8th International Workshop on Climate Informatics
September 19-21, 2018
CI 2018 Hosted by the National Center for Atmospheric Research in Boulder, CO

AN AI APPROACH TO DETERMINING TIME OF EMERGENCE OF CLIMATE CHANGE

Elizabeth A. Barnes[1], Chuck Anderson[2], Imme Ebert-Uphoff[3]

***Abstract*—We investigate a new approach to quantify the time of emergence of climate change using artificial neural networks (ANNs). The basic idea is to use artificial intelligence methods to predict, given a global (or regional) temperature map averaged over a particular year, the year to which the temperature map belongs. We use NCAR's CESM model data from 1920 to 2100 for that purpose. We expected that the ANN prediction model would perform badly for early years (due to large internal variability), and that accuracy would improve drastically in the 21st century when climate change creates a strong, dominant signal, which may then indicate a time of emergence. Performing this prediction experiment for three different regions, namely the entire globe, the US, and a central region of the US, yielded very interesting first results, which are presented and discussed here.**

I. INTRO

The time of emergence (ToE) of climate change is defined as the time at which the signal of climate change clearly stands out over the noise of natural climate variability, making it an important quantity for climate projections, risk assessments and policy decisions. Many different ideas have emerged of how to define and quantify ToE [1], most of which are based on strictly statistical analyses. Examples include using a particular signal-to-noise ratio cutoff [2], applying the t-statistic or Kolmogorov-Smirnov test to test for significant differences between time periods [3], [4], or utilizing the standard error of the regression to estimate the lead-time required for a linear trend to emerge from natural variability [5].

Motivated by recent advances in using artificial intelligence (AI) to autonomously detect complex patterns in many different settings (ranging from face recognition to detection of extreme weather events [6], [7]), we propose a very different approach to defining the ToE of climate change. Specifically, we ask the following two questions, "In which year does climate change become apparent to an AI?" and "Which spatial patterns are detected as dominant signatures of this climate change signal?" To address these questions we train several artificial neural networks (ANNs) to predict the year corresponding to an annual-mean temperature field taken from a climate model simulation, specifically from 35 ensemble members of NCAR's CESM-LE model. These 35 simulations are all runs from the same model (the CESM model) under identical forcings for 1920-2100. The only difference between ensemble members is that each is started with slightly different initial conditions, and these tiny differences lead to large differences in internal climate variability, or climate "noise". Thus, one can think of each ensemble member as one possible evolution of earth's climate, and in this context, the observed evolution of earth's climate as yet another ensemble member. Due to this experimental setup, the climates of all ensemble members evolve independently from one another except for their similarities in the forcings imposed (e.g. greenhouse gas emissions). Thus, the AI is trained to distinguish the climate change signal (in the form of predicting the year) from that of background climate noise.

II. DATA

We use annual-mean global surface temperature data from 35 ensemble members of the Community Earth System Model–Large Ensemble (CESM-LE) model developed by NCAR [8]. The model is run under identical forcings (RCP 8.5) from 1920-2100, providing $181 \times 35 = 6,335$ annual temperature maps, representing 181 different years. The resulting maps have a resolution of 0.94° in latitude and 1.25° in longitude. We reduce the resolution by a factor of 5 in both latitude and longitude, by averaging temperature values over disjoint 5x5 groups of grid points. The resulting maps have a coarse resolution of 4.71° in lat and 6.25° in longitude. Each map is "labeled" with its year.

Corresponding author: Elisabeth A. Barnes, eabarnes@atmos.colostate.edu. [1]Atmospheric Science. [2]Computer Science. [3]Electrical and Computer Engineering. All at Colorado State University, Fort Collins, CO, USA.

8th International Workshop on Climate Informatics
September 19-21, 2018
Hosted by the National Center for Atmospheric Research in Boulder, CO

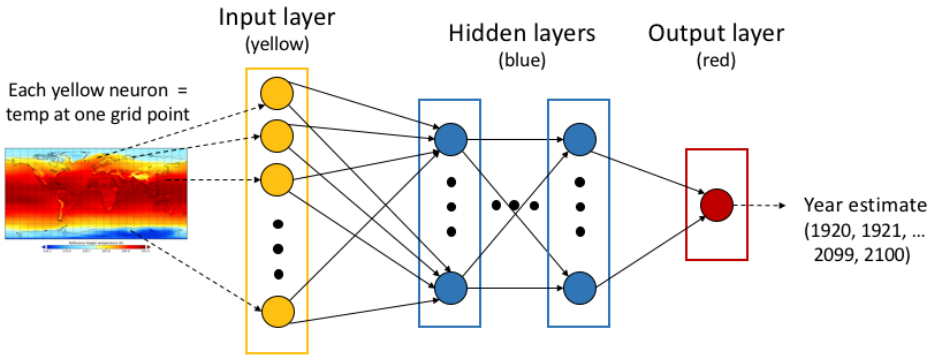

Fig. 1. Schematic of Neural Network Architecture

III. METHOD

We first feed the training data, consisting of annual maps with year labels from a subset of ensemble members, into a neural network, and train the network *to estimate the year*, given the temperature map. This is followed by the testing phase, where we evaluate how well the system estimates the year when given maps from the test set of the remaining ensemble members.

Fig. 1 shows the architecture of the neural network employed here. (For an excellent introduction to neural networks, see this online book [9].) Fig. 1 shows the architecture of the neural network employed here. (For an excellent introduction to neural netoworks, see this online book [9].) This neural network consists of several layers that each have one or more computational units. Each unit of the input layer (yellow) represents the temperature of one grid point of the input map. The input layer is followed by 0, 1 or more hidden layers (blue) that contain auxiliary units. The output layer (red) consists of a single unit representing the estimated year. In this architecture, each unit in one layer receives input from all units in the preceding layer, and its value is determined as a weighted sum of all the incoming values (from the preceding layer), followed by the non-linear *tanh* transformation. The output layer does not use the non-linear transformation. This architecture is often referred to as a fully-connected network, with a linear output unit. We trained the neural network to minimize the squared error in the predicted year using Møller's Scaled Conjugate Gradient algorithm [10]. Note that when the number of hidden layers is 0 then this model is a simple linear regression model, with the output being a linear weighted combination of the values at the grid points. We include this regression model as baseline method for comparison.

The motivation of our ANN method is that studying the estimation error for the different years should provide insights into when the climate change signal becomes so significant that the network can accurately estimate the true year. Furthermore, by analyzing the reasoning of the network we can learn about the dominant spatial patterns of climate change that the AI utilizes to improve the estimation.

IV. EXPERIMENTS

We conducted experiments for a large variety of different set-ups that differ by 1) region; 2) preprocessing of the data; 3) ANN model used; and 4) number of training iterations used. We first define these parameters and the options considered in the experiments.

1) Regions: Experiments are conducted using maps of three different regions, namely spanning the entire globe; the US (lat=[24°,60°],lon=[235°,290°]); and a central region of the US (lat=[36°,48°],lon=[253°,272°]). Options: *global*, *US* or *regional*.

2) Preprocessing: For each region we use the temperature maps in two different forms, once in its original form and once with the regional annual mean subtracted, i.e. we calculate the average of the annual temperature over the considered region and subtract that value for all points in the map. Subtracting the regional means prevents the models from solely relying on the mean temperature increase for the year estimates. Options: *original* or *mean removed*.

3) Model: Models differ by how many hidden layers there are, and the number of units in each layer. They are denoted as arrays, where each number denotes the number of units in one hidden layer. For example, '[5]' means there is one hidden layer with 5 units, '[10,10,10]' means there are 3 hidden layers with 10 units each. The number '0' means that there is no hidden layer, i.e. the model is a linear regression

8th International Workshop on Climate Informatics
September 19-21, 2018
Hosted by the National Center for Atmospheric Research in Boulder, CO

model. Options: 0, [5], [10], [10,10], [10,10,10,10,10].

4) Number of training iterations: The number of iterations used to train a neural network is an important parameter. Too few iterations in the training can prevent the model from converging to a good solution. Too many iterations can lead to overfitting, which results in poor generalization, i.e. the model may look great on training data, but performs poorly on test data. The optimal number of iterations strongly depends on the number of free parameters in the system - more free parameters require more iterations. Options: 100, 1,000 or 2,000.

Measuring Performance: To evaluate performance we split the CESM data into two disjoint sets, a training set and a test set. The data of 28 ensemble members (5,068 maps) is used for training of the neural network, while the remaining 7 ensemble members (1,267 maps) are set aside for its subsequent testing. Setting aside entire ensemble members for testing ensures that *the model cannot simply guess the year of a map based on having been trained on nearby year maps of the same ensemble member*. We perform this entire procedure 5 times. Namely, the 35 ensemble members are put into random order at the very beginning, then divided into five groups of seven consecutive ensemble members. Each of the five groups serves as test data for one experiment, and the remaining ensemble members are used for training in that experiment. Finally, the performance measure for a given set-up is then defined as the RMSE error of the year estimates over all samples of the five test data sets, which contain all maps for all ensemble members exactly once. Thus the RMSE is taken over all 6,335 maps.

V. RESULTS

We trained networks for all combinations of the four parameters described above. Table I shows the best linear and neural network models as judged by the RMSE for the five testing sets - for global, US and regional data, and when using original and regional mean removed data. The results in Table I show that both linear models and neural networks perform surprisingly well for the global maps, even if the regional mean is removed. In fact, the RMSE for the best models is only 3-4 years, i.e. both types of models are generally only a few years off in their estimates. As one would expect, the accuracy of the estimates diminishes quickly when the map size is reduced. Namely, the neural networks (linear models) achieve an RMSE of only 11 (15) years for US maps and only 20 (22) years for regional maps, so estimates are off by one or two decades. Furthermore, the neural networks consistently outperform the linear model by a small margin for global and regional maps and by a larger margin for US maps.

Category	Model	# iter.	RMSE Train	**RMSE Testing**
GLOBAL - Original				
Best linear	0	1,000	2.54	4.41
Best NN	[10,10]	2,000	1.62	3.78
GLOBAL - Mean removed				
Best linear	0	1,000	2.54	4.40
Best NN	[10,10]	1,000	2.33	3.83
US - Original				
Best linear	0	2,000	14.92	15.17
Best NN	[10,10]	1,000	9.70	11.23
US - Mean removed				
Best linear	0	100	14.92	15.17
Best NN	[10,10,10,10,10]	1,000	10.18	11.37
REGIONAL - Original				
Best linear	0	2,000	22.24	22.27
Best NN	[5]	1,000	19.91	20.11
REGIONAL - Mean removed				
Best linear	0	2,000	22.24	22.27
Best NN	[5]	1,000	19.91	20.16

TABLE I
BEST LINEAR AND BEST NEURAL NETWORK MODEL FOR DIFFERENT SCENARIOS , BASED ON RMSE TESTING ERROR.

Fig. 2 shows the typical RMSE in each category separated by year. The plot confirms the trends in Table I, but additionally shows for which years the models perform well. Using global data the models perform consistently well, but with a noticeable drop in error starting around the year 2000. Using US data, the models perform badly in early years, but fairly well for later years, with the error for the original data dropping a few years earlier, and more significantly, than for the case with mean removed. Finally, using the regional data, the models become fairly useless when the mean is removed. In fact, further investigation of the model for Regional Mean Removed reveals that the model does not pick up any significant signal in the early years, and instead predicts a constant year near 2000, i.e. it completely avoids the lower boundary. A similar effect occurs at the upper boundary. Thus, the RMSE reported in Table I does not tell the full story, as it makes it seem as if the models for US and regional perform with an error of similar orders of magnitude (11 vs. 20 years error), when indeed these models are extremely different - the regional model performs poorly throughout, while the US one only performs poorly early on, then latches onto the climate change

8th International Workshop on Climate Informatics
September 19-21, 2018
Hosted by the National Center for Atmospheric Research in Boulder, CO

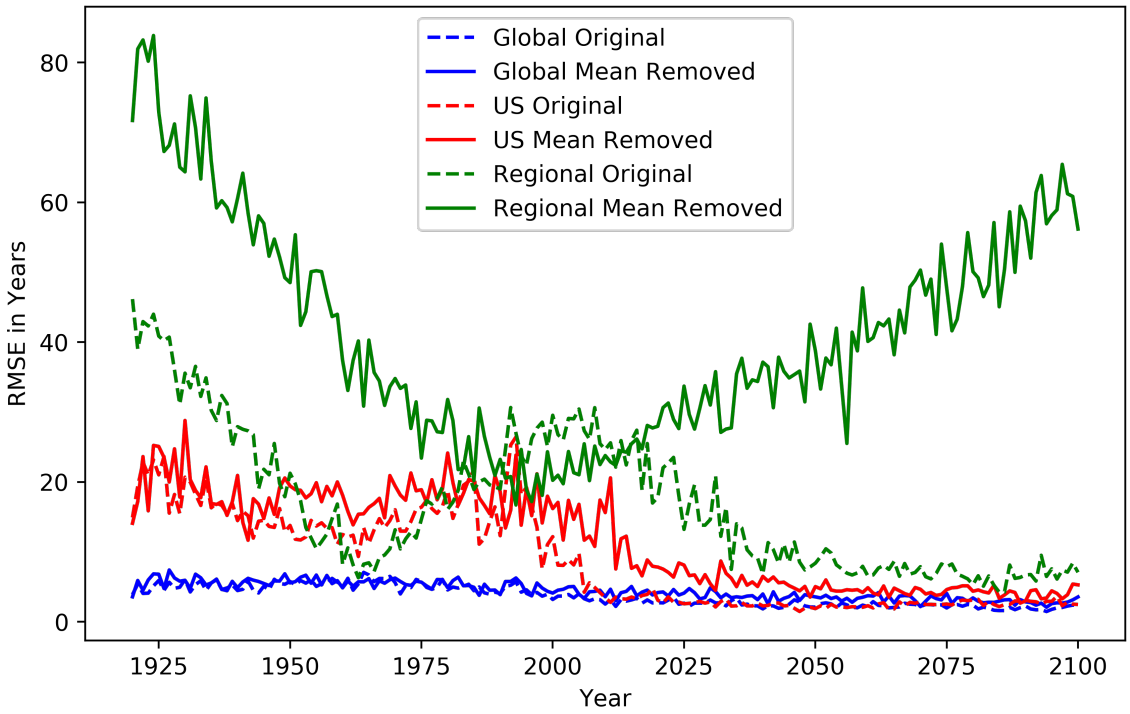

Fig. 2. Typical RMSE Testing - separated by years - for global, US and regional maps, with and without mean removed.

signal. In summary, we find that it should be possible to define ToE based on both Global and US data error curves for original and mean removed (each yielding a different year and a different interpretation), but not for the regional data with mean removed.

VI. Conclusions and Future Work

The fact that the models (even the linear ones) tended to perform well even when the spatial mean was removed suggests that there are spatial patterns, or "fingerprints" of climate change, being utilized by the models. Thus, the next step of our analysis is to identify and study the spatial patterns the various models use for their predictions, e.g. what are the key spatial features that make it so easy for the models to estimate the year of the global maps with such high accuracy? In addition, we will work toward defining a time of emergence of the climate signal from these results.

References

[1] IPCC, *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Intergovernmental Panel on Climate Change 2013, Cambridge University Press, Cambridge, 2013, pg. 283.

[2] E. Hawkins and R. Sutton, "Time of emergence of climate signals," *Geophysical Research Letters*, vol. 39, p. L01702, Jan. 2012.

[3] G. Zappa, B. J. Hoskins, and T. G. Shepherd, "Improving Climate Change Detection through Optimal Seasonal Averaging: The Case of the North Atlantic Jet and European Precipitation," *Journal of Climate*, vol. 28, pp. 6381–6397, May 2015.

[4] I. Mahlstein, G. Hegerl, and S. Solomon, "Emerging local warming signals in observational data," *Geophysical Research Letters*, vol. 39, p. L21711, Nov. 2012.

[5] J. Li, D. W. Thompson, E. A. Barnes, and S. Solomon, "Quantifying the lead time required for a linear trend to emerge from natural climate variability," *Journal of Climate*, vol. 30, no. 24, pp. 10179–10191, 2017.

[6] Y. Liu, E. Racah, J. Correa, A. Khosrowshahi, D. Lavers, K. Kunkel, M. Wehner, W. Collins, *et al.*, "Application of deep convolutional neural networks for detecting extreme weather in climate datasets," *arXiv preprint arXiv:1605.01156*, 2016.

[7] A. McGovern, K. L. Elmore, D. J. Gagne, S. E. Haupt, C. D. Karstens, R. Lagerquist, T. Smith, and J. K. Williams, "Using artificial intelligence to improve real-time decision-making for high-impact weather," *Bulletin of the American Meteorological Society*, vol. 98, no. 10, pp. 2073–2090, 2017.

[8] J. Kay, C. Deser, A. Phillips, A. Mai, C. Hannay, G. Strand, J. Arblaster, S. Bates, G. Danabasoglu, J. Edwards, *et al.*, "The community earth system model (cesm) large ensemble project: A community resource for studying climate change in the presence of internal climate variability," *Bulletin of the American Meteorological Society*, vol. 96, no. 8, pp. 1333–1349, 2015.

[9] M. A. Nielsen, *Neural networks and deep learning*. Determination Press, 2015. Free online book. Available at http://neuralnetworksanddeeplearning.com/.

[10] M. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural networks*, vol. 6, no. 4, pp. 525–533, 1993.