4th International Workshop on Climate Informatics
September 25-26, 2014
Hosted by the National Center for Atmospheric Research in Boulder, CO

ROBUSTNESS AND SYNTHESIS OF EARTH SYSTEM MODELS (ESMS): A MULTI-TASK LEARNING PERSPECTIVE

André R Gonçalves[1,2], Vidyashankar Sivakumar[2], Soumyadeep Chatterjee[2], Devashish Kumar[3], Snigdhansu Chatterjee[2], Auroop Ganguly[3], Vipin Kumar[2], Stefan Liess[2], Pradeep Ravikumar[4], Arindam Banerjee[2]

***Abstract*—Current evaluations of GCMs, as detailed by the IPCC AR5, show that different GCMs often illustrate different skills for different tasks. The IPCC uses an average or multi-model ensemble (MME) for future projections. In this work, we consider evaluating and combining GCMs for multiple tasks from a multi-task learning (MTL) perspective. MTL is an approach to machine learning that learns multiple problems simultaneously thus leading to a better model by taking advantage of the commonality among the tasks [1], [2]. But unlike traditional MTL, for our problem the task relationships are not known initially. Further, each GCM has multiple runs corresponding to different initial conditions which need to be suitably considered. In this work, we present approaches to multi-task sparse structure learning (MSSL) which estimate task relationships along with learning suitable multi-model combinations for each task, and explore ideas for handling multiple initial condition runs for GCMs.**

I. MOTIVATION

Earth system models develop plausible projections for the future climate and provide a framework for understanding climate science and developing risk-informed plans for adaptation and mitigation. While the IPCC assessment reports continue to use multi-model averages, especially for global and continental aggregate projections, studies have demonstrated the usefulness of skill selected model subsets for stakeholder-relevant studies at local to regional scales. Examining the robustness of these models at multiple spatial and temporal aggregate scales becomes crucial to inform decisions and policy. However, a key question arises as to how models may be evaluated or indeed falsified in the context of decadal to century scale projections. In addition, ensemble runs of multiple emissions greenhouse gas emission scenarios, multiple models and multiple initial conditions, provide assessments of uncertainties in future emissions, variability owing to lack of understanding of physics, and intrinsic natural variability of the climate system. The ability to synthesize these ensembles into an envelope of plausible futures to inform decision makers remains an outstanding grand challenge in climate science.

In this work, we consider evaluating and combining Global Climate Models (GCM) for multiple tasks from a multi-task learning (MTL) perspective. Unlike traditional MTL, the relationships between tasks considered in the context of climate are unknown, e.g., does a model with high skill in characterizing regional land sinks for the carbon cycle also do well in characterizing precipitable water, etc. Further, each GCM has multiple initial condition runs which need to be suitably considered. In this work, we present approaches to multi-task sparse structure learning (MSSL) which estimate task relationships along with learning suitable multi-model combinations for each, and explore ideas for handling multiple initial condition runs for GCMs.

II. MULTI MODEL COMBINATION

Different GCM models have a high variance in their forecast of the future climate variables which turns out to produce uncertainty in further analysis based on these predictions. Combining multiple GCMs outputs may lead to a more accurate estimate of a model property through the provision of a larger sample size and significantly reduce forecast variability [3]. Thus, climate scientists have been interested in proper ways of combining GCM outputs, reducing forecast variability without loss of physical significance.

Also, the models considered can either be multiple runs of the same model with different initial conditions or runs from different models. The former only characterizes the uncertainty associated with internal climate variability, whereas the second also includes the impact of model differences.

Corresponding author: André R Gonçalves (andre@cs.umn.edu) University of Campinas, Brazil[1], University of Minnesota[2], Northeastern University[3], University of Texas, Austin[4].

One common multi model approach for prediction mentioned in the IPCC AR5 is to use the average of all models, which could either be the arithmetic mean or the weighted mean. For the second case, some measure of model's ability to simulate the observed climate variable is necessary.

Studies have been conducted to investigate the impact of multi model combination both in terms of prediction accuracy and multi model variability. We refer the interested readers to [4], [5] for further details.

III. Multitask Learning

Multitask Learning (MTL) [2], [1] is a machine learning paradigm which seeks to improve the generalization capacity of a learning task (algorithm/method) by using information from other related tasks.

MTL can benefit from knowing the underlying structure relating the tasks while learning parameters for each task. In situations where some of the tasks may be highly dependent on each other, the strategy of isolating each task does not exploit the potential information one may acquire from other related tasks.

We propose a Multitask Sparse Structure Learning (MSSL) algorithm [6] for GCMs outputs combination. MSSL performs coefficients learning for all locations simultaneously and by using related tasks information can provide more accurate predictions on individual locations. Further, it encourages similar locations, including neighbors, to have similar model coefficients, which for some climate variables is more physically plausible. The main difference between MSSL and the previous multitask learning methods is that MSSL does not assume any dependency structure among locations, but learns it from the data instead.

IV. Evaluation

In a preliminary analysis we consider the problem of GCM outputs combination for land surface temperature prediction in South America. The detailed experimental setup can be found in [6].

We consider 4 baselines for comparison: **(1) Multi-Model Average (MMA)** is the current technique used by Intergovernmental Panel on Climate Change (IPCC); **(2) Best GCM** uses the predicted outputs of the best GCM in the training phase (lowest RMSE); **(3) Ordinary Least Squares (OLS)** regression for each geographic location; and **(4) Multi Model Regression with Spatial Smoothing (S^2M^2R)** is the model recently proposed by [7] which incorporates spatial smoothing using the graph Laplacian over a grid graph.

Table I reports the average and standard deviation RMSE of all 250 geographical locations. While MMA has the highest RMSE, MSSL has the smallest RMSE in comparison to the baselines. The performance of OLS and S^2M^2R are very similar.

MMA	Best GCM	OLS	S^2M^2R	MSSL
1.621	1.410	0.866	0.863	**0.780**
(±0.020)	(±0.037)	(±0.037)	(±0.067)	(±0.039)

TABLE I. Mean and standard deviation of RMSE over all locations for MSSL and the baseline algorithms. MSSL performs best in predicting temperature over South America.

Averaging the model outputs as done by MMA, reduces prediction accuracy. On the other hand MSSL performs better once it learns the right weight combination on the model outputs and incorporates spatial smoothing by learning the task relatedness.

Acknowledgments

The research was supported by NSF Expeditions in Computing grant - IIS-1029711. The work was also supported by NSF grants IIS-0916750, CNS-1314560, and by NASA grant NNX12AQ39A. A.B. acknowledges support from IBM and Yahoo. A.R.G. was supported by Science without Borders grant from CNPq/Brazil. P.R. acknowledges the support of NSF via IIS-1447574. Access to computing facilities by University of Minnesota Supercomputing Institute (MSI).

References

[1] J. Baxter, "A Bayesian Information Theoretic Model of Learning to Learn via Multiple Task Sampling," *Machine Learning*, vol. 28, no. 1, pp. 7–39, 1997.

[2] R. Caruana, "Multitask learning," *Machine Learning*, vol. 28, no. 1, pp. 41–75, 1997.

[3] R. Knutti, G. Abramowitz, M. Collins, V. Eyring, P. J. Gleckler, B. Hewitson, and L. Mearns, "Good practice guidance paper on assessing and combining multi model climate projections," in *IPCC Expert Meeting on Assessing and Combining Multi Model Climate Projections*, 2010.

[4] A. P. Weigel, R. Knutti, M. A. Liniger, and C. Appenzeller, "Risks of model weighting in multimodel climate projections," *Journal of Climate*, vol. 23, pp. 4175–4191, 2010.

[5] C. Tebaldi and R. Knutti, "The use of the multi-model ensemble in probabilistic climate projections," *Philosophical Transactions of the Royal Society A*, vol. 365, no. 1857, pp. 2053–2075, 2007.

[6] A. R. Gonçalves, P. Das, S. Chatterjee, V. Sivakumar, F. J. Von Zuben, and A. Banerjee, "Multitask Sparse Structure Learning," in *23rd ACM CIKM*, 2014.

[7] K. Subbian and A. Banerjee, "Climate Multi-model Regression using Spatial Smoothing," in *SDM*, 2013.