a. Intellectual Focus

Although applied mathematics has long been at the core of modeling processes in fluids, there have long been calls to get more statisticians involved in these problems. Because of the enormous interest in climate change, it is not surprising that many recent public statements in this regard relate to climatology. The academic review of the University of East Anglia Climate Research Unit expressed surprise that the researchers had not worked more closely with statisticians (Washington Post, April 15, 2010). A recent workshop on climate extremes, sponsored by UNESCO and the World Climate Research Program, expressed as one of its final conclusions that it is imperative for the climate science community to develop incentives for statisticians to get involved in climate science. The need for further statistical input into physical oceanography has long been recognized (Panel on Statistics and Oceanography, 1994) and the US EPA has demonstrated its need for better statistical methods by supporting environmental statistics centers, two of which were directed by PI’s of this project (Peter Guttorp, National Research Center for Statistics and the Environment; Michael Stein, Center for Integrating Statistical and Environmental Science). NSF’s long support of the Geophysical Statistics Project and the success of its programs, especially its postdocs, is a testament to the fertility of this intersection of disciplines.

The statistical challenges within just climatology or atmospheric chemistry or oceanography could provide more than enough intellectual fodder for a network. However, considering the atmospheric and oceanic sciences as a whole will allow us to examine the interdependencies between the multitude of processes that influence the biosphere. For example, in order to understand climate fully, it is essential to understand weather, atmospheric chemistry, oceans, ice, and terrestrial processes, particularly with respect to the carbon and water cycles. Biological processes and human activities are both affected by and affect climate on local (urban heat islands), regional (land use patterns) and global scales (greenhouse gas emissions), so that a comprehensive study of climate change also requires hydrology, forestry, ecology, agriculture, economics and public policy to name just a fraction of the relevant disciplines.

The need for statistical models for spatial dependence in meteorology goes back at least to the classic work by Gandin (1963) on objective analysis, or what statisticians might call simple kriging or best linear prediction. Despite this long history, the challenge to develop appropriate statistical models for processes in the atmosphere and oceans remains because of the immense difficulty of accurately capturing dynamic aspects of spatio-temporal processes, especially nonlinear and non-stationary behavior. For time scales on which dynamics matter, spatio-temporal statistics requires far more than combining approaches for time series and purely spatial processes.

In the best tradition of statistical research, the scientific problems in atmospheric and oceanic sciences generate a host of challenges in statistical theory, methods and computation. We see a number of common statistical themes across the scientific problems we plan to study.

- Multivariate models for nonstationarity/nonlinearity/non-Gaussianity. Stationary Gaussian processes are the workhorse of models for processes indexed in continuous space or space-time. However, from the strong non-Gaussian character of precipitation over short time scales to the nonstationarity in space of air pollution in urban regions to the nonstationarity
in time induced by climate change, it is clear that more sophisticated models are needed. For many problems, dependencies between processes (e.g., precipitation and temperature or NOx and ozone) are important. Although there has been much work in the univariate setting, including the pioneering work of Sampson and Guttorp (1992), much more needs to be done on, for example, modeling processes that are nonstationary in both space and time and non-stationarity in the cross-dependencies of multivariate processes.

• Comparing numerical model output to observations is fundamental to evaluating their effectiveness. This seemingly simple problem is in fact quite challenging due, in large part, to the fact that model output and observations rarely refer to identical quantities. In particular, in situ monitors measure processes over very small spatial scales whereas model output at best gives area averages over grid boxes. The relationship of remotely sensed measurements to the quantities of interest is often quite indirect and highly nonlinear.

• Uncertainty quantification. In order to assess the impacts of climate change, extreme weather or air pollution, it is imperative to develop realistic estimates that take account of all of the major sources of uncertainties. Uncertainty assessment when interpolating spatio-temporal processes is already a challenge. Uncertainty assessment when extrapolating to situations that are outside our present experience is particularly difficult. Because climate change predictions require a severe extrapolation in both time and in atmospheric conditions, providing realistic uncertainties for climate projections is especially treacherous, as one needs to take account of the uncertainties due to scenarios, global models, regional models, and impact effects. Quantifying uncertainties due to model misspecification is a particular challenge, but one that cannot be ignored since it may be the largest source of uncertainty.

• From air quality standards to severe weather events, extreme values play a major role in the analysis of environmental data. Statistical inference of extremes when there are strong spatial and temporal dependencies is a notoriously difficult problem both theoretically and practically. The need for new approaches is apparent, especially in settings when a single weather event may produce the annual maximum in some quantity of interest (i.e., precipitation, ozone level) over an extended region.

• All of the previous items create immense computational challenges in addition to the methodological and theoretical challenges. For large, irregularly sited observation networks, even fitting stationary Gaussian process models exactly is not feasible. Extending likelihood-based (including Bayesian) methods to nonstationary, non-Gaussian settings will require the development of new approximations and new algorithms. Parallel processing will be essential to handle the largest problems.

The Network on Statistical Methods in Atmospheric and Oceanic Sciences will coordinate statistical research based on issues in atmospheric and oceanic science; develop a cadre of PhD students, postdocs and young faculty with substantial experience in doing crossdisciplinary work with statisticians and physical, biological and social scientists; develop new statistical tools for
space-time processes relevant to atmospheres and oceans; and organize a series of workshops and summer schools to disseminate these ideas. In addition, the network will sponsor an electronic journal entitled Statistical Methods in Atmospheric and Oceanic Sciences.

The goal of the network is to become the intellectual focus for research on statistical methodology in atmospheric and oceanic sciences, broadly interpreted. International cooperation is therefore extremely important. The network web pages should be the central electronic repository for the latest research in the field, as well as maintaining a calendar of relevant activities, links to data sets and analyses, and to the electronic journal.

The inter-connections between the 12 institutions will ensure that all the scientific aspects of the work: statistical methodology, and analysis of atmospheric and oceanic processes, is relevant and correct. The expertise at the 12 institutions complements each other, and brings synergy. The research and educational activities of this proposal simply could not be done by the individual collaborators with their own grants separately; a group effort is required. NCDC will assist getting access to relevant data, the oceanographers and atmospheric scientists in the team from NCAR, UC, UW, UNC-Wilmington, Duke, NCSU and other institutions will work closely with the statisticians on some of the most relevant and challenging scientific problems in atmospheric and oceanic sciences by developing the appropriate statistical methodology. We will enhance and advance science with the input and expertise brought by the partnership and synergy among the statisticians and atmospheric and oceanic scientists at the 12 institutions. Accordingly, we have carefully developed a governance and management plan (c) to promote this effective collaboration and team synergy.

b Scientific Activities
b1 Research areas
b1.1 Spatial and spatio-temporal extremes in climate and weather

Extreme temperature, wind, and precipitation events may cause loss of life, injury, property damage, and threaten the existence of some species. Thus, understanding and predicting the spatial and temporal variability and trends of extreme weather events is crucial for the protection of socio-economic well-being. The recent report of the government’s Climate Change Science Program (Climate Change Science Program, 2008) states that the greatest impacts of climate change on society and wildlife will be experienced through changes in extreme weather events as global temperatures increase (van Vliet and Leemans, 2006).

Two applications in which extreme value theory could be of particular importance are forest fires and droughts/heat waves. Spatio-temporal methods (e.g. Brillinger et al. (2006); Den et al. (2009); Preisler et al. (2008)) for forest fire risk analysis area has so far not utilized methodology from extreme value theory. In the forest fire context, increasing temperatures could lead to an increase in the number of ignitions, an increase in the length of the fire season, and an increase in the amount of severe fire weather (Flannigan et al. (2005); Podur et al. (2002); Flannigan et al. (2009); Wotton et al. (2010)). Some additional challenges with quantifying extremes in this context is the need for homogenization of data from long records, incorporating information about changes in suppression activities, and fire management strategies. Given the challenges with climate predictions, we will investigate what is the best way to accommodate weather variables to
evaluate impacts under future climate scenarios; a sensible approach may be to focus on assessing how large a change in certain weather variables would lead to specific forestry vulnerabilities.

The statistical theory of extremes for univariate time series can still be problematic to apply to complex forms of extreme weather events such as droughts or heat waves. Although such events are clearly extreme, their definitions tend to be rather nebulous (Meze-Hausken (2008); Robinson (2001); Wilhite (2000)). Nevertheless, climate change predictions are dire for both drought (Dai, 2010) and heat waves (Meehl and Tebaldi, 2004). For heat waves, the primary impact of concern is mortality (Gosling et al., 2009). For drought, the economic impacts can be substantial both in the developing and in the developed world (Wilhite, 2000), with famine being a more disastrous consequence in the developing world (Glantz, 1987).

Such extreme events involve clustering, both in a meteorological and in a statistical sense. Starting with Coles et al. (1994), attempts have been made to apply extreme value theory through extensions of the point process approach to formally model clusters at high levels (e.g., via Markov chain models with bivariate extreme value distributions; Smith et al. (1997)). A simpler approach, based on only univariate extreme value theory, has recently been proposed as an alternative (Furrer et al., 2010).

One of the challenging issues in spatial extreme value modeling is the need for techniques in high dimensions, since most of the multivariate extreme value theories only work well for low-dimensional extremes. We propose an innovative and general statistical framework based on a nonparametric Dirichlet process (DP) copula. The proposed DP mixture model has generalized extreme value (GEV) marginal distributions with spatially varying parameters and the observations are spatially-correlated even after conditioning on the spatially varying parameters. Bayesian nonparametric methods avoid dependence on parametric assumptions by working with probability models on function spaces.

A Dirichlet prior for a random $m$-vector $Z$ has random point masses $\theta_1, \ldots, \theta_K$ ($K$ possibly infinite) independently drawn from some base measure $H$ and probabilities assigned to these masses that come from a Dirichlet distribution, a distribution on the set of $K$ nonnegative numbers adding to 1. These probabilities can be represented in terms of independent $\text{Beta}(1, \nu)$ random variables $V_1, \ldots, V_{K-1}$ (Sethuraman (1994)), so that the Dirichlet prior can be characterized by $H$, $\nu$ and $K$. The Dirichlet distribution is a special case of what is known as a stick-breaking prior.

In order to make this wide class of nonparametric priors useful for our spatial context, we need to index it by space. More generally, we can attempt to introduce dependencies on time or other covariates (leading to nonparametric regression models). Most of the (rather recent) literature in this area follows the ideas in MacEachern (1999), who considered allowing $V = (V_1, V_2, \ldots)$ or $\theta = (\theta_1, \theta_2, \ldots)$ to follow a stochastic process defined over the domain. An application to spatial modelling is developed in Gelfand et al. (2005) by allowing $\theta$ to be drawn from a random field (a Gaussian process). Other spatial extensions are introduced by Griffin and Steel (2006), Reich and Fuentes (2007), Dunson and Park (2008), and An et al. (2009).

Using $s$ to indicate spatial location and $t$ time throughout this proposal, the spatial DP copula models through the process of interest $Y$ in terms of a latent process $Z$. Specifically, for times $t = 1, \ldots, T$ and locations $s_1, \ldots, s_m$, denote by $f_Z$ and $F_Z$ the joint density and joint cumulative
distribution function (cdf), respectively, of $Z = (Z(s_1), \ldots, Z(s_m))$. Then for an $m$-dimensional base measure $H$ and a nugget variance $\tau^2$ (and $K = \infty$), take $f_Z = \sum_{i=1}^{\infty} p_i N_m(Z|\theta_i, \tau^2 I_m)$, where $N_m(.)|\lambda, \Sigma)$ denotes the $m$-variate normal density with mean vector $\lambda$ and covariance $\Sigma$, the $\theta_i = (\theta_i(s_1), \ldots, \theta_i(s_m))$ are independently drawn from $H$ and the $p_i$'s follow the Dirichlet prior and hence are defined by independent $V_i \sim \text{Beta}(1, \nu)$. In the simplest case, $H$ itself is multivariate normal, but other choices are possible.

For $F(s)$ the cdf of $Z(s)$, $T(s) = F_s(Z(s)) \sim \text{Unif}(0, 1)$. The copula $C_Z$ for the distribution function of $Z(s_1), \ldots, Z(s_m)$ is (conditioning on the $\theta_i$ components), $C_Z(u_1, \ldots, u_m) = F_Z(F_{s_1}^{-1}(u_1), \ldots, F_{s_m}^{-1}(u_m))$, where $u_1, \ldots, u_m \in [0, 1]^m$. Then, $Y(s) \sim G^{-1}(T(s)) \sim G$. $G$ is the cdf of the standard Fréchet distribution. Instead of a standard Fréchet, we generalize the marginal distributions to be Generalized Extreme Value (GEV) distributions with space-dependent parameters by incorporating a change of variable in the likelihood function. The multivariate distribution of $Y$ is $F_Y(y_1, \ldots, y_m) = C_Z(G(y_1), \ldots, G(y_m))$.

Another approach to temporal extremes of space-time processes is to assign a spatial prior for the GEV-parameters for annual or seasonal extremes over a network of station while treating the stations as conditionally independent. However, considering, for example, temperature data, it is quite common in upper latitudes for extremely cold weather to arise from Arctic air masses in a high pressure situation. Hence there is a tendency for annual minima to appear simultaneously at several stations. Although the max-stable approach (Kabluchko et al., 2009) has theoretical grounds for being applicable to the case of spatial block maxima, it does not allow an explicit likelihood expression (Padoan et al., 2010), and statistical analysis has proceeded in an ad hoc fashion using a composite likelihood approach in which the likelihood is approximated by a product of bivariate densities. Based on the meteorological description above, the appropriate likelihood (as long as separated maxima can be considered independent) would be the product of conditional densities of the nonextreme sites, given the values at the extreme sites. Calculating these densities can of course be a daunting task in itself, but we propose using the approximation due to Heffernan and Tawn (2004), appropriately extended to the situation at hand. A comparison will then be made with the composite likelihood approach.

**b1.2. Multivariate spatial models**

**Circular models for winds**: With the current push for cleaner energy, many are setting their eyes on harnessing the power of the wind. The siting of wind turbines is critical to the network’s effectiveness. Researchers are currently trying to pinpoint locations where it would be beneficial for these turbines to be placed. To do that, they need to be able to model the wind speed, direction, and duration at different sites. Dependence in speed across sites is critical as it is the distribution of energy generated by the entire network and not at individual sites that is of greatest interest for electric utilities.

Another example of vector fields in space-time is the wind fields generated by a hurricane. Residents living along the coastal area of the southeast United States and Gulf Coast are presented with many hazards during a landfalling hurricane. With populations in these areas increasing, it is imperative that as storms approach the coastline we have the means to give these citizens accurate
and reliable information about possible landfall locations and conditions. Storm winds, torrential rain, and spawned tornadoes each can harm both life and property, but the single largest threat to coastal areas is the storm surge. With homes and businesses being either right at or just a few feet above sea level, this inundation of water pushed by the landfalling storm can quickly take lives and destroy property.

Accurate forecasts for storm surges depend upon accurate modeling of wind forcings. At present, there is no standard model adopted for the specific purpose of improving storm surge forecasts. There have been some models and methods developed to assist with modeling hurricane wind fields. Depperman (1947), Holland (1980) and DeMaria (1992) each presented models that have been termed as axis-symmetric. These models are based upon a cyclostrophic wind balance and place the key dependence on the distance a location is from the storm circulation center. They do not describe the asymmetrical structure of the winds within hurricanes. Winds within the northeast quadrant of the storm are typically stronger than those in other locations within a storm. The asymmetry can be attributed to friction, environment, vertical shear, etc. (Chen and Yau (2003), Ross and Kurihara (1992), Shapiro (1983), and Wang and Holland (1996)). We propose here a novel statistical framework to model and map asymmetrical wind vectors to be used as improved forcing fields for storm surge forecast, providing the corresponding uncertainty analysis. This approach can be applied to other vector fields.

The common statistical modelling approach for wind vectors decomposes the wind fields into the u (North-South) and v (East-West) components (Cartesian representation). Modeling u in hurricanes is challenging because it displays heavy tails, increased variability, and a shorter spatial range near the storm center (nonstationarity). Joint modeling of u and v is also complicated because their correlation varies dramatically in different parts of the spatial domain. In contrast, in hurricanes, log wind speed and wind direction vary smoothly in space and have a simpler joint relationship. Therefore, we propose to model hurricane wind fields using polar coordinates.

We assume that the response in grid cell $i = 1, \ldots, n$ is a vector defined by its cardinal components $(U_i, V_i)$. We transform to polar coordinates, i.e., vector length $\omega_i = \sqrt{U_i^2 + V_i^2} > 0$ and angle $\theta_i = \arctan(U_i, V_i) \in [0, 2\pi)$. To justify a Gaussian model, we further transform to the log vector length $y_i = \log(\omega_i) \in \mathbb{R}$. We specify the model for $(y_i, \theta_i)$ in terms of the true log vector length $\mu_{1i}$ and true angle $\mu_{2i}$. The log vector length is assumed to be $y_i \sim N(\mu_{1i}, \sigma_1^2)$. The mean is decomposed as $\mu_{1i} = X_i^T \beta_1 + g(\mu_{2i}) + \delta_{1i}$, where $X_i^T \beta_1$ represents the contribution to the mean by spatial covariates, $g(\mu_{2i}) = \sum_{k=1}^{M} a_k \sin(k\mu_{2i}) + \sum_{k=1}^{M} b_k \cos(k\mu_{2i})$ captures the relationship between vector direction and log vector length, and $\delta_{1i}$ is a spatial effect.

We model $\theta_i$ by extending the wrapped normal (WN) distribution to the spatial setting. The WN distribution is the symmetric and unimodal distribution obtained by wrapping a normal distribution on the real line around a circle. The density is $f(\theta_i) = \sum_{j=-\infty}^{\infty} \phi(\theta_i | 2\pi j + \mu_{2i}, \sigma_2^2)$, where $\phi(\cdot | m, s^2)$ is the $N(m, s^2)$ density function. Similar to the wind speed model, we decompose the mean angle as $\mu_{2i} = X_i^T \beta_2 + \delta_{2i}$, where $X_i^T \beta_2$ is the covariate effect and $\delta_{2i}$ is a spatial term.

When observations are on a grid, a simple and computationally convenient model for the spatial random effects $\delta_{1i}$ and $\delta_{2i}$ is a proper conditionally autoregressive (CAR) prior (Banerjee et al.,
2004). The CAR covariance is specified through spatial adjacencies. Let \( i \sim j \) indicate that cells \( i \) and \( j \) are spatial neighbors and \( m_i \) be the number of spatial neighbors of cell \( i \). The CAR model for the log vector lengths \( \delta_{1i} \) is defined through the full conditional distribution of \( \delta_{1i} \) given \( \delta_{1j} \) at all other cells with \( j \neq i \). The full conditional distribution is Gaussian with \( E[\delta_{1i}|\delta_{1j}, j \neq i] = \rho_1 \sum_{j \sim i} \delta_{1j}/m_i \) and \( V[\delta_{1i}|\delta_{1j}, j \neq i] = \tau_i^2/m_i \). Although winds, especially in a hurricane, are not divergence-free, it will be worthwhile to investigate whether these models produce realistic levels of divergence.

**Multivariate spatial nonparametric models**: We propose new nonparametric multivariate spatial models that avoid specifying a Gaussian distribution for spatial random effects of interest in atmospheric and oceanic sciences. Here we introduce for the first time an extension of the stick-breaking prior to the multivariate spatial setting that is general enough to allow for nonstationarity and nonseparability in the spatial cross-dependency between the physical processes of interest. Nonseparability refers to allowing the cross-dependence among the variables of interest to be space-dependent. The modeling framework proposed here is computationally efficient because it avoids inverting large matrices and calculating determinants, which often hinders the spatial analysis of large data sets.

The spatial distribution of a stochastic process \( Z(s) \) is modeled using an extension of the stick breaking prior that directly describes nonstationarity. The nonparametric spatial model assigns a different prior distribution to the stochastic process at each location, i.e., \( F_s(Z) \). The distributions \( F_s(Z) \) are unknown and smoothed spatially. The coordinate \( s \) is in \( D \subseteq \mathbb{R}^d \). To simplify notation we will let \( d = 2 \). Extending the stick-breaking prior to depend on location, the prior for \( F_s(Z) \) is the potentially infinite mixture

\[
F_s(Z) = \sum_{i=1}^{M} p_i(s) \delta(X(\phi_i)) = \sum_{i=1}^{M} K_i(s) V_i \prod_{j<i} [1 - K_j(s) V_j] \delta(X(\phi_i)),
\]

where \( p_i(s) = V_i(s) \prod_{j=1}^{i-1} (1 - V_j(s)) \), and \( V_i(s) = K_i(s) V_i \). The weight function, \( K_i \), is a spatial kernel with compact support, centered at \( \phi_i \), with bandwidth parameter \( \epsilon_i \). The bandwidth parameter is modeled as a spatial function, by making \( \epsilon_i \) a function of \( \phi_i \). By modeling \( \epsilon_i \) as a spatially varying parameter, we allow the spatial dependency structure to be different in different regions. \( V_i \sim \text{Beta}(1, \nu) \), and the spatial process \( X(\phi_i) \) has a zero mean-Gaussian process prior with covariance \( \Sigma \).

If the knots are uniformly distributed across space, the bandwidth parameters are not space-dependent functions, and the prior distributions for the knots, bandwidth parameters and \( V_i \)'s are independent. The corresponding covariance (integrating out \( V_i, \phi_i, \epsilon_i \)) is then stationary, assuming the kernel is not spatially dependent. Allowing the bandwidth to be space-dependent will make it nonstationary. The conditional covariance will be in general nonstationary, even when the kernels are not space-dependent.

To extend this to a multivariate setting, we explain the cross spatio-temporal dependency between \( p \) stochastic processes, \( Z_1(s), \ldots, Z_p(s) \), by having a representation for each \( Z_k(s) \),
\[ F_k(Z_k) = \sum_{i=1}^{M} K_{i,k}(s)V_{i,k} \prod_{j<i}[1 - K_{j,k}(s)V_{j,k}]\delta(X_k(\phi_i)), \]

where the knots are shared across the \( p \) space-time processes and the process \((X_1(\phi_i), \ldots, X_p(\phi_i))\) has a multivariate normal prior with covariance \( \Sigma^{(i)} \) that depends on space via the knot locations.

By allowing \( \Sigma^{(i)} \) to depend on \( i \), and therefore change with location, we obtain a cross covariance between the \( Z_k \) processes that varies with space (nonstationarity), and it is in general nonseparable.

We allow not only the magnitude of the cross-dependency structure to vary across space but also its sign, thus, it could be negative in some areas and positive in others. Most of the multivariate models, in particular separable models for the covariance, would not allow this effect.

**Paleoclimatology**: Scientific studies of climate systems and, in particular, the dynamics of the climate systems require knowledge of how climate variables have evolved over space and time. While climate (computer) models provide valuable information about aspects of both historic and future climate, there still is a need to study directly measured climate variables, preferably via instrumental records. Not only can such records be used to validate climate models, they also can be used to directly infer features and dependencies in climate systems. Instrumental records are limited in terms of both spatial coverage and temporal extent. As a result, there is a need to supplement directly measured climate variables with proxies of historical climate.

Mosley-Thompson, Thompson, Calder and Craigmile have studied the relationship between ice core-derived proxies of precipitation and the North Atlantic Oscillation index (Mosley-Thompson et al., 2005; Calder et al., 2008), characterizing the spatially-varying relationships between ice core-derived proxies and particular drivers of climate. Our current work is addressing the evolution of climate variables over different temporal and spatial scales simultaneously. Such multiscale approaches are commonplace in certain areas of statistical climatology (Madden, 1986; Wikle et al., 1999), but have seen more limited use in paleoclimate applications (e.g., Rutherford et al. (2005) and Mann et al. (2007), consider proxy records over just two different temporal scales). Wavelet-based approaches have proven useful in the multiscale analyses of climate and other applications (e.g., Whitcher and Craigmile (2004); Whitcher et al. (2005)) and will be applied in the ice core paleoclimatology setting.

There is also an interest in developing methodology for multivariate proxy series. Such multivariate series arise by making multiple measurements at each “date” in the proxy (for example, measuring oxygen isotope ratios and dust concentrations in ice core-derived proxies). While it is commonplace to “stack” (sum/average) different measurements taken from the same proxy (e.g., Mosley-Thompson et al. (2005)), it is less common to model and explain the relationships between the different measurements themselves, as well as to simultaneously relate these measurements with climate variables. Different measurements derived from the same ice core are known to capture different climate features (for example, dust is thought to capture climate extremes, Lambert et al. (2008)).

**b1.3. Regional climate models**

**Comparing regional climate models to data**: Regional climate models (Rummukainen, 2010)
operate on spatial scales that make them applicable to many impact studies. In order to use them in that context, the first task is to operate a regional model forced by observations (usually in the form of reanalysis). The most common approach (e.g. Kjellström et al. (2005) is to compare the model output to a gridded data product (usually a meteorological reanalysis). If the gridded data has been calculated with a standard error, the simplest comparison is to look at the difference between the fields, standardized by the standard error. However, the regional model (at least when run using a global climate model for boundary conditions) does not produce output that is comparable to weather data. Rather, it produces a distribution of data that is comparable to the distribution of weather.

The shift function (Doksum, 1974) is often useful for visualization of differences between observations and model output. When looking over a large region and doing individual tests (such as Kolmogorov-Smirnov tests for goodness of fit) for each grid square, a spatial plot of p-values is quite common. This plot can be accompanied by a histogram of p-values to look for relatively small deviations from the uniform null distribution, which may not be very obvious in the spatial pattern. Of interest here is particularly to investigate the performance of the multiple tests in the presence of nonstationarity (as is expected from climate change). We will study level and power performance in the presence of nonstationarity in mean and variance. Another approach is to set down a spatial model where the prior expectation is that nearby p-values (or test statistics) are similar.

Guttorp and Xu (2011) compared minimum temperatures at the Stockholm observatory to a regional climate model (RCA3) from the Swedish Meteorological and Hydrological Institute and found that the regional model was too warm, both in terms of daily minimum and daily mean temperatures. This discrepancy is likely because each grid square is an average over open air, forest and water land types, whereas the data are collected in open air. In order to do a better comparison, we will use data from the Swedish synoptic network and use this to predict grid square values, which we will then compare to the open air part of RCA3, forced with the ERA reanalysis (ERA40 and ERA-INTERIM). We will also use the statistical downscaling method by Berrocal et al. (2010) to do a direct comparison of downscaled temperature to station measurements.

A current Norwegian study (Orskaug et al., 2010) of regional modeling of precipitation, based on output from the HARMAD model and gridded data from the Norwegian Meteorological Office, found that the model (when forced by ERA40, which has known difficulties with orographic precipitation, Nikulin et al. (2010)) could reproduce the lower quartile of the precipitation distribution, but was underestimating high quantiles. We will implement the method developed in b1.4 below to estimate a spatially dependent shift function calibration of the downscaled ERA40 that will make the values closer to the observed values, as well as methods to estimate precipitation occurrence and precipitation amounts (given occurrence) on grid squares from station data.

**Ensemble methods**: Atmosphere-ocean general circulation models (AOGCMs, or simply GCMs) couple an atmospheric model with an ocean model and have become an invaluable component in the study of climate and climate change. However, projections of future climate are inherently uncertain, resulting from the natural variability of the climate system, a lack of knowledge about trajectories of future emissions of greenhouse gases and aerosols, and the response of
the global climate system to any given set of future emissions. There are additional uncertainties associated with climate models, including the parametric uncertainty resulting from approximations to processes that exist below the spatial scale resolved by the grid of the climate model, and such that different parameter settings produce realistic simulations of past and current climate but significantly different future simulations under the same scenario, and the structural uncertainty, which is generated by the fundamental choices of each modeling approach (what is or is not included, what kind of grid is chosen, what kind of representation is chosen for particular processes), such that say, the NCAR model is a different model from the UK MetOffice model. As more emphasis is put on climate projections on regional and even local scales, downscaling (via statistical approaches or dynamical approaches such as regional climate models) introduces another source of uncertainty resulting from the appropriate spatial resolution. Of course, these various issues are interrelated and understanding how they impact the output of climate simulations and ultimately the information that is derived from them for decision making is a difficult task.

Utilizing collections or ensembles of climate model output generated by varying initial conditions, forcings, or physical parameterizations or from entirely different models or from some combination of all of these have become the typical mode for exploring uncertainty in climate models. Quantifying uncertainty based on these ensembles has been the subject of much recent research and there has been some success (Tebaldi and Knutti, 2007). However, there is still much work to be done (Knutti et al., 2010).

There continues to be a need for the development of philosophical and statistical frameworks for combining information from the climate model output making up an ensemble. The essential Bayesian hierarchical approach put forward by Tebaldi et al. (2005) has provided the foundation for a great deal of research (e.g., Stainforth et al. (2007); Furrer et al. (2007); Smith et al. (2009); Tomassini et al. (2007)). Alternative approaches are also being developed (e.g., Rougier et al. (2010)) and the issue continues to generate discussion at the intersection of climate science and statistics.

b1.4. Modelling air pollution

Due to the strong dependence on weather conditions in their formation, ozone and other pollutants levels may be sensitive to climate change (Seinfeld and Pandis, 2006). In this project we introduce models to characterize this complex relationship between pollution and weather. There is also great interest in studying the potential effect of climate change on pollution levels, and how this change may affect public health (Bernard et al. (2001); Haines and Patz (2004); Knowlton et al. (2004); Bell et al. (2007)). The modeled relationship between pollution and weather will allow us to quantify potential changes in levels of ozone and other pollutants under climatic changing conditions. This type of work is needed to address the impact of climate change on emission control strategies designed to reduce air pollution, and to quantify the public health effects of climate change.

We implement our approach in a hierarchical Bayesian framework that consists of two main stages. In the first stage, we introduce a statistical model for characterization of the weather data (or climate projections). In the second stage, we model concentrations of ozone as a function of
weather. For computation, we use a directional Bayesian approach (obtaining estimates at each stage: interim posteriors in one stage become priors in the next stage). In the first stage of our hierarchical framework, using data obtained from the National Climatic Data Center, we will obtain the posterior predictive distribution of the weather variables at the locations of interest for Stage 2 using a Bayesian kriging approach (Handcock and Stein (1993); Gilks et al. (1995)).

Let \( y_i \) be the observed eight-hour maximum ozone for day \( t_i \) at location \( s_i \). Our interest lies in estimating the conditional density of \( y_i \) as a function of \( s_i \) and covariates \( X_i = (X_{i1}, ..., X_{ip})' \).

Given our interest in extreme events, we model \( y_i \)'s conditional density via its quantile function \( q(\tau|X_i, s_i) \), which is defined so that \( P\{y_i < q(\tau|X_i, s_i)\} = \tau \in [0, 1] \). We model \( q(\tau|X_i, s_i) \) as

\[
q(\tau|X_i, s_i) = X_i' \beta(\tau, s_i)
\]

where \( \beta(\tau, s_i) = (\beta_1(\tau, s_i), ..., \beta_p(\tau, s_i))' \) are the spatially-varying coefficients for the \( \tau^{th} \) quantile level. Jointly modeling all quantile levels, rather than simply the mean, as a function of the covariates provides a flexible model for the conditional distribution as a function of the covariates. Also, since the covariates effects vary spatially, the shape of the conditional response density also varies spatially.

The quantile curves are modeled using a finite basis expansion \( \beta_j(\tau, s_i) = \sum_{m=1}^M B_m(\tau) \alpha_{jm}(s_i) \), where \( B_m(\tau) \) is a known basis function of \( \tau \) and \( \alpha_{jm}(s_i) \) are unknown coefficients. Collecting terms with common basis functions gives \( X_i \beta(\tau, s_i) = \sum_{m=1}^M B_m(\tau) \theta_m(X_i, s_i) \), where \( \theta_m(X_i, s_i) = \sum_{j=1}^p X_{ij} \alpha_{jm}(s_i) \). The processes \( \beta_j(\tau, s_i) \) must be constructed so that \( q(\tau|X_i, s_i) \) is non-decreasing in \( \tau \) for all \( X_i \) to give a proper quantile function. We use Bernstein basis polynomials \( B_m(\tau) = \binom{M}{m} \tau^m (1-\tau)^{M-m} \). An attractive feature of these basis functions is that if \( \theta_m(X_i, s_i) \geq \theta_{m-1}(X_i, s_i) \) for all \( m > 1 \), then \( q(\tau|X_i, s_i) = X_i' \beta(\tau, s_i) \) is an increasing function of \( \tau \). This reduces the complicated monotonicity constraint to a sequence of simple constraints. These constraints are sufficient, but not necessary, to ensure an increasing function.

To specify a prior for the \( \alpha_{jm}(s_i) \) to ensure monotonicity, we assume that \( X_{i1} = 1 \) for the intercept and the remaining covariates are scaled so that \( X_{ij} \in [0, 1] \). Since the constraints are written in terms of the difference between adjacent terms, we reparameterize to \( \delta_{j1}(s_i) = \alpha_{j1}(s_i) \) and \( \delta_{jm}(s_i) = \alpha_{jm}(s_i) - \alpha_{j(m-1)}(s_i) \) for \( j > 1 \). We ensure the quantile constraint by introducing latent unconstrained variable \( \delta^*_m(s_i) \) and taking

\[
\delta_{jm}(s_i) = \begin{cases} 
\delta^*_jm(s_i), & \delta^*_1m(s_i) + \sum_{j=2}^p I(\delta^*_jm(s_i) < 0)\delta^*_jm(s_i) \geq 0 \\
0, & \text{otherwise}
\end{cases}
\]

To encourage the conditional density functions to vary smoothly across space we will initially model the \( \delta^*_jm(s) \) as independent (over \( j \) and \( m \)) Gaussian spatial processes, although we will investigate the consequences of this independence assumption.

We analyze maximum daily 8-hour average ozone concentrations measured at 631 locations (Figure 1c) in the Eastern US during the summers (June-August) of 1997-2005 (470,239 total observations). We apply the spatial quantile approach presented here to calibrate the climate models. Figure 1a illustrates the nonlinear relationship between temperature and ozone. We also use output
Figure 1: Panels (a) plots ozone (pooled over space and day) by daily average temperature. Panel (b) is the calibration plot of the temperature quantiles for Georgia. Panel (c) gives the probability that the three-year (2041-2043) average of the fourth-highest daily maximum 8-hour average ozone concentrations exceeds 75 ppb for the future climate scenario.

from the GFDL deterministic atmospheric computer model (AM2.1; The GFDL Global Atmospheric Model Development Team (2004)). To take into account the change of spatial support, we use block kriging to obtain the output of the deterministic model at the point locations at which we have ozone values. Figure 1b plots the sample quantile function for the observed and modeled temperature and wind speed for all days in 1997-1999 in Georgia. The distribution of daily average temperature agrees quite well below the median, whereas the modeled temperature has a heavier right tail than the observed temperature. We use the approach given by (1) and (2) to calibrate and transform the model output to better reflect the weather data. Figure 1b shows the calibrated temperatures. Model outputs for 2041-2045 with large temperatures are reduced to resolve the discrepancy between observed and modeled 1997-1999 data. The transformation could be allowed to change across space to account for spatial variation in the calibration.

Using our model for the relationship between meteorology and ozone, we can easily simulate ozone under a variety of climate scenarios. As an illustration, we use the calibrated GFDL projected temperature, wind speed, and cloud cover for 2041-2045. Figure 1c depicts the probability of the three-year (2041-2043) average of the fourth-highest daily maximum 8-hour average ozone concentration being greater than 0.075 ppm, the current EPA standard for ozone. Exceedence probabilities are greatest in the Northeast and Midwest.

b1.5 Ocean temperature profiles

A problem of continuing interest in oceanography is the relationship between temperature and depth in the ocean (Rappold et al., 2007). This relationship is complicated by the phenomenon of mixed layer depth (Rappold et al., 2007) which introduces a severe nonlinearity. Moreover, the mixed layer depth occurs at different depths for different locations at different times. Statistical contributions to investigate this phenomenon have, to date, been very limited (Rappold et al., 2007).
One way to approach the problem is through space-time functional data analysis, i.e., at each $(s, t)$ there is a function which connects depth, $x$, to temperature $Y$, up to error, i.e., $Y(x; s, t) = g(x; s, t) + \epsilon(x; s, t)$. It is expected that the functional curves will be similar for locations close to each other and for times close to each other.

Suppressing time for the moment, we envision two sampling settings. In the first, we collect (depth, temperature) pairs at a collection of depths for a set of locations. In the second, we sample the levels; i.e., at a specified depth, we observe temperatures for a set of locations (which may vary with depth). In either setting, we can anticipate some analysis of variance (ANOVA) structure. That is, the locations are grouped into “populations” which are defined as different geographic regions within the ocean. Typically, the regions are too far apart to allow a single common spatial model. So, an ANOVA model seems appropriate and it is of interest to compare the temperature/depth relationships across the regions.

In the first setting the individual curves will only be sparsely observed so that we can not hope to learn about fine detail in the relationship; instead we model the curves as random mean square continuous realizations of a suitable process model. The second setting is even more problematic. There is still a conceptual individual level curve but the sampling design encourages us to introduce modeling at the depth level. Again, we utilize random process realizations. Work with functional ANOVA using process realizations is limited thus far (e.g., Kaufman and Sain (2010)). It is more customary to use basis representations such as splines, building on the seminal work of Ramsay and Silverman (2006).

We briefly describe our proposed Bayesian nonparametric approaches to handle both sampling settings to allow more flexible specifications than Gaussian process realizations. In fact, we work with Dirichlet process mixtures (see section b1.1) to explain the observations, envisioning that for individuals (i.e., spatial locations) within a population (i.e., within a region), individual level temperature vs. depth curves vary around a population mean temperature vs. depth curve. This can be done directly with the spatial Dirichlet process (Gelfand et al., 2005). However an alternative version is to introduce a hierarchical Dirichlet process (Teh et al., 2006). This allows each population distribution to be random, hence of course the associated population mean, but in fact allows for direct comparison of the distributions across populations as well as comparison of other population features. This specification accommodates the the first sampling scenario. To handle the second, we propose to add a nesting component to the hierarchical Dirichlet process to allow for additional local individual variability associated with the sampling at the specified depth levels. Population level distributions and their features can still be compared.

**b.2 Workshops and summer schools**

**b.2.1 Workshops**

We will organize annual workshops at either SAMSI or NCAR on statistical topics in atmospheric and oceanic sciences. The tentative schedule for the first three years is given here.

2011-12 Comparing numerical model output to data. Techniques for evaluation of numerical models require comparison of distributional properties of station data to those of the model outputs. The workshop will go in to detail of the interpretation of grid square values, and study ways
of downscaling models to station level as well as upscaling station data to model level. Special problems arising when comparing model output to remote sensing data will also be addressed.

2012-13 Multivariate nonstationary/non-Gaussian processes. This workshop will consider the conceptual, mathematical and computational challenges in multivariate statistical models that are nonstationary (in space and/or time) and/or non-Gaussian and their application to atmospheric and oceanic sciences. Specific applications include nonstationarity in time due to climate change and nonstationarity in pollution and weather due to local influences (e.g., urban areas, large point sources). We will also consider multivariate models for non-stationary cross-dependence structures, for instance in wind vectors or in speciated particulate matter.

2013-14 Extremes. This will be a follow-up to the very successful 2010 BIRS workshop on Extreme events in weather and climate (http://temple.birs.ca/~10w5016/). The BIRS workshop developed a research agenda for studying extreme climate and weather events and this meeting will look at what progress has been made in three years, update the research agenda, and look for new approaches to these problems. We will attempt to hold this workshop at BIRS as well.

b.2.2 Summer schools
The network will offer three kinds of summer schools. First, there will be summer schools for undergraduate students, aimed at interesting mathematical sciences majors in statistics and its applications in geophysics. Second, we will organize summer schools to bring graduate students and postdocs in both geophysics and statistics up to speed in crossdisciplinary work. The structure of these summer schools will be separate lectures in the morning for the two groups, and joint computer work, where groups are formed with participants from both areas, in the afternoons.

The third kind of summer school is aimed at geophysical scientists, and are called Sharp Statistical Tools for Geophysicists. The idea here is to present some of the methods available to the modern statisticians (such as space-time models, extreme value theory, multivariate statistics, and generalized linear models) that would not have been included in the program of study for a typical atmospheric and/or oceanic sciences department.

b.2.3 Conferences
As a closing event for this round of network funding, the network will volunteer to organize the 13th triennial International Meeting on Statistical Climatology, the foremost venue for interaction between statisticians and climate scientists (see http://cccma.seos.uvic.ca/ims/home.shtml for details about this series of conferences).

b.3 Exchange of researchers
The main purpose of the funding for this proposal is to allow intense and substantial collaboration within the network. We are budgeting for students and postdocs spending extended periods of times at any nodes of the network. In addition, there is funding for researchers at various nodes to visit other nodes. We will develop a simple web application form for this, and while funding is available, the decisions will be made between two node directors. If a proposed visit needs more funding than is already allocated to the node, the management committee will consider the request. In these cases, particular emphasis will be on nodes that do not have substantial research funding of their own.
b.4 Electronic journal

We will investigate the feasibility of starting an electronic journal Statistical Methods in Atmospheric and Oceanic Sciences. The publisher would likely be Berkeley Electronic Press.

b.5 Outreach activities

At the network web pages (see also section g) we will present nontechnical summaries of our work, links to appropriate statistical and geophysical sites, and present short video clips that can be used by media. Students from the University of Washington Department of Technical Communication will assist in this work as part of their own portfolio development. This will include web pages visualizing some of the data analyses and the uncertainties associated with climate events, audiovisual presentations aimed at non-technical audiences, and highlights of the research findings of the team.

In order to encourage interest in statistics among high school students, it is important to demonstrate the usefulness of statistics in dealing with important societal problems. We will develop a few lessons on statistical issues in environmental science, for example, on how to look at trends in extremes. It is anticipated that these lessons would be particularly useful for AP Statistics classes for the several weeks between the AP Statistics test and the end of the school year. We will first develop the classes together with a high school teacher in the Seattle school district. If the class is successful, we will provide online materials as well as talking points and videotaped lectures for public use. This can eventually be coordinated with the American Statistical Association Center for Statistics Education curricular web pages for K-12 statistics education. The curriculum unit will be presented at the United States Conference On Teaching Statistics (USCOTS; http://www.causeweb.org/uscots/) held biennially at the Ohio State University.

The University of Chicago Math Department has long run successful programs for Chicago Public School students (the Young Scholars Program, http://www.math.uchicago.edu/ysp) and for teachers (the Seminars for Elementary Specialists And Mathematics Educators, http://www.math.uchicago.edu/sesame). We have entered into discussions with Paul Sally, the director of these programs, as to ways we could contribute by, for example, giving lectures, providing content and, if a graduate student or postdoc were interested, teaching a class. The NCSU group will work on K-12 outreach hrough the Kenan Fellows for Curriculum and Leadership Development Program (http://www.ncsu.edu/kenanfellows) and the Science House (http://www.science-house.org/). The PI, who is currently a member of both programs, plans to engage public school teachers in the creation and development of curriculum and teaching resources that bring some of cutting-edge research in environmental spatial statistics into the hands of students.

c. Governance and Management

c1. Management group

The management group, consisting of the hub directors Fuentes, Guttorp and Stein, will direct the grant activities and communicate at least weekly using a Skype connection. Among their tasks is to coordinate the monthly steering committee meetings and to plan the annual meetings...
at NCAR or SAMSI. Another important aspect of the steering committee work is to maintain close
contact with the NCAR and NCDC groups to ensure that the physical aspects of the work are
appropriate and founded in solid science. A third task is to make changes in the network structure
in consultation with the steering committee.

c2. Steering committee

The steering committee will consist of the management group and the node directors and will
have monthly cybermeetings to coordinate activities. One of its most important tasks is planning
the biannual network wide videoconference meetings. The committee will also ensure that graduate
students and postdocs are deeply involved in the scientific as well as the methodological aspects
of the work, and help them organize joint cyberactivities. A third task is to advise the management
group regarding changes in the network structure, such as adding and removing nodes as needed
as the research evolves. Suggestions for such changes may come from the management group,
from node members, or from individuals or groups interested in joining the network. The network
website will encourage interested parties to make inquiries about joining the network. We will
also make direct contacts with individuals at schools with large populations of underrepresented
minorities to seek out new additions to the network.

c3. Hubs and nodes

The network has three hubs. The primary hub is at North Carolina State University (Montserrat
Fuentes, Statistics (director); Brian Reich, Statistics; Fred Semazzi and Lian Xie, Marine, Earth
and Atmospheric Sciences). One auxiliary hub is at the University of Washington (Peter Gut-
torp (director); Dennis Lettenmaier, Civil Engineering Eric Salath, Climate Policy; Paul Sampson,
Statistics; John M. Wallace, Atmospheric Sciences). The other auxiliary hub is at the University
of Chicago (Michael Stein (director); Ian Foster, Computer Science and Computation Institute and
Argonne National Laboratory; Liz Moyer, Geophysical Sciences).

Hubs will have nodes. Each node receiving funding from the network will designate a node
director, who is responsible for the reporting of activities to the management group (section f). One
such node, with particular standing in the network, is the National Center for Atmospheric Research
( NCAR; Richard Katz (director); Eric Gilleland, Research Applications Laboratory; Doug Nychka,
and Steve Sain, Institute for Mathematics Applied to Geosciences; Claudia Tebaldi, Institute for
Mathematics Applied to Geosciences and Climate Central), which will host visits from students,
postdocs and senior researchers (see attached letter of commitment). NCAR should be thought of
as a node connected to all other network nodes.

The NCSU hub has nodes at Duke (Alan Gelfand, Statistics (director); Wenhong Li, Earth
and Ocean Sciences; Prasad Kasibhatla, Environmental Chemistry; Gabi Hegerl, Earth and Ocean
Sciences (and Climate Systems Science, University of Edinburgh); M. Susan Lozier, Physical
Oceanography), SAMSI (Richard Smith (director)), UNC-Wilmington (Frederick Bingham, Physics
and Physical Oceanography (director)) and the National Climatic Data Center (Dongsoo Kim (di-
rector; Ross Vose).

The University of Washington has nodes at the Pacific Institute for Mathematical Sciences
(Charmaine Dean, Statistics, Simon Fraser University (director); Francis Zwiers, Pacific Climate
Impact Center) and at San Diego University (Barbara Bailey, Mathematics and Statistics (director); Richard Levine, Computational Statistics).

The University of Chicago has nodes at the Ohio State University (Catherine Calder, Statistics (director); Peter Craigmile, Statistics) and Purdue University (Hao Zhang, Statistics (director); Bo Li, Statistics).

c4. Hub meetings
The research group at each hub and its nodes will hold regular meetings to update each other on activities and products, and discuss ideas and joint work.

c5. Network meeting
Twice a year the group will have a videoconference to discuss current activities, establish cross-network connections and add new research problems for consideration. Once a year, the entire group will meet in person at NCAR or SAMSI. At this meeting, each of the nodes will present their work. The group will discuss new approaches and adapt the research program, if necessary, to new and promising directions. To the extent possible, we will coordinate these annual meetings with relevant NCAR or SAMSI activities such as workshops, summer schools, or conferences. These meetings (two virtual, one in person) are essential to maintaining the coherence of the network.

c6. Scientific advisory group
We anticipate setting up a small advisory group (with one or two geophysicists and one or two statisticians). This group will participate in the annual meeting and advise the research team as to promising directions, as well as communicating with the management team between annual meetings. Possible names from participating institutions include Chris Bretherton (UW), Mark Berliner (OSU), Jim Berger (Duke/UC), Noel Cressie (OSU) and Joe Tribbia (NCAR). Members of the advisory group will be chosen in consultation with the relevant NSF program officer and will likely include at least one individual from an institution outside the network. However, we believe that individuals from network institutions will be easier to engage actively in the advisory process.

c7. Connection to other networks
Guttorp is technical director of the Nordic Network on Statistical Approaches to Regional Climate Models for Adaptation (http://www.nrcse.washington.edu/NordicNetwork/). There are already several interactions between this network and the University of Washington, with research collaborations and visits going both directions. We expect substantial and increasing interaction between the two networks.

Guttorp and Dean are among four co-PIs for the Pacific Institute for Mathematical Sciences (PIMS) Collaborative Research Group on Environmetrics (http://www.pims.math.ca/scientific/collaborative-research-groups/environmetrics-2007-2010), which has specialized in climate research, particularly in the area of forest fires. The continued interaction will be enhanced by the Canadian nodes of the University of Washington hub. A supporting letter from PIMS is attached.

The proposed network will therefore have very natural ties with these groups. In particular, we expect to continue the exchange of researchers and graduate students that is already ongoing.
Letters of commitment from both groups are attached to the proposal. In addition, the statistical climatology groups in Exeter and Southampton have expressed interest in participating in the network.

Michael Stein is a co-PI on the recently NSF-funded Center for Robust Decision Making on Climate and Energy Policy, which studies the integration of climate and economic models to assess the consequences of various economic policies related to energy and climate. Two of the lead participants in this project, Ian Foster and Liz Moyer, are members of the Chicago hub and look forward to supporting cooperation between these two efforts.

d. Broadening Participation Plan

The recruitment of promising undergraduates to statistics in the atmospheric and oceanic sciences will be multi-pronged. Individual undergraduates will be approached at each node and invited to participate in the research group. Potential undergraduate activities are described in section e1 below. In addition, we will use summer schools (see section b2.2) to focus recruitment towards especially promising juniors in underrepresented groups.

The PI will give some talks and short courses in Spanish in Hispanic countries to broaden the participation of underrepresented geographic and ethnic groups. We will also try to get underrepresented groups involved in this type of work using our contacts in traditionally black and hispanic universities. Postdoc and graduate student recruiting will emphasize underrepresented groups as is normally done in university hiring. The team will be also active in K-12 educational efforts (see section b5). The PI is a senior leader of the NSF ADVANCE program at NCSU (http://www.ncsu.edu/odi/advance/) promoting diversity. The skills learned through that program will be applied to the management of this project.

e. Human Resources Development Plan

e1. Undergraduate students

The group will involve undergraduate students in the research project. Stein, Fuentes and Guttorp all have experience with undergraduate research supervision. There are a variety of possible areas of undergraduate work, such as data download and management, data visualization, public communication, web page design, code documentation and verifying the reproducibility of computations reported on in publications. Some senior undergraduates may be able to visit NCAR, where the statistics group has experience with advising undergraduates. Part of the recruitment of undergraduates will be the summer schools outlined in section b2.2.

e2. Selection and mentoring of graduate students and postdocs

Graduate students with interest in atmospheric or oceanic applications will be selected by node or hub directors in conjunction with their graduate program coordinators. Postdocs will be selected by the appropriate node or hub director after national advertising. The graduate research assistants and postdocs will meet weekly with their respective node directors, who will supervise the postdocs’ and students’ research progress. The postdocs and graduate students will also participate in the biannual group cybermeetings and in the annual network meetings, and will be encouraged to make presentations at scientific meetings. For further details, see the mentoring plan.
e3. Jointly taught graduate courses

We anticipate teaching joint graduate courses on space-time statistics, statistical methods in atmospheric and oceanic sciences and in wavelet analysis. These courses will be available to graduate students at any node, and will also be made available to the students in the PIMS consortium at UBC Vancouver, UBC Okanagan and Simon Fraser University. Further courses may be developed as the network expands.

e4. Videomeetings

We will hold biannual video meetings of the entire group with web document access (and joint real-time editing capability) using collaborative software such as WebEx. Not only will the participants see and hear each other, it is also possible to share (in real time) and jointly edit a variety of types of documents, presentations, graphics, etc. The facilities for these meetings are already available at each of the hubs and at several nodes. This type of communication is vastly superior to conference calls and email communications (although such communication forms can complement the video meetings). In order to collaborate on writing, web application sharing tools will be used, and we will also use Skype for simple small audio-visual group communications. Many of these tools are freely available, and for those that are not, the management group will coordinate purchases by nodes.

f. Evaluation Plan

The steering committee is charged with the ongoing evaluation of the network success. Measures of success include: undergraduate projects completed; graduate students recruited and graduated; postdocs recruited; papers submitted and published resulting from network activities; presentations at scientific meetings; internode visits and resulting collaborations; software products released to the public and their subsequent use; and visitors to the research and outreach parts of the web pages. In tracking publications and presentations at scientific meetings, we will monitor the disciplines of the journals/meetings (e.g., statistical versus geophysical) and whether the author teams are multidisciplinary. All graduating students and departing postdocs will be interviewed about the quality of their experience and asked for suggestions on how it might have been improved.

In order to enable the steering committee to keep track of network activities, each hub director will report annually all hub activities, publications and presentations. Anyone who makes an extended visit (more than a week, say) to another institution will be required to submit a brief statement prior to the visit on the goals of the visit and a followup statement after the visit on how well these goals were met. A web form will be developed to facilitate collecting this information. All attendees of workshops, summer schools or conferences will be asked to submit comments on the meeting and suggestions for improving future meetings. The steering committee will also keep track of the international exchanges that have been enabled by the network and discuss and plan new opportunities for exchanges in both directions. In addition, postdoc mentors and graduate student advisors are expected to follow the career paths of their students and report to the hub director every two years during the life of the network. The steering committee will keep track of the international exchanges that have been enabled by the network and discuss and plan new opportunities
for exchanges in both directions.

The management team is charged with producing the annual report for the NSF, in accordance with the details specified in the request for proposals. All students, postdocs and others who participated materially in the network (e.g., beyond attending a single workshop or meeting) will be required to provide a written summary of their work in the past year and their plans for the coming year. These summaries will be an important input for the annual reports, with special attention paid to how well the goals for each year were subsequently met. Another key task of the management team is to seek out opportunities to add new nodes to the network. Nodes that would increase the intellectual and/or ethnic diversity of the network will be a particular focus of this effort. In addition, the management team will regularly evaluate the performance of all existing nodes. Any problems that are identified should be addressed forthrightly and if they cannot be adequately resolved, nonperforming nodes should be dropped from the network.

g. Dissemination Plan

We will disseminate our results to the scientific community through the usual channels: publications in peer-reviewed statistical and geophysical journals, and presentations at national and international conferences. We plan to conduct annual technical workshops (see section e5 for details). Presentations will be made available on the web. We are currently planning presentations or sessions at the annual Joint Statistical Meetings and the annual TIES (The International Environmetrics Society) meetings, at the Statistical Climatology Meeting 2013 in Korea, and at the International Statistical Institute Meetings in Ireland in 2011 and in Hong Kong in 2013. We will strongly encourage all personnel to attend meetings and present papers and posters in relevant subject matter disciplines such as meetings of the American Geophysical Union.

Because there is little standard software available that properly accounts for spatial correlation in modeling atmospheric and oceanic processes, we propose a software development initiative. We will develop free user-friendly software compatible with the free software package R, to implement the new spatial models.

We will use the Northwest Research Center for Statistics and the Environment (NRCSE) web service. The network will have its own home page in the NRCSE web pages (see our previous project pages at www.nrcse.washington.edu/ClimateStats). Among other information, we will post technical reports, appropriate presentations by group members (if possible in streaming format, otherwise at least the slides) and well-documented source code for the software developed. An open data policy will be part of the outreach (see part h).

h. Data Policy

To the extent possible all data sets used in network funded research will be made publicly available on the network web pages or its ftp server. Every effort will be made to supply appropriate metadata. Likewise, open source code will be made available to enable other researchers to reproduce network funded research. Undergraduates will have the task of verifying that data and code posted are in fact sufficiently well documented to make the research truly reproducible and the data and software products effectively usable by relevant audiences.
References


23


Rutherford, S., M. Mann, T. Osborn, R. Bradley, K. Briffa, M. Hughes, and P. Jones (2005). Proxy-
based northern hemisphere surface temperature reconstructions: Sensitivity to method, predictor
network, target season, and target domain. Journal of Climate 18, 2308–2329.

Sampson, P. D. and P. Guttrop (1992). Nonparametric estimation of nonstationary spatial covari-


Shapiro, L. (1983). The asymmetric boundary layer flow under a translating hurricane. Journal of


and decision-support relevance in climate predictions. Philosophical TransactionsRoyal Society
of London Series A 365, 2145–2162.


The GFDL Global Atmospheric Model Development Team (2004). The new GFDL global atmo-
sphere and land model AM2-LM2: Evaluation with prescribed SST simulations. Journal of
Climate 17, 4641–4673.


