5th International Workshop on Climate Informatics
September 24-25, 2015
Hosted by the National Center for Atmospheric Research in Boulder, CO

TRACKING SEASONAL PREDICTION MODELS

Timothy DelSole[1,2], Claire Monteleoni[3], Scott McQuade[3], Michael K. Tippett[4,5], Kathleen Pegion[1,2], J. Shukla[1,2]

***Abstract*—A machine learning algorithm for combining predictions is applied to seasonal predictions of the NINO3.4 index from six coupled atmosphere-ocean models. The algorithm adaptively tracks a dynamic sequence of "best experts" and produces a probability that a particular expert is best. Averaging based on this probability effectively yields a multi-model prediction. The algorithm gives seasonal predictions that are more skillful than any individual model and better than the multi-model mean.**

I. MOTIVATION

Today, many institutions routinely predict the monthly averaged weather. The question arises as to how to *combine* these forecasts to produce a superior forecast product. Numerous studies have shown that removing the model labels and pooling the resulting forecasts often produces accurate and probabilistically reliable forecasts that are at least as good, or better, than more sophisticated combination methods [1], [2], [3], [4], [5].

Here we explore a machine learning algorithm that tracks a sequence of estimates of the best model at each time step. The algorithm, due to Monteleoni and Jaakkola [6], has been applied previously to *climate* simulations and shown to produce better hindcasts, in a mean square sense, than predicting with the average over model predictions [7]. Here we show that the same algorithm also produces good *seasonal* predictions.

II. METHOD

Consider M model predictions of some observable. Let i_t be a discrete random variable that identifies the "best" model at time t. Let θ_t denote the set of *all* observations and predictions up to and including time t. One can derive a Bayesian estimate of i_t based on all data up to but not including t; i.e., $p(i_t|\theta_{t-1})$. This estimate involves assumptions about how the best model changes with time. At the lowest level, the algorithm assumes a first order Markov system characterized by a transition probability $p(i_{t+1}|i_t;\alpha)$, where α parameterizes the one-step transition probabilities (i.e., degree of non-stationarity). The desired distribution can be derived recursively from Bayes theorem as

$$p(i_{t+1}|\theta_t) = \frac{\sum_{i_t} p(i_{t+1}|i_t;\alpha)e^{-L(i_t,t)}p(i_t|\theta_{t-1})}{Z_t}, \quad (1)$$

where Z_t is a normalization constant and $L(i_t,t)$ is the negative log-likelihood of the data at time t given that the best model is i_t. The scheme is initialized using $p(i_1|\theta_0) = 1/M$.

Following [7], the likelihood $L(i_t,t)$ is assumed to be proportional to the squared difference between observation and the prediction by model i_t at time t. Also, the transition probability follows the `Fixed-share` algorithm [8] in that the best model at the next step has probability $1-\alpha$ of remaining the best, and probability $\alpha/(M-1)$ of transitioning to another model.

If α is chosen far from the level of non-stationarity appropriate for the system that generated the data, then the algorithm will perform poorly [6]. Monteleoni and Jaakkola proposed the `Learn-`α algorithm for learning the value of this parameter simultaneously with learning the best model. The algorithm effectively performs an ensemble of updates, each member using a fixed value of α. The resulting α-ensemble is treated as another Bayesian estimation problem, except now the problem is to identify the best value of α. This latter algorithm assumes a self-transition probability of 1 and hence has no further parameters to specify. A pseudo-code for the entire algorithm has been published [7]. The cumulative loss (i.e., the cumulative squared error) of the above algorithm can be bounded relative to the cumulative loss of the fixed-α algorithm that uses the optimal value of α, which can only be known in hindsight (i.e. after having seen all the data) [6]. This bound does not depend on any distributional assumptions about the forecast errors.

III. EVALUATION

The above algorithm was applied to seasonal hindcasts from state-of-the-art coupled atmosphere-ocean models in the North American Multi-model Ensemble (NMME)

[1]George Mason University, Fairfax, VA. [2]Center for Ocean-Land-Atmosphere Studies, George Mason University, Fairfax, VA. [3]George Washington University, Washington, DC. [4]Columbia University, New York, NY. [5] Center of Excellence for Climate Change Research, Department of Meteorology, King Abdulaziz University, Jeddah, Saudi Arabia. Corresponding author: Timothy DelSole, tdelsole@gmu.edu.

[9]. We analyze models that currently are used for forecasting and a period for which complete hindcasts are available from the IRI portal, namely hindcasts during 1982-2010 from: National Centers for Environmental Prediction (CFSv2), Canadian Centre for Climate Modelling and Analysis (CanCM3 and CanCM4), the Geophysical Fluid Dynamics Laboratory (CM2p1-aer04), National Aeronautics and Space Administration (NASA), and a joint collaboration between the Center for Ocean-Land-Atmosphere Studies, University of Miami, and the National Center for Atmospheric Research (CCSM3). Anomalies were computed with respect to the climatological mean for the periods before and after 1998 separately, for reasons discussed in [10]. A model hindcast is the mean of 6 ensemble members from that model. We consider monthly mean NINO3.4 hindcasts. The verification was obtained from version 2 of the NOAA Optimum Interpolation SST [11].

Figure 1a shows that for 2.5 month lead the Learn-α algorithm produces a smaller cumulative squared error ("loss") than any individual model and than the multi-model model mean. The algorithm can perform better than the multi-model mean because it assigns larger weights to better models. For example, inspection of the weights (fig. 1b) reveals that the algorithm favors different models, or model clusters, for different years up until the 1990s, but thereafter gradually gives most weight to the single model that has the least loss (CanCM3).

We observe no improvement over the multi-model mean for leads greater than 3 months (see fig. 1c). One reason might be that NINO3.4 predictions have a strong seasonal character, a fact that is not explicitly taken into account by the algorithm. Moreover, for lead times greater than one month, the performance of previous month's forecast is not available because the verification needed to compute a loss still lies in the future. Therefore, the update at time t is based on the loss at t-minus-lead. We are currently exploring ways to account explicitly for seasonal variations and to use short lead errors to inform weights at long leads.

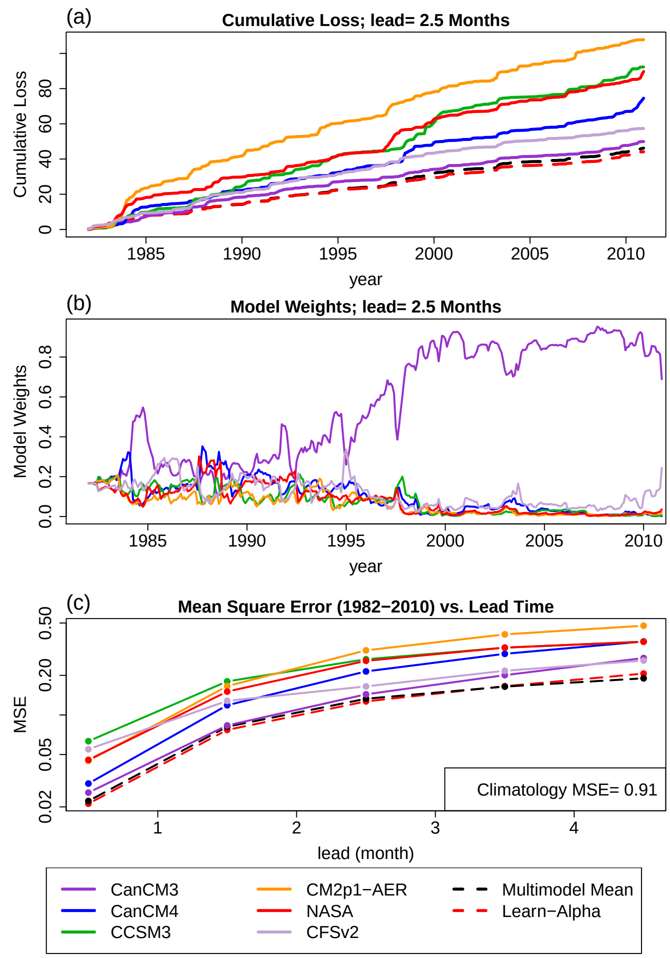

Fig. 1. (a) Cumulative loss (i.e., cumulative squared error) for hindcasts of monthly mean NINO3.4 at lead 2.5 months by individual NMME models (solid), multi-model mean (black dashed), and Learn-α (red dashed). (b) The model weights determined by Learn-α. (c) Mean square error versus lead time. The color legend is indicated in the bottom panel.

REFERENCES

[1] V. V. Kharin and F. W. Zwiers, "Climate predictions and multimodel ensembles," *J. Climate*, vol. 15, pp. 793–799, 2002.

[2] P. Peng, A. Kumar, H. van den Dool, and A. G. Barnston, "An analysis of multimodel ensemble predictions for seasonal climate anomalies," *J. Geophys. Res.*, pp. 18–1 to 18–12, 2002. doi:10.1029/2002JD002712.

[3] R. Hagedorn, F. J. Doblas-Reyes, and T. N. Palmer, "The rationale behind the success of multi-model ensembles in seasonal forecasting - I. Basic concept," *Tellus A*, vol. 57, no. 3, pp. 219–233, 2005.

[4] T. DelSole, "A Bayesian framework for multimodel regression," *J. Climate*, vol. 20, pp. 2810–2826, 2007.

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

[6] C. Monteleoni and T. Jaakkola, "Online learning of nonstationary sequences," in *NIPS'03: Advances in Neural Information Processing Systems*, vol. 16, 2003.

[7] C. Monteleoni, G. A. Schmidt, S. Saroha, and E. Asplund, "Tracking climate models," *Statistical Analysis and Data Mining*, vol. 4, no. 4, pp. 372–392, 2011.

[8] M. Herbster and M. K. Warmuth, "Tracking the best expert," *Machine Learning*, vol. 32, pp. 151–178, 1998.

[9] B. P. Kirtman et. al, "The North American Multi-Model Ensemble (NMME): Phase-1 seasonal to interannual prediction, Phase-2 toward developing intra-seasonal prediction," *Bull. Am. Meteor. Soc.*, vol. 95, pp. 585–601, 2014.

[10] A. G. Barnston and M. K. Tippett, "Predictions of Nino3.4 SST in CFSv1 and CFSv2: A Diagnostic Comparison," *Clim. Dyn.*, vol. 41, pp. 1–19, 2013.

[11] R. W. Reynolds, N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, "An improved in situ and satellite SST analysis for climate," *J. Climate*, vol. 15, pp. 1609–1625, 2002.