Preview of Award 1451954 - Final Project Report

Cover

Federal Agency and Organization Element to Which Report is Submitted:	4900
Federal Grant or Other Identifying Number Assigned by Agency:	1451954
Project Title:	EAGER: Collaborative Research: Learning Relations between Extreme Weather Events and Planet-Wide Environmental Trends
PD/PI Name:	Claire Monteleoni, Principal Investigator
Recipient Organization:	George Washington University
Project/Grant Period:	09/01/2014 - 08/31/2017
Reporting Period:	09/01/2016 - 08/31/2017
Submitting Official (if other than PD\PI):	Claire Monteleoni Principal Investigator
Submission Date:	12/01/2017
Signature of Submitting Official (signature shall be submitted in accordance with agency specific instructions)	Claire Monteleoni

Accomplishments

* What are the major goals of the project?

Our project is motivated by the problem of understanding and predicting climate extremes, especially at regional scales, which is currently a Grand Challenge in climate science (World Climate Research Programme, 2013) In particular, we are interested in improving our understanding and prediction of extreme events, learning relations between climate change trends and extreme events, and developing machine learning algorithms to do so.

* What was accomplished under these goals (you must provide information for at least one of the 4 categories below)?

Major Activities:

(1) Modeling tails or extremes of response variables of interest (such as precipitation) can be done in a couple of different ways: (A1) focusing on the tails explicitly (i.e., max or min over samples), or (A2) focusing on high or low quantiles. When working with finite samples, advances on (A1) has continued to stay challenging due to technical reasons. Hence we focused on making advances on (A2) with finite samples based on the formalism of quantile regression. Quantile regression aims at modeling the conditional median and quantiles of a response variable given certain predictor variables. In several real world settings including climate sciences, one needs to consider the problem of quantile regression in high dimensions where the number of predictor variables is much higher than the number of samples available for parameter estimation. We investigated such a high-dimensional quantile regression problem (Sivakumar and Banerjee, ICML 2017), illustrated the key geometric and statistical aspects of the problem under the assumption that the parameter vector is structured, e.g., sparse, group sparse, etc., and characterized the estimation error of such models.

(2) We investigated a novel approach for using regularized regression for finding a predictive relation between a response, e.g., Texas heat waves, and a field of other variables, e.g., sea surface temperature (DelSole and Banerjee, J. Climate, 2017). Prior work on such modeling ran into two key issues: (a) a direct application of a regularized regression model like LASSO or ridge regression in completely unaware of the strong spatial correlation in the field of covariates and usually leads to uninterpretable non-spatial variable selection/weighting, and (b) the amount of training samples available is rather limited to learn a stable model with skill. In our work, we address both of these concerns. First, our regularized regression is not based on the native covariates but based on a spectral representation which automatically captures spatial smoothness. Second, the training was done based on climate model outputs, so the training data is quite large and evaluated on real observations where the model was illustrated to have strong skill.

(3) Predicting changes in climate extremes requires predicting climate change itself. In this period, we have made advancements in predicting natural variability. Specifically, we showed that models without interactive ocean circulations could predict certain temperature patterns skillfully on multi-year time scales (Srivastava and DelSole 2017). This finding has important implications for understanding the role of ocean dynamics in decadal predictability. Both accomplishments exploited regularization techniques developed in other aspects of this project.

(4) We provided an algorithm based on a hidden Markov random field (MRF), whose structure is a geospatial lattice over the globe, to improve climate model ensemble predictions by sequentially incorporating climate observation data with climate model simulation output, while simultaneously learning the level of both temporal and spatial non-stationarity.

(5) The project also advanced the understanding of light-weight machine learning algorithms for processing large data sets. We analyzed the scalability of the widely-used k-means algorithm to large data sets, such as those used in our study of climate extremes. To do so, we turned to stochastic variants of k-means, e.g. online k-means and mini-batch k- means variants, and analyzed their convergence.

Specific Objectives:

Significant Results:

Key outcomes or Other achievements:

*** What opportunities for training and professional development has the project provided?**

The project supported training of Abhishek Srivastava, PhD student at GMU.

The project provided training for Cheng Tang, PhD student at GWU.

The project provided training for Vidyashankar Sivakumar, PhD student at UMN.

*** How have the results been disseminated to communities of interest?**

The work on “Statistical Seasonal Prediction Based on Regularized Regression” has been published in the Journal of Climate (DelSole and Banerjee, 2016).

The work on “Decadal Predictability Without Ocean Dynamics” has been published in the Proceedings of the National Academies (Srivastava and DelSole, 2017).

The work on “High Dimensional Structured Quantile Regression” has been published in the International Conference on Machine Learning (Sivakumar and Banerjee, 2017).

Papers resulting from the project were also published in the Conference on Artificial Intelligence and Statistics proceedings (Tang and Monteleoni, 2017), the Ecology Law Quarterly (Glicksman, Markell, and Monteleoni, 2017), and a book, “Large-Scale Machine Learning in the Earth Sciences” (McQuade and Monteleoni, 2017).

T. DelSole presented research supported by this project as an invited speaker at (1) the Workshop on Atlantic Climate Variability: Dynamics, Prediction and Hurricane Risk, Columbia U., NY, (2) the Joint Statistical Meeting, Baltimore, MD, and (3) Fourth Sante Fe Conference on Global and Regional Climate Change, Sante Fe, NM, (see products list for more details).

The project helped support public presentations by T. DelSole on the Science of Climate Change at (1) Greenspring Retirement Community, Springfield, VA, (2) Climate Science Roundtable (organized by Sierra Club and Science Museum of Virginia), (3) The Fairfax 100, Fairfax, VA.

The Project helped support C. Monteleoni’s trip to present a talk on the work on the project, “Spatiotemporal online learning with expert advice, with applications to climate science and finance, “ at the Workshop on Machine Learning for Spatiotemporal Forecasting, at the Neural Information Processing Systems (NIPS) Conference, Barcelona, Spain, 2016.

Products

Books

Book Chapters

Inventions

Journals or Juried Conference Papers

A. Srivastava and T. {D}el{S}ole (2017). Decadal predictability without ocean dynamics. *Proceedings of the National Academy of Sciences*. 114 (9), 2177--2182. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Cheng Tang and Claire Monteleoni (2017). {Convergence rate of stochastic k-means}. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*. 54 1495--1503. Status = PUBLISHED; Acknowledgment of Federal Support = No ; Peer Reviewed = Yes

R. L. Glicksman and D. L. Markell and C. Monteleoni (2017). Technological Innovation, Data Analytics, and Environmental Enforcement. *Ecology L. Q.*. 44 41-88. Status = PUBLISHED; Acknowledgment of Federal Support = No ; Peer Reviewed = Yes

S. McQuade and C. Monteleoni (2017). Spatiotemporal Global Climate Model Tracking. *Large-Scale Machine Learning in the Earth Sciences, Srivastava, Nemani, Steinhaeuser (Eds.), Chapman & Hall/CRC*. . Status = PUBLISHED; Acknowledgment of Federal Support = No ; Peer Reviewed = Yes

T. DelSole and A. Banerjee (2017). Statistical Seasonal Prediction Based on Regularized Regression. *Journal of Climate*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

V. Sivakumar and A. Banerjee (2017). High-Dimensional Structured Quantile Regression. *International Conference on Machine Learning (ICML)*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

X. Yan and T. {D}el{S}ole and M. K. Tippett (2017). What Surface Observations Are Important for Separating the Influences of Anthropogenic Aerosols from Other Forcings?. *J. Climate*. (11), 4165--4184. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes

Licenses

Other Conference Presentations / Papers

Other Products

Other Publications

Patents

Technologies or Techniques

Thesis/Dissertations

Websites

Participants/Organizations

What individuals have worked on the project?

Name	Most Senior Project Role	Nearest Person Month Worked
Monteleoni, Claire	PD/PI	1
Tang, Cheng	Graduate Student (research assistant)	1

Full details of individuals who have worked on the project:

Claire Monteleoni
Email: cmontel@gwu.edu
Most Senior Project Role: PD/PI
Nearest Person Month Worked: 1

Contribution to the Project: Led the GW research activities leading the the 3 publications reported in this reporting period.

Funding Support: George Washington University, and another NSF grant: 1650080

International Collaboration: No
International Travel: Yes, Spain - 0 years, 0 months, 5 days

Cheng Tang
Email: tangch@gwu.edu
Most Senior Project Role: Graduate Student (research assistant)
Nearest Person Month Worked: 1

Contribution to the Project: Performed the research and first-authored a paper at AISTATS 2017.

Funding Support: 2 months on this grant at half effort. One semester she was supported by George Washington University as a TA. For several months she was supported as a research visitor at a different university. Partial support on NSF-1650080.

International Collaboration: No
International Travel: No

What other organizations have been involved as partners?
Nothing to report.

What other collaborators or contacts have been involved?
Nothing to report

Impacts

What is the impact on the development of the principal discipline(s) of the project?

This project has advanced climate prediction capabilities on seasonal and multi-year time scales. Specifically, our work has shown that (1) empirical models can predict annual mean temperatures skillfully a year or two in advance, and (2) simple regularized regression methods can produce skillful predictions of monthly mean temperatures in North America.

Further, we developed an approach for quantile regression in high-dimensional settings which can operate even when the number of samples is smaller than the number of covariates. The approach will be useful for future work in predictive modeling of tails and extremes of climate variables of interest.

Progress also included developing a hidden Markov random field (MRF)-based algorithm to sequentially incorporate climate observation data with climate model simulation output, while simultaneously learning the level of both temporal and spatial non-stationarity. The project also advanced the understanding of light-weight machine learning algorithms for processing large data sets.

What is the impact on other disciplines?

Advances in statistical prediction on seasonal and multi-year time scales are crucial to the study of sustainability.

What is the impact on the development of human resources?

In the present reporting periods, three PhD students were trained.

What is the impact on physical resources that form infrastructure?
Nothing to report.

What is the impact on institutional resources that form infrastructure?
Nothing to report.

What is the impact on information resources that form infrastructure?
Nothing to report.

What is the impact on technology transfer?
Nothing to report.

What is the impact on society beyond science and technology?

The threat of climate change is one of the greatest challenges currently facing society. Research progress resulting from this project is uniquely poised for societal impact by improving predictions of climate change on seasonal and multi-year time scales. Resulting discoveries will be valuable to communities and decision makers. This project has contributed to an environmental law article.

Changes/Problems

Changes in approach and reason for change

The work on Decadal Predictability took more time than expected, leaving less time than expected for developing the methodology for identifying nonlinear or non-Gaussian relations in data. However, the former work lead to a paper published in the Proceedings of the National Academy of Sciences that breaks new ground by presenting a view of decadal predictability at odds with the prevailing idea that the ocean circulation carries the memory.

Actual or Anticipated problems or delays and actions or plans to resolve them
Nothing to report.

Changes that have a significant impact on expenditures
Nothing to report.

Significant changes in use or care of human subjects
Nothing to report.

Significant changes in use or care of vertebrate animals
Nothing to report.

Significant changes in use or care of biohazards
Nothing to report.