Improving Seasonal Forecasting Using Probabilistic Deep Learning

Donald D. Lucas
Baoxiang Pan, Gemma Anderson, André Goncalves, Céline Bonfils, and Jiwoo Lee

28 March 2023
Deep learning is becoming competitive with dynamical seasonal forecast systems

Skill in predicting October-March average precipitation initialized from July

Observational period 1982-2012

Which map is from NOAA’s North American Multi-Model Ensemble (NMME), and which is from deep learning of climate models?
This presentation provides an overview of our recent JAMES paper.

RESEARCH ARTICLE
10.1029/2021MS002766

Special Section:
Machine learning application to Earth system modeling

Key Points:

Improving Seasonal Forecast Using Probabilistic Deep Learning

Baoxiang Pan¹, Gemma J. Anderson¹, André Goncalves¹, Donald D. Lucas¹, Céline J. W. Bonfils¹, and Jiwoo Lee¹

¹Lawrence Livermore National Laboratory, Livermore, CA, USA
Our JAMES paper was motivated by a project to improve seasonal predictions of water resources in California.

Persistent climate patterns steer storms toward or away from California. E.g., the “Ridiculously Resilient Ridge”

Can deep learning exploit low frequency signals in the climate system to improve seasonal predictions?
Two recent papers suggested deep learning applied to climate simulations may be useful

[Geophysical Research Letters]

Diagnosing Secular Variations in Retrospective ENSO Seasonal Forecast Skill Using CMIP5 Model-Analogs

Hui Ding, Matthew Newman, Michael A. Alexander, and Andrew T. Wittenberg

Step 1: Search from existing climate simulations whose SST is close to observed SST
Step 2: Apply their following season’s states as analog forecasts

[Deep learning for multi-year ENSO forecasts]

Yoo-Geun Ham, Jeong-Hwan Kim, and Jing-Jia Luo

Step 1: Take predictor–predictand pairs from climate simulations.
Step 2: Build a convolutional neural network regression model.
Step 3: Apply few observations for fine–tuning.
Step 4: Apply the fine–tuned model for forecast.

[Anomaly correlation maps and conditional probability plots]
Two recent papers suggested deep learning applied to climate simulations may be useful

**Geophysical Research Letters**

**Diagnosing Secular Variations in Retrospective ENSO Seasonal Forecast Skill Using CMIP5 Model-Analogs**

Hui Ding, Matthew Newman, Michael A. Alexander, and Andrew T. Wittenberg

**Key Points**
- Seasonal tropical Indo-Pacific

**Steps**
1. Search from existing climate simulations whose SST is close to observed SST
2. Apply their following season's states as analog forecasts

**Anomaly correlation**

- NMME-model analog (SST)
- NMME-model analog (precip)
- CMIP5 Best-10 analog (SST)
- CMIP5 Best-10 analog (precip)

**Deep learning for multi-year ENSO forecasts**

Yoo-Geun Ham, Jeong-Hwan Kim & Jing-Jia Luo

**Nature** 573, 568–572(2019) | [Cite this article]

**Steps**
1. Take predictor–predictand pairs from climate simulations.
2. Build a convolutional neural network regression model.
3. Apply few observations for fine-tuning.
4. Apply the fine-tuned model for forecast.
Deep learning (DL) is a promising approach to help improve climate predictions

- Deep neural networks can discover patterns or “features” automatically that humans cannot
- Can keep memory effects and rich spatial structure
- Observational data can be used more effectively (“data-driven” approach)
- Can fully leverage pre-existing data
- Offers huge speed up in time and resources
- Complements traditional methods
Probabilistic approaches are required for seasonal forecasts

**Dynamical forecasts**
- Initialize from different states
- Incorporate multiple models
- Limited sample size due to computational costs

**Deep learning forecasts**
- Initialize at different points in continuous runs to learn seasonal dependencies
- Train on multiple climate models
- Use generative/variational methods to represent distributions
We developed the *Conditional Generative Forecasting (CGF)* system for seasonal forecasts

**CGF is built using**

*Convolutional neural networks (CNNs)*
- From computer vision
- Designed for maps and images

*Autoencoders / U-Nets*
- Effective at extracting low-dimensional patterns and features
- Non-linear generalization of Empirical Orthogonal Functions

*Conditional variational methods*
- Approximate probability distribution function (PDF) of the input / output data
- Enable drawing plausible samples from the PDF

**Inputs:** maps of ocean heat content in July (8 layers, 5-100 m)  
**Outputs:** maps of Oct-March average precipitation or air temperature
Autoencoders are related to Empirical Orthogonal Functions

- EOFs are a technique used for extracting spatial and spatiotemporal patterns.
- EOFs are also known as Principal Component Analysis (PCA)

Consider a set of monthly mean sea level pressure (SLP) maps

EOFs characterize ~80% of the variability in the North Atlantic with 3 modes instead of 968 grid cells (22 lat by 44 lon)

North Atlantic Oscillation

EOF1 (53.3%)
EOF2 (16.5%)
EOF3 (11.6%)
EOF analysis is a form of unsupervised learning

Example EOF analysis

- Trained EOFs on 350 images from the Olivetti face database
- Each image has 64 x 64 pixels (4096 total pixels)
- Retained 100 EOFs (explain 57.7% of the variance)
EOF analysis is a form of unsupervised learning

Example EOF analysis

- Trained EOFs on 350 images from the Olivetti face database
- Each image has 64 x 64 pixels (4096)
- Retained 100 EOFs (explain 57.7% of the variance)
Autoencoders are related to Empirical Orthogonal Functions

EOFs

- The “O” in EOF implies the components in the reduced space are linearly related.
- Independent components are available for interpretation.
- Fast to train and relatively few samples needed, but many components may be needed for noisy data.
Autoencoders are related to Empirical Orthogonal Functions

Autoencoders

- Neural networks transform the data.
- Latent space features can be non-linearly related.
- Interpretability can be challenging.
- Many samples needed and slow to train.
- Easy to augment latent space with additional information.
- Autoencoders with linear activation functions == EOFs.
Conditional Generative Forecasting (CGF) model

Encoding

Ocean heat ($X$)
Input: $8 \times 90 \times 180$ [~130,000]
Latent space: $128 \times 6 \times 12$ [~9,000]

Precipitation -or- Surface temperature ($Y$)
Input: $1 \times 90 \times 180$ [~16,000]
Latent space: $128 \times 6 \times 12$ [~9,000]

Embedding

GCM information ($M$) is concatenated into the latent space with a learned embedding vector.

Probabilistic Forecasts

The latent space $z$ is represented by a Gaussian distribution that depends on $X$ and $M$, and a mapping function that relates $Y$ to $z$. 
Conditional Generative Forecasting (CGF) model

Encoding

*Ocean heat (X)*
Input: 8 x 90 x 180  [~130,000]
Latent space: 128 x 6 x 12  [~9,000]

*Precipitation -or- Surface temperature (Y)*
Input: 1 x 90 x 180  [~16,000]
Latent space: 128 x 6 x 12  [~9,000]

Embedding

GCM information (M) is concatenated into the latent space with a learned embedding vector

Probabilistic Forecasts

The latent space $z$ is represented by a Gaussian distribution that depends on $X$ and $M$, and a mapping function that relates $Y$ to $z$. 
**Conditional Generative Forecasting (CGF) model**

**Encoding**

*Ocean heat (X)*

Input: $8 \times 90 \times 180$ [~130,000]
Latent space: $128 \times 6 \times 12$ [~9,000]

*Precipitation -or- Surface temperature (Y)*

Input: $1 \times 90 \times 180$ [~16,000]
Latent space: $128 \times 6 \times 12$ [~9,000]

**Embedding**

GCM information (M) is concatenated into the latent space with a learned embedding vector

**Probabilistic Forecasts**

The latent space $z$ is represented by a Gaussian distribution that depends on $X$ and $M$, and a mapping function that relates $Y$ to $z$. 
CGF used 52,201 simulated years from 30 climate models

- Built using control + pre-1982 historical simulations
- Tested using 30 years of historical simulations (1982-2012)
Evaluation Strategy

Dynamical Forecasts

- North American Multi-Model Ensemble (NMME)
- Forecast anomalies of precipitation and 2-m temperature
- Four models, 10 ensemble members each (CCSM4, CESM1, CanCM4, GFDL CM2.5)

“Observations”

- For comparison: precipitation from the Global Precipitation Climatology Project (GPCP)
- For comparison: temperature from ECMWF Reanalysis v5 (ERA5)
- For CGF input: ocean heat from ECMWF Ocean Reanalysis (ORAS5)

Metrics to assess forecast skill

- Deterministic:
  1. anomaly correlation coefficient
  2. normalized root mean square skill
- Probabilistic:
  1. area under ROC curve
  2. continuous ranked probability score
Evaluation Strategy

Dynamical Forecasts
- North American Multi-Model Ensemble (NMME)
- Forecast anomalies of precipitation and 2-m temperature
- Four models, 10 ensemble members each (CCSM4, CESM1, CanCM4, GFDL CM2.5)

“Observations”
- For comparison: precipitation from the Global Precipitation Climatology Project (GPCP)
- For comparison: temperature from ECMWF Reanalysis v5 (ERA5)
- For CGF input: ocean heat from ECMWF Ocean Reanalysis (ORAS5)

Metrics to assess forecast skill
- Deterministic:
  1. anomaly correlation coefficient
  2. normalized root mean square skill
- Probabilistic:
  1. area under ROC curve
  2. continuous ranked probability score
Evaluation Strategy

Dynamical Forecasts

- North American Multi-Model Ensemble (NMME)
- Forecast anomalies of precipitation and 2-m temperature
- Four models, 10 ensemble members each (CCSM4, CESM1, CanCM4, GFDL CM2.5)

“Observations”

- For comparison: precipitation from the Global Precipitation Climatology Project (GPCP)
- For comparison: temperature from ECMWF Reanalysis v5 (ERA5)
- For CGF input: ocean heat from ECMWF Ocean Reanalysis (ORAS5)

Metrics to assess forecast skill

- Deterministic:
  1. anomaly correlation coefficient
  2. normalized root mean square skill
- Probabilistic:
  1. area under ROC curve
  2. continuous ranked probability score
CGF is competitive with dynamical seasonal forecast systems

Forecast skill for 2-m temperature for CGF\textsubscript{CanESM}

- *Conditioned* on a single GCM (CanESM)
- Anomaly correlation coefficients for October-March average initialized from July ocean heat

CGF\textsubscript{CanESM} for CanESM simulations  
CGF\textsubscript{CanESM} for observations  
CanCM4 for observations

---

model simulations vs. observations  
deep learning vs. dynamical system
CGF is competitive with dynamical seasonal forecast systems

Forecast skill for precipitation for CGF_{CanESM}

- *Conditioned* on a single GCM (CanESM)
- Anomaly correlation coefficients for October-March average initialized from July ocean heat

CGF_{CanESM} for CanESM simulations

CGF_{CanESM} for observations

CanCM4 for observations

---

model simulations vs. observations

deep learning vs. dynamical system
CGF is competitive with dynamical seasonal forecast systems

Single model ensemble forecasts for CGF_{CanESM}

- 2-m air temperature
- Spatially averaged ACC

- precipitation
- Spatially averaged ACC

- Ensemble size

Deep learning vs. dynamical

Dynamical size limit

CGF_{CanESM} for CanESM simulations

CGF_{CanESM} for observations

CanCM4 for observations
CGF is competitive with *multi-model ensemble* dynamical seasonal forecast systems

Forecast skill for precipitation for CGF ensemble

- Considers multiple models and multiple initializations per model
- Anomaly correlation coefficients for October-March average initialized from July ocean heat

**CGF_{Ens} for real-world forecast**

- Thirty dynamical models, ten ensemble members each

**CGF_{M} for real-world forecast**

- Optimal dynamical model ($\tilde{M}$) with 300 realizations

**NMME for real-world forecast**

- Four dynamical models, ten ensemble members each
CGF’s latent space can be interpolated, sampled, and explored

**t-SNE for CGF precipitation**
(t-distributed stochastic neighbor embedding)

https://douglasduhaime.com/posts/visualizing-latent-spaces.html
Summary and conclusions

• Deep learning can help with climate modeling and seasonal forecasting.
  • e.g., autoencoders are non-linear generalizations of Empirical Orthogonal Function analysis
  • generative models can capture distributions of data for probabilistic models and ensembles

• We built a deep learning-based Conditional Generative Forecast (CGF) model.
  • Uses conditional variational autoencoder technology.
  • Skillful in predicting Oct-Mar precipitation and temperature initialized from July ocean heat content (relative to dynamical ensemble forecasts).
  • Incorporates probability distributions of the initial state and multi-model ensembles.

• Work is on-going to better understand sources of predictability at the seasonal time scale using deep learning systems.
  • E.g., using saliency maps and analysis

• See Pan et al (2021) in JAMES for related work using Generative Adversarial Networks to bias correct climate projections.
Architecture of CGF

(a) Structure of the Conditional Generative Forecasting (CGF) model

(b) Residual block: $R_C$

(c) Transposed residual block: $T_C$

(d) Neural network operators

- Convolution
- BatchNorm
- ReLU
- Max Pooling
- Plus
- Catenate
- Transpose convolution
- Reparameterization
- Embedding