Climate Informatics: Accelerating Discovering in Climate Science with Machine Learning

The goal of climate informatics, an emerging discipline, is to inspire collaboration between climate scientists and data scientists, in order to develop tools to analyze complex and ever-growing amounts of observed and simulated climate data, and thereby bridge the gap between data and understanding. Here, recent climate informatics work is discussed, along with some of the field's remaining challenges.

The impacts of present and potential future climate change pose important scientific and societal challenges. Scientists have observed changes in temperature, sea ice, and sea level, and attributed those changes to human activity. It is an urgent international priority to improve our understanding of the climate system—a system characterized by complex phenomena that are difficult to observe and even more difficult to simulate. Despite the increasing availability of computational resources, current analytical tools have been outpaced by the ever-growing amounts of observed climate data from satellites, environmental sensors, and climate-model simulations. Computational approaches will therefore be indispensable for these analysis challenges. The goal of the fledgling research discipline, *climate informatics*, is to inspire collaboration between climate scientists and data scientists (machine learning, statistics, and data mining researchers), and thus bridge the gap between data and understanding. Research on climate informatics will accelerate discovery and answer pressing questions in climate science.

Machine learning is an active research area at the interface of computer science and statistics. The goal of machine learning research is to develop *algorithms*, automated techniques, to detect patterns in data. Such algorithms are critical to a range of technologies including Web search, recommendation systems, personalized Internet advertising, computer vision, and natural language processing. Machine learning also benefits the natural sciences, such as biology; the interdisciplinary bioinformatics field has facilitated many discoveries in genomics and proteomics. The impact of machine learning on climate science has the potential to be similarly profound.

Here, we focus specifically on challenges in climate modeling; however, there are myriad collaborations possible at the intersection of these two fields. Recent work reveals that collaborations with climate scientists also generate interesting new problems for machine learning.[1] To broaden the discussion, we propose challenge problems for climate informatics, some

CLAIRE MONTELEONI
George Washington University
GAVIN A. SCHMIDT
NASA Goddard Institute for Space Studies
SCOTT MCQUADE
George Washington University

of which we discuss in detail elsewhere,[2] and review recent successes in climate informatics. Climate scientists and machine learning, data mining, and statistics researchers discuss these topics at the Climate Informatics Workshop (an annual event we launched in 2011). Our prior work, including a survey, provides further discussion of related work on climate informatics.[1-3] Additional resources are available on the climate informatics wiki (http://sites.google.com/site/1stclimateinformatics).

Climate Modeling

Climate scientists use *climate models*, large-scale mathematical models run as computer simulations, to understand and predict the climate. Geophysical experts, including climate scientists and meteorologists, encode data from observed processes into highly complex, nonlinear mathematical models. As shown in Figure 1, each general circulation model (GCM) includes numerous individual climate-process models, such as cloud formation, rainfall, wind, ocean currents, and radiative heat transfer through the atmosphere. Results emergent from the models, such as the sensitivity of the climate to increasing greenhouse gases, are crucial to researchers trying to project and forecast the Earth's climate.[4] (It is important to note that whereas in machine learning, data mining, and statistics, the term *model* typically means data-driven, in climate modeling this term refers to a system of mathematical models that are based on scientific first principles).

Global climate modeling efforts began in the 1970s. The models have become more complex as computational resources have evolved. Currently, about 25 laboratories across the world support almost 50 climate models, forming the basis of climate projections (predictions) assessed by the Intergovernmental Panel on Climate Change (IPCC), which was established by the United Nations in 1988 and received the 2007 Nobel Peace Prize (shared with former Vice President Al Gore). Climate scientists initially developed the Coupled Model Intercomparison Project version 3 (CMIP3) archive to support the IPCC Fourth Assessment Report.[5] Researchers have used the archive in more than 500 publications, and it's a rich source of climate simulation output. The CMIP5 project continues the tradition of making global climate-model predictions easier to use, and it will be quite significant in future IPCC reports.

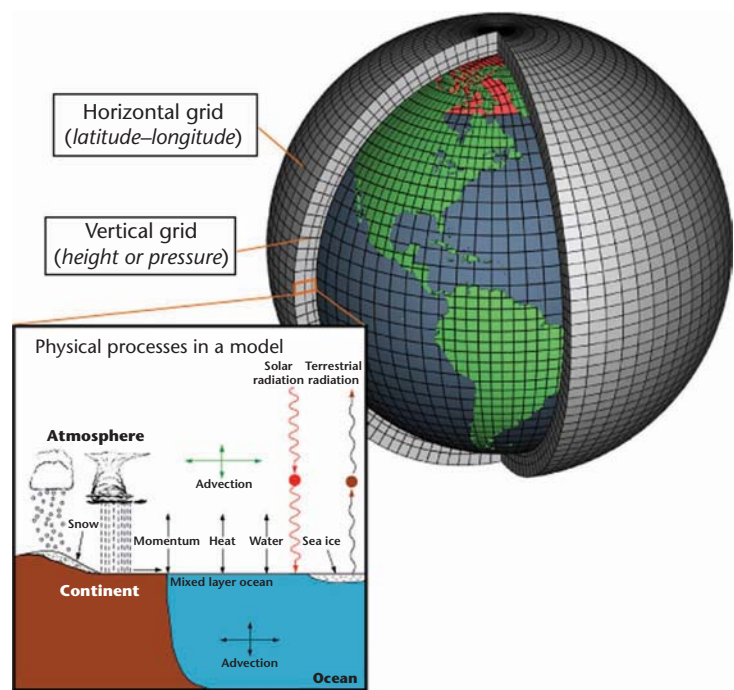The *multi-model ensemble* (MME), the ensemble of climate models that informs the

Figure 1. A global climate model discretization, and a selection of included physical processes. The US National Oceanic and Atmospheric Administration *(*NOAA) website offers this information and much more concerning climate change (http://celebrating200years.noaa.gov/breakthroughs/climate_model/welcome.html).

IPCC, has high variance over model predictions, for a variety of reasons. Different teams of scientists designed each model based on scientific first principles, which led to differences in scope, discretization assumptions, the included science, and coding errors. Even though the different models' predictions vary greatly, some climate scientists observed that overall, the average prediction (over multiple quantities, performance metrics, and time periods) is more consistent than any one model.[6,7] There has been growing interest, in the climate-modeling community, in better ways to combine MME predictions, as well as methods to assess the "skill" of a single climate model (quantifying the model's accuracy over a naive prediction). Researchers attempting to rank or weight models must show that the choices are meaningful for the specific context. One approach supported by the climate science community is the *perfect model* assumption. In this framework, researchers assume one model to be the "truth," then, over a calibration interval, they evaluate prediction methods trained on simulated observations generated by the "true" model. Scientists discussed these issues at an IPCC Expert Meeting on Assessing and Combining Multi-Model Climate Projections.[8,9]

Tracking Climate Models

Our research applies machine learning algorithms to the problem of tracking the multi-model ensemble.[1,3,9] Our results on temperature data (observed and predicted temperature anomalies averaged over global, regional, annual, and monthly scales) show that our algorithm produces predictions that nearly match, and sometimes surpass, the results of the best model for the entire observation sequence. This is significant, because only in hindsight can one determine the best model for the whole observation sequence. We used online learning algorithms with the goal of making both real-time and future predictions. Moreover, our research shows that the naive "batch" approach has disadvantages due to the nonstationary nature of the observations and the relatively short history of model prediction data.

Climate scientists use temperature anomalies to express both the climate-model predictions and the true observations. A temperature anomaly is the difference between the observed temperature and the temperature at the same location at a fixed, benchmark time. To put it another way, anomalies are measurements of temperature change. Climate scientists use temperature anomalies because, while temperatures vary widely over geographical location, temperature anomalies typically vary less. For example, in a particular month it might be 80°F in New York, and 70°F in Toronto, but the anomaly from the benchmark time might be 1°F in both places. Thus, variance is lower when researchers average temperature anomalies over many geographic locations, than when they use absolute temperatures. Figure 2 shows climate model simulation runs, and observation data, averaged over many geographical locations, and many times in a year, yielding one value for a global mean temperature anomaly per year. In this case, researchers baselined the benchmark over the period 1951–1980 (one can convert between benchmark eras by subtracting a constant). The figure shows the climate model predictions we used as input to our global annual experiments, where the thick red line is the mean prediction over all models, in both plots. The thick blue line indicates the true observations.

We obtained our results on Tracking Climate Models (TCM)[1] by applying the Learn-α algorithm,[10] which tracks a shifting sequence of temperature values with respect to "expert" predictions, which we used to represent the climate models.[1] In our previous work,[10] bringing a view from probabilistic graphical models to bear on traditional algorithms for "online learning with expert advice,"[11,12] we re-derived such algorithms as Bayesian updates of a Hidden Markov Model (HMM), in which the identity of the current best expert is the hidden variable. This allowed us to introduce an algorithm that *learns* the switching rate between best experts, while simultaneously performing the original prediction task.[10] When we apply the Learn-α algorithm to the climate-model setting, the algorithm learns hierarchically, based on a set of generalized HMMs, where the hidden variable is the current best climate model's identity.

We ran experiments on NASA's historical temperature data (http://data.giss.nasa.gov/gistemp), averaged annually and globally, from 1900 through 2009, as well as the corresponding predictions of 20 different climate models per year (from the CMIP3 archive at www-pcmdi.llnl.gov/ipcc/about_ipcc.php). These model simulations started from an approximately stable climate in the 19th century, and were stepped forward using estimates of changes in the external drivers of climate change (greenhouse gases, volcanoes, atmospheric particulates, land-use changes, and so on). However, the model dynamics self-generate the month-to-month and year-to-year variability. The GCM output wasn't informed by observations, therefore it's valid to run historical experiments using the GCM ensemble predictively on historical data. We also ran experiments using climate model predictions through the year 2098, in order to harness the climate models' future predictions. Of course, there's no observation data in the future with which we could evaluate the machine learning algorithms. So, to achieve this goal, we ran future simulations using the scientific community's "perfect model" assumption; we fixed one climate model, then used its predictions as the quantity to learn based only on the remaining 19 climate models' predictions (and repeated this process 10 times).

We also ran experiments at higher spatial and temporal granularity. We used hindcasts of the IPCC global climate models and the analogous true observations, over specific geographical regions corresponding to several continents, at monthly and annual time scales. The predicted quantity was still a temperature anomaly. However, the data
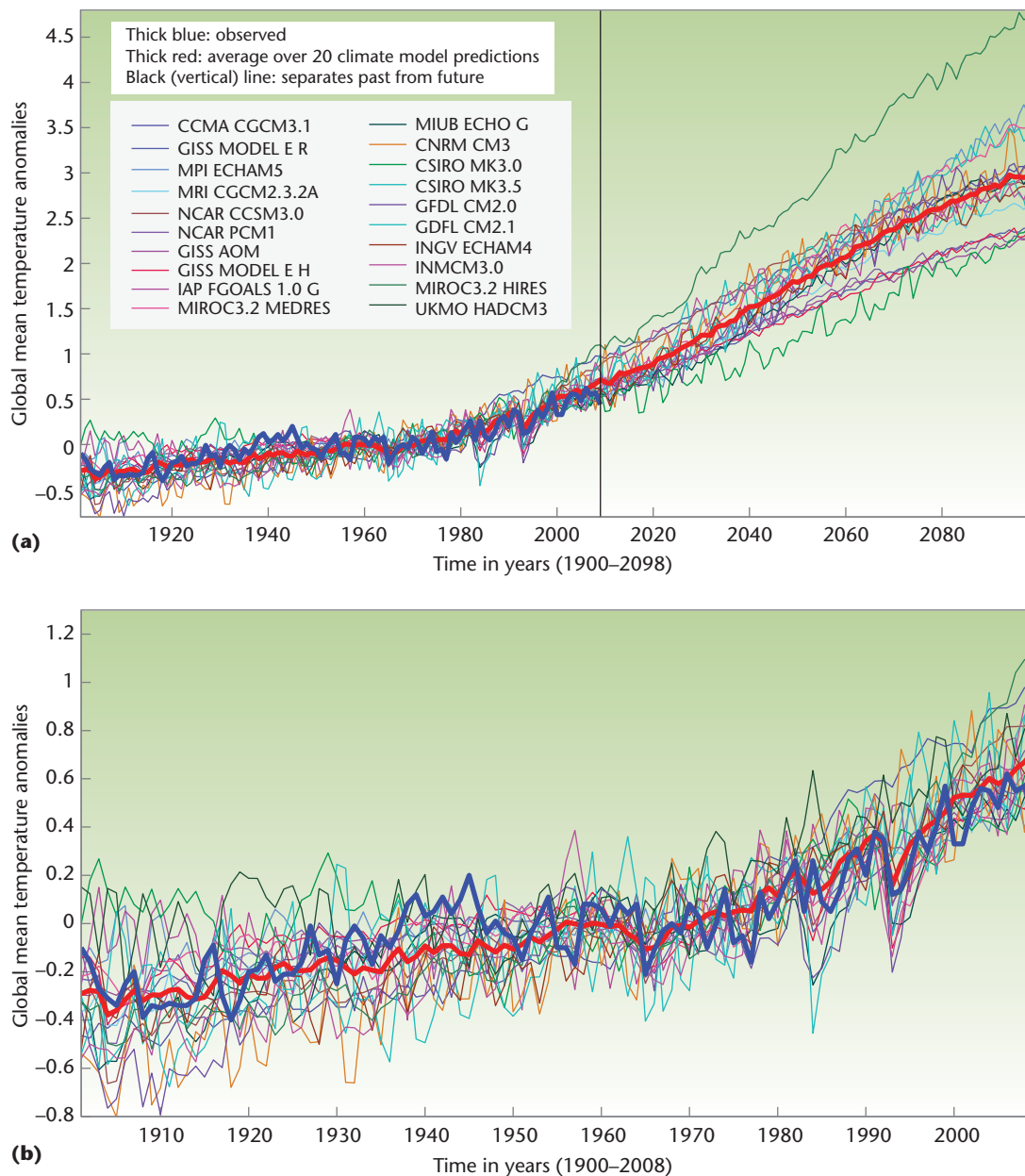
(a)

Thick blue: observed
Thick red: average over 20 climate model predictions
Black (vertical) line: separates past from future

Legend:
- CCMA CGCM3.1
- GISS MODEL E R
- MPI ECHAM5
- MRI CGCM2.3.2A
- NCAR CCSM3.0
- NCAR PCM1
- GISS AOM
- GISS MODEL E H
- IAP FGOALS 1.0 G
- MIROC3.2 MEDRES
- MIUB ECHO G
- CNRM CM3
- CSIRO MK3.0
- CSIRO MK3.5
- GFDL CM2.0
- GDFL CM2.1
- INGV ECHAM4
- INMCM3.0
- MIROC3.2 HIRES
- UKMO HADCM3

(b)

Figure 2. Global mean temperature anomalies. (a) Climate model predictions through 2098, with observations through 2008. The black vertical line separates past (hindcasts) from future predictions. (b) Here, we zoom in on observations and model predictions through 2008. The legends refer to both figures.

was averaged over a smaller geographical region than the whole globe; in particular, we ran experiments for latitude-longitude boxes corresponding to Africa, Europe, and North America. In addition to annual experiments, we also ran experiments using monthly averages in each of the regions. NASA provided observed data (http://data.giss.nasa.gov/gistemp) and the Koninklijk Nederlands Meteorologisch Instituut (KNMI) Climate Explorer (http://climexp.knmi.nl) provided the model-prediction data. Both model and observation data spanned from January 1900 through October 2010 (1,330 months). We also used monthly regional model predictions through the year 2098 to run six future simulations on 2,376 months (starting in 1900).

In every experiment we ran, Learn-α had a lower mean annual prediction error than the current default practice in climate science, which is to average over all the climate model predictions. Furthermore, Learn-α surpassed the best climate model's performance in all but two experiments (historical global annual

and historical monthly Africa). Even then, Learn-α nearly matched the performance of the best climate model. Similarly, Learn-α surpassed (batch) least-squares linear regression in all but two experiments (a global annual future simulation and a monthly future simulation for North America) and, again, its performance was still close. Learn-α's outperformance of batch linear regression on almost all experiments suggests that the data's nonstationary nature, coupled with the limited amount of historical data, poses challenges to a naive batch algorithm. Further experiments with a variety of different batch-learning algorithms would test this hypothesis (we recently achieved encouraging results using sparse matrix completion, an unsupervised batch technique).[13]

The plots in Figure 3 show squared error between predictions and (simulated) observations, from 1900–2098, on a future simulation using global annual temperature anomalies. We plot Learn-α's learning curve against the best and worst climate models' performance (from the remaining ensemble, computed in hindsight), and the performance of the average prediction over the ensemble of remaining climate models. Learn-α successfully predicts one climate model's predictions up to the year 2098, which is notable, because future predictions vary widely among the climate models. We ran 10 future simulations with global annual temperature anomalies, each with a different climate model providing the simulated observations. In each simulation, Learn-α suffers less prediction error than the mean over the remaining models' predictions, on 75–90 percent of the years. Figure 2 shows a marked fan-out among the model predictions that increases into the future. Over time, the model predictions' mean performance diverges from most individual model trajectories. In the historical global annual experiment, Learn-α suffers less prediction error than the model predictions' mean for more than 75 percent of the years.

Geospatially Tracking Climate Models

Previous work provided techniques to combine the predictions of the multi-model ensemble, at various geographic scales, by considering each geospatial region as an independent problem.[1,7] However, climate patterns across the globe often vary significantly and concurrently, so assuming that each geospatial region is independent could limit the performance of these previous approaches. We therefore extended our work on the TCM algorithm as follows:[1]

- We used a richer modeling framework that accounts for GCM predictions at higher geospatial resolutions.
- Using a nonhomogeneous HMM, we modeled the neighborhood influence among geospatial regions.
- We ran experiments to validate these extensions' effectiveness.

We proposed a new algorithm: Neighborhood-augmented Tracking Climate Models (NTCM).[3] This algorithm extends the TCM algorithm to operate in a setting where the GCM's predictions are assessed at a higher spatial resolution. NTCM takes into account regional neighborhood influences when it forms predictions. The NTCM algorithm is fully described elsewhere,[3] and differs from TCM in two main ways:

- We modified the Learn-α algorithm to include influence from a geospatial region's neighbors in how the algorithm updates the weights over experts (the multi-model ensemble of GCMs' predictions in that geospatial region).
- Our master algorithm runs multiple instances of this modified Learn-α algorithm simultaneously, each on a different geospatial region, and uses their predictions to make a combined global prediction.

We modified the time-homogeneous HMM that generates the TCM algorithm[1,10] in order to model the neighborhood influence.[3] We instead use dynamically updated (nonhomogeneous) transition dynamics (the probabilistic model of how the best climate model's identity changes over time). These dynamics depend on a geospatial neighborhood scheme: the set of nearby regions that influence the region in question. Researchers can use a variety of shapes and sizes to define the neighborhood scheme.

We first used a simple neighborhood scheme in which the four immediately adjacent regions (north, south, east, and west) are the geographical region's possible neighbors.[3] We ran experiments with our algorithm on historical data, using temperature observations and GCM hindcasts. We obtained historical
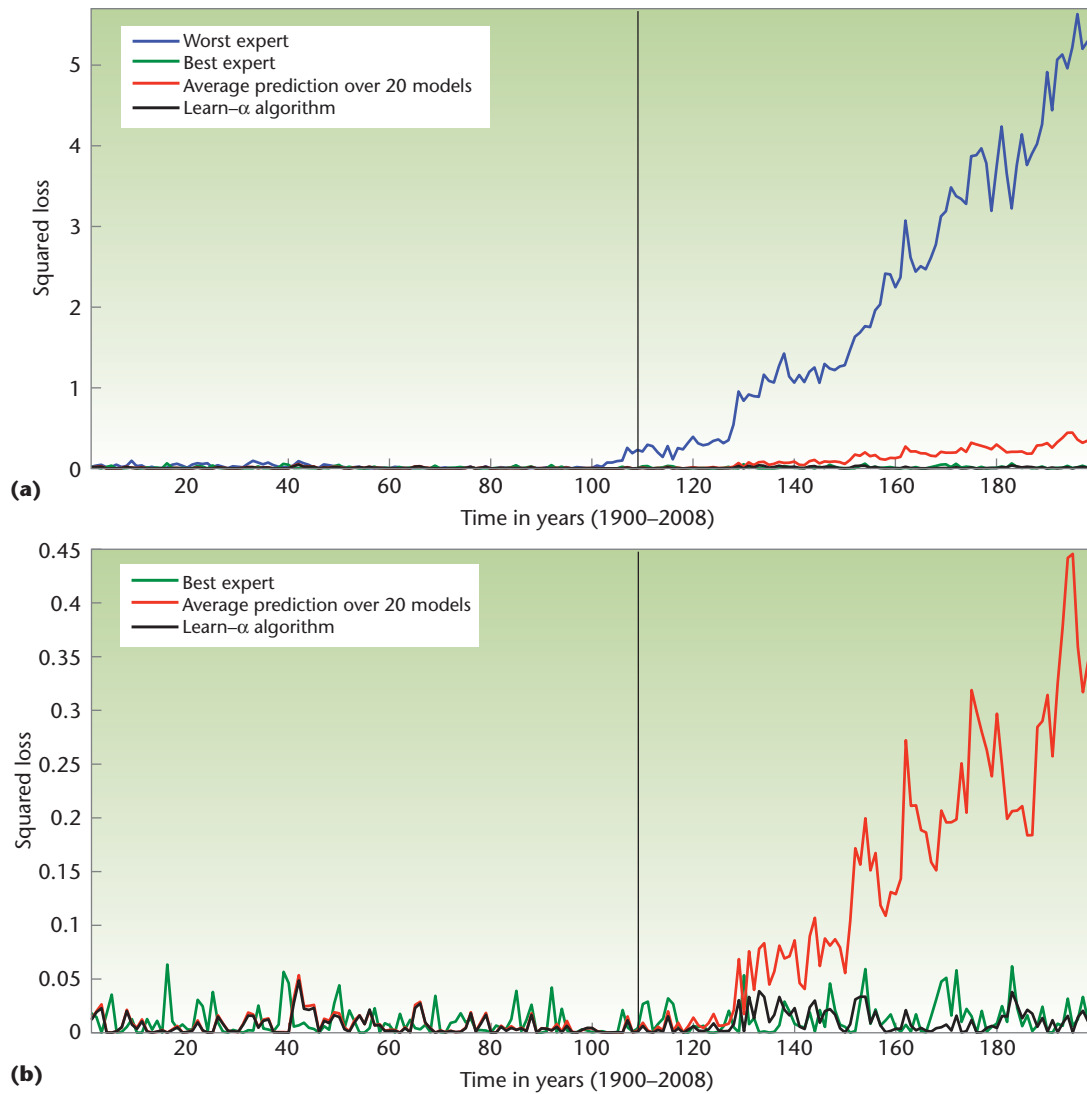
(a)

(b)

Figure 3. (a) The squared error between predictions and (simulated) observations, from 1900–2098, on a future simulation using global annual temperature anomalies. This shows the algorithm tracking the predictions of one climate model using the predictions of the remaining 19 as input, with no true temperature observations. The simulated observations are from the best-performing climate model from the whole ensemble (previously computed on historical data). The black vertical line separates past from future. (b) Here, the graph zooms in on the *y* axis.

climate model data from the CMIP3 archive (www-pcmdi.llnl.gov/ipcc/about_ipcc.php) using the Climate of the 20th Century Experiment (20C3M). Figure 4 compares the new algorithm's performance over time: NTCM (indicated in red and blue) using 45-degree square regions, versus global Learn-α (as in the original TCM algorithm,[1] indicated in black) in a graph illustrating cumulative annual prediction error. This graph indicates that, for most years, and in particular for years later in the time-sequence, the NTCM algorithm's cumulative global prediction error was less than that of the global Learn-α algorithm

used in TCM, with NTCM's $\beta = 1$ variant (full neighborhood influence) obtaining lower prediction error than that of the $\beta = 0$ variant (no neighborhood influence).

Challenge Problems for Climate Informatics

Climate scientists are working on many different kinds of problems for which machine learning, and other computer science expertise, could potentially have a big impact. Here, we provide a brief description of a few examples (with a discussion of related work in the literature) that typify these ideas, although any
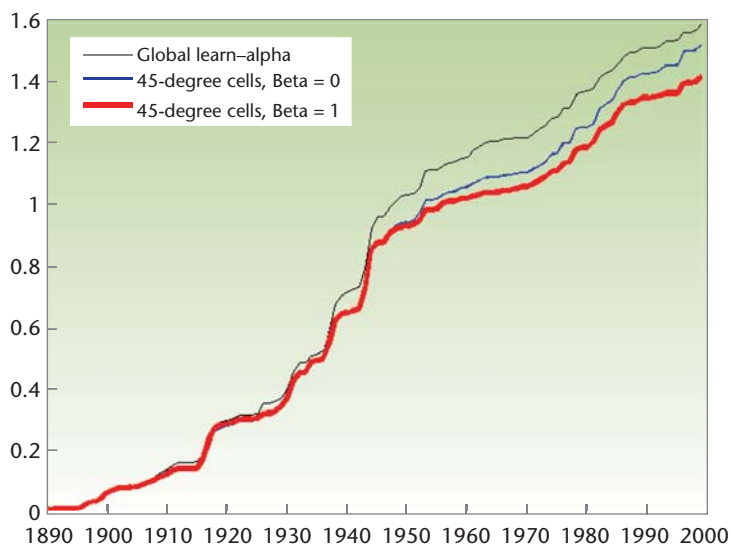
Figure 4. Comparison of the Neighborhood-augmented Tracking Climate Models (NTCM) algorithm's performance over time. The cumulative annual prediction error of NTCM is shown, using 45-degree square cells, compared to Learn-α's prediction error.

specific implementation mentioned shouldn't be considered the last word.

Improving Multi-Model Ensemble Predictions

As previously discussed, researchers have developed and are improving multiple climate models; currently there are about 25 centers across the globe, many with multiple modeling groups. Each model shares some basic features with some of the other models, but generally researchers design and independently implement unique models. In coordinated "Model Intercomparison Projects" (MIPs)—most usefully, the Coupled MIP (CMIP3, CMIP5), the Atmospheric Chemistry and Climate MIP (ACCMIP), and the PaleoClimate MIP (PMIP3)—modeling groups attempt to perform analogous simulations with similar boundary conditions, but with multiple models. These multi-model ensembles offer the possibility to assess what features are robust across models. They also facilitate the study of the roles of internal variability, structural uncertainty, and scenario uncertainty in assessing predictions at different time and space scales. Finally, MIPs provide multiple opportunities for model-observation comparisons. Questions of interest include the following:

- Are there "skill" metrics for present or past model simulations that are useful for future predictions?

- Are there weighting strategies that maximize predictive skill? How would researchers explore this?

Weather and seasonal forecasts also raise these questions, but because of the long time scales involved in climate prediction, they're more difficult for climate researchers to address.[8,14] Our research provides an example of how researchers can use the MME to predict climate change.[1,3,9,13]

Parameterization Development

Climate models need to be able to model the relevant physics at all scales, even those finer than any finite model can currently resolve. Examples include cloud formation, turbulence in the ocean, land surface heterogeneity, ice floe interactions, and chemistry on dust particle surfaces, to name a few. Typically, scientists, using physical intuition and limited calibration data, dealt with these phenomena by using parameterizations (physically coherent approximations of the bulk effects) that attempted to capture the phenomenology of a specific process, and its sensitivity in terms of the (resolved) large scales. As observational data become more available, and direct numerical simulations of key processes become more tractable, the potential for machine learning and data mining techniques to help define and automate new parameterizations and frameworks is increasing. For example, some researchers have used neural network frameworks to develop atmospheric radiation models for use in GCMs.[15]

Paleo Reconstructions

Understanding how climate varied in the past, before the onset of widespread instrumentation, is of great interest to climate scientists. The climate changes seen in the Paleo record dwarf those in the 20th century, and hence could provide insight into the significant changes we expect this century. However, Paleo data are even sparser than instrumental data, and they aren't usually directly commensurate with the instrumental record. Paleo records (such as water isotopes, tree rings, pollen counts, and so on) could indicate climate change by proxy, but often nonclimatic influences affect their behavior, and sometimes their relationships to more standard variables (such as temperature or precipitation) are nonstationary or convolved. Scientists face an enormous challenge in bringing together disparate, multiproxy evidence to discover

38

COMPUTING IN SCIENCE & ENGINEERING

large-scale patterns of climate change—or, in contrast, in building models with enough "forward modeling" capability that they can use the proxies directly as modeling targets.[16]

Data Assimilation and Initialized Decadal Predictions

The main way in which sparse observational data is used to construct complete fields is through *data assimilation*. This field of research concerns how to update physics-based models with observed data, and includes such technology as the ensemble Kalman filter.[17] Data assimilation is a staple of weather forecasts, and the various re-analyses in the atmosphere and ocean. In many ways, this is the most sophisticated use of the combination of models and observations, but its use in improving climate predictions is still in its infancy. For weather time scales, this works well. For longer-term forecasts (seasons to decades) the key variables are in the ocean, not the atmosphere; climate scientists have yet to fully develop a climate model initalization in which the evolution of ocean variability models the real world in useful ways.[18,19] Climate scientists find the early results intriguing, if not convincing, and many more examples are slated to come online in the new CMIP5 archive.[20]

Advances in data assimilation could also benefit other areas of computer science. In robotics, for example, when a robot uses a physics-based model for the dynamics governing its movement, often it must also incorporate information gleaned from the onboard sensors, and, in some cases, additional real-time instructions from a human controller.

Developing and Understanding Perturbed Physics Ensembles (PPE)

The spread among different model predictions from different modeling groups is one way to measure the models' "structural uncertainty." However, we can't consider these models a controlled random sample from the space of all plausible models. An approach that leads to a more accurately characterized ensemble is to take a single model, and vary multiple (uncertain) parameters within the code, generating a family of similar models that nonetheless sample a good deal of the intrinsic uncertainty that arises when we choose any specific set of parameter values. Researchers have successfully used these Perturbed Physics Ensembles (PPEs)

in the Climateprediction.net and Quantifying Uncertainty in Model Predictions (QUMP) projects to generate controlled model ensembles that they can systematically compare to observed data, and then make inferences.[21,22] However, designing such experiments and efficiently analyzing sometimes thousands of simulations is a challenge, but one which is increasingly going to be attempted.

We hope that this article encourages future work, not only on some of the challenge problems proposed here, but also on new problems in climate informatics. A huge and varied amount of climate data is available, providing a rich and fertile playground for future machine learning and data mining research. Even exploratory data analysis could prove useful for accelerating discovery. There are countless collaborations possible at this intersection of climate science and machine learning, data mining, and statistics. We strongly encourage future progress on a range of emerging problems in climate informatics. ᴄꜱᴇ

References

1. C. Monteleoni et al., "Tracking Climate Models," *Statistical Analysis and Data Mining*, vol. 4, no. 4, 2011, pp. 372–392.
2. C. Monteleoni et al., "Climate Informatics," *Computational Intelligent Data Analysis for Sustainable Development; Data Mining and Knowledge Discovery Series*, CRC Press, 2013, pp. 81–126.
3. S. McQuade and C. Monteleoni, "Global Climate Model Tracking Using Geospatial Neighborhoods," *Proc. 26th AAAI Conf. Artificial Intelligence,* AAAI, 2012, pp. 335–341.
4. G.A. Schmidt et al., "Present Day Atmospheric Simulations Using GISS ModelE: Comparison to In-Situ, Satellite and Reanalysis Data," *J. Climate*, 2006, vol. 19, no. 2, pp. 153–192.

5. S. Solomon et al., *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change,* Cambridge Univ. Press, 2007.

6. T. Reichler and J. Kim, "How Well Do Coupled Models Simulate Today's Climate?," *Bull. Am. Meteorological Soc.,* vol. 89, no. 3, 2008, pp. 303–311; http://dx.doi.org/10.1175/BAMS-89-3-303.

7. C. Reifen and R. Toumi. "Climate Projections: Past Performance No Guarantee of Future Skill?" *Geophysical Research Letters,* vol. 36, no. 13, 2009; doi:10.1029/2009GL038082.

8. R. Knutti et al., *Expert Meeting on Assessing and Combining Multi Model Climate Projections: Good Practice Guidance Paper on Assessing and Combining Multi Model Climate Projections,* meeting report, Intergovernmental Panel on Climate Change (IPCC), 2010; www.ipcc-wg2.gov/meetings/EMs/IPCC_EM_MME_GoodPracticeGuidancePaper.pdf.

9. C. Monteleoni, S. Saroha, and G. Schmidt, "Can Machine Learning Techniques Improve Forecasts?" *Intergovernmental Panel on Climate Change (IPCC) Expert Meeting on Assessing and Combining Multi-Model Climate Projections,* presentation, IPCC, 2010.

10. C. Monteleoni and T. Jaakkola, "Online Learning of Non-Stationary Sequences," *Proc. Advances in Neural Information Processing Systems,* 2003; http://books.nips.cc/papers/files/nips16/NIPS2003_LT04.pdf.

11. N. Littlestone and M.K. Warmuth, "The Weighted Majority Algorithm," *Proc. IEEE Symp. Foundations of Computer Science,* IEEE CS, 1989, pp. 256–261.

12. M. Herbster and M.K. Warmuth, "Tracking the Best Expert," *Machine Learning,* vol. 32, no. 2, 1998, pp.151–178.

13. M. Ghafarianzadeh and C. Monteleoni, "Climate Prediction via Matrix Completion," *Proc. 27th Conf. Artificial Intelligence, Late-Breaking Papers Track,* AAAI, 2013, pp. 35–37.

14. C. Tebaldi and R. Knutti, "The Use of the Multi-Model Ensemble in Probabilistic Climate Projections in Probabilistic Climate Projections," *Philosophical Transactions of the Royal Soc. A,* vol. 365, no. 1857, 2007, pp. 2053–2075.

15. V.M. Krasnopolsky and M.S. Fox-Rabinovitz, "Complex Hybrid Models Combining Deterministic and Machine Learning Components for Numerical Climate Modeling and Weather Prediction," *Neural Networks,* vol. 19, no. 2, 2006, pp. 122–134; doi:10.1016/j.neunet.2006.01.002.

16. G.A. Schmidt, A. LeGrande, and G. Hoffmann, "Water Isotope Expressions of Intrinsic and Forced Variability in a Coupled Ocean-Atmosphere Model," *J. Geophysical Research,* vol. 112, no. D10, 2007; doi:10.1029/2006JD007781.

17. G. Evensen, "Sequential Data Assimilation with a Nonlinear Quasi-Geostrophic Model Using Monte Carlo Methods to Forecast Error Statistics," *J. Geophysical Research-All Series,* vol. 99, no. C5, 1994, pp. 10143–10162.

18. N.S. Keenlyside et al., "Advancing Decadal-Scale Climate Prediction in the North Atlantic Sector," *Nature,* vol. 453, no. 7191, 2008, pp. 84–88.

19. 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. 769–799.

20. G.A. Meehl et al., "Global Climate Projections," *Climate Change 2007: The Physical Science Basis,* tech. report, S. Solomon et al., eds., Cambridge Univ. Press, 2007; www.ipcc.ch/publications_and_data/ar4/wg1/en/ch10.html.

21. R. Knutti et al., "Constraining Climate Sensitivity from the Seasonal Cycle in Surface Temperature," *J. Climate,* vol. 19, no. 17, 2007, pp. 4224–4233.

22. J.M. Murphy et al., "A Methodology for Probabilistic Predictions of Regional Climate Change from Perturbed Physics Ensembles," *Philosophical Trans. Royal Soc. A,* vol. 365, no. 1857, 2007, pp. 1993–2028.

Claire Monteleoni *is an assistant professor of computer science at George Washington University. Her research interests include machine learning algorithms and theory for problems including learning from data streams, raw (unlabeled) data, private data, and climate informatics. Monteleoni has a PhD in computer science from MIT. Contact her at cmontel@gwu.edu.*

Gavin A. Schmidt *is the deputy director at the NASA Goddard Institute for Space Studies. His main research interest lies in understanding climate variability, both its internal variability and its response to external forces. Schmidt has a PhD in applied mathematics from University College London. Contact him at gavin.a.schmidt@nasa.gov.*

Scott McQuade *is a lead sensors systems engineer at MITRE Corporation and a doctoral candidate in computer science at George Washington University. His research interests include online learning, spatial learning, and algorithm development. McQuade has an MS in computer science from George Washington University. Contact him at mcquade@gwmail.gwu.edu.*

cn *Selected articles and columns from IEEE Computer Society publications are also available for free at http://ComputingNow.computer.org.*