samsi
NSF·Duke·NCSU·UNC

Climate Program Remote Sensing Workshop
February 12-14, 2018

Posters

Jenny Brynjarsdottir, Case-Western Reserve University
Jonathan Hobbs, **Amy Braverman**, and **Lukas Mandrake**, JPL

"Optimal Estimation versus MCMC for CO2 retrievals"

The Orbiting Carbon Observatory-2 (OCO-2) collects infrared spectra from which atmospheric properties are retrieved. OCO-2 operational data processing uses Optimal Estimation (OE), a state-of-the-art approach to inference of atmospheric properties from satellite measurements. One of the main advantages of the OE approach is computational efficiency, but it only characterizes the first two moments of the posterior distribution of interest. Here we obtain samples from the posterior using a Markov Chain Monte Carlo (MCMC) algorithm, and compare this empirical estimate of the true posterior to the OE results. We focus on 600 simulated soundings that represent the variability of physical conditions encountered by OCO-2 between November 2014 and January 2016. We treat the two retrieval methods as ensemble and density probabilistic forecasts, where the MCMC yields an ensemble from the posterior and the OE retrieval result provide the first two moments of normal distribution. To compare these methods we apply both univariate and multivariate diagnostic tools and proper scoring rules. The general impression from our study is that when compared to MCMC, the OE retrieval performs reasonably well for the main quantity of interest, the column averaged CO2 concentration XCO2, but not for the full state vector X which includes a profile of CO2 concentrations over 20 pressure levels, as well as several other atmospheric properties.

Sandy Burden
University of Wollongong

"Functional ANOVA comparison of CO2 Flux Predictions"

There are presently at least nine different flux-inversion (FI) models that produce spatially detailed CO2 flux field predictions based on XCO2 retrievals obtained from the OCO-2 Mission. This ensemble of predictions is a valuable resource for understanding FI models and for investigating and reducing prediction uncertainty. However, summarizing and evaluating the ensemble is not straightforward. FI models are frequently based on the same, or similar, sources of information, and hence their output may not be independent. For inference it is crucial to account for this dependence to avoid underestimating prediction uncertainty. This poster demonstrates the use of Functional ANOVA for comparing predicted spatial fields from multiple FI models, taking into consideration shared assumptions, parameters and/or data. Since the predictions are modeled as observed realizations from an underlying smooth random field with a common basis, the approach also reduces the amount of data required for analysis and facilitates comparison of multiple, potentially distributed, spatiotemporal fields.

Isabelle Grenier and **Bruno Sanso**
University of California, Santa Cruz

"Modelling Precipitation Levels in California due to Atmospheric Rivers"

Atmospheric Rivers are elongated regions in the atmosphere that transport water vapor out of the tropics. In California, these are responsible for the heavy rainfalls we observe during the winter. Due to climate change, we expect the number and the intensity of atmospheric rivers to increase. The goal of our research is to model the precipitation levels due to atmospheric rivers to assess the impact of climate change on the water supply in California. We _rst developed a low resolution model which aggregates precipitation over California at the monthly level. The covariates include the number of atmospheric rivers observed, a seasonality factor and the maximum integrated water vapor transport recorded during the month. The average prediction error for the winter months is between 10% and 30%. However, the accuracy is much lower for months out of the peak rain season. Our future work will focus on increasing the time and spatial resolution to increase the predictive accuracy.

Otto Lamminpaa
Finnish Meteorological Institute

"Dimension Reduction for Remote Sensing of Atmospheric Methane Profiles"

Determining the density profiles of trace gases from measured absorption spectra is an ill-posed inverse problem, in which the measurement typically contains limited amount of information. We consider ground based Fourier transform infrared spectrometer (FTIR, part of TCCON network) solar absorption measurements from FMI Arctic Research Centre, to invert atmospheric methane (CH4) density profiles. This problem is computationally costly, which motivates the development of a dimension reduction scheme. In this study, we use Bayesian framework adaptive MCMC to characterize the full posterior distribution of the solution and the related uncertainties. As a main result, we present a dimension reduction method based on splitting the problem into informative and non-informative subspaces.

Huikyo Lee, **Krysztof Gorski**, and **Brian Wilson**
Jet Propulsion Laboratory

"Multi-Resolution Investigation of Climate Models using High-End Computing Resources: a Parallel Version of the Regional Climate Model Evaluation System Powered by Healpix"

While systematic, multi-model experimentation and evaluation have been undertaken for years (e.g., the CMIP5), the development and application of infrastructure for systematic, observation based evaluations of spatial patterns in key climate variables simulated with various spatial resolutions are less mature, owing in part to the needed advances and synergies in both climate and data sciences. One of the main challenges in using existing analysis tools is to carry out the multiresolution investigation of climate models. Given this, the principal science objective of my work is to provide quantitative and robust evaluations of spatial patterns simulated by climate models across multiple scales: comparison of spatial features at coarse (e.g. 100 km) and _ne scales (e.g. 1-10 km) separately between observations and models. Model evaluation also critically rests on data science and technology infrastructure, including access to datasets, storage, and computation. The vast amounts of model and observational data at high resolution required by the model evaluation process have to be brought together in a high-performance, service-based cyberinfrastructure to support large-scale Earth science analytics{ a challenge that the current study will address. I introduce the Jet Propulsion Laboratory's Regional Climate Model Evaluation System (RCMES) powered by the Hierarchical Equal Area isoLatitude Pixelization (HEALPix) as a web-based service for evaluating climate models at different spatial resolutions. The unique capabilities of HEALPix include open-source libraries to facilitate the handling and distribution of massive datasets at different resolutions using parallel computing, and fast and robust analysis of spatial patterns from observational and model datasets regridded into HEALPix pixels, which have been widely used by astronomers and planetary scientists. Both RCMES and HEALPix are open-source software toolkits with a broad user base. We will maximize the utility of RCMES optimized with its parallel processing capabilities as a service through high-end computing (HEC) resources. Our preliminary result indicates that RCMES enhanced with HEALPix could contribute to the ESGF computing application program interface (API) in order to enhance the visibility and utilization of NASA satellite observations in CMIP6.

Pulong Ma and **Emily Kang**
University of Cincinnati

"Dynamic Fused Gaussian Process for Massive Sea Surface Temperature Data from MODIS and AMSR-E Instruments"

Sea surface temperature (SST) is a key climate and weather measurement, which plays a crucial role in understanding climate systems. Massive amount of SST datasets can be collected from satellite instruments each day with the advance of new remote-sensing technologies. However, these data are often sparse, irregular, and noisy. In addition, different instruments will produce SST data with incompatible supports and distinct error characteristics. For instance, the Moderate Resolution Imaging Spectroradiometer (MODIS) is able to produce SST data at 9km spatial resolution each day, whose quality is subject to weather conditions such as cloud; the Advanced Microwave Scanning Radiometer-Earth Observing System (AMSR-E) is able to produce SST data at 25km spatial resolution each day, whose quality is subject to radio frequency interference. Statistical methods for combining different sources of remote-sensing data will give much more accurate uncertainty analysis. In this article, we propose a Dynamic Fused Gaussian Process (DFGP) model, which extends the Fused Gaussian Process (FGP) in Ma and Kang (2017) to a spatio-temporal model that enables fast statistical inference such as smoothing and filtering for massive datasets. The change-of-support problem is also explicitly addressed in DFGP when statistical inference is made based on different sources of data whose spatial resolutions are incompatible. We also develop a stochastic Expectation-Maximization (EM) algorithm to allow fast parameter estimation in a distributed computing environment. The proposed DFGP is applied to a total of 3.5 million SST datasets in a one-week period in tropical Pacific Ocean area from MODIS and AMSR-E instruments.

Pulong Ma and **Emily Kang**, University of Cincinnati; and
Amy Braverman and **Hai Nguyen**, Jet Propulsion Laboratory

"Spatial Statistical Downscaling for Constructing High-Resolution Nature Runs in Global Observing System Simulation Experiments"

Observing system simulation experiments (OSSEs) have been widely used as a rigorous and cost effective way to guide development of new observing systems, and to evaluate the performance of new data assimilation algorithms. Nature runs (NRs), which are outputs from deterministic models, play an essential role in building OSSE systems for global atmospheric processes because they are used both to create synthetic observations at high spatial resolution, and to represent the \true" atmosphere against which the forecasts are verified. However, most NRs are generated at resolutions coarser than actual observations. Here, we propose a principled statistical downscaling framework to construct high-resolution NRs via conditional simulation from coarse-resolution numerical model output. We use nonstationary spatial covariance function models that have basis function representations. This approach not only explicitly addresses the change-of-support problem, but also allows fast computation with large volumes of numerical model output. We also propose a data-driven algorithm to select the required basis functions adaptively, in order to increase the exibility of our nonstationary covariance function models. In this article we demonstrate these techniques by downscaling a coarse-resolution physical NR at a native resolution of 1-degree latitude by 1.25-degree longitude of global surface CO2 concentrations to 655,362 equal-area hexagons.

Florian Schafer
Caltech

“Compression, Inversion, and Approximate PCA of Dense Kernel Matrices at Near-Linear Computational Complexity”

Many popular methods in machine learning, statistics, and uncertainty quantification rely on priors given by smooth Gaussian processes, like those obtained from the Matern covariance functions. Furthermore, many physical systems are described in terms of elliptic partial differential equations. Therefore, implicitely or explicitely, numerical simulation of these systems requires an efficient numerical representation of the corresponding Green's operator. The resulting kernel matrices are typically dense, leading to (often prohibitive) O (N2) or O (N3) computational complexity. In this work, we prove rigorously that the dense N_N kernel matrices obtained from elliptic boundary value problems and measurement points distributed approximately uniformly in a d-dimensional domain can be Cholesky factorised to accuracy _ in computational complexity O (N log2(N) log2d (N=_) in time and O (N log(N) logd (N=_) in space. For the closely related Matern covariances we observe very good results in practise, even for parameters corresponding to non-integer order equations. As a byproduct, we obtain a sparse PCA with near-optimal low-rank approximation property and a fast solver for elliptic PDE. We emphasise that our algorithm requires no analytic expression for the covariance function. Our work connects the probabilistic interpretation of the Cholesky factorisation, the screening effect in spatial statistics, and numerical homogenisation. In particular, results from the game theoretic approach to numerical analysis (“Gamblets") allow us obtain rigorous error estimates.

Massimo Stella, Fondazione Bruno Kessler; **Sanja Selakovic**, University of Utrecht; **Alberto Antonioni**, Universidad Carlos III de Madrid; and **Cecilia S. Andreazzi**, Fiocruz Foundation

“Multi-Layer Ecological Data Processing for Modelling Pathogen Spread: the Ecomultiplex Model”

Multiple routes of transmission for many diseases are investigated separately despite their potential interplay. As a unifying framework for understanding parasite spread through interdependent transmission paths, we present the "ecomultiplex" model, where multi-layer ecological data about predator-prey and parasite-host interactions are processed, combined and represented as a spatially embedded multiplex network. We adopt this framework for designing and testing potential control strategies for parasite spread in two empirical host communities. We base our simulations on the distributed spread of the parasite between multiplex layers. Our results show that the ecomultiplex network model is an efficient and low data-demanding method for identifying which species

promote parasite spread, offering mechanistic interpretation of preliminary empirical findings and opening new insights in designing efficient control strategies for parasite containment.

Joaquim Teixeira, **Jon Hobbs**, and **Michael Gunson**
Jet Propulsion Laboratory

“The OCO-2 Retrieval Algorithm: Sensitivity to Choice of Prior Covariances”

We present the results of an investigation into the sensitivity of the OCO-2 retrieval algorithm, to choices made in setting the retrieval's prior covariance matrix, using a Monte Carlo framework. The OCO-2 retrieval algorithm is an implementation of Bayes' Rule, and the prior covariance weights the cost function towards the prior mean. We wish to understand the effect of different prior covariance matrices on the retrieved CO2 profile. After constructing a set of alternative prior covariances (by manipulating lag correlation over the elements the CO2 profile vector, the variance of the total column average, and the level-by-level variances) we run Monte Carlo simulations with these alternatives across a set of marginal distributions chosen to represent different geophysical conditions. We observe that the choice of the prior covariance matrix can have a substantial effect on the retrieved CO2 profile, while leaving the total column average unchanged. These effects are invariant across different choices of marginal distributions used to generate synthetic \true" state vectors. We see that the choice of lag correlation in the CO2 profile, and standard deviation are the main drivers of these effects.

Michael Turmon, **Jonathan Hobbs**, **JT Reager**, **C. David**, **J. Famiglietti**
Jet Propulsion Laboratory

“Uncertainty Propagation for a Large Scale Hydrological Routing Model”

Hydrological routing models use river connectivity information to propagate the localized lateral inflows of surface and subsurface water runoff into downstream flows. The resulting modeled flows can be used for planning and risk analysis, which has motivated the determination of standard errors for flows. We describe computational tradeoffs among several approaches for determination of streamflow uncertainties, which generally correspond to different assumptions about the spatial/temporal covariance of inflows from runoff. We introduce a "reach random effects" model to account for large-scale error correlation, as may be caused by spatially-correlated errors in precipitation forcing. We describe implementation of uncertainty propagation using RAPID (David et al. 2011) applied over the 650,000 reaches of the Western Contiguous United States covered by the NHDPlus network. Finally, we observe that new space missions should provide novel remote-sensing observations of flows at sparsely-sampled points in the river network. We use the accessibility of the full space-time flow covariance to understand the constraints on network flows offered by these new observations.