A Statistics Program at NCAR, Project Summary

A grand scientific challenge for this century is to understand the complex interrelationships among the atmosphere, ocean, land processes, biosphere and human activities that define the Earth System. Coupled with this effort is the need to predict changes to our environment and to translate such changes into more immediate economic and societal impacts. There is a corresponding challenge for statistical science to tackle the large and complex data sets that are now the norm in many areas of science and engineering and to leverage the rich structure and prior knowledge afforded by traditional numerical models.

The Geophysical Statistics Project (GSP) at the National Center for Atmospheric Research (NCAR) provides a unique opportunity for research that addresses the challenges to the geosciences and statistics. The plan of research and a three-part visitor program will enable GSP to bridge the geosciences and the statistical research communities. A *Postdoctoral Fellows* (PF) program, that has been successful though past support, will continue to train recent statistics Ph Ds in interdisciplinary collaboration and statistical research for application in the geosciences. A *Thesis-In-Residence* program will support the visit of graduate students to NCAR to incorporate substantial geophysical applications as part of their dissertation research. A *Statistics Chautauqua* will be a month-long program primarily for young statistics faculty and will involve a series of lectures and interactions with NCAR scientists on collaborative research in the geosciences and statistics. Each of these programs will help statistical researchers target geophysical applications. Moreover, GSP will also use its vantage point at NCAR and the visitor programs to spur new statistical research that is motivated by geophysical questions and data.

Intellectual Merit

The statistical research includes contributions to the analysis of spatial and spatio-temporal data, the design and analysis of computer experiments, filtering, the statistics of extremes, data mining, and statistical computing. Many of these methods will be adapted to large data sets and publicly available software will be developed to implement new methods for spatial spatial data and extremes. The intellectual contributions to applications in the geosciences will be equally broad and include the introduction of stochastic modeling for sub-grid scale processes, estimating the climatology for extreme events, the construction of spatially coherent weather generators and statistical improvements to methods of data assimilation. An important element throughout this research is being able to quantify the uncertainty in a derived method or result.

Broader Impacts

This collaborative research will participate in the broader science initiatives at NCAR and in so doing align individual projects with the larger challenges facing the geosciences. Through the training of young researchers, a vigorous visitor program, regular sponsored workshops and alliances with other research centers and academic departments, this work will reach a wide community and serve as model for collaboration among statisticians and researchers in the geosciences. A broader impact will be the creation of an infrastructure for the statistics research community to address applications in the geosciences. The long term results will be the solution of some of the difficult problems in the geosciences through the transfer and development of new statistical methodology.

Project Description

1 Results From Prior NSF support

A Statistics Program at the National Center for Atmospheric Research (7/1999-6/2004) DMS-9815344 has provided significant support for the Geophysical Statistics Project (GSP), a statistical research and training group embedded within the National Center for Atmospheric Research (NCAR — NCAR has a staff of more than a 1000 members including several hundred Ph.D. scientists and receives more than half of its funding from the NSF Division of Atmospheric Sciences.). GSP has been a successful model for making contributions to atmospheric and oceanographic sciences through the development and application of new statistical methodology. This has been achieved by supporting a critical mass of statistics visitors at NCAR, training young researchers, and maintaining close collaboration with NCAR geophysical scientists.

The long-term impact on statistics has been substantial, not only contributing a large body of published research and software, but also through an extended network of university researchers who have participated in GSP as visitors and postdoctoral fellows (PF). GSP's impact on geophysical researchers has been through *statistics by example*. Successful collaboration with NCAR scientists has provided case studies where new statistical methods have tackled problems and questions that otherwise would not have been attempted. Scientific applications identified through GSP have also suggested new statistical models and theory.

1.1 Scientific Accomplishments

Because of the emphasis on collaborative research, GSP scientific results have appeared in leading statistical, geophysical, and environmental statistical publications. A noteworthy feature of this research is that the majority of publications are co-authored by both geophysical and statistical scientists. In order to succinctly report this wide range of research; project publications, submitted manuscripts and technical reports have been organized into some broad categories. This outline is complemented by more detailed highlights of some key scientific contributions.

Statistical methodology and theory
Time series analysis [8],[7],[22],[26],[27],[28],[43],[48],[57],[55],[59],[81],[80],[88],[102],[109],[110],[116],[114];
Spatial statistics [24],[31],[100],[63],[69],[71],[76],[93],[45];
Space/time processes [16],[91],[96],[94];
Nonlinear time series and dynamical systems [6],[83];
Nonparametric regression [18],[42],[22],[44],[72],[73],[90],[104],[106],[107],[108],[115];
Bayesian methodology [29],[30];
Analysis of functional data [52],[53].

Geophysical Applications
Analysis of the climate record [46],[65],[77],[78],[105],[111];
Time series analysis [1],[17],[32],[38],[39],[47],[56],[70],[75],[74],[86],[99],[89],[103];
Spatial fields [20],[23],[49],[58];
Validation and statistical analogs of numerical models [2],[4],[5],[3],[19],[62],[98];
Extreme events [37],[35],[41],[50],[60],[61];
Climate change [15],[54],[14],[51],[68],[113],[112];
Forecasting and data assimilation [9],[10],[12],[21],[79],[84],[85],[101],[82].

Overviews and surveys [11],[13],[25],[33],[34],[36],[64],[66],[67],[95],[97],[92].

1.2 Research Highlights

The following sampling of GSP research accomplishments helps emphasize the substantial collaboration among the postdoctoral fellows/visiting scientists and resident NCAR scientists. These highlights are meant to be illustrative; they only constitute a subset of GSP research activities.

Statistics for large spatial datasets A challenge to spatial statistics is to scale methods from the moderate-size datasets considered in the statistics literature to the large and substantive records assembled to study the atmosphere and oceans. As a testbed problem, GSP has constructed spatially and temporally complete monthly meteorological records for the coterminous U.S. based on historical meteorological data ([31]). The work implemented local covariance models that blended sample correlations with a global covariance model, successfully handled a large number of spatial locations (8000-12000) over many time points (1200 months) and provided cross-validated measures of uncertainty of the in-filled estimates. Motivated by large geophysical datasets, two related projects are to adapt multi-resolution models for nonstationary ([69]) and non-Gaussian fields ([57]).

Statistical analogs of atmospheric numerical models The size and complexity of output produced by state-of-the-art geophysical models often require statistical descriptions to understand and summarize the overall behavior and specific aspects of the simulated system. Cloud-resolving models (CRM) are fine-scale convective models for the atmosphere that can create and evolve individual clouds. Extreme value distributions were used to model maximum vertical velocity ([61]) in CRM output as a stochastic representation of processes that are important for large-scale atmospheric general circulation models (GCM). One result was a quantification of the impact of wind exceedances on the total mass flux in the CRM domain. Also based on model output, multi-resolution methods have been designed to extract features representing the organization of convection (e.g. squall lines) ([62]), providing an objective classification of the types of convection in a model simulation. Current GCMs rely on heuristic schemes to overlap cloud layers and this may introduce biases in the amount of absorbed and reflected radiation in the atmosphere. As an alternative, a statistically based cloud overlap model was estimated from CRM experiments and was found to give better simulation of the radiative fluxes ([98]).

At a larger spatial scale, a combination of nonlinear time series models and clustering algorithms have been used to summarize the dynamical properties of the large-scale atmospheric circulation from a long integration of the NCAR Community Climate Model (CCM0) ([83]). The statistical model quantified the system's nonlinear behavior and identified regimes in phase-space whose dwell times were found to be consistent with the timescales of atmospheric motions.

Data assimilation and forecasting Data assimilation refers to combining information from a geophysical model with observations to give an improved estimate of the state. This process lies at the heart of operational numerical weather prediction and has formal relationships to nonlinear filtering. Current work in the atmospheric sciences community is on a (Monte Carlo based) Ensemble Kalman Filter (EnKF). GSP statistical contributions include extensions to non-Gaussian distributions for the state ([9],[10],[99]). By applying multivariate statistical theory, it was possible to analyze and better characterize the dependence of EnKF performance on ensemble size ([101]).

Aircraft are substantially affected by the presence of clear air turbulence. This localized atmospheric phenomenon is very difficult to predict using conventional weather forecast models. A successful approach trained a flexible discriminant model on pilot reports of turbulence using turbulence indices derived from a numerical weather prediction model as predictors ([84],[85]). These statistical forecasts outperform other operational approaches currently tested at NCAR.

Analysis of long-term climate series Long-term variation in solar irradiance can be confounded with possible climate change due to human activities, consequently, it is important to investigate the relationship between solar flux and surface temperatures. A multi-resolution analysis was applied to several proxy measures of solar flux and to proxies for temperature to determine the relationship among these climate time series ([70]). One main result indicates evidence for the forcing of temperature by solar flux on a timescale consistent with the Gleissberg cycle (period of $\approx$ 85 years). Another influence on temperature records are large, explosive volcanic eruptions that inject substantial amounts of sulfur dioxide into the lower stratosphere. A multiprocess state-space model was estimated and was able to identify and quantify the impulse-like volcanic signal in proxy records, thus distinguishing those temperature proxy series that are more sensitive to aerosol loading ([1],[56]).

1.3 Postdoctoral and Graduate Training

The highest priority of the project has been to recruit and train new recipients of Ph.D.'s in statistics and probability. GSP has been able to attract high quality individuals from nationally-ranked programs in statistics and has placed the majority of GSP fellows in university positions or in other research environments (the second line in the list below indicates the fellow's subsequent position). One milestone that indicates the growing maturity of GSP is the presence of Tebaldi as a Project Scientist and Gilleland as an Associate Scientist on on the permanent NCAR scientific staff.

Former postdoctoral fellows and placement

Craig Johns, Statistics, U. of California - Davis, 1999; GSP 8/1999- 7/2001
Assistant Professor, Dept. of Mathematics, U. of Colorado - Denver.

Philippe Naveau, Statistics, Colorado State U., 1998; GSP, 7/1998- 12/2000
Assistant Professor, Dept. of Applied Mathematics, U. of Colorado - Boulder

Claudia Tebaldi, Statistics, Duke U., 1997; GSP, 12/1997-6/2000
Project Scientist, Environmental and Societal Impacts Group, NCAR.

Enrica Bellone, Statistics, U. Washington, 2000; GSP 9/2000- 9/2002
Research Associate, Dept. of Civil and Environmental Engineering, Imperial College, London.

Brandon Whitcher Statistics, U. Washington, 1998; GSP 9/2000- 9/2002
Senior Researcher, Research Statistics Unit, GlaxoSmithKline, London.

Hee-Seok Oh: Statistics, Texas A&M, 1999; GSP 9/2000- 8/2002
Assistant Professor, U. of Alberta.

Thomas Bengtsson: Dept. of Mathematics, U. of Missouri, 1999; GSP 8/2000-8/2003
Jerzy Neyman (visiting) Assistant Professor, U. of California at Berkeley.

Current postdoctoral fellows and research interests

Sarah Streett, Statistics, Colorado State U.; 9/2000- *Non-Gaussian time series and spatial processes.*

Reinhard Furrer, Statistics, Swiss Federal Institute of Technology, 9/2002- *Spatial statistics, filtering.*

Jarrett Barber, Statistics, North Carolina State U., 9/2002- *Misaligned data, nonstationary fields.*
(75% appointment Duke U.)

Uli Schneider, Applied Mathematics, CU - Boulder, 7/2003- *Perfect sampling, extremes.*

Tomoko Matsuo, UCAR/State U. of New York - Stonybrook 7/2003- *Data assimilation.*

Dorin Drignei, Statistics, Iowa State U., 1/04- *Analysis of computer experiments.*

GSP-supported graduate students and research interests

Eric Gilleland, Colorado State U., 9/1999 - 12/2002 *Air Quality, space-time models, extremes.*
Associate Scientist, Research Applications Program, NCAR

Curtis Storley, Colorado State U., 6/2003 - *Data mining, feature tracking.*

Dan Cooley, CU - Boulder, 5/2003 - *Paleoclimate, extremes.*

1.4 Outreach

Workshops
GSP-sponsored workshops covered statistics relevant to applications in the geophysical sciences and provided an introduction to some outstanding problems in the geosciences. Each workshop was attended by more than 50 participants divided among graduate students, post docs, and senior researchers in statistics and the atmospheric sciences.

Statistics for Large Datasets [24-26 July 2000] included speakers on scientific visualization and data mining as well as statistics for geophysical and environmental problems. It was jointly sponsored by the National Research Center for Statistics and the Environment, University of Washington.

Spatio-Temporal Modeling [1-6 June 2003] brought together senior and young researchers in applied mathematics, statistics and environmental sciences and was jointly sponsored by the Statistical and Applied Mathematical Sciences Institute.

Statistical software
GSP has been active in developing statistical packages suited to the analysis of geophysical or spatial data. The base URL for GSP software is `http://www.cgd.ucar.edu/stats/software.shtml` and the main postings are:

Fields An R package focused on the analysis of spatial data including large datasets and nonstationary covariance models and simulation.

Waveslim A package in R for wavelet analysis.

RadioSonde [R,S] functions to ingest common radiosonde data files and optionally create SKEW-T/log p plots and wind profiles.

DI: Design Interface A graphically-based interactive program in S-PLUS©that allows one to edit and evaluate spatial designs.

Snackbar [R,S] routines for reading netCDF, Fortran, and C binary files.

extRemes toolkit A package in R for extremes.

Visiting scientists to GSP
To foster connections with the university statistics community, GSP also maintains a shorter term visitor program. Senior visitors have been chosen with the postdoctoral fellows in mind, as they often serve as additional mentors. A typical short-term visitor gives one or more seminars to GSP and local statistics groups and meets individually with each GSP post doc. GSP has supported several longer term visitors from local statistics group to visit on regular basis; e.g. R. Jones and S. Sain (U. Colorado) and T. Lee (Colorado State U.). A number of statistics graduate students have also visited GSP in conjunction with their thesis research or as interns.

Partial visitor list:

Senior visitors and faculty: *W. Eddy*, Carnegie Mellon U.; *D. Ruppert*, Cornell U.; *J. Berger, M. Clyde, D. Higdon, M. West*, Duke U.; *A. Grady, D. Holland*, EPA; *Z.-Q. Lu*, Hong Kong U.; *T.-H. Li*, IBM; *N. Cressie*, Iowa State U.; *A. Chedin*, IPSL-LMD Ecole Polytechnique; *A. Braverman*, JPL; *I. Jolliffe*, King's College; *A. Kaplan*, Lamont-Doherty; *L. Sparling*, NASA/Goddard; *G. Huerta*, New Mexico State U.; *M. Genton*, North Carolina State U.; *M. Berliner*, Ohio State U.; *D. Cox*, Rice U.; *J. Red-Horse*, Sandia National Labs; *T. Hastie*, Stanford U.; *S. Morgenthaler*, Swiss Fed. Inst. of Technology; *R. Eubanks*, Texas A&M; *G. Nason*, U. of Bristol; *P. Bickel, J. Rice*, U. of California - Berkeley; *R. Shumway*, U. of California - Davis; *J. Corcoran*, U. of Colorado - Boulder; *K. Kafadar, S. Sain*, U. of Colorado, Denver; *R. Lund*, U. of Georgia; *B. Bailey*, U. of Illinois; *F. Turkman*, U. of Lisbon; *C. Wikle*, U. of Missouri. *W. Dunsmuir*, U. of New South Wales; *J. Stroud*, U. Pennsylvania; *S. Marron, R. Smith*, U. of North Carolina; *T. Haas*, U. of Wisconsin.
Students: *R. Buchberger, M. Eschenberg*, Colorado State U.; *C. Paciorek*, Carnegie Mellon U.; *E. Gilleland, C. Storlie*, Colorado State U.; *R. Paolo*, Duke U.; *P. Abbit*, Iowa State U.; *M. Vrac*, IPSL-LMD Ecole Polytechnique. *A. Nail, K. Madsen*, North Carolina State U.; *B. Fournier*, Swiss Fed. Inst. of Technology; *Y. Munoz*, Texas A&M; *L. Welty*, U. Chicago; *U. Schneider, D, Cooley*, U. of Colorado; *V. Bulaevskaya*, U. of Minnesota; *P. Caragea*, U. of North Carolina; *B. Das*, U. of Washington.

2 Overview: *A Statistics Program at NCAR*

A grand scientific challenge for this century is to understand the complex interrelationships among the atmosphere, oceans, land processes, biosphere, and human activities that define the Earth System. Coupled with this effort is the need to predict and quantify changes to our environment and to translate such changes into more immediate economic and societal impacts. Hallmarks of this research are complex numerical models describing geophysical processes and societal impacts. Complementing modeling activities are a rich variety of observational and experimental data that are used to suggest new model components, evaluate existing models, and are used for prediction (in concert with numerical models). The exponential increase in computing resources and the proliferation of satellite-based observing systems have supported an unprecedented growth in both numerical modeling and observational datasets. However, despite the promise of these vast new modeling tools and massive geophysical data records, there remains the challenge of integrating models across many different scales and fusing the information afforded by observations with models. Statistical science will be indispensable for integrating models, interpreting diverse types of geophysical data, assimilating data into numerical models, and developing rigorous measures of analysis uncertainty.

> *The most fruitful areas for growth of the sciences are those between established fields. ... It is these boundary regions of science that offer the richest opportunities to the qualified investigator. (Norbert Wiener)*

A grand challenge for statistical science is to tackle the large and complex datasets that are now the norm in many areas of science and engineering and to leverage the rich structure and prior knowledge afforded by traditional numerical models. New methodology in many areas of statistics including spatio-temporal models, Bayesian statistics for high-dimensional problems, extremes, data mining, and experimental design can be motivated by the need to scale statistical tools to massive datasets and to fashion statistical models that enhance the collaboration of statisticians with other scientists. We assert that the geosciences and, in particular, a focus related to the atmosphere and the Earth System, are a fertile testbed to pursue new statistical methodology.

The Geophysical Statistics Project (GSP) at the National Center for Atmospheric Research (NCAR) provides a unique opportunity for research that addresses the challenges to the geosciences and statistics. NCAR's mission is to lead large, multidisciplinary research projects that would be difficult for individual faculty or universities to pursue. The scientific program at NCAR, supported by several hundred Ph.D. scientists, is characterized by extensive experimental and observational datasets and mature numerical models. It is precisely this scientific context that motivates the challenges faced by statistical science.

We outline a plan of research and a three-part visitor program that will enable GSP to bridge the geosciences and the statistical research communities. In addition to a *Postdoctoral Fellows* (PF) program, GSP will reach graduate students through a *Thesis-In-Residence* program and young faculty through a *Statistics Chautauqua*[1]. This proposal builds on past successful programs and adds new elements that broaden GSP's connections and outreach. There will also be an expansion in GSP's scientific involvement at NCAR by nesting the statistical research within broad strategic initiatives of the center. The participation of GSP in these initiatives represents a new level of integration of statistics at NCAR and this alignment makes it possible to contribute to the grand challenge problems of the Earth system *and* statistical science.

3 Statistical Research

The geosciences offer a rich context for statistical research and methodology. Geophysical processes often possess complex interactions at many different temporal and spatial scales and are associated with extensive observational datasets and numerical models. Coupled with the influence of human activities on the environment and the vulnerability of society to extreme events, an added level of complexity is

[1] *Chautauqua*: A community assembly for educational purposes patterned after the summer cultural and educational center in Chautauqua, NY.

layered onto the physical and biological systems. From this perspective, we propose projects that are diverse in both geophysical application and statistical research.

3.1 Spatial Processes

Statistical methodology for spatial data is a focus for GSP. Nearly every project has a spatial component and nearly every member of GSP achieves some expertise in this area. Geophysical fields are nonstationary and so a primary area of research is developing flexible, computationally efficient representations for nonstationary covariance functions. This work will focus on representations using multi-resolution bases, such as wavelets ([69],[133],[119]), and process convolution approaches ([136]). To fix concepts, let $\boldsymbol{f}$ be a spatial field on a regular grid and $\boldsymbol{f} = W\boldsymbol{a}$ be an expansion where the columns of W are basis functions and $\boldsymbol{a}$ are random coefficients with covariance D. A useful decomposition for the covariance matrix for $\boldsymbol{f}$ is WDW^T. A multi-resolution approach would consider a wavelet basis for W and focus on sparse structure for D. In contrast, process convolution builds structure into W, often with stationary specifications for D. A fundamental area of statistical research is to understand the kinds of processes that can be approximated by these models and balance the trade-off between covariance model complexity and the accuracy of spatial predictions.

Another primary area of research is to scale statistical methods to handle large spatial prediction problems where observations occur at irregular locations. One approach is to introduce sparsity in the covariance matrix by tapering with a compactly supported (positive definite) kernel (e.g. [128]). Given the correct choice of taper, this approach can be justified based on matching the tail behavior of the spectral density of the covariance function ([164]). This analysis suggests that, under certain assumptions, spatial process estimators and classical kernel estimators will be asymptotically equivalent. Thus we see the potential to unify these two approaches for function estimation.

The modeling of non-Gaussian random fields faces the basic challenge of formulating useful departures from the multivariate normal. A flexible approach is the use of a latent Gaussian field to drive non-Gaussian behavior. For example, it has been effective to model the patchiness of daily precipitation occurrence (rain/no rain) over space by thresholding a Gaussian process. The introduction of nonstationary covariance models depending on covariates (such as season, elevation or aspect in the case of precipitation) will yield an important class of models for binary fields.

Modeling forecast error covariance functions The National Centers for Environmental Prediction (NCEP) regularly archives its weather forecasts along with the observations. The comparison of *forecast* with *observed* is a way to assess the forecast skill. Although the forecast (or background) error covariance matrix is a main ingredient of the variational methods used in operational weather forecasting, this forecast verification dataset has not been closely studied or modeled. To understand how the forecast error covariance depends on the predicted atmospheric flow, we will use the partial derivatives of the predicted wind vector field (the strain tensor) as a covariate for anisotropic covariance models. This analysis may suggest statistical models for the error covariance function as a function of the atmospheric state and lead to adaptive background covariances for data assimilation.

Electric field dynamics in the upper atmosphere The variability of the upper atmospheric electrodynamics is not well determined because of the sparsity of observations and the rapid temporal variation of the ionosphere relative to measurements ([143],[120]). For example, direct measurements of the (two-dimensional) electric field at a particular time are only sampled by single satellite tracks. A statistical problem is to determine a two-dimensional covariance function based on single one-dimensional transects of many independent realizations of the field. Preliminary work ([49]) based on principal components with missing data ([130]) has produced physically interpretable results. The goal is to produce a complete covariance function using a multi-resolution basis and to incorporate this model into the Assimilative Mapping of the Ionosphere Electrodynamics procedure (AMIE)([155]).

Soil properties The analysis of soil properties is important for assessing the impact of climate change on agriculture (through crop models) and for the land component of a climate model ([131]). One aspect of crop model research is to quantify the variability in soil characteristics across space to understand biases in aggregation. Another issue is to relate bulk properties of the soil to the water holding capacity (the pedotransfer function) ([152]). A useful Bayesian model for the pedotransfer function is a multivariate regression model with additive components whose priors are based on Matern families of processes. This approach is more flexible than most parametric versions in soil science and also allows for generating ensembles of depth profiles as a way of characterizing the uncertainty in soil properties.

3.2 Space-Time Processes

Spatial fields that evolve over time are ubiquitous in the physical sciences. Although there is some work on space-time covariance functions ([123]) and more recent ideas on Matern-like families ([163]) a process-oriented approach considers a hierarchical model (HM) that allows for the blending of physical dynamics with stochastic components estimated from data (or model output) ([167],[91],[96],[93],[16]). For example, if $\boldsymbol{x}_t$ is the field at time t; $\boldsymbol{x}_t = g(\boldsymbol{x}_{t-1}, \theta) + \boldsymbol{u}_t$. Here g incorporates basic physical dynamics of the system along with possible dependence on parameters and $\boldsymbol{u}_t$ is a stochastic process that can evolve according to an autoregressive process: $\boldsymbol{u}_t = W\boldsymbol{a}_t$ and $\boldsymbol{a}_t = G\boldsymbol{a}_{t-1} + \boldsymbol{\epsilon}_t$. The next level of the hierarchy considers models for θ, G and other parameters and additional layers can add still more structure. At the bottom of this hierarchy, physical information can be incorporated using informative priors on some of the parameters, leveraging the Bayesian paradigm. Thus, hierarchical models build complex behavior through a series of simpler models and become a powerful statistical method when developed using (approximate) physical theory and prior knowledge. A basic statistical issue is to balance the complexity of the HM with the ability to estimate hidden components of the model from limited data. This is not only an issue of model selection and parsimony but also has an impact on the efficiency of Markov Chain Monte Carlo sampling of high-dimensional posterior distributions.

Stochastic subgrid-scale parameterizations A fundamental problem in the construction of models for the atmosphere and ocean is that they are limited in spatial and temporal resolution. Processes that occur at smaller scales can not be modeled directly and are accounted for by a *parametrization* based on the larger scales of the model state. For example, in a General Circulation Model (GCM) individual thunderstorms (convection) can not be resolved and so the important vertical mixing and transport of heat must be inferred through a parametrization. We will explore stochastic parametrizations where a random component incorporates the indeterminacy of the subgrid-scale processes based on the large-scale states. This project builds from a growing interest in introducing stochastic elements into large-scale models for the atmosphere ([154],[141]). A series of experiments with a cloud-resolving model (CRM)([148]) will be used to fit a statistical model that describes the spatio-temporal relationship between the large-scale conditions and the resulting convective processes and cloud formation. To produce a compact stochastic description of this process, the (cloud) fields will be represented using a hierarchical model where the large-scale state variables enter as covariates and the physical relationships are synthesized with stochastic modeling. The goal will be an easily computed stochastic description of the convective process as a function of the GCM state.

Inferring CO_2 fluxes from observations To model the present and future climate it is necessary to have an understanding of the carbon cycle including the release and absorption of CO_2. Although it is difficult to measure sources and sinks for carbon directly, they can be inferred by solving an inverse problem based on carbon concentration measurements ([151]). A statistical framework for this problem has two components. An observational model relates CO_2 fluxes to concentrations at given locations and times and a state equation describes the temporal and spatial structure in the fluxes. The observation equation depends on an atmospheric transport model taking observed (analysis) winds as input and advecting CO_2; one important issue is to include uncertainty in the transport operator. The state equation will have a HM structure along with a regression component based on seasonality and land cover. This project will explore the potential of HMs to yield estimates and companion measures of uncertainty for a spatio-temporal field and quantify the capability of the current network for monitoring CO_2. An important extension is to include other constituents, such as CO, to improve estimates of transport.

Weather generators and spatially coherent downscaling Stochastic weather generators play an important role in the assessment of climate change. Their function is to simulate realistic daily meteorology (based on a particular climate) to be used as inputs for agricultural, hydrologic, health effects, or energy models ([147],[146],[145]). A useful model for daily meteorology is an autoregressive model that is conditioned by precipitation occurrence ([126]). This research will improve these models by adding flexible, nonparametric transformations of weather variables that can account for the non-Gaussian distributions of meteorology and an observation-driven model for precipitation occurrence that can simulate more realistic runs of wet and dry days ([110]). The temporal model will also be extended to simulate spatially coherent meteorology by adding spatially dependent shocks into the autoregressive component and dependent binary fields for the occurrence process. Part of the statistical challenge is to extrapolate the parameters and transformations in the model smoothly across space to locations where data is not available. The result will be a stochastic model that can be used for regional assessment and will accurately reproduce weather extremes and variability.

3.3 Design and Analysis of Computer Experiments

Sophisticated numerical models simulating complicated processes are now the norm in many fields and often represent a synthesis of the state of knowledge of a particular system. The design and analysis of computer experiments (DACE) focuses on space-filling type designs for the input space and the use of stochastic process priors to model the response surface of the codes to the inputs ([157],[124]). Interpolation or very small measurement errors make the estimators more sensitive to covariance misspecification and extrapolation and so cross-validation techniques need to be blended with more formal Bayesian formulations to give robust prediction uncertainties.

A generalization of the traditional designs for computer experiments is to consider both inputs and outputs as being functional in value. For a climate experiment, the time-varying inputs of greenhouse gases, aerosols and other forcings that drive the climate system are functional inputs and the resulting spatial pattern of temperature change is a functional output. One design for functional inputs is to expand the inputs in a low-dimensional set of basis functions and then apply a conventional design to the coefficients. One constraint is to reproduce qualitative features in the input functions (e.g. monotonicity, positivity), which may not be possible considering linear subspaces.

Typically, researchers develop a family of computer models varying in accuracy, run time and size. Given limited computing resources and the need to make model predictions at many different input combinations, the statistical problem is to combine the use of "cheap" models with "expensive" ones while minimizing the prediction error. We will follow a framework of tuning cheap models to reproduce the results of expensive runs and then using the cheap model as part of the extrapolation to predict results under different input combinations (e.g. [118],[140]).

Designed experiments for different emission scenarios An emission scenario refers to the temporal and spatial pattern of greenhouse gases and aerosols emitted into the atmosphere and depends on projections of population growth, economic and societal changes, and possible mitigating actions taken by governments ([144]). The statistical problem is to prescribe a concise set of emission scenarios with the goal of accurately determining the future climate under a wide range of other possibilities. Part of this task is to incorporate results of past model experiments as part of the design criterion as well as natural constraints on the space of possible scenarios.

Climate model interpolation The number of climate model experiments using fully coupled Atmosphere Ocean General Circulation Models (AOGCM)s with different emission scenarios is limited due to computational resources. Based on a set of AOGCM experiments and an intermediate-size climate model ([160]) a "cheap" versus "expensive" model strategy will be developed to predict under a variety of other emission scenarios.

3.4 Particle Filtering and Data Assimilation

A basic filtering problem is to estimate the state of an evolving system based on noisy or incomplete observations. A well-known example for the atmosphere is numerical weather prediction (NWP). The solution usually involves two steps: 1) updating the distribution of the state of the system based on a new observation and 2) propagating the distribution forward in time using a dynamical model. Particle filters ([125]) are discrete approximations to the distributions in steps 1) and 2) and work well for low-dimensional systems and stochastic system dynamics. However, these techniques do not extend easily to NWP where the system is high-dimensional (e.g. $d = 10^7$) and the dynamics are largely deterministic but nonlinear. As an alternative, the geosciences have developed ensemble Kalman Filters (EnKF) as stable methods of filtering ([127],[117],[139]). The EnKF is surprisingly efficient even for very small ensemble sizes (e.g. n=20), but there are several heuristic adjustments required to make these algorithms function. Understanding why the approximations and tuning parameters work will provide guidance for automatically tailoring the filter to new problems. Another outstanding statistical problem is to estimate the stability and performance of the filter for different ensemble sizes and different numbers of observations given assumptions about the system dynamics (e.g., its stationary distribution). It is common that stable filters tend to yield narrower spreads in the forecast distributions when compared to the true state of the system ([134]). Thus, another challenge is to construct stable and efficient filters while modifying the forecast distribution to achieve more accurate inferences.

Blending non-Gaussian and Gaussian filters Many nonlinear systems exhibit non-Gaussian, possibly multi-modal distributions for the state when sampled at large time intervals. In low-dimensional systems it is possible to make substantial improvements by using a filter based on mixtures of Gaussian distributions ([10]). However, in high-dimensional systems, it is not possible to consider an arbitrary mixture distribution and the non-Gaussian and EnKF filters need to be blended effectively. One approach is based on localization. Usually, updating a single observation strongly influences just a few components of the state vector. A non-Gaussian update is proposed for this set of state components and the remainder are updated using the EnKF. These ideas will be implemented in a mesoscale atmospheric model for thunderstorms ([165]), where we expect bimodal forecast distributions for some of the state variables.

Adaptive algorithms for forecasting Numerical weather prediction is a cyclical procedure where observations are assimilated and forecasts are made at regular intervals. The sequential nature of this process suggests the possibility of statistical estimates for algorithm parameters that adapt to the current weather regime or season. One approach is a sequential Bayesian procedure where the algorithm tuning parameters are updated along with the state ([125]). There are several statistical problems that arise from the nonlinearity of the system and the discrete approximation of the distributions. Some of these include discounting the forecast distribution and balancing the amount of past information with the variation in the system. Another application is to adjust for data quality and model bias through a statistical process control model whereby the contribution of the data or model forecasts could be down-weighted.

3.5 Scientific Data Mining

Data mining includes flexible statistical methods for pattern recognition, classification and regression analysis in the context of large datasets. Although these methods are usually applied to commercial, demographic, or health-related databases, a similar viewpoint can be useful in the physical sciences when one must resort to semi-empirical models to identify or classify nonlinear phenomena (e.g. [137],[161]). Two areas of focus in this proposal are image analysis and robust smoothing.

Identifying coherent structures in turbulent flows The classical nonlinear equations for fluid flow are well known to simulate coherent structures such as vortices (hurricanes, ocean eddies) and jets (the Gulf Stream) that cannot be understood directly from the physical equations. Moreover, three-dimensional turbulent flow, appropriate for the atmosphere's surface layer and charged particle flows, yields coherent behavior (vortex sheets) that is even difficult to visualize. The statistical challenge is to extract such coherent structures from the turbulent flow. Quantifying these structures can be useful at a theoretical level to derive scaling laws ([166]) or to create indices that compare structures found in observed data with those in model simulations. Research on isolating structures in images will include the use of multi-resolution templates ([142],[90]), mathematical morphology ([159]), and adaptive tracking algorithms ([138],[158]) for studying temporal evolution.

Quality-checking geophysical datasets The historical records for surface meteorological variables and satellite data products are large and often contain spurious values. This research will develop flexible and robust smoothers to flag large residuals. One promising method is wavelet thresholding based on the "pseudo data" idea from robust regression ([73],[107]). A modeling issue is balancing the amount of local smoothing against the outlier distribution. A sharp peak in the field could be attributed either to nonstationary behavior of the process or to an outlier. Some approaches for the choice of local smoothing parameters are to use covariates based on physical information from other sources or, based on temporal information, accumulate a reference distribution for the average properties of the field.

3.6 Extremes

Key statistical issues are often driven by the tail behavior of a distribution rather than the central tendency. Extremes play an essential role in assessing impacts from environmental and meteorological factors because of the vulnerability of society to rare but extreme events. Extreme value theory, while being well developed in univariate and bivariate settings, offers new areas of research in extending methods to spatial and spatio-temporal fields. Classically, maxima have been modeled by the Generalized Extreme Value Distribution (GEV) and exceedances above a threshold by the Generalized Pareto Distribution (GPD). Identifying how the parameters of these distributions are linked to climatic factors (natural and anthropogenic forcings) is fundamental to understanding the range of scenarios of future changes in the frequency and intensity of extreme events (such as heavy precipitation, droughts and high temperatures) ([135],[153]). The use of Bayesian methods is a natural way of improving the inference for parameters by combining different complex climatic covariate relationships. Although Monte Carlo sampling for posterior distributions has already been used for extreme value statistical modeling ([121]), to our knowledge, *perfect* sampling ([132]) has not yet been explored to improve the inference quality. This approach can be also used for Bayesian model selection. Choosing between competing extreme models greatly facilitates the investigation of causes and predictor variables for extreme events.

Extreme events in hourly precipitation Understanding the risk of flooding and the construction of flood plain boundaries in an urban area depends on the local climatology of extreme precipitation. A common statistic for extreme rainfall rates is the return level. In statistical terms, this quantity is a quantile from a GEV distribution fitted to observed data and so depends on the uncertainty in estimating this distribution. For a given area the return level is a function of space, season, and other covariates and its variation needs to be accomodated when used in hydrologic models that determine flood plains. One area of research would be to combine extreme value theory with spatial models for the tail parameters and other covariates to produce rigorous estimates of the return levels. A key contribution will be the propagation of the uncertainty in the extreme value distributions into the final determination of flood plain boundaries. The Colorado Front Range will be used as a study area because of the proximity of urban areas to flood prone drainages.

Spatial scaling of extremes A challenge in transferring the results of climate change projections to tangible societal impacts is relating the spatially averaged meteorological variables simulated by an AOGCM to their values at a point (e.g. [150]). Current climate models are restricted in their resolution to grid cells that are 100km to 300km on a side. The statistical problem is to infer the distribution of point extremes based on the grid cell averages simulated by the model. This question is posed under the assumption that the climate model yields a perfect simulation, in particular with regard to its capacity to generate variables with heavy tail distributions. A companion issue is how well (imperfect) climate projections reproduce the extremes observed in our current climate. Matching extreme distributions between models and data must be done in a regression (covariate) setting in order to adjust for the different timing of large-scale variation such as the EL Nino/Southern Oscillation or the North Atlantic Oscillation. Regardless of the strength of the connection between grid cell average and the climate at point locations, it is crucial to characterize the uncertainty in this relationship in order to produce uncertainty estimates when these climate scenarios are subsequently used in impact models.

3.7 Statistical Computing

Bayesian hierarchical models and other related modeling frameworks are rapidly becoming mature and flexible tools for handling complex geophysical processes. In contrast to the richness of the statistical methodology, there is a deficit in principles for adapting methods to large and interesting problems.

Geophysics and other areas where data are measured over time and space create a unique set of computational problems because the statistical models must have a global extent which cannot readily break into small, independent pieces. The interaction among different components can add greatly to the computational burden.

Computing for large spatial datasets The theory described in Section 3.1 suggests that deliberately introducing sparsity in a covariance matrix through positive definite tapering will not introduce much error in the spatial estimate. We will couple this idea with the use of fast multiplication of covariance matrices to give an algorithm that accomplishes the spatial prediction for large and irregularly spaced spatial datasets (10^4 to 10^5 locations). Part of the efficiency is gained by registering all locations to a fine spatial grid and taking advantage of the fast Fourier transform or the discrete wavelet transform for regular arrays of points. (Although each spatial location is discretized to a grid point not all grid points need to have observations.) Another research topic is the fast simulation of spatial fields on grids and conditional sampling. The algorithms will build on circulant embedding ideas ([162],[168]) to handle the simulation of fields with long correlation scales. Finally, some important extensions of this work are devising approximate likelihoods using sparse methods and dealing with nonstationary covariances using mixtures of stationary fields or multi-resolution approximations.

Blending near-surface ocean winds A long-term project to test the transfer of Bayesian hierarchical models to a substantial geophysical application is *GibbsWinds*, a value-added surface wind dataset for the tropical Indian and western Pacific ocean, based on *QSCAT* satellite observations and NCEP reanalysis data ([30]). The goal is to use NCAR's state-of-the-art massively parallel computing architecture to tackle a large problem without compromising the integrity of the statistical analysis. Some areas of research include dividing the computation into manageable chunks without introducing artifacts, distributing computation among many nodes, and developing strategies for managing the irregular datastreams being ingested by the computational core and the resulting output streams from the MCMC algorithms.

4 Program structure

The need for close collaboration among scientists and statisticians to vitalize and enrich statistical science and to make advances on scientific problems is undisputed. Overall we believe that GSP is a successful model for achieving these goals in the geosciences primarily because of the location at NCAR *and* support for distinctly statistical research. In March, 2002 GSP underwent a site review by an independent panel convened by NSF-Division of Mathematical Sciences. Part of the review panel's conclusion ([149]) was *"The Panel concurs that, by almost any measure the GSP has been a resounding success ..."*. While we believe the overall structure of GSP is sound, in this proposal we outline two opportunities to broaden the program to graduate students and young faculty.

The vision for GSP's future is a small group of permanent NCAR staff with a much larger group of statistical researchers and students visiting on terms from a few days to semesters to multiple years. A permanent presence is important for the continuity of long-term projects while an emphasis on visitors is an efficient model to reach the larger statistical community. The philosophy for GSP is careful mentoring of young researchers and providing a high level of support for computing and data access. This is necessary to achieve nontrivial and novel scientific results, gain collaborative experience with other scientists, and still allow young statistical scientists to build their own research programs. The following sections outline the composition of GSP and the program to implement its research.

4.1 Program Members

GSP staff The Section Head (Nychka) and a new Scientist (*see* Section 4.5) will share daily oversight along with mentoring and project management. Support staff will consist of two Associate Scientists to provide computing, data, and algorithm support and an Administrative Assistant.

Graduate students Two to four graduate students will be supported in the thesis-in-residence program that typically lasts an academic semester. In addition, one to three graduate students will participate in GSP who are based in the local statistics groups: Colorado State University (CSU), University of Colorado(CU) - Boulder, and CU - Denver.

Postdoctoral Fellows On average, five statistics PFs will be on staggered terms from two to three years.

Faculty visitors The Statistics Chautauqua program will bring in the largest contemporaneous number of visitors to GSP and is expected to entail 3-5 faculty members and 3-5 graduate students and/or statistics post docs from other statistics departments or centers. GSP will also continue to encourage a regular visiting schedule (e.g. one day each week) by faculty at local statistics groups (S. Sain *CU - Denver*, T. Lee and R. Davis *CSU*, P. Naveau *CU - Boulder*).

4.2 Thesis-In-Residence for Statistics Ph.D. Students

This program is designed for Ph.D. graduate students who have finished coursework and are researching and writing dissertations. Our premise is that by taking advantage of the uninterrupted time and resources at NCAR, the student can incorporate substantial geophysical applications into his/her thesis work that would be difficult given the usual time and amount of support for thesis research. The goal is to broaden both the scope and depth of the thesis while minimizing the extra time this might take at the student's home department. This will be possible because many of the ongoing GSP projects have obvious applications for a wide range of statistical methodology. In addition, the GSP facilities for computing and access to geophysical data are already well-suited for statistical research. The direct benefits to GSP are the transfer of new statistical methodology to NCAR science projects. Some indirect benefits are the possibility of recruiting outstanding students as PFs and also entraining the students' advisers into longer term collaboration. Graduate student visitors also provide the opportunity of mentoring and interaction with GSP PFs. While there is no substitute for the student's thesis adviser, GSP's focus on geophysical problems provides a larger group to support the student's research. GSP has already successfully hosted several Ph.D. students in this manner including Christian Schoelzel and Christoph Gebhardt, U. of Bonn (paleoclimate reconstruction) and Barnali Das, U. of Washington (modeling of nonstationary spatial fields).

This opportunity would be launched with a short article in *AMSTAT News* and subsequently advertised there and in the *IMS Bulletin*. Promising students will also be identified from GSP workshop participants and through ongoing collaborations with statistics faculty.

4.3 Post Doctoral Fellows

The core of GSP has been a critical mass of three to six statistics PFs on staggered terms ranging from six months to three years (terms less than two years are due to members securing permanent research and teaching positions outside of NCAR). Indeed most of the research reported in Section 1 comes from their efforts. Feedback from past PFs has highlighted the productive interaction and support among colleagues at the same junior level. Our plan is to continue this kind of group with two modifications: a broader plan for recruiting promising statistical Ph.D.'s and more flexibility through shared appointments with other centers and universities.

At the inception of GSP, nearly ten years ago, post doctoral appointments in Statistics were rare and GSP largely enjoyed an open field in recruitment, especially in the area of geoscience applications. We must now address maintaining a group of statistics post docs in an environment where competition for talented post docs is keen and direct entry into tenure-track positions can siphon off some of the best new graduates.

Increased visibility: We believe that the richness, diversity and societal relevance of the scientific program at NCAR will always attract some statistical graduates for PF positions. Rather than periodic notices for positions, GSP will recruit throughout the calendar year with the possibility of attracting strong students finishing "out of cycle", e.g. completing a degree in early Fall. Several PFs became acquainted with GSP/NCAR as students through sponsored workshops and we will continue to use this activity for visibility and recruitment. Finally, through exposure from the Thesis-in-residence and

Summer Chautauqua programs (outlined below), we expect that promising Ph.D. students will be more inclined to consider NCAR/GSP post doc opportunities.

Alliances: The PF positions will also emphasize the option to combine the NCAR appointment with other activities and other institutions. Some examples include:

- Sharing joint post doc positions with other centers (SAMSI, Center for Integrating Statistics and Environmental Science (CISES), Space-Time Aquatic Resources Modeling and Analysis Program (STARMAP), Los Alamos National Lab (LANL)).
- Joint appointments or significant visits to other statistics programs (e.g. Colorado State University, University of North Carolina - Chapel Hill, University of California - Berkeley, and University of Washington).
- Gaining teaching experience through one of the local statistics groups (CU - Boulder, - Denver and CSU).

As examples, GSP has already been successful incorporating joint positions with other statistics departments (Jarrett Barber, 25% NCAR/GSP and 75% Duke U.; Craig Johns, 50% NCAR/GSP and 50% CU - Denver), providing teaching experience within the PF appointment (Reinhard Furrer, CU - Boulder, 9/02-5/03) and extended visits to statistics departments (Claudia Tebaldi, University of Washington, 4/00-5/00). Appended are letters of support from these institutions. Some common interests are filtering and data assimilation (SAMSI, CISES), impacts of climate change (CISES, STARMAP), analysis of computer experiments (LANL) and statistical computing (CISES, LANL).

4.4 A Statistics Chautauqua

The GSP Statistics Chautauqua program will be a deliberate concentration of visitors to GSP for a month-long period in the summer. The composition of the visitors will be broad, ranging from statistics graduate students, to young researchers (e.g. post docs) from other centers and universities, to both junior and senior statistics faculty. However, the targeted group for this program are young statistics faculty interested in expanding their research into geophysical applications. The goal is to create a community environment that facilitates the transfer of statistics to geophysical problems at NCAR and the engagement of statistical researchers with substantive geophysical problems. The month-long term of this program recognizes that substantial progress on scientific problems requires some understanding of the basic science and often the need for several meetings with the scientific team simply to define the statistical problems. We also recognize that the serendipitous and informal interactions among researchers in a stimulating professional environment are perhaps the most important way to spur creative interdisciplinary research. Part of the format of this program is to include tutorial talks on the scientific background necessary to appreciate geophysical problems. For example, although the equations describing the dynamics of the atmosphere and ocean are fundamental to the geosciences, few statisticians are familiar with these concepts. Access to NCAR's extensive data archives will also be addressed. Another important component is fostering interaction among the visitors and GSP members themselves. This includes informal lunchtime seminars on research areas, a reading/discussion group on articles in a particular area and expository lectures on emerging areas of statistics.

This program will be advertised in the *AMSTAT News*, the *IMS Bulletin* and other statistics newsletters. Preference will be given to underrepresented groups in statistics, to young faculty members and to graduate student/adviser pairs.

4.5 Program Management

Leadership The overall project leadership will come from the GSP Section Head (Doug Nychka), the four co-PIs, and a new scientist in GSP. The new scientist, supported by this proposal, will be at the equivalent rank to a university Associate Professor. His/her addition will provide time for both GSP scientists to manage collaboration between GSP and NCAR initiatives, maintain the continuity of small projects and serve on NCAR committees. This activity will be new for GSP and is necessary to maximize the impact of statistical science at NCAR. It is also a reflection of the maturity of GSP moving beyond a statistics research program into a full-fledged scientific section at NCAR.

Oversight and guidance The program will maintain an external advisory panel consisting of researchers in statistics and the atmospheric sciences. This advisory group will convene annually to assess the overall direction of research and the effectiveness of the training components. The format of this panel meeting is a site visit: review materials are submitted beforehand, all members of GSP present their work and the panel produces a written report. Current members are:

Peter Guttorp (University of Washington)
Sallie Keller-McNulty (Los Alamos National Laboratory)
Gerald R. North (Texas A&M University)
John Rice (University of California, Berkeley)
Andrew Solow (Woods Hole Oceanographic Institution)
Daniel S. Wilks (Cornell University)
Chris Wikle (University of Missouri)
Brani Vidakovic (Georgia Institute of Technology)

In addition, a GSP internal advisory panel represents the science divisions of NCAR.

Mentoring The roles of GSP permanent staff are purposefully designed to maximize the mentoring of GSP PFs and visiting students. This is in contrast to more traditional supervision where direction of a postdoctoral student is an activity typically added to a faculty member's teaching, research, and administrative duties. In the past, the GSP project leader has held regular weekly meetings with each PF and attended most of the meetings between GSP PFs and NCAR scientists. Equally important, it is expected that the PFs will also get guidance and support from the scientists with whom they collaborate. We have found that NCAR scientists are generous with their time in working with younger researchers.

Project selection and life cycle Projects selected early in a PF's term will typically reflect interests and skills derived from his/her thesis research. However, it is the goal of the program to always emphasize scientific merit and quality on par with the statistical research supported by NSF-Division of Mathematical Sciences. We will follow a confidence/challenge approach for the PF's suite of projects. Well defined and incremental projects are identified early in the PF's term to give him/her confidence in interdisciplinary research and some immediate research products. The next phase is to identify longer-term projects that carry higher risk but also higher payoff both in statistical methodology and scientific impact and may involve new areas of statistical expertise.

Nearly every successful project at GSP has its own unique story of inception and of match-making. However, the presence of GSP as an embedded group within NCAR provides some regular modes for project identification: attendance of the regular seminars hosted by the scientific divisions and annual division retreats, interaction with non-statistics post docs and graduate students, the GSP advisory panels, and the involvement of GSP staff in NCAR strategic initiatives. Any potential project will be vetted through the co-PIs for its overall value and the quality of mentoring from the involved scientists. Closing or continuing a project is also based on the established project selection criterion ([129]). These decisions will involve the external advisory panel and will consider the impact to the overall GSP program.

4.6 Integration Within NCAR

In the past two years NCAR has created Strategic Initiatives that cut across traditional divisional boundaries and pursue integrated scientific programs of international prominence. Three NCAR initiatives and a fourth initiative that is under review are briefly described along with their connections to this proposal.

Weather and Climate Impact Assessment will provide leadership for U.S. participation in the Intergovernmental Panel on Climate Change and includes characterizing the uncertainties in the assessment process, understanding the role of extremes in societal impacts, and designing emission scenarios and climate model experiments. The projects on soils, weather generators, and the sections on Design and Analysis of Computer Experiments (3.3), and Extremes (3.6) contribute to these goals.

Data assimilation addresses the need for a national program to pursue research on the integration of geophysical models with observations and to support NCEP and other operational forecasting centers through technology transfer. The project on forecast errors and the section on Filtering and Data Assimilation (Section 3.4) provide statistical foundations to this effort and infuse new ideas from the statistical community.

Biogeosciences recognizes the importance of the biosphere on the Earth System and is focused on the perturbations to the atmosphere's chemistry by human activities. Understanding the fluxes of carbon at the global scale is central to this initiative and the project on inferring CO_2 fluxes is a collaborative effort with scientists in this initiative.

Models and Methods for Multiscale Geophysical Processes (*under review*) draws on statistical models, the theory of turbulence, experimental data and numerical models to characterize multiscale features through stochastic models. The projects on stochastic parameterizations and identifying coherent structures in turbulent flows are part of this initiative.

4.7 Visibility and Outreach

Workshops GSP will cosponsor a workshop broadly related to the interface of statistics and the geosciences at least every other year of the project. Previous GSP workshops have been very successful with both tutorial- and research-oriented talks and ample time for discussion. Possible future workshop topics include: Data Assimilation, Design and Analysis of Computer Experiments, Blending Stochastic and Deterministic models, Scientific Data Mining, Statistics for Large Datastreams, and Relating Climate to Human Health. Some potential alliances include SAMSI, the PRIMES and STARMAP programs at CSU, CISES and LANL. Internal to NCAR, GSP will collaborate with the Geophysical Turbulence Program and the Data Assimilation Initiative on joint workshops and summer schools.

Presentations Program members will present research at statistical meetings and departmental colloquia. In particular, the permanent GSP members will regularly visit university statistics departments to increase awareness of GSP accomplishments and encourage applications to the visitor programs. An important mode of visibility is through contributed and invited sessions at the Joint Statistical Meetings (JSM), Interface, the Young Researchers Conference, and other conferences where GSP members present their current research. An effort will be placed on having at least one session (either special contributed or invited) every year at JSM that is a venue for GSP research.

Internet resources GSP will maintain an extensive home page (`www.cgd.ucar.edu/stats`) with links to technical reports, program members, current projects, software and datasets and other statistical and geophysical sites. The external version of the homepage will give prominence to the visitor programs.

Statistics software and numerical models The program will make available statistical software related to GSP projects and participate in the development of at least four important public resources. The research on spatial models will be implemented in a new spatial analysis package synthesizing the ideas implemented in the `fields` and `GeoR` packages with extensions to large datasets and space-time models. The `wgen` weather generator package will also be extended to a spatial context. Research in extremes will be transferred into the `extRemes` toolkit including extensions to spatial analysis. The Data Assimilation Research Testbed (`DART`) will disseminate a wide range of numerical models and methods to test ideas of data assimilation.

References

GSP articles published or in review

[1] C. Ammann and P. Naveau. Multi-decadal periodicity in tropical explosive volcanism and its influence on climate. *Geophysical Research Letters*, 30(5):1210–1029, 2003.

[2] B. A. Bailey, L. M. Berliner, W. Collins, D. W. Nychka, and J. T. Kiehl. Neural networks: Cloud parameterizations. In L. M. Berliner, D. Nychka, and T. Hoar, editors, *Studies in the Atmospheric Sciences*, volume 144 of *Lecture Notes in Statistics*, pages 97–116, New York, 2000. Springer-Verlag.

[3] B. A. Bailey, L. M. Berliner, W. Collins, D. W. Nychka, and J. T. Kiehl. Spatio-temporal neural network modeling of cloud cover. *Technometrics*, 2003. In review.

[4] B. A. Bailey and S. C. Doney. Quantifying the effects of noise on biogeochemical models. In *Computing Science and Statistics*, volume 32 of *Proceedings of the 32nd Symposium of the Interface*, pages 447–453, 2000.

[5] B. A. Bailey, S. C. Doney, and I. D. Lima. Quantifying the effects of dynamical noise on the predictability of a simple ecosystem model. *Environmetrics*, 2003. In press.

[6] B. A. Bailey, D. W. Nychka, and S. Ellner. A central limit theorem for local Lyapunov exponents. *Stochastic Processes and their Applications*, 2003. In review.

[7] T. Bengtsson and J. Cavanaugh. A simulation-based model selection criteria for the state-space framework. *Journal of Computational and Graphical Statistics*, 2002. In review.

[8] T. Bengtsson and J. Cavanaugh. State-space discrimination using Kullback information measures. *Technometrics*, 2002. In review.

[9] T. Bengtsson and D. W. Nychka. Adaptive methods in numerical weather. In *Proceedings of the 1st Spanish Workshop on Spatio-Temporal Modeling of Environmental Processes.*, Benicassim, Castellon (Spain), 2001.

[10] T. Bengtsson, C. Snyder, and D. W. Nychka. Toward a nonlinear ensemble filter for high-dimensional systems. *Journal of Geophysical Research, Atmosphere*, 2002. In review.

[11] L. M. Berliner. Hierarchical Bayesian modeling in the environmental sciences. *Allgemeines Statistisches Archiv, Journal of the German Statistical Society*, 84:141–153, 2000.

[12] L. M. Berliner. Monte Carlo based ensemble forecasting. *Statistics and Computing*, 11:269–275, 2001.

[13] L. M. Berliner. Physical-statistical modeling in geophysics. *Journal of Geophysical Research*, 108(D24), 2003. 8776, doi: 10.1029/2002JD002865.

[14] L. M. Berliner. Uncertainty in climate change. *Statistical Science*, 2003. To appear.

[15] L. M. Berliner, R. A. Levine, and D. J. Shea. Bayesian climate change assessment. *Journal of Climate*, 13:3805–3820, 2000.

[16] L. M. Berliner, R. F. Milliff, and C. K. Wikle. Bayesian hierarchical modeling of air-sea interaction. *Journal of Geophysical Research*, 108(C4):1–18, 2003. 3104, doi:10.1029/2002JC001413.

[17] L. M. Berliner, C. K. Wikle, and N. Cressie. Long-lead prediction of pacific SSTs via Bayesian dynamic modeling. *Journal of Climate*, 13:3953–3968, 2000.

[18] D. J. Cummins, T. G. Filloon, and D. Nychka. Confidence intervals for nonparametric curve estimates: Toward more uniform point-wise coverage. *Journal of the American Statistical Association*, 96:233–246, 2001.

[19] J. M. Davis, D. W. Nychka, and B. A. Bailey. A comparison of the regional oxidant model with observed data. *Atmospheric Environment*, 34:2413–2423, 2000.

[20] S. C. Doney, D. M. Glover, M. Fuentes, and S. McCue. Mesoscale variability of Sea-viewing Wide Field-of-view Sensor (SeaWiFS) satellite ocean color: Global patterns and spatial scales. *Journal of Geophysical Research, Oceans*, 108(C2), 2003. 3024, doi:10.1029/2001JC000843.

[21] R. M. Errico, L. Fillion, D. W. Nychka, and Z.-Q. Lu. Some statistical considerations associated with the data assimilation of precipitation observations. *Quarterly Journal of the Royal Meteorological Society*, 126(562):339–360, 2000. Part A.

[22] Y. Fan and B. Whitcher. A wavelet solution to the spurious regression of fractionally differenced processes. *Applied Stochastic Models in Business and Industry*, 2003. Forthcoming.

[23] M. Fuentes, S. C. Doney, D. M. Glover, and S. J. McCue. Spatial structure of the SeaWiFS ocean color data in the North Atlantic Ocean. In L. M. Berliner, D. Nychka, and T. Hoar, editors, *Studies in the Atmospheric Sciences*, volume 144 of *Lecture Notes in Statistics*, pages 153–171, New York, 2000. Springer-Verlag.

[24] R. Furrer. Covariance estimation under spatial dependence. *Journal of Multivariate Analysis*, 2003. In review.

[25] R. Gençay, F. Selçuk, and B. Whitcher. *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics.* Academic Press, San Diego, September 2001. ISBN: 0-12-279670-5.

[26] R. Gençay, F. Selçuk, and B. Whitcher. Asymmetry of information flow between fx volatilities across time scales. *Econometrica*, 2002. In review.

[27] R. Gençay, F. Selçuk, and B. Whitcher. Robustness of systematic risk across time scales. *Journal of International Money & Finance*, 2002. In review.

[28] G. K. Grunwald and R. H. Jones. Markov models for time series with mixed distribution. *Environmetrics*, 11:327–339, 2000.

[29] T. J. Hoar. Bayes meets big iron. In *Proceedings of the 1999 Joint Statistical meetings*, Journal of the American Statistical Association, 1999.

[30] T. J. Hoar, R. F. Milliff, D. W. Nychka, C. K. Wikle, and L. M. Berliner. Wind realizations from a Bayesian hierarchical model: Computation and preparation for atmosphere-ocean research applications. *Journal of Computational and Graphical Statistics*, 2003. In review.

[31] C. Johns, D. Nychka, T. Kittel, and C. Daly. Infilling sparse records of spatial fields. *Journal of the American Statistical Association*, 2003. accepted.

[32] R. H. Jones and A. L. Jones. Testing skill in earthquake predictions. *Seismological Research Letters*, 74(6), 2003. In press.

[33] A. Kaplan, M. A. Cane, and Y. Kushnir. Reduced space approach to the optimal analysis interpolation of historical marine observations: Accomplishments, difficulties, and prospects. In *WMO Guide to the Applications of Marine Climatology*, volume WMO/TD-1081, Geneva, Switzerland, 2003. World Meteorological Organization. In press.

[34] R. W. Katz. Sir Gilbert Walker and a connection between El Niño and statistics. *Statistical Science*, 17:97–112, 2002.

[35] R. W. Katz. Stochastic modeling of hurricane damage. *Journal of Applied Meteorology*, 41:754–762, 2002.

[36] R. W. Katz. Techniques for estimating uncertainty in climate change scenarios and impact studies. *Climate Research*, 20:167–185, 2002.

[37] R. W. Katz, M. B. Parlange, and P. Naveau. Statistics of extremes in hydrology. *Advances in Water Resources*, 25:1287–1304, August-December 2002. doi:10.1016/S0309-1708(02)00056-8.

[38] R. W. Katz, M. B. Parlange, and C. Tebaldi. Stochastic modeling of the effects of large-scale circulation on daily weather in the southeastern U.S. *Climatic Change*, 2003. In press.

[39] T. Kittel, P. Thornton, J. A. Royle, and T. N. Chase. Climates of the Rocky Mountains: Historical and future patterns. In J. Baron, editor, *Rocky Mountain Futures: An Ecological Perspective.*, pages 59–82, Covelo, CA, 2002. Island Press.

[40] T. Kupper, J. D. Berset, R. Etter-Holzer, R. Furrer, and J. Tarradellas. Concentrations and specific loads of polycyclic musks in sewage sludge originating from a monitoring network in Switzerland. *Chemosphere*, 2003. accepted.

[41] J. Lee and R. B. Lund. Trends in extreme United States temperatures. *Journal of the American Statistical Association*, 2003. In review.

[42] T. C. M. Lee and H.-S. Oh. Automatic polynomial wavelet regression. *Statistics and Computing*, 2003. In review.

[43] T.-H. Li and H.-S. Oh. Wavelet spectrum and its characterization property for random processes. *IEEE Trans. on Information Theory*, 48:2922–2937, 2002.

[44] Z.-Q. Lu. Multivariate local polynomial fitting for Martingale nonlinear regression models. *Annals of the Instititute of Statistical Mathematics*, 51(4):691–706, December 1999.

[45] Z.-Q. Lu, L. M. Berliner, and C. Snyder. Experimental design for spatial and adaptive observations. In L. M. Berliner, D. Nychka, and T. Hoar, editors, *Studies in the Atmospheric Sciences*, volume 144 of *Lecture Notes in Statistics*, pages 65–78, New York, 2000. Springer-Verlag.

[46] R. Lund, L. Seymour, and K. Kafadar. Temperature trends in the United States. *Environmetrics*, 12:673–690, 2001.

[47] R. B. Lund and J. Reeves. Detection of undocumented changepoints – a revision of the two-phase regression model. *Journal of Climate*, 17:2547–2554, 2002.

[48] R. B. Lund, Q. Shao, and I. V. Basawa. Parsimonious periodic time series modeling. *Australian & New Zealand Journal of Statistics*. In review.

[49] T. Matsuo, A. D. Richmond, and D. W. Nychka. Modes of high-latitude electric field variability derived from DE-2 measurements: Empirical Orthogonal Function (EOF) analysis. *Geophysical Research Letters*, 29(7), 2002. 1107, doi:10.1029/2001GL014077.

[50] G. A. Meehl, C. Tebaldi, and D. W. Nychka. Changes in frost days in simulations of 21st century climate. *Climate Dynamics*, 2003. In review.

[51] G. A. Meehl, W. M. Washington, C. Ammann, J. M. Arblaster, T. M. L. Wigley, and C. Tebaldi. Combinations of natural and anthropogenic forcings and 20th century climate. *Journal of Climate*, 2003. In review.

[52] W. Meiring. Seasonal variation in stratospheric ozone levels, a functional data analysis study. In L. M. Berliner, D. Nychka, and T. Hoar, editors, *Studies in the Atmospheric Sciences*, volume 144 of *Lecture Notes in Statistics*, pages 79–96, New York, 2000. Springer-Verlag.

[53] W. Meiring. Ozonesonde mid-latitude stratospheric ozone variability, with emphasis on the quasi-biennial oscillation: A functional data analysis approach. *Journal of the American Statistical Association*, 2003. In review.

[54] J. F. B. Mitchell, D. J. Karoly, et al. Detection of climate change and attribution of causes. In J. Houghton et al., editors, *Climate Change 2001: The Scientific Basis.*, pages 695–738. Cambridge University Press, 2001. L. M. Berliner (Contributing Author).

[55] P. Naveau. Almost sure relative stability of the maximum of a stationary sequence. *Advances in Applied Probability*, 35(3):721–736, 2003.

[56] P. Naveau, C. Ammann, H.-S. Oh, and W. Guo. An automatic statistical methodology to extract pulse-like forcing factors in climatic time series: Application to volcanic events. In A. Robock, editor, *Volcanism and the Earth's Atmosphere*, Geophysical Monographs, 2003. In press.

[57] P. Naveau, P. Brockwell, and D. W. Nychka. Dependence structure of random wavelet coefficients in terms of cumulants. In A. Cohen, C. Rabut, and L. Schumaker, editors, *Curve and Surface Fitting*, pages 343–350, Nashville, TN, 1999. Vanderbilt University Press. ISBN 0-8265-1357-3.

[58] P. Naveau, R. Furrer, and P. Keckhut. The spatio-temporal influence of the vortex on Arctic total column ozone variability. In J. Mateu, D. Holland, and W. Gonzalez-Manteiga, editors, *The ISI International Conference on Environmental Statistics and Health*, pages 131–139. Universidade de Santiago de Compostela, 2003.

[59] P. Naveau and M. Genton. The general skewed Kalman filter. *Journal of Multivariate Statistics*, 2003. In review.

[60] P. Naveau, R. W. Katz, and M. Moncrieff. Extremes and climate: an introduction and a case study. *Notes de l'Institut Pierre Simon Laplace*, 11:4–22, 2001. ISSN 1626-8334.

[61] P. Naveau and M. Moncrieff. A statistical formulation of convective mass fluxes. *Quarterly Journal of the Royal Meteorological Society*, 129(592):2217–2233, July 2003. Part A.

[62] P. Naveau, M. Moncrieff, J.-I. Yano, and X. Wu. Exploratory statistical analysis of tropical oceanic convection using discrete wavelet transforms. In L. M. Berliner, D. Nychka, and T. Hoar, editors, *Studies in the Atmospheric Sciences*, volume 144 of *Lecture Notes in Statistics*, pages 117–132, New York, 2000. Springer-Verlag.

[63] P. Naveau and H.-S. Oh. Polynomial wavelet regression for images with irregular boundaries. *IEEE Transactions on Image Processing*, 2003. In press.

[64] P. Naveau and D. Paillard. Briefing note on statistical downscaling methodologies for WP3. In *The EC BIOCLIM Project (2000-2003) 5th Euratom Framework Programme*, 2002.

[65] P. Naveau, M. Vrac, M. Genton, A. Chedin, and E. Diday. Two statistical methods for improving the analysis of large climatic data sets: General skewed Kalman filters and distributions of distributions. In *Proceedings of The Fourth European Conference on Geostatistics for Environmental Applications*, Barcelona, 2002. geoENV2002.

[66] D. Nychka. Challenges in understanding the atmosphere. *Journal of the American Statistical Association*, 95:972–975, 2000.

[67] D. W. Nychka. Challenges in understanding the atmosphere. In A. Raftery, M. Tanner, and M. Wells, editors, *Statistics in the 21st Century*, pages 199–206, New York, NY, 2001. Chapman and Hall/CRC.

[68] D. W. Nychka and C. Tebaldi. Comment on "calculation of average, uncertainty range and reliability of regional climate changes from AOGCM simulations via the 'reliability ensemble averaging' (REA) method". *Journal of Climate*, 16(5):883–884, 2002.

[69] D. Nychka, C. K. Wikle, and J. A. Royle. Multiresolution models for nonstationary spatial covariance functions. *Statistical Modelling: An International Journal*, 2:315–331, 2002.

[70] H.-S. Oh, C. Ammann, P. Naveau, D. Nychka, and B. Otto-Bliesner. Multi-resolution time series analysis applied to solar irradiance and climate reconstructions. *Journal of Atmospheric and Solar-Terrestrial Physics*, 65:191–201, 2003.

[71] H.-S. Oh and T.-H. Li. Estimation of global temperature fields from scattered observations by a spherical-wavelet-based spatially adaptive method. *Journal of the Royal Statistical Society: Series B*, 2003. In press.

[72] H.-S. Oh, P. Naveau, and G. Lee. Polynomial boundary treatment for wavelet regression. *Biometrika*, 88:291–298, 2001.

[73] H.-S. Oh, D. W. Nychka, T. M. Brown, and P. Charbonneau. Period analysis of variable stars by robust smoothing. *Journal of the Royal Statistical Society*, In press, 2003. Series C (Applied Statistics).

[74] F. Parisi and R. Lund. Seasonality and return periods of land-falling Atlantic hurricanes. *Australian & New Zealand Journal of Statistics*, 42:271–282, 2000.

[75] M. B. Parlange and R. W. Katz. An extended version of the Richardson model for simulating daily weather variables. *Journal of Applied Meteorology*, 39:610–622, 2000.

[76] J. A. Royle. Multivariate spatial models. In L. M. Berliner, D. Nychka, and T. Hoar, editors, *Studies in the Atmospheric Sciences*, volume 144 of *Lecture Notes In Statistics*, pages 23–44, New York, 2000. Springer-Verlag.

[77] B. D. Santer, T. M. L. Wigley, J. S. Boyle, D. J. Gaffen, J. J. Hnilo, D. W. Nychka, D. E. Parker, and K. E. Taylor. Statistical significance of trends and trend differences in layer-average atmospheric temperature time series. *Journal of Geophysical Research, Atmosphere*, 105:7337–7356, 2000.

[78] E. E. Small, L. C. Sloan, and D. W. Nychka. Changes in surface air temperature caused by desiccation of the Aral sea. *Journal of Climate*, 14:284–299, 2001.

[79] G. Sneddon. A statistical perspective on data assimilation in numerical models. In L. M. Berliner, D. Nychka, and T. Hoar, editors, *Studies in the Atmospheric Sciences*, volume 144 of *Lecture Notes in Statistics*, pages 7–21, New York, 2000. Springer-Verlag.

[80] S. B. Streett, R. A. Davis, and W. T. M. Dunsmuir. Maximum likelihood estimation for an observation driven model for Poisson counts. *Methodology and Computing in Applied Probability*, 2003. In review.

[81] S. B. Streett, R. A. Davis, and W. T. M. Dunsmuir. Observation driven models for Poisson counts. *Biometrika*, 2003. To appear.

[82] C. Tebaldi. Looking far back vs. looking around enough: operational weather forecasting by "spatial" composition of recent observations. In *Proceedings of the 16th Conference on Probability and Statistics in the Atmospheric Sciences*, AMS Annual Meeting, Orlando, FL, January 2002.

[83] C. Tebaldi, G. Brantstator, and D. W. Nychka. Looking for nonlinearities in the large scale dynamics of the atmosphere. In *Proceedings of the First SIAM International Conference on Data Mining*, Chicago, IL, April 2001.

[84] C. Tebaldi, D. W. Nychka, B. Brown, and B. Sharman. Predicting clear-air turbulence. In L. M. Berliner, D. Nychka, and T. Hoar, editors, *Studies in the Atmospheric Sciences*, volume 144 of *Lecture Notes in Statistics*, pages 133–151, New York, 2000. Springer-Verlag.

[85] C. Tebaldi, D. W. Nychka, B. G. Brown, and B. Sharman. Flexible discriminant techniques for forecasting clear-air turbulence. *Environmetrics*, 13(8):859–878, 2002.

[86] C. Tebaldi, M. West, and A. F. Karr. Statistical analyses of freeway traffic flows. *Journal of Forecasting*, 21(2):39–68, 2002.

[87] M. Vrac, E. Diday, P. Naveau, and A. Chedin. Mixture of distributions of distributions. In *Proceedings de la Société Francophone de Classification*, 2001.

[88] B. Whitcher. Wavelet-based estimation for seasonal long-memory processes. *Technometrics*, 2003. In review.

[89] B. Whitcher, S. D. Byers, P. Guttorp, and D. Percival. Testing for homogeneity of variance in time series: Long memory, wavelets and the Nile river. *Water Resources Research*, 38(5):doi: 10.1029/2001WR000509, 2002.

[90] B. Whitcher, J. B. Weiss, D. W. Nychka, and T. J. Hoar. Stochastic multi-resolution models for turbulence. In M. Akritis and D. Politis, editors, *Recent Advances and Trends in Nonparametric Statistics*, volume forthcoming, 2003.

[91] C. K. Wikle. Hierarchical space-time dynamic models. In L. M. Berliner, D. Nychka, and T. Hoar, editors, *Studies in the Atmospheric Sciences*, volume 144 of *Lecture Notes in Statistics*, pages 45–64, New York, 2000. Springer-Verlag.

[92] C. K. Wikle. Hierarchical models in environmental science. *International Statistical Review*, 71:181–199, 2003.

[93] C. K. Wikle and L. M. Berliner. Combining information across spatial scales. *Technometrics*, In review, 2002.

[94] C. K. Wikle, L. M. Berliner, and R. F. Milliff. Hierarchical bayesian approach to boundary value problems with stochastic boundary conditions. *Monthly Weather Review*, 131:1051–1062, 2003.

[95] C. K. Wikle and N. A. Cressie. Space-time statistical modeling of environmental data. In H. T. Mowrer and R. G. Congalton, editors, *Quantifying Spatial Uncertainty in Natural Resources: Theory and Applications for GIS and Remote Sensing*, pages 213–236, Chelsea, MI, 2000. Ann Arbor Press.

[96] C. K. Wikle, D. Nychka, R. F. Milliff, and L. M. Berliner. Spatio-temporal hierarchical Bayesian blending of tropical ocean surface wind data. *Journal of the American Statistical Society*, 96:382–397, 2001.

[97] C. K. Wikle and J. A. Royle. Spatial statistical modeling in biology. In *Encyclopedia of Life Support Systems*. EOLSS Publishers Co. Ltd., 2002.

GSP manuscripts

[98] E. Bellone, W. Collins, and X. Wu. Statistical models for cloud overlap and amount derived from a cloud-resolving model. 2003. Manuscript.

[99] T. Bengtsson, R. F. Milliff, R. H. Jones, and D. W. Nychka. A state-space model for ocean drifter motions dominated by inertial oscillations. 2003. Manuscript.

[100] B. Das. *Global Covariance Modeling: A Deformation Approach to Anisotropy.* Ph.D. dissertation, University of Washington, Seattle, WA, 2000. Department of Statistics.

[101] R. Furrer, T. Bengtsson, and J. Cavanaugh. Evaluation of forecast and analysis covariance in Kalman filter variants. 2003. Manuscript.

[102] R. Gençay, F. Selçuk, and B. Whitcher. Scale invariance and time-frequency analysis with wavelet hidden markov trees. 2002. Manuscript.

[103] S. Jain and B. Whitcher. Wavelet analysis of climate-related flood variability. 2003. Manuscript.

[104] T. C. M. Lee and H.-S. Oh. Robust penalized splines. 2003. Manuscript.

[105] P. Naveau, C. Ammann, and H.-S. Oh. A statistical methodology to extract volcanic signals in climatic time series. 2002. Manuscript.

[106] H.-S. Oh and T. C. M. Lee. Hybrid wavelet shrinkage: a method for boundary correction. 2003. Manuscript.

[107] H.-S. Oh, T. C. M. Lee, and D. W. Nychka. Wavelet thresholding for data with outliers. 2003. Manuscript.

[108] H.-S. Oh and D. W. Nychka. A robust selection method of the smoothing parameter for smoothing splines. 2003. Manuscript.

[109] S. B. Streett, R. A. Davis, T. Rydberg, and N. Shephard. The Cbin model for counts: testing for common features in the speed of trading, quote changes, limit and market order arrivals. 2003. Manuscript.

[110] S. B. Streett, D. W. Nychka, and R. A. Davis. Observation driven models for precipitation occurrence. 2003. Manuscript.

[111] C. Tebaldi. Influence of large scale circulation measures on precipitation at local stations in the south-east U.S.. NRCSE Technical Reports Series 55, National Research Center for Statistics and the Environment, 2000.

[112] C. Tebaldi, D. W. Nychka, and L. O.Mearns. An exercise in simple pattern scaling: how regional aggregation impacts uncertainty in the scaled signals. 2003. Manuscript.

[113] C. Tebaldi, R. L. Smith, D. W. Nychka, and L. O. Mearns. Quantifying uncertainty in projections of regional climate change: a Bayesian approach to the analysis of multimodel ensembles. 2003. Manuscript.

[114] B. Whitcher and P. Craigmile. Multivariate spectral analysis using Hilbert wavelet pairs. 2003. Manuscript.

[115] B. Whitcher, J. B. Weiss, D. W. Nychka, and T. J. Hoar. A multiresolution census algorithm for calculating vortex statistics. 2003. Manuscript.

[116] B. Whitcher and C. K. Wikle. Spectral analysis of atmospheric time series using Hilbert wavelet pairs. 2002. Manuscript.

Other references

[117] J. L. Anderson and S. L. Anderson. A Monte-Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127:2741–2758, 1999.

[118] M. Bayarri, J. Berger, D. Higdon, M. Kennedy, A. Kottas, R. Paulo, J. Sacks, J. Cafeo, J. Cavendish, C. Lin, and J. Tu. A framework for validation of computer models. In D. Pace and S. Stevenson, editors, *Proceedings of the Workshop on Foundations for V&V in the 21st Century*. International Society for Modeling and Simulation, 2002.

[119] K. Chou, S. Golden, and A. Willsky. Multiresolution stochastic models, data fusion, and wavelet transforms,. *Signal Processing*, 34(3):257–282, 1993.

[120] M. V. Codrescu, T. J. Fuller-Rowell, J. C. Foster, J. M. Holt. and S. J. Cariglia. Electric field variability associated with the Millstone Hill electric field model. *Journal of Geophysical Research*, 105(A3):5265–5274, 2000.

[121] S. Cole. *An Introduction to Statistical Modeling of Extreme Values.* Springer, New York, NY, 2001.

[122] N. Cressie. *Statistics for Spatial Data, revised edition.* Wiley, New York, NY, 1993.

[123] N. Cressie and H. C. Huang. Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Stat. Assoc.* 94:1330–1340, 1999.

[124] C. Currin, T. Mitchell, M. Morris, and D. Ylvisaker. Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association (Theory and Methods)*, 86(416):953–963, 1991.

[125] A. Doucet, N. Freitas, and N. Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Series in Statistics. Springer-Verlag, New York, NY, 2001.

[126] D. S. Wilks. Simultaneous stochastic simulation of daily precipitation, temperature and solar radiation at multiple sites in complex terrain. *Agricultural and Forest Meteorology*, 96(1):85–101, 1999.

[127] G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99 C5(10):143–162, 1994.

[128] G. Gaspari and S. E. Cohn. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Met. Soc.*, 125: 723–757, 1999.

[129] Geophysical Statistics Project. A project selection checklist. In `www.cgd.ucar.edu/stats/projectsel.shtml`

[130] G. James, T. Hastie, and C. Sugar. Principal component models for sparse functional data. *Biometrika*, 87(3):587–602, 2000.

[131] A. Gijsman, S. Jagtap, and J. Jones. Wading through a swamp of complete confusion: how to choose a method for estimating soil water retention parameters for crop models. *European Journal of Agronomy*, 18:75–105, 2002.

[132] P. J. Green and D. J. Murdoch. Exact sampling for Bayesian inference: towards general purpose algorithms (with discussion). In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics 6*, pages 301–321. Oxford University Press, 1999.

[133] H.-C. Huang, Cressie N., and Gabrosek J. Fast, resolution-consistent spatial prediction of global processes from satellite data. *Journal of Computational & Graphical Statistics*, 11(1):63–88, 2002.

[134] T. M. Hamill, C. Snyder, and R. E. Morss. A comparison of probabilistic forecasts from bred, singular vector and perturbed observation ensembles. *Monthly Weather Review*, 128:1835–1851, 2000.

[135] G. Hegerl, F. Zwiers, P. Stott, and V. Kharin. Detectability of anthropogenic changes in temperature and precipitation extremes. *Journal of Climate*, 2003. Submitted.

[136] D. Higdon, J. Swall, and J. Kern. Non-stationary spatial modeling. In *Bayesian Statistics 6*, New York, NY, 1999. Oxford University Press.

[137] T. Hinke, J. Rushing, H. S. Ranganath, and S. J. Graves. Techniques and experience in mining remotely sensed satellite data. *Artificial Intelligence Review*, 14:503–531, 2000.

[138] K. L. Hodges. A general method for tracking analysis and its applications to meteorological data. *Monthly Weather Review*, 122:2573–2586, 1994.

[139] P. L. Houtekamer and H. L. Mitchell. Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126:796–811, 1998.

[140] M. Kennedy and A. O'Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87:1–13, 2000.

[141] A. J. Majda, I. Timofeyev and E. Vanden Eijnden. A mathematical framework for stochastic climate models. to appear in *Comm. Pure Applied Math.*

[142] Y. Mallet, D. Coomans, J. Kautsky, and O. De Vel. Classification using adaptive wavelets for feature extraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:1058–1066, 1997.

[143] T. Matsuo, A. D.Richmond, and K. Hensel. High-latitude ionospheric electric field variability and electric potential derived from DE-2 plasma drift measurements: dependence on IMF and dipole tilt. *Journal of Geophysical Research*, 108:10.1029/2002JA009429, 2003.

[144] L. Mearns, M. Hulme, T. Carter, R. Leemans, M. Lal, P. Whetton, L. Hay, R. N. Jones, R. Katz, T. Kittel, J. Smith, and R. Wilby. Climate scenario development. In J. T. H. et al., editor, *Climate Change 2001: The Scientific Basis*, pages 739–768, Providence, RI, 2001. Cambridge University Press.

[145] L. Mearns and C. Rosenzweig. Formation of climate change scenarios incorporating changes in daily climate variability and application to crop models. In W. Howe and A. Henderson-Sellers, editors, *Assessing Climate Change: The Story of the Model Evaluation Consortium for Climate Assessment*, volume Chap. 15. Harwood Academic Publishers, 1997.

[146] L. O. Mearns. Methodological issues concerning the determination of the effects of changing climatic variability on crop yields. In *Climate Change and International Impacts*, volume Special Publication No. 59, pages 123–146, Madison, WI, 1995. American Society of Agronomy:ASA.

[147] L. O. Mearns, C. Rosenzweig, and R. Goldberg. The effect of changes in daily and interannual climatic variability on CERES-wheat: A sensitivity study. *Climatic Change*, 32:257–292, 1996.

[148] M. Moncrieff. Organized convective systems: Archetipa, dynamical models, mass and momentum flux theory, and parameterization. *Quarterly Journal of the Royal Meteorological Society*, 118:819–850, 1992.

[149] NSF Site Review Panel. Report of the NSF site visit panel to the geophysical Statistics Program at the National Center for Atmospheric Research. Technical report, `http://www.cgd.ucar.edu/stats/pub/Panel03/NSFsite.report.pdf`, 2002.

[150] T. J. Osborn and M. Hulme. Development of a relationship between station and gridbox rainday frequencies for climate model validation. *Journal of Climate*, 10: 1885–1908, 1997.

[151] S. Pacala, G. Hurtt, D. Baker, P. Peylin, R. Houghton, R. Birdsey, L. Heath, E. Sundquist, R. Stallard, P. Ciais, P. Moorcroft, J. Caspersen, E. Shevliakova, B. Moore, G. Kohlmaier, E. Holland, M. Gloor, M. Harmon, S. Fan, J. Sarmiento, C. Goodale, D. Schimel, and C. Field. Consistent land- and atomosphere-based U.S. carbon sink estimates. *Science*, 292(5525):2316–2320, 2001.

[152] Y. Pachepsky, W. Rawls, and D. Timlin. The current status of pedotransfer functions: their accuracy, reliability and utility in field- and region-scale modeling. In D. Corwin and T. Loague, K.M.and Ellsworth, editors, *Assessment of Non-Point Source Pollution in the Vadose Zone*, volume Geophysical Monograph 108, pages 223–224, Washington, DC, 1999. American Geophysical Union.

[153] T. N. Palmer and J. Risnen. Quantifying the risk of extreme seasonal precipitation events in a changing climate. *Nature*, 415:512–514, 2002.

[154] T. Palmer. A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate predictions *Quarterly Journal of the Royal Meteorological Society*, 127: 279–304, 2001.

[155] A. D. Richmond and Y. Kamide. Mapping electrodynamic features of the high-latitude ionosphere from localized observations: Technique. *Journal of Geophysical Research*, 93(A6): 5741–5759, 1988.

[156] D. Ruppert, M. P. Wand, and R. J. Carroll. *Semiparametric Regression.* Cambridge University Press, 2003.

[157] J. Sacks, W. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4:409–435, 1989.

[158] I. A. Sadarjoen and F. H. Post. Detection, quantification, and tracking of vortices using streamline geometry. *Computers and Graphics*, 24:333–341, 2000.

[159] J. Serra. *Image Analysis and Mathematical Morphology.* Academic Press, 1982.

[160] A. P. Sokolov and P. H. Stone. A flexible climate model for use in integrated assessments. *Climate Dynamics*, 14:291–303, 1998.

[161] P. Smyth, K. Ide and M. Ghil. Multiple Regimes in Northern Hemisphere Height Fields via Mixture Model Clustering *Journal of the Atmospheric Sciences*,56:3704–3723, 1999.

[162] M. Stein. Fast and exact simulation of fractional Brownian surfaces. *Journal of Computational & Graphical Statistics*, 11(3):587–599, 2002.

[163] M. Stein. *Personal Communication*, 2003

[164] M. L. Stein. *Interpolation of Spatial Data.* Springer Series in Statistics. Springer-Verlag, New York, 1999.

[165] J. Sun and N. A. Crook. Dynamical and microphysical retrieval from doppler radar observations using a cloud model and its adjoint. Part I: Model development and simulated data experiments. *Journal of the Atmospheric Sciences*, 54:1642–1661, 1997.

[166] J. B. Weiss and J. C. McWilliams. Temporal scaling behavior of decaying two-dimensional turbulence. *Physics of Fluids A*, 5:608–621, 1993.

[167] C. K. Wikle, L. M. Berliner and N. A. Cressie. Hierarchical Bayesian space–time models. *Environmental and Ecological Statistics*, 5:117–154, 1998.

[168] A. T. A. Wood and G. Chan. Simulation of stationary Gaussian process in [0; 1] d. *Journal Computational Graphical Statistics*,3:409–432, 1994.