

A Multitask Learning View on the Earth System Model Ensemble

André R. Gonçalves | University of Campinas and University of Minnesota

Fernando J. Von Zuben | University of Campinas

Arindam Banerjee | University of Minnesota

Earth system models (ESMs) are the primary mechanisms for obtaining projections under different climate change scenarios. Researchers use ensembles of climate models to gain better accuracy and reduce uncertainty. A multitask learning-based method can build ESM ensembles for all regions jointly to improve predictions for individual ones.

Projections of future climate variables such as temperature, precipitation, and atmospheric pressure are fundamental to obtaining a picture of how the climate will evolve in the near and long term; we then can infer how these changes will affect life on Earth. For example, an increase in global temperatures is expected to raise sea levels, which will have a potentially large impact on coastal areas, with consequences such as increased flooding exposure. Regional climate change could also alter forests, crop yields, and water supplies. Human health, fauna equilibrium, and many other types of ecosystems will be impacted by these changes.

Based on the projected variables, we can also forecast other, more complex climate phenomena. Information about future sea surface temperature, for example, can be used to project El Niño-Southern Oscillation (ENSO),¹ which involves fluctuating ocean temperatures in the equatorial Pacific. ENSO is known to provoke variations in regional climate patterns around the globe, including hurricanes in the US southeast, droughts in the Brazilian and Indian northeast, and warmer winters in the Canadian and Alaskan northwest.

Table 1. Earth system models (ESMs) used in experiments.

ESM	Origin	Citation
BCC_CSM1.1	Beijing Climate Center, China	L. Zhang et al., "Projections of Annual Mean Air Temperature and Precipitation over the Globe and in China during the 21st Century by the BCC Climate System Model BCC CSM1.0," <i>Acta Meteorologica Sinica</i> , vol. 26, no. 3, 2012, pp. 362–375.
CCSM4	National Center for Atmospheric Research (NCAR), US	W. Washington et al., "The Use of the Climate-Science Computational End Station (CCES) Development and Grand Challenge Team for the Next IPCC Assessment: An Operational Plan," <i>J. Physics</i> , vol. 125, no. 1, 2008; http://iopscience.iop.org/1742-6596/125/1/012024 .
CESM1	National Center for Atmospheric Research (NCAR), US	Z.M. Subin et al., "Boreal Lakes Moderate Seasonal and Diurnal Temperature Variation and Perturb Atmospheric Circulation: Analyses in the Community Earth System Model 1 (CESM1)," <i>Tellus A</i> , vol. 64, 2012; www.tellusa.net/index.php/tellusa/article/view/15639 .
CSIRO	Commonwealth Scientific and Industrial Research Organization, Australia	H.B. Gordon et al., <i>The CSIRO Mk3 Climate System Model</i> , tech. paper 60, CSIRO Atmospheric Research, 2002.
HadGEM2	Met Office Hadley Centre, UK	W. Collins et al., "Development and Evaluation of an Earth-System Model, HadGEM2," <i>Geoscientific Model Development Discuss</i> , vol. 4, 2011, pp. 997–1062.
IPSL	Institut Pierre-Simon Laplace, France	J. Dufresne et al., "Climate Change Projections Using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5," <i>Climate Dynamics</i> , 2012; www.lmd.jussieu.fr/~jldufres/publi_ipslcm5/Smi/ipsl-cm5.pdf .
MIROC5	Atmosphere and Ocean Research Institute, Japan	M. Watanabe et al., "Improved Climate Simulation by MIROC5: Mean States, Variability, and Climate Sensitivity," <i>J. Climate</i> , vol. 23, no. 23, 2010, pp. 6312–6335.
MPI-ESM	Max Planck Institute for Meteorology, Germany	V. Brovkin et al., "Evaluation of Vegetation Cover and Land-Surface Albedo in MPI-ESM CMIP5 Simulations," <i>J. Advances in Modeling Earth Systems</i> , 2013; http://onlinelibrary.wiley.com/doi/10.1029/2012MS000169/abstract .
MRI-CGCM3	Meteorological Research Institute, Japan	S. Yukimoto, Y. Adachi, and M. Hosaka, "A New Global Climate Model of the Meteorological Research Institute: MRI-CGCM3: Model Description and Basic Performance," <i>J. Meteorological Soc. Japan</i> , vol. 90, 2012, pp. 23–64.
NorESM	Norwegian Climate Centre, Norway	M. Bentsen et al., "The Norwegian Earth System Model, NorESM1-M-Part 1: Description and Basic Evaluation," <i>Geoscientific Model Development</i> , vol. 5, 2012, pp. 2843–2931.

Such projections are performed by computer simulations based on mathematical models that attempt to emulate dynamical, physical, and biogeochemical processes relevant to the climate system. These *Earth system models* (ESMs) are complex mathematical representations of the major climate system components (atmosphere, land surface, ocean, and sea ice) and their interactions. Given certain initial conditions (socioeconomic scenarios, for example), the models simulate all interactions among climate system components and produce predictions for a variable of interest. Owing to many sources of uncertainty (which we'll discuss in the next section), the projections can have a high variability, thus making it difficult to perform inferences.

Earth System Model Uncertainties and Multimodel Ensemble

The forecasts of future climate variables produced by ESMs have high variability due to three sources of uncertainty: future anthropogenic emissions of greenhouse gases, aerosols, and other natural forcings ("emission uncertainties"); imprecision due to incomplete understanding of climate systems ("model uncertainties"); and the existence of inherent internal climate variability itself ("initial condition uncertainties"). In this work, we focus on reducing model uncertainties and producing more reliable projections.

Climate science institutes from various countries (see Table 1 for a few examples) have proposed

several ESMs, differing slightly from each other in the way they model climate processes that aren't fully understood. Consequently, each ESM can produce different projections for the same climate variable and initial condition. To give a sense of the variability among ESMs, Figure 1 shows South American mean temperature anomalies for the period of 1901 to 2000 that 10 different ESMs produced. A well-accepted approach of addressing model uncertainty is the concept of multimodel ensemble,² in which instead of relying on a single ESM, projection is performed based on a set of produced simulations.

There's still no consensus on the best method of combining ESMs outputs.² The simplest approach is to assign equal weights to all ESMs, then perform an arithmetic mean. Other approaches suggest assigning different weights to individual ESMs,^{3,4} with the weights specifically chosen to reflect the skill levels of the ESMs. The latter is the focus of our approach, in which weights for each geographical location are estimated via a least square fitting.

The primary novelty of our methodology is that it jointly solves all least square problems in a multitask learning (MTL) fashion, allowing the exchange of information among related geographical locations.

Multitask Learning

Learning for multiple tasks, such as regression and classification, simultaneously arises in many practical situations. For example, in spam filtering, the problem of learning a personalized filter (classifier) can be treated as a single supervised learning task involving data from multiple users; in finance forecasting, models for simultaneously predicting the value of many possibly related indicators are required; and in multilabel classification, where the problem is usually split into binary classification problems for each label, the classifiers' joint synthesis can be beneficial.

The most common strategy to deal with multiple tasks is to perform the learning procedure for each task independently. However, in situations where the tasks might be related to each other, the strategy of isolating each task won't exploit the potential information we could acquire from other related tasks. This is the main motivation behind MTL,^{5,6} which seeks to improve a learning task's generalization capability by exploiting task commonalities. To allow information exchange, MTL usually employs a shared representation. MTL's benefits over traditional independent learning are

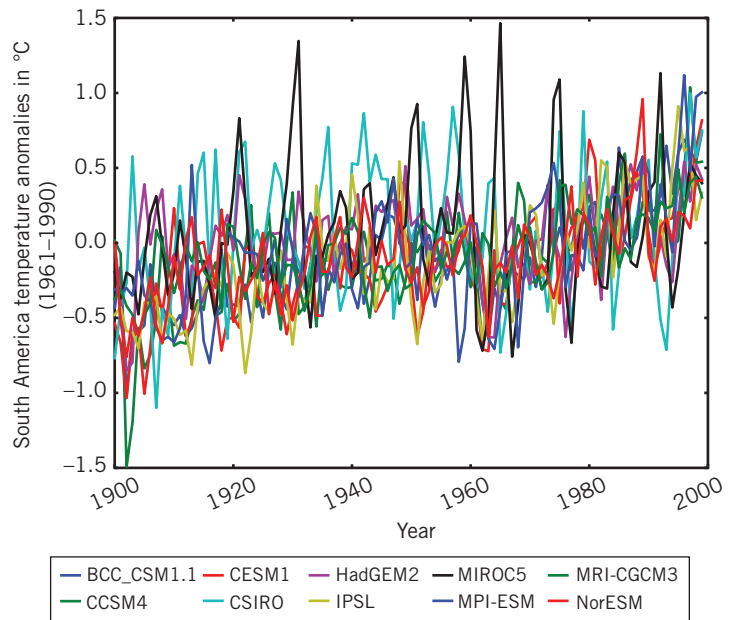


Figure 1. South American land mean temperature anomalies in °C for the 10 Earth system models (ESMs) described in Table 1. Note the high variability among them.

supported in many experimental and theoretical works.^{7–10} Figure 2 illustrates MTL's joint learning performed in contrast to traditional single-task (independent) learning.

Many MTL methods^{7,11} make the assumption that all tasks are related. However, this might not hold for some applications—in fact, sharing information with unrelated tasks can be detrimental to task performance.¹² A fundamental step is to estimate the true relationship structure among tasks, thus promoting proper information sharing among related tasks while we avoid using information from unrelated tasks.

Recently, researchers have explicitly focused on modeling task relationships.^{13–15} In particular, we proposed a hierarchical Bayesian model in which task dependencies are explicitly estimated from the data, in addition to each task's parameters (weights).¹⁵ Unlike other MTL methods, our approach—called multitask sparse structure learning (MSSL)—measures tasks' relationships in terms of partial correlation, which has a more meaningful interpretation in terms of conditional independence than ordinary correlation. Additionally, MSSL uses efficient optimization methods for estimating the graph dependence and task-specific parameters, which makes MSSL suitable for problems with a large number of tasks, such as ESM ensembles.

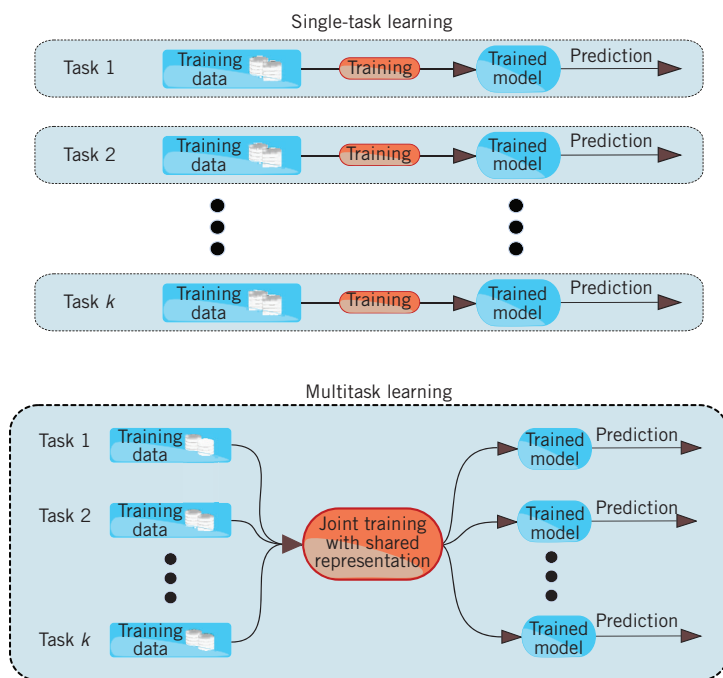


Figure 2. Differences between multitask learning (MTL) and traditional single-task learning. In MTL, the learning process involves all tasks and is performed jointly, allowing for information exchange.

In MSSL, task dependencies are represented by a weighted graph structure, where edges indicate that the connected nodes are related (conditionally dependent). The weight associated with each edge indicates how strong is the dependence between the two tasks. The procedure to estimate the parameters has two steps:

1. Estimate the weighted dependence graph based on current task-specific parameters.
2. Learn all task parameters jointly by sharing information with related tasks determined by the graph computed in Step 1.

These steps iterate until a convergence criterion is met (for example, graph dependence and weights don't change in consecutive iterations). As each step solves a convex optimization problem, an alternating minimization process guarantees the convergence to a local minimum.¹⁶

In Step 1, a structure learning problem must be solved, and for this we used recent advances—for the experiments, for example, we used an alternating direction method of multipliers (ADMM).¹⁷ Step 2 requires the solution of a penalized quadratic programming problem, which we solved with the established proximal gradient descent-based

fast iterative shrinkage-thresholding algorithm (FISTA).¹⁸

We can then interpret the problem of ESM ensembles for several distinct geographical locations as an instance of MTL. Weighting the ESMs according to their relative performance in a control period is a linear regression task. By using MSSL,¹⁵ we're assuming that geographical locations with similar model skill patterns will present similar ESM output values. We make an equivalent assumption for the weights, enforcing the idea that similar geographical locations tend to have similar weights.

Case Study: Temperature Forecasting in South America

To demonstrate the benefits of MSSL, we consider the problem of combining multiple ESMs for land surface temperature prediction in South America. Being that South America is the world's fourth-largest continent, covering approximately 12 percent of the Earth's land area, its climate varies considerably. The Amazon River basin in the north has the typical hot wet climate suitable for the growth of rain forests. The Andes Mountains, on the other hand, remain cold throughout the year. The desert regions of Chile are the driest portions of South America.

For our simulations, we used temperature outputs from 10 CMIP5 ESMs,¹⁹ with a single run for each one (see Table 1 for the list). We obtained global observation data for surface temperature (in degrees Celsius) from the Climate Research Unit (CRU; www.cru.uea.ac.uk). We aligned the data from the ESMs and CRU observations to have the same spatial and temporal resolution, using publicly available climate data operators (CDOs; <https://code.zmaw.de/projects/cdo>). For all the experiments, we used a 2.5° × 2.5° grid over latitudes and longitudes in South America and monthly mean temperature data along 100 years, to 2000, with records starting from 16 January 1901. In total, we considered 1,200 time points (monthly data) and 250 spatial locations over land.

Methodology

We considered the following four baselines for comparing and evaluating MSSL performance:

- average model, the current technique used by the Intergovernmental Panel on Climate Change (IPCC), which gives equal weight to all ESMs at every location;

Table 2. Mean and standard deviation of the root-mean-square error over all locations.

Average (IPCC)	Best ESM	OLS	S ² M ² R	<i>r</i> -MSSL
1.621	1.410	0.866	0.863	0.780
(±0.020)	(±0.037)	(±0.037)	(±0.067)	(±0.039)

- best ESM, which uses the predicted outputs of the best ESM in the training (control) phase (lowest root-mean-square error [RMSE]);
- linear regression, an ordinary least squares (OLS) regression for each geographical location; and
- multimodel regression with spatial smoothing (S²M²R),²⁰ which can be seen as a special case of MSSL with local spatial dependence assumption.

We'll refer to these baselines and MSSL as the “models” and the constituent ESMs as “submodels.” We used the residual-based version of the MSSL algorithm, referred to as *r*-MSSL, in the experiments. (More details appear elsewhere.¹⁵)

We considered a moving window of 50 years of data for training and the next 10 years for testing, resulting in five train/test sets. The decadal climate prediction problem has been a focus of study in the recent past.²¹

Results

Table 2 reports the mean and standard deviation RMSE in degrees Celsius for all 250 geographical locations. While the average model has the highest RMSE, MSSL has the smallest RMSE in comparison to the baselines. The performance of OLS and S²M²R are similar. Values in boldface indicate statistical significance ($p < 0.05$).

Figure 3 shows the RMSE per geographical location for the average model, best ESM, OLS and *r*-MSSL. As previously mentioned, South America has a diverse climate, and not all of the ESMs are designed to take it into account and capture it. Hence, averaging the model outputs as done by IPCC produces less accurate predictions when compared to the other methods. Moreover, *r*-MSSL performs better possibly due to a proper definition of weights in the combination of the model outputs through the incorporation of spatial smoothing by learning the task relationship. The neighborhood dependence assumption of S²M²R neglects possible long-range dependence among geographical areas while OLS completely

ignores dependence information. In particular, the average model isn't performing well on the west coast and southern part of South America (parts of Peru, Chile, and Argentina, which experience different climate conditions).

Figure 4 shows the relationship structure estimated by *r*-MSSL among the geographical locations. The regions connected by blue lines were identified as being dependent on each other. We can immediately see that locations in the northwest part of South America are densely connected. This area has a typical tropical climate and comprises the Amazon rainforest, which is known to have a hot and humid climate throughout the year with low temperature variation.²² The cold climates that occur in the southernmost parts of Argentina and Chile are clearly highlighted. Such areas have low temperatures throughout the year, but there are large daily variations.

We can make another important observation about South America's west coast, ranging from central Chile to Venezuela, and passing through Peru, which has one of the driest deserts in the world. These areas are located to the left of the Andes Mountains and are known to exhibit an arid climate. The average model didn't perform well on this region compared to *r*-MSSL—we can see the long lines connecting these coastal regions, which might explain the improvement in terms of RMSE reduction achieved by *r*-MSSL. The algorithm uses information from related locations to enhance its performance in these areas.

The lack of connecting lines in central Argentina can be explained by the temperate climate, which presents a greater range of temperatures when compared to tropical climates and can have extreme climatic variations. This is a transition area between the cold southernmost region and the hot and humid central area of South America. It also comprises Patagonia, a semiarid scrub plateau that covers nearly the entire southern portion of mainland, whose climate is strongly influenced by the South Pacific air current. Due to high variability, it becomes harder to provide accurate temperature predictions.

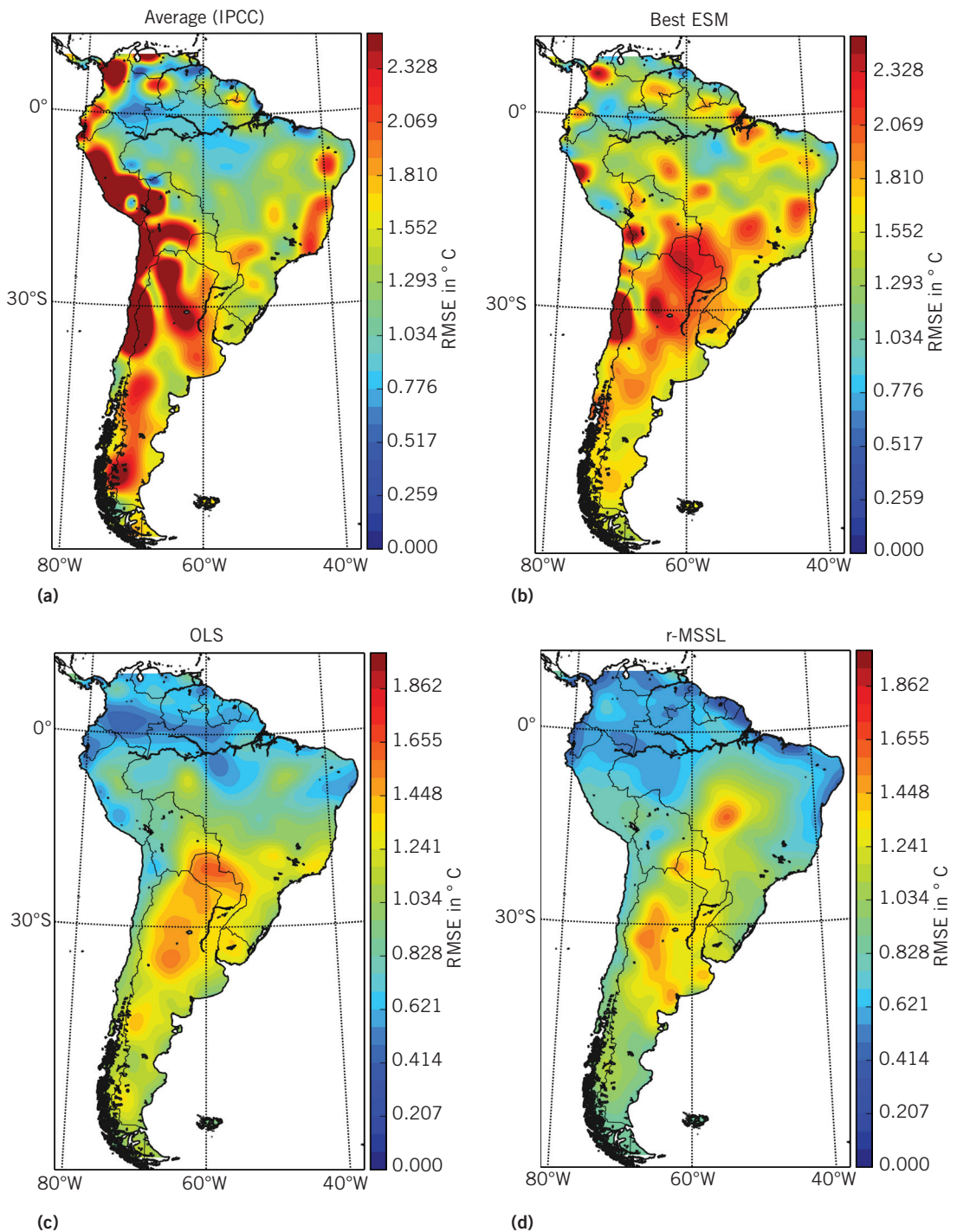


Figure 3. Root-mean-square error (RMSE) per location for each one of the four approaches: (a) the average model, (b) the best ESM, (c) the ordinary least squares (OLS), and (d) the *r*-MSSL. Because S^2M^2R produced almost the same RMSE as OLS, it's not shown here.

We also observe that locations within Chile and Argentina are less densely connected to other parts of South America. A possible explanation could

be that Chile includes the Atacama Desert, a dry region located to the west of the Andes, and Argentina, especially the southern part, experiences heavy

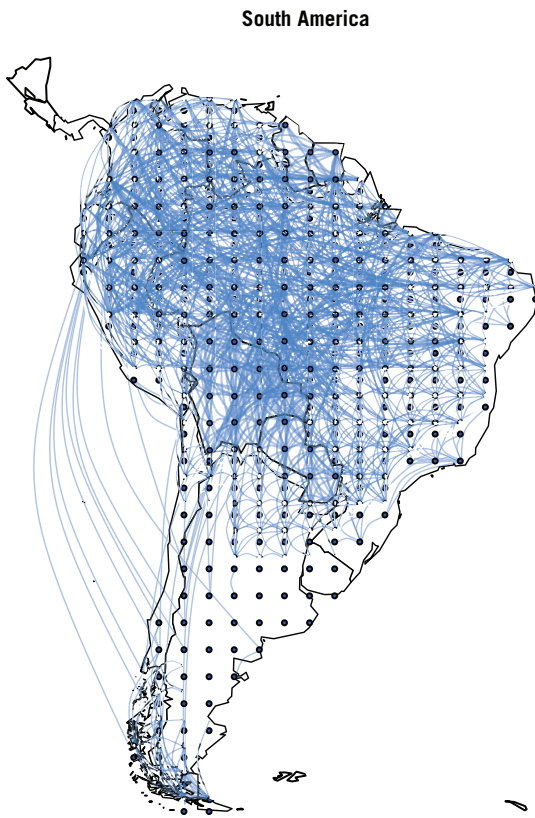


Figure 4. Relationships between geographical locations estimated by the r -MSSL algorithm. The blue lines indicate that connected locations are dependent on each other.

snowfall, in contrast to the hot and humid rain forests or the dry and arid deserts on the west coast. Both these regions experience climate conditions that are disparate from the northern rain forests and from each other. The task dependencies estimated from the data reflect this disparity. Figure 5 presents the dependency structure using a chord diagram. Each point on the circle's periphery is a location in South America and has associated with it a learning task to predict temperature at that location. The locations are arranged serially on the periphery according to the respective countries. We can immediately see that the locations in Brazil are strongly connected to parts of Peru, Colombia, and parts of Bolivia. These connections are interesting, as these parts of South America comprise the Amazon rain forest.

Handling other sources of uncertainty will be the focus of future research steps. For example, uncertainty due to anthropogenic forcings (which depend on socioeconomic factors including

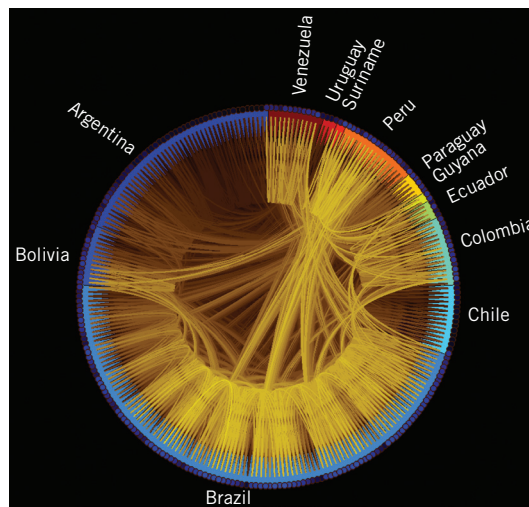


Figure 5. Chord graph representing the structure estimated by the r -MSSL algorithm. Each point on the circle's periphery is a location in South America and has associated with it a learning task to predict temperature at that location. Lighter connections indicate stronger dependence between locations.


global geopolitical agreements to control pollutant emissions) is usually handled by running the ESM for future scenarios of distinct amounts of gas emission. Consequently, we can generate multiple runs of the same ESM and incorporate strategies of feature and structure grouping²³ into the MSSL algorithm to properly deal with these runs. ■

Acknowledgments

We thank Vidyashankar Sivakumar, Pujya Das, and Soumyadeep Chatterjee for their contributions in early stages of this work. Gonçalves and Von Zuben thank CNPq for financial support. Banerjee was supported by US National Science Foundation grants IIS-1447566, IIS-1422557, CCF-1451986, CNS-1314560, IIS-0953274, and IIS-1029711, NASA grant NNX12AQ39A, and gifts from Yahoo and IBM.

References

1. G.J. Van Oldenborgh, S. Philip, and M. Collins, "El Niño in a Changing Climate: A Multi-model Study," *Ocean Science*, vol. 1, no. 2, 2005, pp. 81–95.
2. A.P. Weigel et al., "Risks of Model Weighting in Multimodel Climate Projections," *J. Climate*, vol. 23, no. 15, 2010, pp. 4175–4191.
3. T.N. Krishnamurti et al., "Improved Weather and Seasonal Climate Forecasts from Multimodel Superensemble," *Science*, vol. 285, no. 5433, 1999, pp. 1548–1550.

4. C. Tebaldi and R. Knutti, "The Use of the Multi-model Ensemble in Probabilistic Climate Projections," *Philosophical Trans. Royal Soc. A*, vol. 365, no. 1857, 2007, pp. 2053–2075.
 5. J. Baxter, "A Bayesian/Information Theoretic Model of Learning to Learn via Multiple Task Sampling," *Machine Learning*, vol. 28, no. 1, 1997, pp. 7–39.
 6. R. Caruana, "Multitask Learning," *Machine Learning*, vol. 28, no. 1, 1997, pp. 41–75.
 7. T. Evgeniou and M. Pontil, "Regularized Multi-task Learning," *Proc. ACM SIGKDD Conf. Knowledge Discovery and Data Mining*, 2004, pp. 109–117.
 8. S. Ben-David and R. Borbely, "A Notion of Task Relatedness Yielding Provable Multiple-Task Learning Guarantees," *Machine Learning*, vol. 73, no. 3, 2008, pp. 273–287.
 9. S. Bickel et al., "Multi-task Learning for HIV Therapy Screening," *Proc. Int'l Conf. Machine Learning*, 2008, pp. 56–63.
 10. J. Chen et al., "A Convex Formulation for Learning a Shared Predictive Structure from Multiple Tasks," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 35, no. 5, 2013, pp. 1025–1038.
 11. A. Argyriou, T. Evgeniou, and M. Pontil, "Multi-task Feature Learning," *Proc. Neural Information Processing Systems (NIPS 07)*, 2007, pp. 41–50.
 12. J. Baxter, "A Model of Inductive Bias Learning," *J. Artificial Intelligence Research*, vol. 12, 2000, pp. 149–198.
 13. Y. Zhang and J. Schneider, "Learning Multiple Tasks with Sparse Matrix-Normal Penalty," *Proc. Neural Information Processing Systems (NIPS 10)*, 2010, pp. 2550–2558.
 14. Y. Zhang and D.-Y. Yeung, "A Convex Formulation for Learning Task Relationships in Multi-task Learning," *Proc. Conf. Uncertainty in Artificial Intelligence*, 2010; http://event.cwi.nl/uai2010/papers/UAI2010_0144.pdf.
 15. A. Gonçalves et al., "Multi-task Sparse Structure Learning," *Proc. ACM Int'l Conf. Information Knowledge and Management*, 2014, pp. 451–460.
 16. A. Gunawardana and W. Byrne, "Convergence Theorems for Generalized Alternating Minimization Procedures," *J. Machine Learning Research*, vol. 6, 2005, pp. 2049–2073.
 17. S. Boyd et al., "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundation and Trends in Machine Learning*, vol. 3, no. 1, 2011, pp. 1–122.
 18. A. Beck and M. Teboulle, "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems," *SIAM J. Imaging Sciences*, vol. 2, no. 1, 2009, pp. 183–202.
 19. K.E. Taylor, R.J. Stouffer, and G.A. Meehl, "An Overview of CMIP5 and the Experiment Design," *Bulletin Am. Meteorological Soc.*, vol. 93, no. 4, 2012, p. 485.
 20. K. Subbian and A. Banerjee, "Climate Multi-model Regression Using Spatial Smoothing," *Proc. SIAM Int'l Conf. Data Mining*, 2013; http://www-users.cs.umn.edu/~banerjee/papers/13/climate_model_spatial.pdf.
 21. D.M. Smith et al., "Improved Surface Temperature Prediction for the Coming Decade from a Global Climate Model," *Science*, vol. 317, no. 5839, 2007, pp. 796–799.
 22. V.A. Ramos, "South America," *Encyclopaedia Britannica Online Academic Edition*, 2014.
 23. P. Danaher, P. Wang, and D.M. Witten, "The Joint Graphical Lasso for Inverse Covariance Estimation across Multiple Classes," *J. Royal Statistical Soc. B*, vol. 76, no. 2, 2014, pp. 373–397.
-
- André R. Gonçalves** is pursuing a PhD in Computer Engineering at the University of Campinas and was a visiting scholar at the University of Minnesota, Twin Cities. His research interests include machine learning, multi-task learning, structure learning, probabilistic graphical models, and Bayesian models. Gonçalves has a MS in Electrical Engineering from the University of Campinas. Contact him at andreric@dca.fee.unicamp.br.
-
- Fernando J. Von Zuben** is a professor in the Department of Computer Engineering and Industrial Automation, School of Electrical and Computer Engineering, University of Campinas. His research interests include computational intelligence, bioinspired computing, multivariate data analysis, and machine learning. Von Zuben has a PhD in electrical engineering from the University of Campinas. He's a senior member of IEEE and a member of the International Neural Network Society. Contact him at vonzuben@dca.fee.unicamp.br.
-
- Arindam Banerjee** is an associate professor in the Department of Computer and Engineering and a Resident Fellow at the Institute on the Environment, University of Minnesota, Twin Cities. His research interests include statistical machine learning, data mining, convex analysis, and optimization and their applications in complex real-world problems. Banerjee has a PhD in electrical and computer engineering from University of Texas at Austin. Contact him at banerjee@cs.umn.edu.
-
-  Selected articles and columns from IEEE Computer Society publications are also available for free at <http://ComputingNow.computer.org>.