with the same forcing scenario (RCP 8.5, which includes historical forcings where available and a "business as usual" projection for future forcings) covering the period 1920–2010. Haugen et al. (2018) used a 50-member ensemble due to Sriver, Forest and Keller (2015) to fit changing seasonal patterns of daily temperature distributions using quantile regression. To obtain uncertainty estimates, Haugen et al. (2018) resampled runs rather than using block bootstrapping of years within runs, thus avoiding any assumptions of near independence or stationarity within runs. The resulting bootstrap uncertainties showed that the large ensemble provided accurate estimates of a wide range of quantiles using a fairly rich set of basis functions for the evolving seasonal pattern of temperature. In particular, the size of the dataset made it possible to estimate even fairly extreme quantiles accurately without appealing to extreme value theory.

3. DATA

As valuable as climate model output is, it is still no substitute for observational data. As in most fields, dataset sizes are exploding in climate science. Here, we just mention two data sources whose scopes are well-suited for advancing both climate science and statistics.

3.1 The Argo Network

The Argo array is a collection of free-floating devices for measuring the temperature and salinity of the ocean (Argo, 2000, Riser et al., 2016). Every 10 days, each float measures pressure, temperature and salinity along the water column from 2000 m to the surface. The Argo array reached its steady-state goal of 3000 floats in 2007 and provides over 100,000 vertical transects of ocean characteristics each year. The vertical resolution of most of the newer floats is 2 m. The Argo array provides the first systematic set of measurements on the state of the upper ocean and, unsurprisingly, has led to thousands of scientific publications (www.argo.ucsd.edu/Bibliography.html). The horizontal resolution of the array is not particularly high, with an approximate density of one float for every $3° \times 3°$ patch of ocean, although the actual sampling locations are highly irregular. The site www.argo.ucsd.edu/Gridded_fields.html lists a number of gridded data products based on Argo data, many of which are at 1° or 0.5° resolution. Thus, space-time interpolation of Argo data is of considerable interest, including statistical approaches such as kriging, which is often called objective analysis or optimal interpolation in the geoscience literature.

Kuusela and Stein (2018) develop models and methods for space-time interpolation of Argo temperature data based on kriging, which requires estimation of a spatio-temporal covariance function. For every space-time location in the interpolation grid, parameters of this covariance function were fit to observations within an appropriate spatial and seasonal window using maximum likelihood. Although this work included several advances over other interpolation schemes used for ocean temperature fields, it is worthwhile to point out some of the important problems this work does not address. First, the form of the local space-time covariance function is just a nugget effect plus an exponential covariance function with separate range parameters for latitude, longitude and time, thus ignoring possible space-time asymmetries in the covariance structure (Gneiting, 2002) and assuming that the exponential form provides a reasonable description for the local variations in both space and time. Second, salinity and its covariation with temperature is ignored. Third, the scheme operates at one pressure level at a time and, thus, does not explicitly model the variation in the vertical dimension. One difficulty with modeling in the vertical dimension is how to account for the constraint that the density of water, which is essentially a function of temperature and salinity, must very nearly monotonically increase with depth. Thus, a standard Gaussian process model for the bivariate process of temperature and salinity does not really make sense when considering the vertical dimension, although there is work on adapting Gaussian processes to include monotonicity constraints (Wang and Berger, 2016, López-Lopera et al., 2018). Finally, and perhaps most intriguingly, Kuusela and Stein (2018) assume that it is appropriate to treat the space-time locations of the observations as nonrandom. However, the floats move with the currents and these currents are themselves largely driven by variations in density, so that where we observe is not independent of what we observe. Gray and Riser (2014) use both the temperature and salinity data and the movements of the floats to infer ocean currents, but still treats the locations of observations as exogenous when modeling temperature and salinity.

In principle, since a complete model for the ocean would determine currents, temperature and salinity, the reconstruction of the ocean state from the Argo data is an inverse problem that is amenable to data assimilation methods. There is a substantial literature on oceanographic data assimilation (Forget et al., 2015, Stammer et al., 2016); for example, Forget et al. (2015) use temperature and salinity measurements from a number of sources including the Argo network to estimate the ocean state, but does not use the motion of the floats. The efficacy of these assimilation methods is limited by the quality of the dynamical models, models for error covariances needed in the data assimilation scheme and computational limitations. Thus, more empirical approaches to state estimation are still commonly used in oceanography and will likely continue to be used for the foreseeable future. Statisticians could and should seek to contribute both to the empirical approaches and to the statistical aspects of oceanographic data assimilation schemes by joining oceanographic research teams.

3.2 Remote Sensing Data

The development of space-time covariance functions for environmental processes has been an active topic in the statistical literature for the last twenty years or so (Cressie and Huang, 1999, Gneiting, 2002, Stein, 2005, Cressie and Wikle, 2011, Sigrist, Künsch and Stahel, 2015). Even when these models have rough physical motivation as in Sigrist, Künsch and Stahel (2015), theoretical considerations do not generally lead to covariance models that are specified up to scale and range parameters and, thus, there is great value in using observational data to learn about degrees of smoothness in space and time and the nature of space-time interactions (Stein, 2005).

To fully explore space-time interactions requires data that is high resolution in both space and time. Historically, environmental data has tended to be high resolution in space or time but not both. Recent geostationary satellites such as GOES-16 and GOES-17 provide images of the Earth at hemispheric scales in 16 wavelength bands that, for some purposes, may be viewed as dense in space and high frequency time (Schmit et al., 2017). For example, GOES-17 provides images every 15 minutes at the hemispheric scale and every 5 minutes for the continental United States with a spatial resolution between 0.5 and 2 km depending on the wavelength band. This density of observations may obviate the need for interpolation of the data and instead lends itself to focusing on the space-time properties of the wide range of environmental quantities that can be inferred from these multivariate spectral images. For a simple example, Kuusela and Stein (2018) showed how local estimates of spatial range parameters along latitudes and longitudes for ocean temperature correspond to known features of the underlying processes, but one can imagine finding subtler properties of space-time interactions from statistical analysis of recent GOES data that would be of scientific interest.

4. NON-GAUSSIAN PROCESSES

It is common to describe space-time dependencies in environmental processes using covariance functions. Covariance functions provide a complete description of the dependencies for Gaussian processes, but may be inadequate for non-Gaussian processes. Precipitation over shorter time scales, with its preponderance of zeroes, is obviously non-Gaussian, but many other climate variables, such as temperature and relative humidity, have distinctly non-Gaussian properties. It would be of great statistical and scientific interest to develop statistical models that provide good descriptions of marginal and joints distributions for such spatial and space-time processes. Some recent efforts in this direction include Wikle (2015), Wallin and Bolin (2015), Xu and Genton (2017). High resolution data will be particularly helpful in making progress in this direction.

A common tactic for developing non-Gaussian process models is to use an unobserved Gaussian process as a building block. This approach was made popular by Diggle, Tawn and Moyeed (1998), which assumes the link function of a generalized linear model is a Gaussian process, and that, conditional on the Gaussian process, the values of the observed process are independent. This independence makes MCMC readily usable for Bayesian inferences under the model, including predictions. However, the conditional independence makes the model unsuitable for many applications. For example, consider modeling the occurrence of precipitation on a given day using logistic regression with the logit of the probability of occurrence at a location being the sum of a Gaussian process and a linear combination of observed covariates. The conditional independence assumption implies that realizations of the model produce speckled patterns of precipitation rather than spatially coherent wet and dry regions. To address this weakness, Olson and Kleiber (2017) use a thresholded Gaussian process model for daily precipitation occurrence. If the Gaussian process is continuous, then the resulting wet regions have the desired spatial coherence. However, likelihood calculations are difficult under this model when there is a large observation network, requiring integrations over orthants of multivariate normal distributions whose dimension equals the number of observations. Olson and Kleiber (2017) carry out inference using approximate Bayesian computations, which only require the ability to simulate from the process, exploiting the fact that simulating a thresholded Gaussian process is straightforward. Stochastic precipitation generators have a long history in the hydrological literature (Wilks and Wilby, 1999, Ailliot et al., 2015). These models often have the properties that simulations can be done efficiently but calculating likelihoods or conditional distributions are challenging, so that approximate Bayesian computations may prove broadly useful for fitting such models to data and calculating predictive distributions of the process at unobserved locations.

The demand for high-resolution gridded maps of daily precipitation that interpolates the observational data is great, so efforts to produce such maps proceed despite the lack of a principled statistical solution to the problem. Di Luzio et al. (2008) and Yatagai et al. (2012) describe ad hoc interpolation schemes based on local weighted averages where the weights make meteorologically motivated use of orographic information, an approach that is called climatologically aware interpolation. Hutchinson et al. (2009) argue that more realistic interpolations are obtained by first interpolating the dichotomous occurrence field and then, separately, interpolating the amount of precipitation at those locations for which the first interpolation indicates occurrence. Both interpolations use kriging, which is a linear procedure, so may be substantially suboptimal for interpolating the dichotomous occurrence field. Chen, Ou and Gong (2010) compare a number of interpolation schemes for directly interpolating all precipitation amounts (including zeroes) and finds ordinary kriging works best among the methods studied, despite its obvious problems for a process that is often 0. None of these works propose a stochastic model for the precipitation process and generally use some kind of cross-validation to assess the accuracy of the method. Cross-validation does not give locally meaningful uncertainties for either the occurrence field or the amount of precipitation, nor does it provide a way of giving uncertainties for total precipitation over a region.

To obtain locally meaningful interpolation uncertainties when one has irregular observations, which is the case for rain gauge networks, some degree of modeling is unavoidable. Left-truncation of a Gaussian process at 0 together with some monotonic transformation of the positive values to match the empirical distribution of precipitation amounts provides a unified model for both occurrence and amounts of precipitation and was initially proposed by Bell (1987). However, Berrocal, Raftery and Gneiting (2008) argued that using two separate Gaussian processes, one to model occurrences and a separate, independent Gaussian process that, after a pointwise transformation, models precipitation amounts, can provide a better description of the spatial structure of daily precipitation fields, although the goal in this work was short-term forecasting, not interpolation. In my view, Gaussian processes as a building block for space-time precipitation models may need to be abandoned to obtain realistic representations of important spatio-temporal features of precipitation events such as cyclonic patterns and squall lines. The conceptual challenges of developing appropriate models are immense and it will be critical to have diagnostic methods to assess how well the models capture complex space-time dependencies (Sun and Stein, 2015).

5. STATISTICS AND STATISTICIANS IN CLIMATE SCIENCE

An often-used quote in climatology is "The climate is what you expect; the weather is what you get." To be accurate, "expect" should be interpreted broadly as the entire multivariate space-time relationship of relevant climate variables and not just their means. Descriptions of this multivariate relationship are generally in terms of probability distributions, so there is intrinsically a large statistical component to climatology. As a consequence, a substantial proportion of climate scientists are well-versed in at least some aspects of statistics, whereas the proportion of statisticians who have more than a passing knowledge of climatology is quite limited. The fact that differential equations are not generally a part of a statistician's education does not help matters. Over the years, I have co-organized a number of workshops, classes and other events meant to bring together climatologists and statisticians and have found it challenging to obtain good balances between the two groups in terms of both numbers and levels of understanding of the other side's discipline. The situation is somewhat improved from a few decades ago, but effective leadership within academic institutions and funding agencies will be essential to make further progress.

The intellectual opportunities in statistical climatology are great and provide ample motivation for the engagement of the best of the statistics profession. Considering the potential impact of climate change on all life on Earth, there may be no scientific area in greater need of statistical input. Forecasting future climate change and its impacts in a way that takes appropriate account of the myriad available sources of information to produce accurate forecasts with realistic uncertainties (IPCC, 2013, Tebaldi and Knutti, 2007, Knutti et al., 2010, Nordhaus, 2018) is a challenge that the statistics community, working in close concert with climate scientists, can and must address.

ACKNOWLEDGMENTS

The author thanks Doug Nychka and Robert Kopp for careful readings of drafts of this paper. This material was based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR) under Contract DE-AC02-06CH11357.

REFERENCES

AILLIOT, P., ALLARD, D., MONBET, V. and NAVEAU, P. (2015). Stochastic weather generators: An overview of weather type models. *J. SFdS* **156** 101–113. MR3338244

ANDERSON, B. and BELL, M. (2009). Weather-related mortality: How heat, cold, and heat waves affect mortality in the United States. *Epidemiology* **20** 205–213.

ARGO (2000). Argo float data and metadata from Global Data Assembly Centre (Argo GDAC).

ARRHENIUS, S. (1896). On the influence of carbonic acid in the air on the temperature on the ground. *Philos. Mag.* **41** 237–276.

ARRHENIUS, S. (1908). *Worlds in the Making*: *The Evolution of the Universe*. Harper & Brothers Publishers, New York.

BALSAMO, G., ALBERGEL, C., BELJAARS, A., BOUSSETTA, S., BRUN, E., CLOKE, H., DEE, D., DUTRA, E., MUÑOZ-SABATER, J. et al. (2015). ERA-Interim/Land: A global land surface reanalysis data set. *Hydrol. Earth Syst. Sci.* **19** 389–407.

BELL, T. L. (1987). A space-time stochastic model of rainfall for satellite remote-sensing studies. *J. Geophys. Res.*, *Atmos.* **92** 9631–9643.

BERLINER, L. M., LEVINE, R. A. and SHEA, D. J. (2000). Bayesian climate change assessment. *J. Climate* **13** 3805–3820.

BERNER, J., ACHATZ, U., BATTÉ, L., BENGTSSON, L., DE LA CÁMARA, A., CHRISTENSEN, H. M., COLANGELI, M., COLEMAN, D. R. B., CROMMELIN, D. et al. (2017). Stochastic parameterization: Toward a new view of weather and climate models. *Bull. Am. Meteorol. Soc.* **98** 565–588.

BERROCAL, V. J., RAFTERY, A. E. and GNEITING, T. (2008). Probabilistic quantitative precipitation field forecasting using a two-stage spatial model. *Ann. Appl. Stat.* **2** 1170–1193. MR2655654 https://doi.org/10.1214/08-AOAS203

CASTRUCCIO, S., MCINERNEY, D. J., STEIN, M. L., LIU CROUCH, F., JACOB, R. L. and MOYER, E. J. (2014). Statistical emulation of climate model projections based on precomputed GCM runs. *J. Climate* **27** 1829–1844.

CHEN, D., OU, T. and GONG, L. (2010). Spatial interpolation of daily precipitation in China: 1951–2005. *Adv. Atmos. Sci.* **27** 1221–1232.

CHENG, L., AGHAKOUCHAK, A., GILLELAND, E. and KATZ, R. W. (2014). Non-stationary extreme value analysis in a changing climate. *Clim. Change* **127** 353–369.

COLLINS, M., BOOTH, B. B. B., BHASKARAN, B., HARRIS, G. R., MURPHY, J. M., SEXTON, D. M. H. and WEBB, M. J. (2011). Climate model errors, feedbacks and forcings: A comparison of perturbed physics and multi-model ensembles. *Clim. Dyn.* **36** 1737–1766.

COOLEY, D. and SAIN, S. R. (2010). Spatial hierarchical modeling of precipitation extremes from a regional climate model. *J. Agric. Biol. Environ. Stat.* **15** 381–402. MR2787265 https://doi.org/10.1007/s13253-010-0023-9

COOLEY, D., CISEWSKI, J., ERHARDT, R. J., JEON, S., MANNSHARDT, E., OMOLO, B. O. and SUN, Y. (2012). A survey of spatial extremes: Measuring spatial dependence and modeling spatial effects. *REVSTAT* **10** 135–165. MR2912374

CORTI, S., PALMER, T., BALMASEDA, M., WEISHEIMER, A., DRIJFHOUT, S., DUNSTONE, N., HAZELEGER, W., KRÖGER, J., POHLMANN, H. et al. (2015). Impact of initial conditions versus external forcing in decadal climate predictions: A sensitivity experiment. *J. Climate* **28** 4454–4470.

CRESSIE, N. and HUANG, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *J. Amer. Statist. Assoc.* **94** 1330–1340. MR1731494 https://doi.org/10.2307/2669946

CRESSIE, N. and WIKLE, C. K. (2011). *Statistics for Spatio-Temporal Data. Wiley Series in Probability and Statistics*. Wiley, Hoboken, NJ. MR2848400

DAVISON, A. C. and HUSER, R. (2015). Statistics of extremes. *Annu. Rev. Stat. Appl.* **2** 203–235.

DAVISON, A. C., PADOAN, S. A. and RIBATET, M. (2012). Statistical modeling of spatial extremes. *Statist. Sci.* **27** 161–186. MR2963980 https://doi.org/10.1214/11-STS376

DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory: An Introduction. Springer Series in Operations Research and Financial Engineering*. Springer, New York. MR2234156 https://doi.org/10.1007/0-387-34471-3

DIGGLE, P. J., TAWN, J. A. and MOYEED, R. A. (1998). Model-based geostatistics. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **47** 299–350. MR1626544 https://doi.org/10.1111/1467-9876.00113

DI LUZIO, M., JOHNSON, G. L., DALY, C., EISCHEID, J. K. and ARNOLD, J. G. (2008). Constructing retrospective gridded daily precipitation and temperature datasets for the conterminous United States. *J. Appl. Meteorol. Climatol.* **47** 475–497.

EDWARDS, P. N. (2011). History of climate modeling. *Wiley Interdiscip. Rev.: Clim. Change* **2** 128–139.

EINMAHL, J. H. J., DE HAAN, L. and ZHOU, C. (2016). Statistics of heteroscedastic extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 31–51. MR3453645 https://doi.org/10.1111/rssb.12099

FAWCETT, L. and WALSHAW, D. (2007). Improved estimation for temporally clustered extremes. *Environmetrics* **18** 173–188. MR2345653 https://doi.org/10.1002/env.810

FERRO, C. A. T. and SEGERS, J. (2003). Inference for clusters of extreme values. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **65** 545–556. MR1983763 https://doi.org/10.1111/1467-9868.00401

FIELD, C., BARROS, V., STOCKER, D., QIN, D., DOKKEN, K., EBI, M., MASTRANDREA, K., MACH, G.-K., PLATTNER, S. et al., eds. (2012). *IPCC*, 2012: *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change*. Cambridge Univ. Press, Cambridge.

FORGET, G., CAMPIN, J. M., HEIMBACH, P., HILL, C. N., PONTE, R. M. and WUNSCH, C. (2015). ECCO version 4: An integrated framework for non-linear inverse modeling and global ocean state estimation. *Geosci. Model Dev.* **8** 3071–3104.

GASPARRINI, A. and ARMSTRONG, B. (2011). The impact of heat waves on mortality. *Epidemiology* **22** 68–73. https://doi.org/10.1097/EDE.0b013e3181fdcd99

GNEITING, T. (2002). Nonseparable, stationary covariance functions for space-time data. *J. Amer. Statist. Assoc.* **97** 590–600. MR1941475 https://doi.org/10.1198/016214502760047113

GRAY, A. R. and RISER, S. C. (2014). A global analysis of Sverdrup balance using absolute geostrophic velocities from Argo. *J. Phys. Oceanogr.* **44** 1213–1229.

HAUGEN, M. A., STEIN, M. L., MOYER, E. J. and SRIVER, R. L. (2018). Estimating changes in temperature distributions in a large ensemble of climate simulations using quantile regression. *J. Climate* **31** 8573–8588.

HOLLMANN, R., MERCHANT, C. J., SAUNDERS, R., DOWNY, C., BUCHWITZ, M., CAZENAVE, A., CHUVIECO, E., DEFOURNY, P., DE LEEUW, G. et al. (2013). The ESA climate change initiative: Satellite data records for essential climate variables. *Bull. Am. Meteorol. Soc.* **94** 1541–1552.

HOSKING, J. R. M., WALLIS, J. R. and WOOD, E. F. (1985). An appraisal of the regional flood frequency procedure in the UK Flood Studies Report. *Hydrol. Sci. J.* **30** 85–109.

HUANG, W. K., STEIN, M. L., MCINERNEY, D. J., SUN, S. and MOYER, E. J. (2016). Estimating changes in temperature extremes from millennial-scale climate simulations using generalized extreme value (GEV) distributions. *Adv. Stat. Climatol. Meteorol. Oceanogr.* **2** 79–103.

HUTCHINSON, M. F., MCKENNEY, D. W., LAWRENCE, K., PEDLAR, J. H., HOPKINSON, R. F., MILEWSKA, E. and PAPADOPOL, P. (2009). Development and testing of Canada-wide interpolated spatial models of daily minimum-maximum temperature and precipitation for 1961–2003. *J. Appl. Meteorol. Climatol.* **48** 725–741.

IPCC (2013). *Climate Change* 2013: *The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge Univ. Press, Cambridge.

JALBERT, J., FAVRE, A.-C., BÉLISLE, C. and ANGERS, J.-F. (2017). A spatiotemporal model for extreme precipitation simulated by a climate model, with an application to assessing changes in return levels over North America. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **66** 941–962. MR3715590 https://doi.org/10.1111/rssc.12212

JUN, M., KNUTTI, R. and NYCHKA, D. W. (2008). Spatial analysis to quantify numerical model bias and dependence: How many climate models are there? *J. Amer. Statist. Assoc.* **103** 934–947. MR2528820 https://doi.org/10.1198/016214507000001265

KATZ, R. W. and BROWN, B. G. (1992). Extreme events in a changing climate: Variability is more important than averages. *Clim. Change* **21** 289–302.

KATZ, R. W., PARLANGE, M. B. and NAVEAU, P. (2002). Statistics of extremes in hydrology. *Adv. Water Resour.* **25** 1287–1304.

KAY, J. E., DESER, C., PHILLIPS, A., MAI, A., HANNAY, C., STRAND, G., ARBLASTER, J. M., BATES, S. C., DANABASOGLU, G. et al. (2015). The Community Earth System Model (CESM) Large Ensemble Project: A community resource for studying climate change in the presence of internal climate variability. *Bull. Am. Meteorol. Soc.* **96** 1333–1349.

KHARIN, V. V. and ZWIERS, F. W. (2005). Estimating extremes in transient climate change simulations. *J. Climate* **18** 1156–1173.

KHARIN, V. V., ZWIERS, F. W., ZHANG, X. and WEHNER, M. (2013). Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Clim. Change* **119** 345–357.

KNUTTI, R., FURRER, R., TEBALDI, C., CERMAK, J. and MEEHL, G. A. (2010). Challenges in combining projections from multiple climate models. *J. Climate* **23** 2739–2758.

KUUSELA, M. and STEIN, M. L. (2018). Locally stationary spatio-temporal interpolation of Argo profiling float data. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **474** 20180400.

LÓPEZ-LOPERA, A. F., BACHOC, F., DURRANDE, N. and ROUSTANT, O. (2018). Finite-dimensional Gaussian approximation with linear inequality constraints. *SIAM/ASA J. Uncertain. Quantificat.* **6** 1224–1255. MR3857898 https://doi.org/10.1137/17M1153157

LORENZ, E. N. (1963). Deterministic nonperiodic flow. *J. Atmos. Sci.* **20** 130–141.

MEARNS, L. O., SAIN, S., LEUNG, L. R., BUKOVSKY, M. S., MCGINNIS, S., BINER, S., CAYA, D., ARRITT, R. W., GUTOWSKI, W. et al. (2013). Climate change projections of the North American Regional Climate Change Assessment Program (NARCCAP). *Clim. Change* **120** 965–975.

NEELIN, J. D., BRACCO, A., LUO, H., MCWILLIAMS, J. C. and MEYERSON, J. E. (2010). Considerations for parameter optimization and sensitivity in climate models. *Proc. Natl. Acad. Sci. USA* **107** 21349–21354.

NORDHAUS, W. (2018). Projections and uncertainties about climate change in an era of minimal climate policies. *Am. Econ. J. Econ. Policy* **10** 333–360.

INSTITUTE OF HYDROLOGY (1975). *Flood Studies Report*. Natural Environment Research Council, London.

OLSON, B. and KLEIBER, W. (2017). Approximate Bayesian computation methods for daily spatiotemporal precipitation occurrence simulation. *Water Resour. Res.* **53** 3352–3372.

OPITZ, T., HUSER, R., BAKKA, H. and RUE, H. (2018). INLA goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles. *Extremes* **21** 441–462. MR3855716 https://doi.org/10.1007/s10687-018-0324-x

PAYNTER, D., FRÖLICHER, T. L., HOROWITZ, L. W. and SILVERS, L. G. (2018). Equilibrium climate sensitivity obtained from multimillennial runs of two GFDL climate models. *J. Geophys. Res., Atmos.* **123** 1921–1941.

PEÑA-ARANCIBIA, J. L., VAN DIJK, A. I. J. M., RENZULLO, L. J. and MULLIGAN, M. (2013). Evaluation of precipitation estimation accuracy in reanalyses, satellite products, and an ensemble method for regions in Australia and South and East Asia. *J. Hydrometeorol.* **14** 1323–1333.

PIERREHUMBERT, R. T. (2010). *Principles of Planetary Climate*. Cambridge Univ. Press, Cambridge. MR2778154 https://doi.org/10.1017/CBO9780511780783

RAHMSTORF, S., CAZENAVE, A., CHURCH, J. A., HANSEN, J. E., KEELING, R. F., PARKER, D. E. and SOMERVILLE, R. C. J. (2007). Recent climate observations compared to projections. *Science* **316** 709–709.

RAMANATHAN, V. (1988). The greenhouse theory of climate change: A test by an inadvertent global experiment. *Science* **240** 293–299.

RIBES, A., ZWIERS, F. W., AZAÏS, J.-M. and NAVEAU, P. (2017). A new statistical approach to climate change detection and attribution. *Clim. Dyn.* **48** 367–386.

RISER, S. C., FREELAND, H. J., ROEMMICH, D., WIJFFELS, S., TROISI, A., BELBEOCH, M., GILBERT, D., XU, J., POULIQUEN, S. et al. (2016). Fifteen years of ocean observations with the global Argo array. *Nat. Clim. Change* **6** 145–153.

RODDA, J. C. and DIXON, H. (2012). Rainfall measurement revisited. *Weather* **67** 131–136.

ROTH, M., JONGBLOED, G. and BUISHAND, A. (2019). Monotone trends in the distribution of climate extremes. *Theor. Appl. Climatol.* **136** 1175–1184.

RUMMUKAINEN, M. (2012). Changes in climate and weather extremes in the 21st century. *Wiley Interdiscip. Rev.: Clim. Change* **3** 115–129.

SAHA, S., MOORTHI, S., PAN, H.-L., WU, X., WANG, J., NADIGA, S., TRIPP, P., KISTLER, R., WOOLLEN, J. et al. (2010). The NCEP climate forecast system reanalysis. *Bull. Am. Meteorol. Soc.* **91** 1015–1058.

SCHMIT, T. J., GRIFFITH, P., GUNSHOR, M. M., DANIELS, J. M., GOODMAN, S. J. and LEBAIR, W. J. (2017). A closer look at the ABI on the GOES-R series. *Bull. Am. Meteorol. Soc.* **98** 681–698.

SEXTON, D. M. H., MURPHY, J. M., COLLINS, M. and WEBB, M. J. (2012). Multivariate probabilistic projections using imperfect climate models part I: Outline of methodology. *Clim. Dyn.* **38** 2513–2542.

SIGRIST, F., KÜNSCH, H. R. and STAHEL, W. A. (2015). Stochastic partial differential equation based modelling of large space-time data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 3–33. MR3299397 https://doi.org/10.1111/rssb.12061

SRIVER, R. L., FOREST, C. E. and KELLER, K. (2015). Effects of initial conditions uncertainty on regional climate variability: An analysis using a low-resolution CESM ensemble. *Geophys. Res. Lett.* **42** 5468–5476.

STAMMER, D., BALMASEDA, M., HEIMBACH, P., KÖHL, A. and WEAVER, A. (2016). Ocean data assimilation in support of climate applications: Status and perspectives. *Annu. Rev. Mar. Sci.* **8** 491–518. https://doi.org/10.1146/annurev-marine-122414-034113

STEIN, M. L. (2005). Space-time covariance functions. *J. Amer. Statist. Assoc.* **100** 310–321. MR2156840 https://doi.org/10.1198/016214504000000854

STEIN, M. L. (2017). Should annual maximum temperatures follow a generalized extreme value distribution? *Biometrika* **104** 1–16. MR3626483 https://doi.org/10.1093/biomet/asw070

STERL, A., SEVERIJNS, C., DIJKSTRA, H., HAZELEGER, W., JAN VAN OLDENBORGH, G., VAN DEN BROEKE, M., BURGERS, G., VAN DEN HURK, B., JAN VAN LEEUWEN, P. et al. (2012). When can we expect extremely high surface temperatures? *Geophys. Res. Lett.* **35** L14703.

SUN, Y. and STEIN, M. L. (2015). A stochastic space-time model for intermittent precipitation occurrences. *Ann. Appl. Stat.* **9** 2110–2132. MR3456368 https://doi.org/10.1214/15-AOAS875

TAYLOR, K. E., STOUFFER, R. J. and MEEHL, G. A. (2012). An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93** 485–498.

TEBALDI, C. and KNUTTI, R. (2007). The use of the multi-model ensemble in probabilistic climate projections. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **365** 2053–2075. MR2317897 https://doi.org/10.1098/rsta.2007.2076

TOUMA, D., MICHALAK, A. M., SWAIN, D. L. and DIFFENBAUGH, N. S. (2018). Characterizing the spatial scales of extreme daily precipitation in the United States. *J. Climate* **31** 8023–8037.

TROY, T. J., KIPGEN, C. and PAL, I. (2015). The impact of climate extremes and irrigation on US crop yields. *Environ. Res. Lett.* **10** 054013.

VARDI, M. Y. (2010). Science has only two legs. *Commun. ACM* **53** 5.

WALLIN, J. and BOLIN, D. (2015). Geostatistical modelling using non-Gaussian Matérn fields. *Scand. J. Stat.* **42** 872–890. MR3391697 https://doi.org/10.1111/sjos.12141

WANG, X. and BERGER, J. O. (2016). Estimating shape constrained functions using Gaussian processes. *SIAM/ASA J. Uncertain. Quantificat.* **4** 1–25. MR3452261 https://doi.org/10.1137/140955033

WESTRA, S., ALEXANDER, L. V. and ZWIERS, F. W. (2013). Global increasing trends in annual maximum daily precipitation. *J. Climate* **26** 3904–3918.

WIKLE, C. K. (2015). Modern perspectives on statistics for spatio-temporal data. *Wiley Interdiscip. Rev.: Comput. Stat.* **7** 86–98. MR3348724 https://doi.org/10.1002/wics.1341

WILKS, D. S. and WILBY, R. L. (1999). The weather generation game: A review of stochastic weather models. *Prog. Phys. Geogr., Earth Environ.* **23** 329–357.

XU, G. and GENTON, M. G. (2017). Tukey g-and-h random fields. *J. Amer. Statist. Assoc.* **112** 1236–1249. MR3735373 https://doi.org/10.1080/01621459.2016.1205501

YATAGAI, A., KAMIGUCHI, K., ARAKAWA, O., HAMADA, A., YASUTOMI, N. and KITOH, A. (2012). APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bull. Am. Meteorol. Soc.* **93** 1401–1415.

YOKOHATA, T., WEBB, M. J., COLLINS, M., WILLIAMS, K. D., YOSHIMORI, M., HARGREAVES, J. C. and ANNAN, J. D. (2010). Structural similarities and differences in climate responses to CO_2 increase between two perturbed physics ensembles. *J. Climate* **23** 1392–1410.

ZWIERS, F. W., ALEXANDER, L. V., HEGERL, G. C., KNUTSON, T. R., KOSSIN, J. P., NAVEAU, P., NICHOLLS, N., SCHÄR, C., SENEVIRATNE, S. I. et al. (2013). Climate extremes: Challenges in estimating and understanding recent changes in the frequency and intensity of extreme climate and weather events. In *Climate Science for Serving Society* (G. Asrar and J. Hurrell, eds.) 339–389. Springer, Dordrecht.