LECTURE 7:
Neural network approaches

ML-4430: Machine learning approaches in climate science

9 June 2021

Weather Forecast Postprocessing

1

- Rasp & Lerch, 2018
- ~~Scher & Messori, 2018~~
- ~~Grönquist et al., 2020~~

Climate Model Parametrisation

2

- Krasnopolsky, Fox-Rabinovitz & Belochitski, 2013
- ~~Rasp, Pritchard & Gentine, 2018~~
- ~~Yuval, O'Gorman & Hill, 2021~~

Data-Driven Climate Modeling

3

- Scher, 2018
- ~~Scher & Messori, 2020~~
- ~~Rasp & Thuerey, 2021~~

ENSO Forecasting

4

- Ham, Kim & Luo, 2019
- ~~Mahesh et al., 2019~~
- ~~Cachay et al., 2021~~

NOVEMBER 2018 RASP AND LERCH 3885

SPECIAL
Waves to Weather (W2W)
COLLECTION

Neural Networks for Postprocessing Ensemble Weather Forecasts

STEPHAN RASP

Meteorological Institute, Ludwig-Maximilians-Universität, Munich, Germany

SEBASTIAN LERCH

Institute for Stochastics, Karlsruhe Institute of Technology, Heidelberg Institute for Theoretical Studies, Karlsruhe, Germany

(Manuscript received 23 May 2018, in final form 14 August 2018)

Main idea ...

- Weather model output typically is biased
  - Trends, wet / dry or warm / cold biases

- To remove this bias we need to infer a relation between the weather forecast and the real observations

- Traditionally, this has been done using regression models

- Here, the authors use neural networks

Data

- Forecast data:
  - THORPEX Interactive Grand Global Ensemble (TIGGE) dataset
  - Forecasts for surface stations in Germany at lead times of 48 h
  - ECMWF 50-member ensemble forecasts initialized at 0000 UTC every day
  - Upscaled onto a 0.58 3 0.58 grid
  - 2 m temperature, plus auxiliary variables

- Observation data:
  - 537 weather stations in Germany
  - 2 m temperature only

- Training: 2007 – 2015 & 2015 only
- Test: 2016

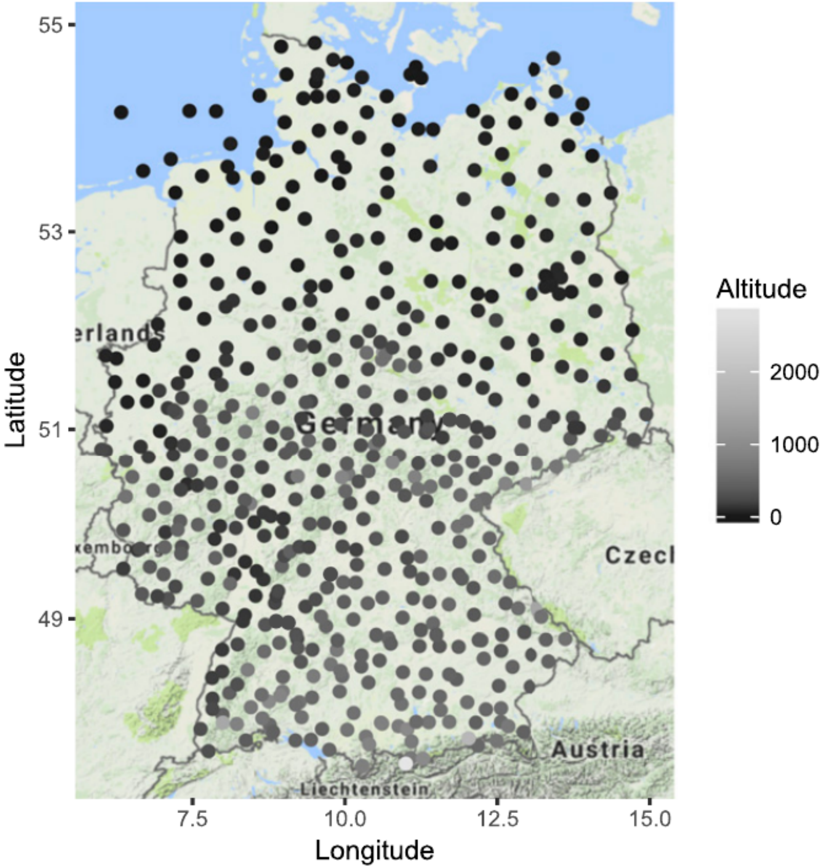


FIG. 2. Locations of DWD surface observation stations. The grayscale values of the points indicate the altitude (m).

TABLE 1. Abbreviations and descriptions of all features.

Feature	Description
Ensemble predictions (mean and std dev)	
t2m	2-m temperature
cape	Convective available potential energy
sp	Surface pressure
tcc	Total cloud cover
sshf	Sensible heat flux
slhf	Latent heat flux
u10	10-m *U* wind
v10	10-m *V* wind
d2m	2-m dewpoint temperature
ssr	Shortwave radiation flux
str	Longwave radiation flux
sm	Soil moisture
u_pl500	*U* wind at 500 hPa
v_pl500	*V* wind at 500 hPa
u_pl850	*U* wind at 850 hPa
v_pl850	*V* wind at 850 hPa
gh_pl500	Geopotential at 500 hPa
q_pl850	Specific humidity at 850 hPa
Station-specific information	
station_alt	Altitude of station
orog	Altitude of model grid point
station_lat	Lat of station
station_lon	Lon of station

$$y_{s,t} \,\Big|\, \mathbf{X}_{s,t}^{\mathrm{t2m}} \sim \mathcal{N}_{(\mu_{s,t}, \sigma_{s,t})}$$

the top part of Table 1 are combined with station-specific features in the bottom part, and aggregated into a vector of predictors $\mathbf{X}_{s,t} \in \mathbb{R}^p$, $p = 42$. Further, we write $\mathbf{X}_{s,t}^{\mathrm{t2m}}$ to denote the vector of predictors that only contains the mean value and standard deviation of the 2-m temperature forecasts.

TABLE 1. Abbreviations and descriptions of all features.

Feature	Description
Ensemble predictions (mean and std dev)	
t2m	2-m temperature
cape	Convective available potential energy
sp	Surface pressure
tcc	Total cloud cover
sshf	Sensible heat flux
slhf	Latent heat flux
u10	10-m *U* wind
v10	10-m *V* wind
d2m	2-m dewpoint temperature
ssr	Shortwave radiation flux
str	Longwave radiation flux
sm	Soil moisture
u_pl500	*U* wind at 500 hPa
v_pl500	*V* wind at 500 hPa
u_pl850	*U* wind at 850 hPa
v_pl850	*V* wind at 850 hPa
gh_pl500	Geopotential at 500 hPa
q_pl850	Specific humidity at 850 hPa
Station-specific information	
station_alt	Altitude of station
orog	Altitude of model grid point
station_lat	Lat of station
station_lon	Lon of station

1. Weather Forecast Postprocessing → Rasp & Lerch, 2018

Model	Description
Raw ensemble	
Benchmark postprocessing methods	
EMOS-gl	Global EMOS
EMOS-loc	Local EMOS
EMOS-loc-bst	Local EMOS with boosting
QRF	Local quantile regression forest
Neural network models	
FCN	Fully connected network
FCN-aux	. . .with auxiliary predictors
FCN-emb	. . .with station embeddings
FCN-aux-emb	. . .with both of the above
NN-aux	One-hidden-layer NN with auxiliary predictors
NN-aux-emb	. . .and station embeddings

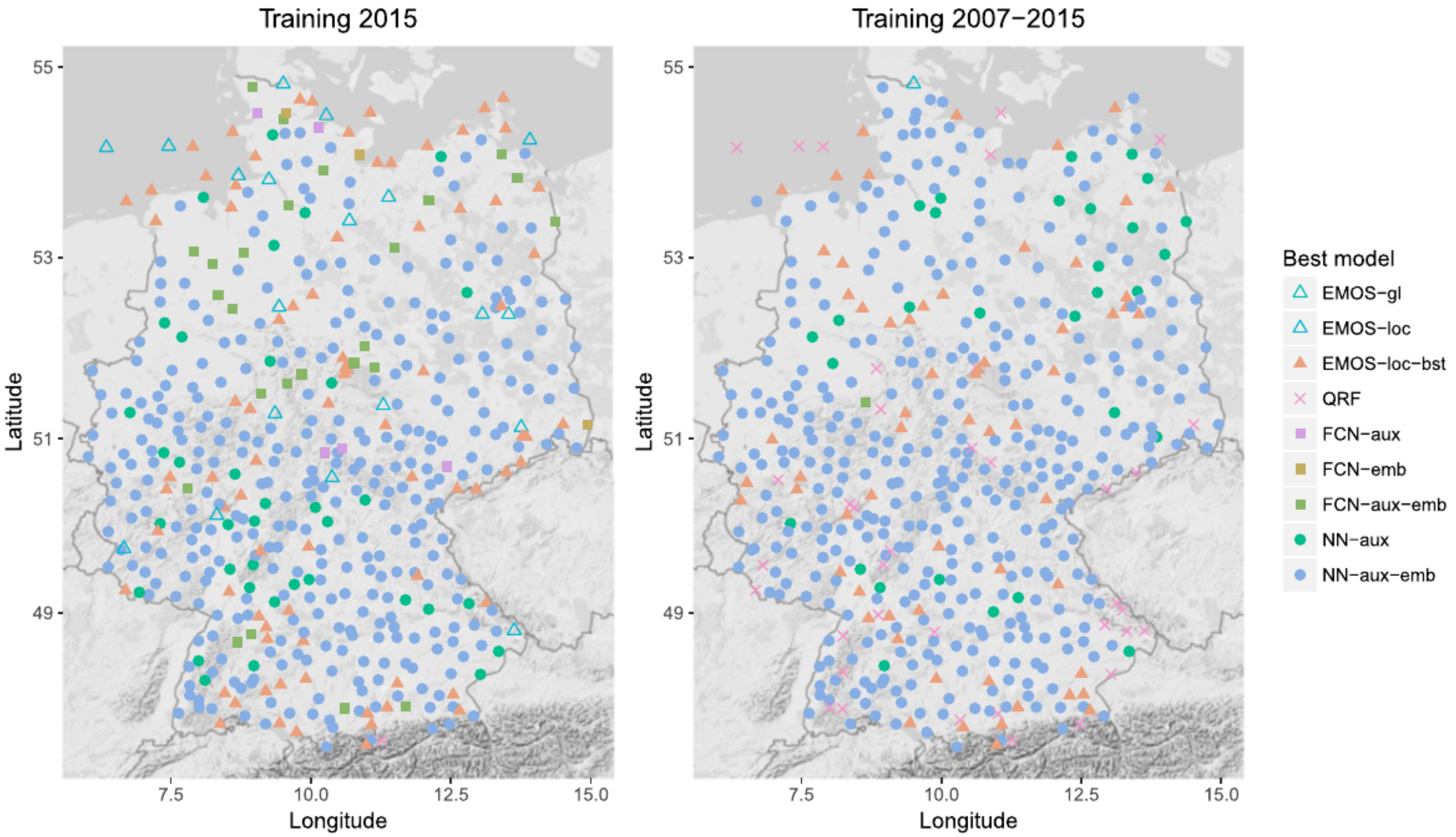


FIG. 4. Observation station locations color coded by the best performing model (in terms of mean CRPS over calendar year 2016) for models trained on data from (left) 2015 and (right) 2007 to 2015. Point shapes indicate the type of model.

1. Weather Forecast Postprocessing → Rasp & Lerch, 2018

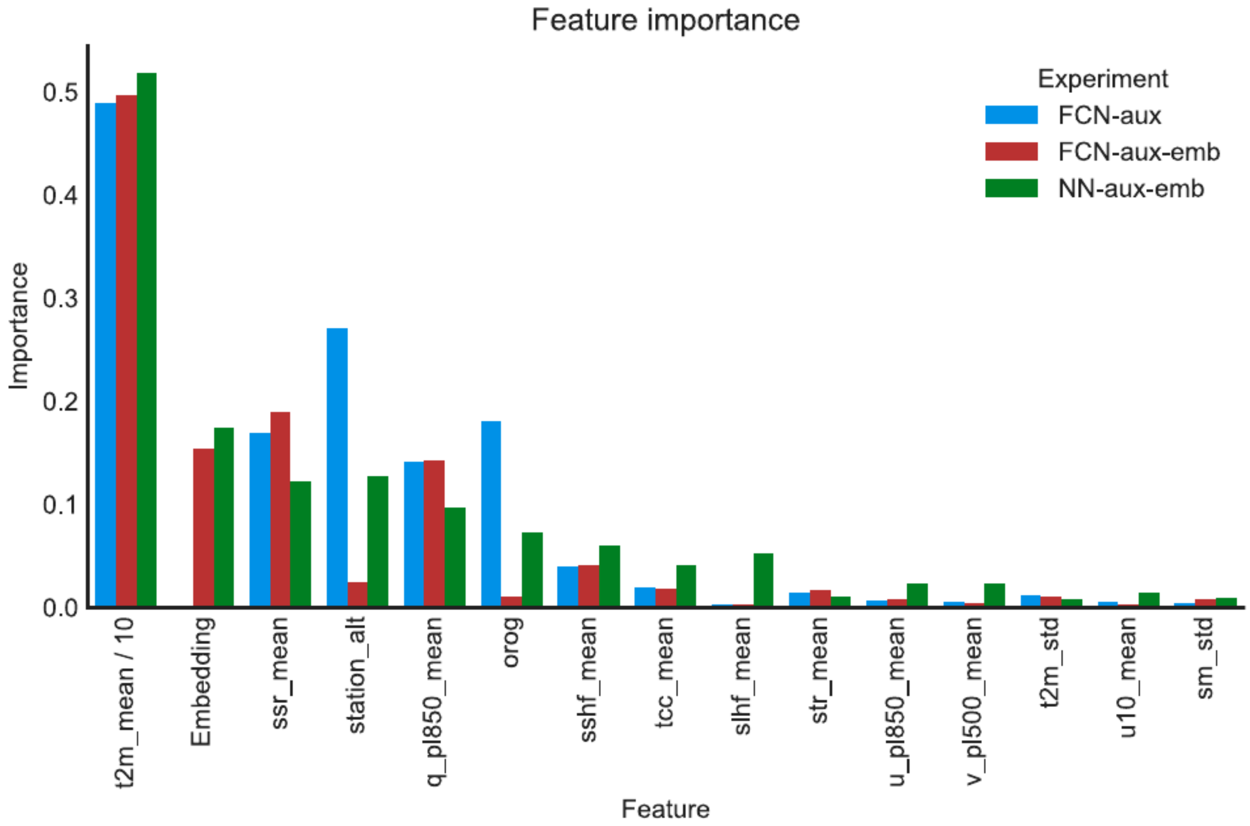


FIG. 5. Feature importance for the 15 most important predictors. Note that the values for t2m_mean are divided by 10. See Table 1 for variable abbreviations and descriptions.

To summarise ...

- Neural networks were used to learn the relation between weather forecast output and observations
- Weather forecast data augmented with additional features of weather stations were embedded for better performance
- Neural network based method outperformed all other state-of-the-art post-processing methods
- Randomisation of inputs allowed to infer the relevance of various features for the final output performance

Received: 25 May 2018 | Revised: 29 August 2018 | Accepted: 01 October 2018 | Published on: 26 November 2018

DOI: 10.1002/qj.3410

RESEARCH ARTICLE

Quarterly Journal of the Royal Meteorological Society RMetS

Predicting weather forecast uncertainty with machine learning

Sebastian Scher | Gabriele Messori

Department of Meteorology and Bolin Centre for Climate Research, Stockholm University, Sweden

Weather forecasts are inherently uncertain. Therefore, for many applications fore-

DEEP LEARNING FOR POST-PROCESSING ENSEMBLE WEATHER FORECASTS

PREPRINT

Peter Grönquist
ETH Zürich
petergro@student.ethz.ch

Chengyuan Yao
ETH Zürich
chyao@student.ethz.ch

Tal Ben-Nun
ETH Zürich
tal.bennun@inf.ethz.ch

Nikoli Dryden
ETH Zürich
nikoli.dryden@inf.ethz.ch

Peter Dueben
ECMWF
peter.dueben@ecmwf.int

Shigang Li
ETH Zürich
shigang.li@inf.ethz.ch

Torsten Hoefler
ETH Zürich
htor@inf.ethz.ch

Weather Forecast Postprocessing

1

- Rasp & Lerch, 2018
- ~~Scher & Messori, 2018~~
- ~~Grönquist et al., 2020~~

Climate Model Parametrisation

2

- Krasnopolsky, Fox-Rabinovitz & Belochitski, 2013
- ~~Rasp, Pritchard & Gentine, 2018~~
- ~~Yuval, O'Gorman & Hill, 2021~~

Data-Driven Climate Modeling

3

- Scher, 2018
- ~~Scher & Messori, 2020~~
- ~~Rasp & Thuerey, 2021~~

ENSO Forecasting

4

- Ham, Kim & Luo, 2019
- ~~Mahesh et al., 2019~~
- ~~Cachay et al., 2021~~

Hindawi Publishing Corporation
Advances in Artificial Neural Systems
Volume 2013, Article ID 485913, 13 pages
http://dx.doi.org/10.1155/2013/485913

Research Article

Using Ensemble of Neural Networks to Learn Stochastic Convection Parameterizations for Climate and Numerical Weather Prediction Models from Data Simulated by a Cloud Resolving Model

Vladimir M. Krasnopolsky,[1,2] Michael S. Fox-Rabinovitz,[2] and Alexei A. Belochitski[3,4]

[1] *National Centers for Environmental Prediction, NOAA, College Park, MD 20740, USA*
[2] *Earth System Sciences Interdisciplinary Center, University of Maryland, College Park, MD 20740, USA*
[3] *Geophysical Fluid Dynamics Laboratory, NOAA, Princeton, NJ 08540, USA*
[4] *Brookhaven National Laboratory, Upton, NY 11973, USA*

2. Climate Model Parametrisation → Krasnopolky, Fox-Rabinovitz & Belochitski, 2013

Main idea ...

- General circulation models are necessarily required to parametrise several aspects of climate dynamics
  - An important parametrisation is cloud dynamics

- Cloud resolving models (CRMs) are typically very high resolution and time consuming

- Here, the authors propose to learn the parametrisation from CRM data using neural networks

Data and models used ...

- TOGA-COARE
  - the international observational experiment in the tropics conducted for the 4-month period from November 1992 to February 1993)
  - horizontal resolution of 1 km
  - 64 or 96 vertical layers
  - time integration step of 5 s

- Integrate CRM over a domain of 256 × 256 km

- Basic plan:
  - Run CRM
  - Make “pseudo-observations” from CRM onto the variable set and resolution of GCMs
  - Train neural networks with pseudo-observations

Neural network setup ...

- Simulation details:
  - CRM run for 120 (model) days
  - Output aggregated to hourly resolution
  - Approx 2800 data points
- 80:20 split for training and test data
  - 2240 data points for training
  - 560 data points for test
- Final choice for number of hidden neurons (HID) is 5

TABLE 1: NN architecture (inputs and outputs) investigated in the paper.

NN architecture	NN inputs		NN outputs			
In : out	T	QV	Q1C	Q2	PREC	CLD
36 : 55	18	18	18	18	1	18

T is temperature, QV is atmospheric moisture—vapor mixing ratio, Q1C: the "apparent heat source," Q2: the "apparent moist sink," PREC: precipitation rates, and CLD: cloudiness. Numbers in the table show the dimensionality of the corresponding input and output parameters. In : Out stand for NN inputs and outputs and show their corresponding numbers.

TABLE 2: The number of fitting parameters (NN weights), N_C, at different values of HID = k (see (3)).

	HID					
	1	2	5	10	15	20
N_C	166	273	594	1129	1667	2199

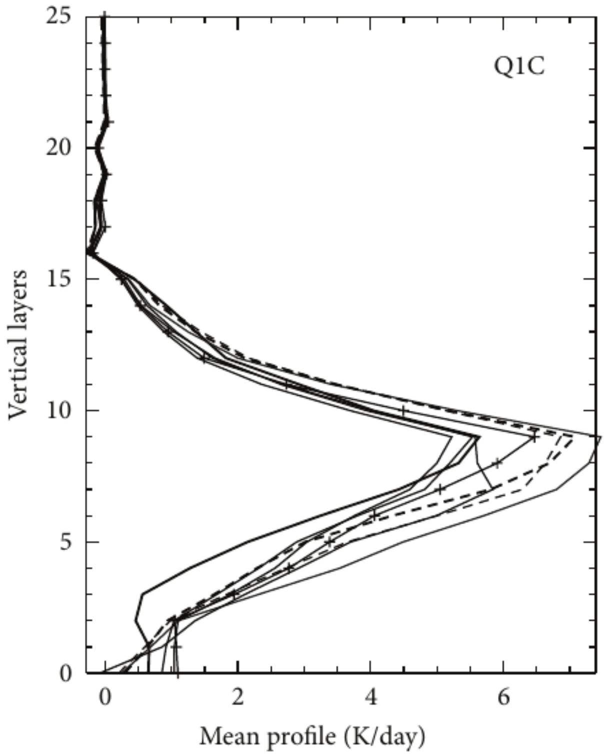


FIGURE 5: Q1C (the apparent heat source from convection) mean profiles on the test set produced by different NN ensemble members. The different curves presented in the figure correspond to different ensemble members; the thick solid line shows the verification data in the test set.

2. Climate Model Parametrisation → Krasnopolky, Fox-Rabinovitz & Belochitski, 2013

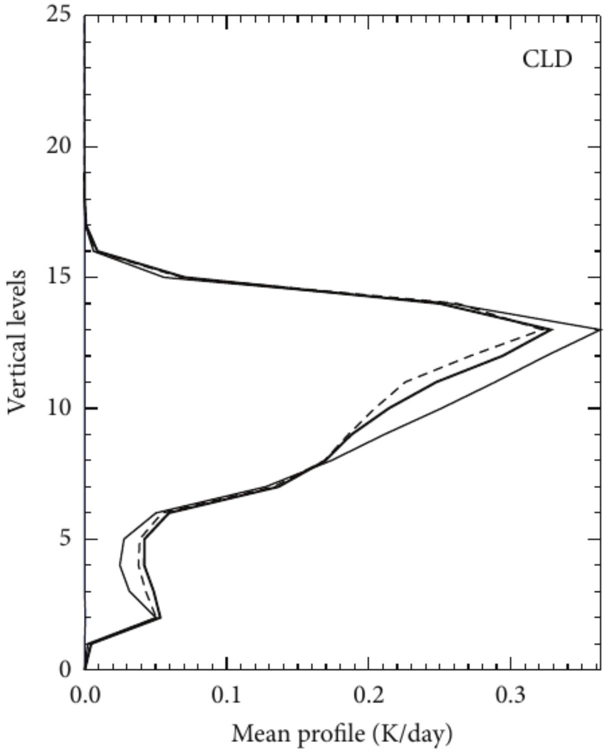


FIGURE 7: Three different mean cloud (CLD) profiles for the TOGA-COARE period: CAM-NN (thick solid), pseudo-observations (dashed), and CAM (solid).

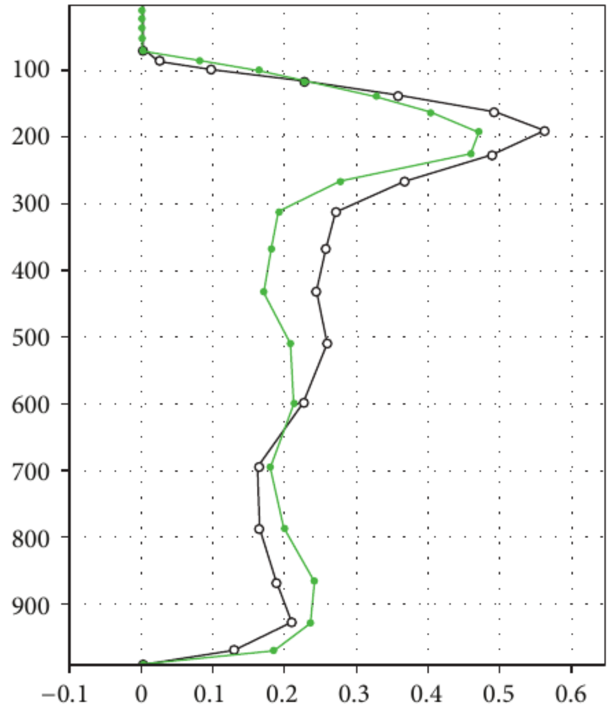


FIGURE 9: Vertical profiles of decadal boreal winter mean CLD for the TOGA-COARE location, in fractions, for the CAM-NN (open circles) and CAM (full circles) runs. Atmospheric pressure in hPa is the vertical coordinate.

2. Climate Model Parametrisation → Krasnopolky, Fox-Rabinovitz & Belochitski, 2013

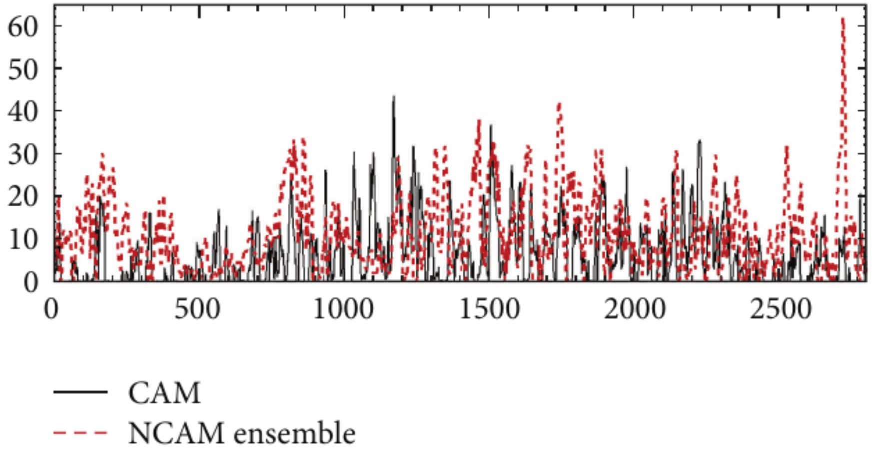


FIGURE 8: Precipitation (PREC, in mm/day) time series: CAM (black solid) and CAM-NN (or NCAM) ensemble mean (red dashed).

To summarise ...

- Cloud dynamics were parametrised using neural networks
- The neural network was trained to learn the relation between CRM and GCM
- Proof-of-concept study demonstrated with the TOGA-COARE CRM and the CAM models
- Results between the neural network parametrisations and the standard parametrisations were comparable

PNAS

Deep learning to represent subgrid processes in climate models

Stephan Rasp[a,b,1], Michael S. Pritchard[b], and Pierre Gentine[c,d]

[a]Meteorological Institute, Ludwig-Maximilian-University, 80333 Munich, Germany; [b]Department of Earth System Science, University of California, Irvine, CA 92697; [c]Department of Earth and Environmental Engineering, Earth Institute, Columbia University, New York, NY 10027; and [d]Data Science Institute, Columbia University, New York, NY 10027

Edited by Isaac M. Held, Geophysical Fluid Dynamics Laboratory, National Oceanic and Atmospheric Administration, Princeton, NJ, and approved August 8, 2018 (received for review June 14, 2018)

The representation of nonlinear subgrid processes, especially

extremes. Further increasing the resolution to a few hundred

Geophysical Research Letters

RESEARCH LETTER

10.1029/2020GL091363

Key Points:

- Neural-network parameterization gives stable simulations that replicate climate of idealized simulation of atmosphere at high resolution
- Separate predictions of the effect of each subgrid process allows physical constraints to be incorporated into the parameterization
- Parameterization with reduced numerical precision can decrease computational demands without affecting the simulated climate

Use of Neural Networks for Stable, Accurate and Physically Consistent Parameterization of Subgrid Atmospheric Processes With Good Performance at Reduced Precision

Janni Yuval[1], Paul A. O'Gorman[1], and Chris N. Hill[1]

[1]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

Abstract A promising approach to improve climate-model simulations is to replace traditional subgrid parameterizations based on simplified physical models by machine learning algorithms that are data-driven. However, neural networks (NNs) often lead to instabilities and climate drift when coupled

Weather Forecast Postprocessing

1

- Rasp & Lerch, 2018
- ~~Scher & Messori, 2018~~
- ~~Grönquist et al., 2020~~

Climate Model Parametrisation

2

- Krasnopolsky, Fox-Rabinovitz & Belochitski, 2013
- ~~Rasp, Pritchard & Gentine, 2018~~
- ~~Yuval, O'Gorman & Hill, 2021~~

Data-Driven Climate Modeling

3

- Scher, 2018
- ~~Scher & Messori, 2020~~
- ~~Rasp & Thuerey, 2021~~

ENSO Forecasting

4

- Ham, Kim & Luo, 2019
- ~~Mahesh et al., 2019~~
- ~~Cachay et al., 2021~~

Geophysical Research Letters

RESEARCH LETTER
10.1029/2018GL080704

Key Points:
- A neural network can emulate the dynamics of a simple general circulation model
- The trained network can successfully forecast the model weather
- The network can produce a realistic representation of the model climate

Toward Data-Driven Weather and Climate Forecasting: Approximating a Simple General Circulation Model With Deep Learning

S. Scher[1]

[1]Department of Meteorology and Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden

Main idea ...

- Proof-of-concept that neural networks can *emulate* the dynamics of a general circulation model

- Simplified model (PUMA) with coarse resolution
  - no seasonal cycle (eternal Northern Hemispheric winter)
  - No orography,
  - horizontal resolution of T21(∼625 km, 32 × 64 grid points when projected on a regular latlon grid)
  - 10 vertical levels
  - no diurnal cycle
  - no ocean
  - time step of 45 min

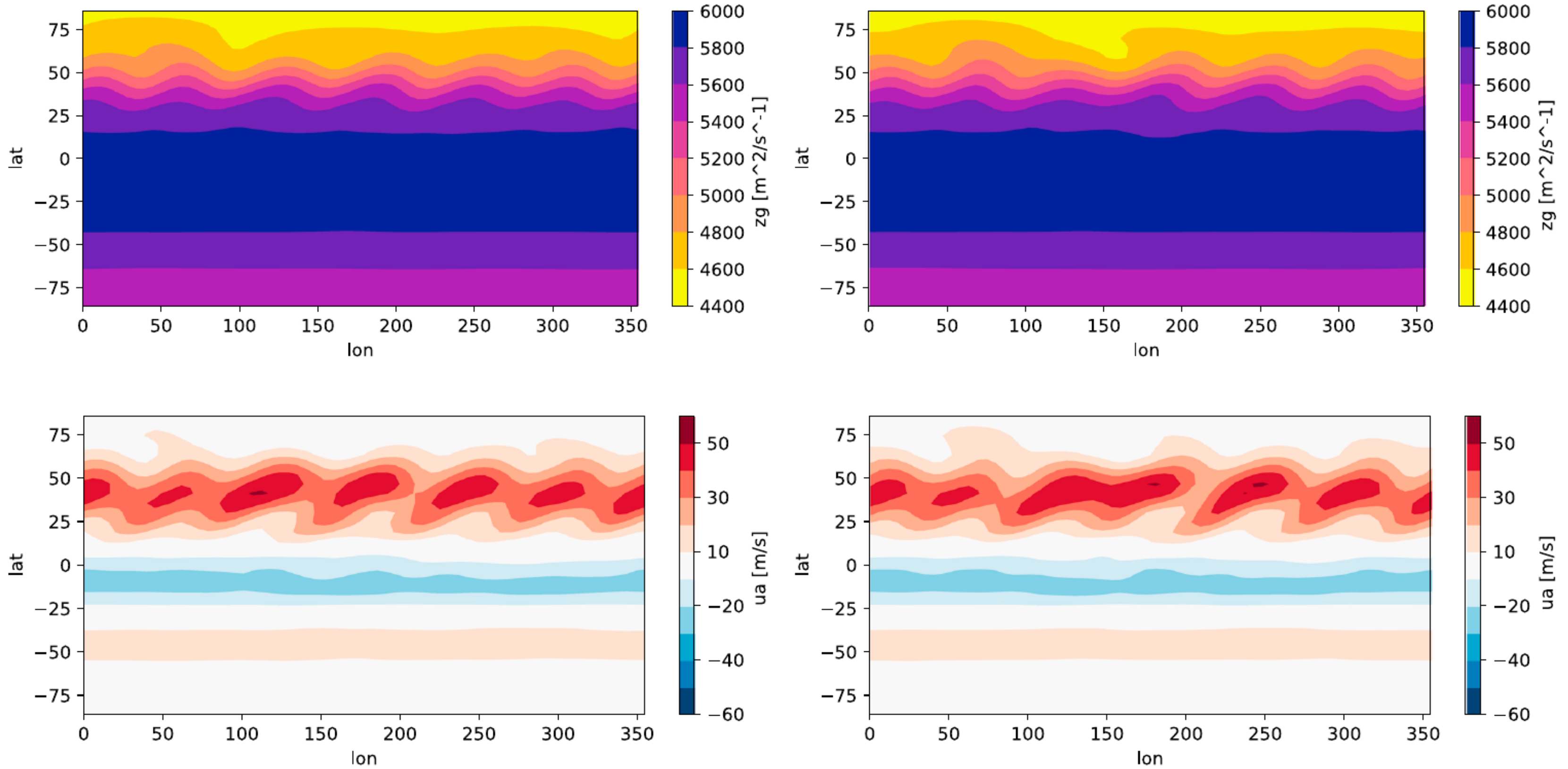


Figure 1. Two model states of the PUMA model, separated by 5 days (from left to right). The upper row shows geopotential at 500 hPa (zg), and the lower row shows zonal wind at 300 hPa (ua).

3. Data-driven Climate Modelling → Scher, 2018

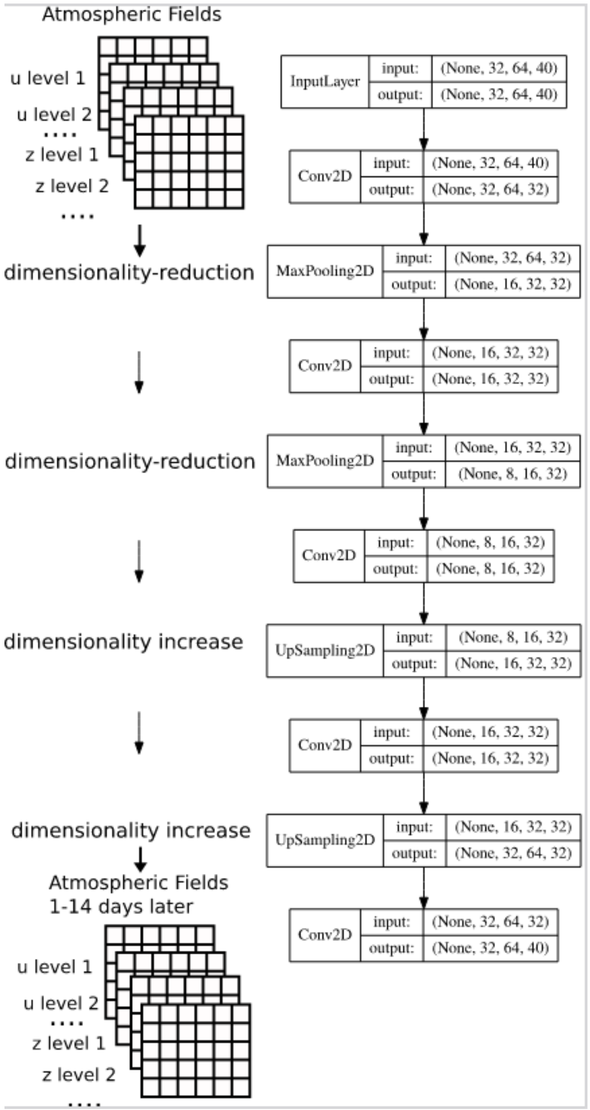
Model setup ...

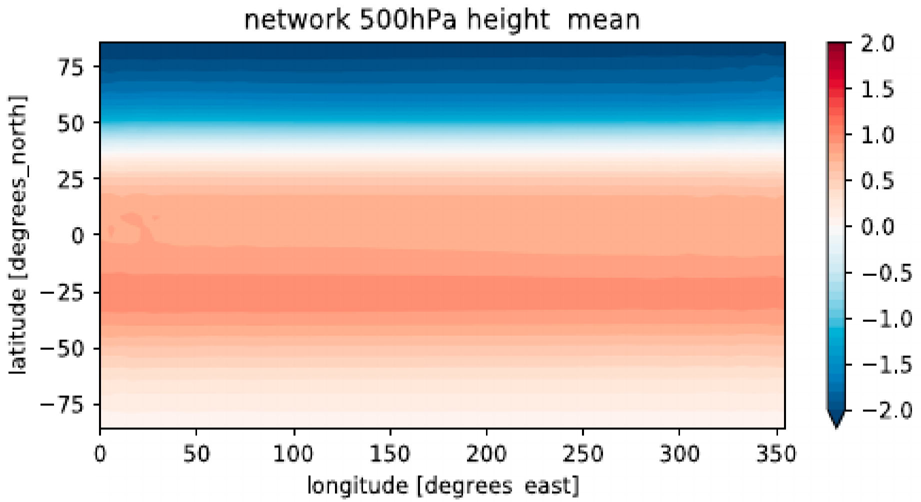
- First 30 years of model run are discarded as *spin-up* years

- 150 years of data are used
  - 100 years for training
  - 20 years for validation
  - 30 years for testing

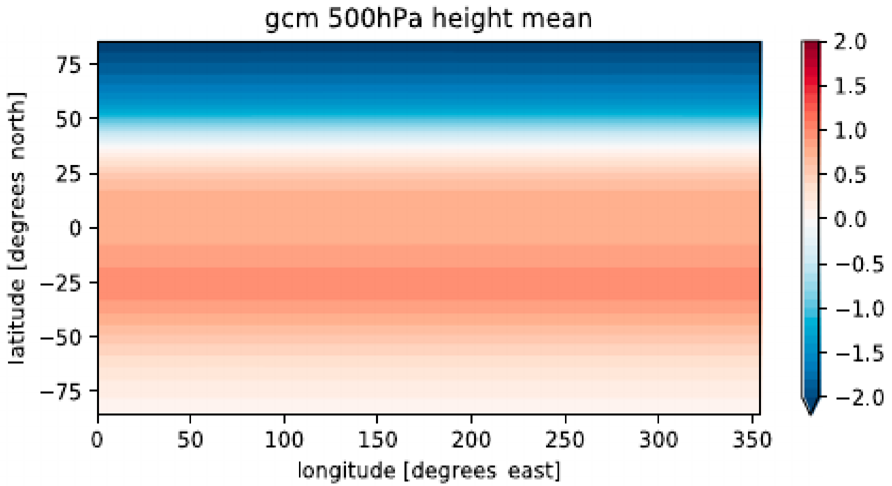
- Autoencoder network architecture
  - Combined with 2D convolutions and max pooling

- High dimensional input and outputs
  - 40 channels × 2, 048 grid points = 81, 920



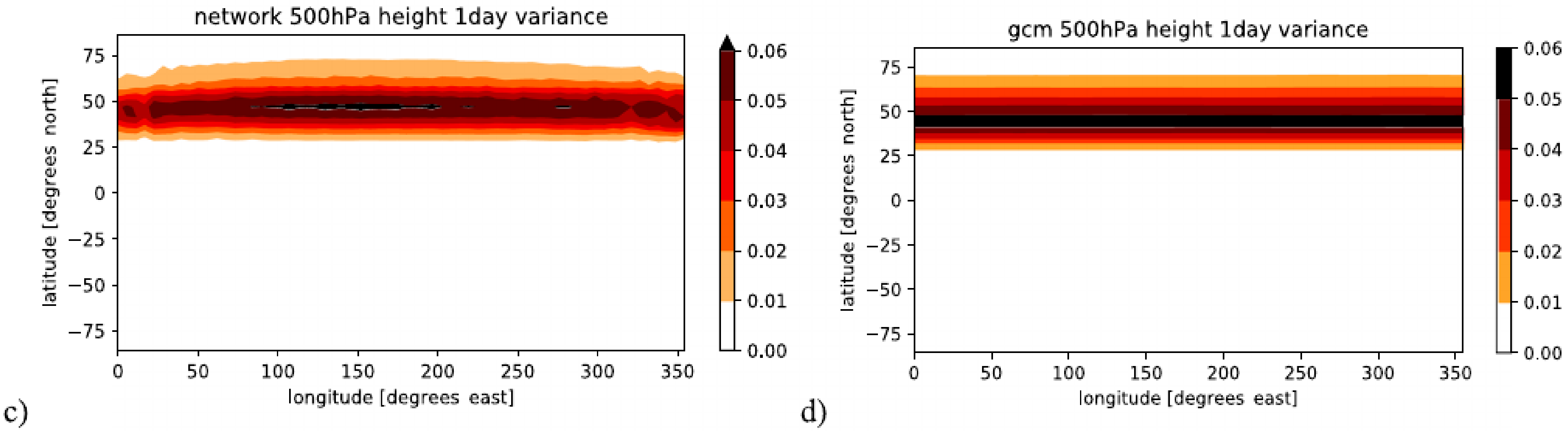


a)

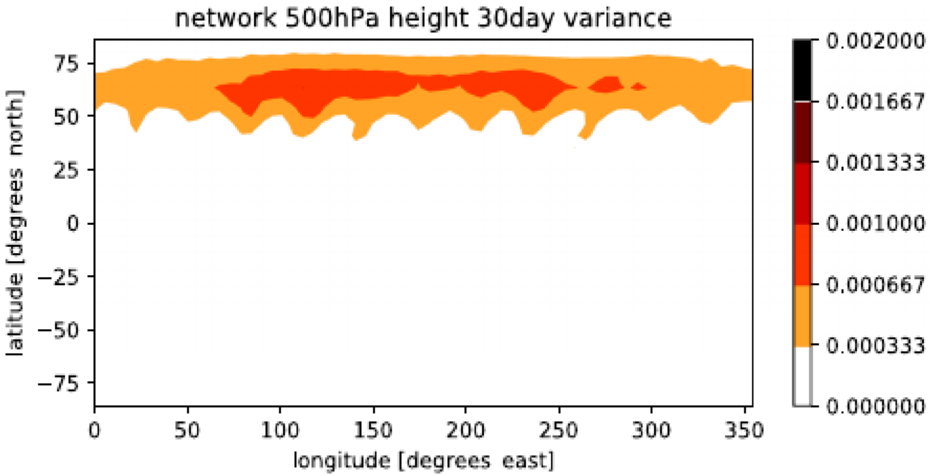


b)

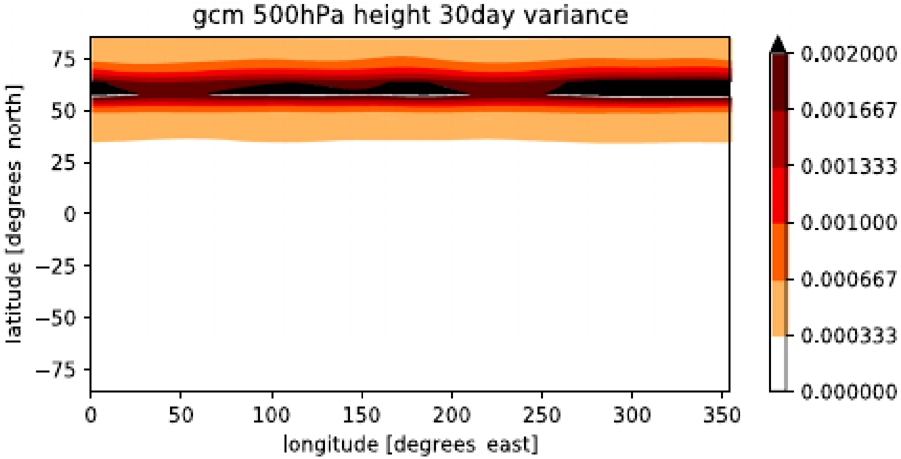
3. Data-driven Climate Modelling → Scher, 2018



3. Data-driven Climate Modelling → Scher, 2018

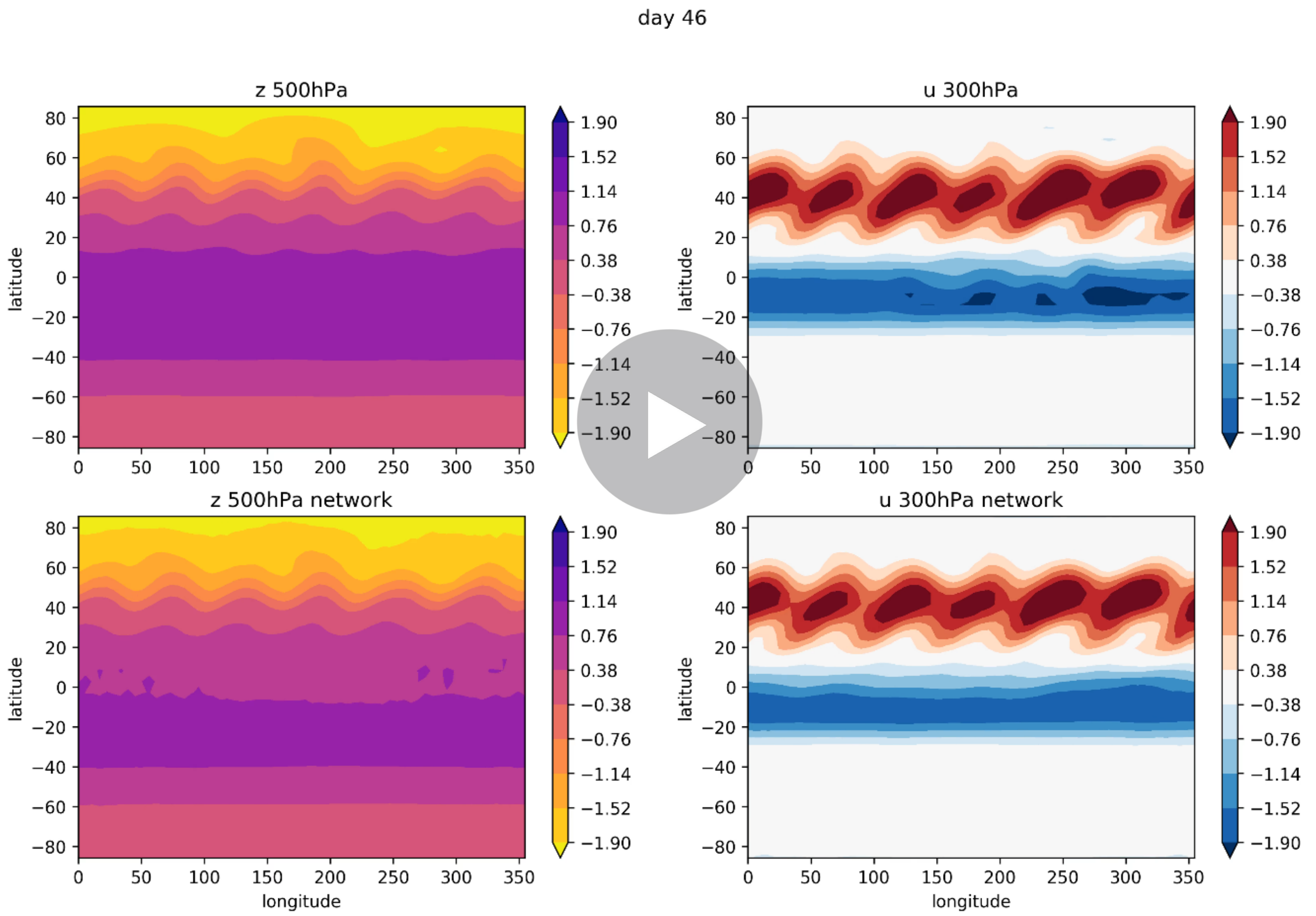


e)



f)

3. Data-driven Climate Modelling → Scher, 2018



3. Data-driven Climate Modelling → Scher, 2018

To summarise ...

- A neural network was trained on the output of a simple atmospeheric climate model
- The goal was to learn to emulate the dynamical behaviour of the atmosphere as modeled in the climate model
- The neural network learns the dynamics successfully
- The results from the neural network are slightly noisy but overall are convincing

JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2020MS002331

Ensemble Methods for Neural Network-Based Weather Forecasts

Sebastian Scher[1] and Gabriele Messori[1,2]

[1]Department of Meteorology and Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden,
[2]Department of Earth Sciences and Centre of Natural Hazards and Disaster Science (CNDS), Uppsala University, Uppsala, Sweden

Key Points:

- We test four different methods to transform a deterministic neural network weather forecasting system into an ensemble forecasting system
- The ensemble mean of all methods is more skilful than a deterministic neural network forecast
- The spread-error correlation of the four methods is comparable to that of numerical weather prediction (NWP) forecasts

Abstract Ensemble weather forecasts enable a measure of uncertainty to be attached to each forecast, by computing the ensemble's spread. However, generating an ensemble with a good spread-error relationship is far from trivial, and a wide range of approaches to achieve this have been explored—

JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE

10.1029/2020MS002405

Key Points:

- A large convolutional neural network is trained for the WeatherBench challenge
- Pretraining on climate model data improves skill and prevents overfitting
- The model sets a new state-of-the-art for data-driven medium-range forecasting

Data-Driven Medium-Range Weather Prediction With a Resnet Pretrained on Climate Simulations: A New Model for WeatherBench

Stephan Rasp[1,2] and **Nils Thuerey**[1]

[1]Department of Informatics, Technical University of Munich, Munich, Germany, [2]Now at ClimateAi, San Francisco, USA

Abstract Numerical weather prediction has traditionally been based on the models that discretize the dynamical and physical equations of the atmosphere. Recently, however, the rise of deep learning

Weather Forecast Postprocessing

1

- Rasp & Lerch, 2018
- ~~Scher & Messori, 2018~~
- ~~Grönquist et al., 2020~~

Climate Model Parametrisation

2

- Krasnopolsky, Fox-Rabinovitz & Belochitski, 2013
- ~~Rasp, Pritchard & Gentine, 2018~~
- ~~Yuval, O'Gorman & Hill, 2021~~

Data-Driven Climate Modeling

3

- Scher, 2018
- ~~Scher & Messori, 2020~~
- ~~Rasp & Thuerey, 2021~~

ENSO Forecasting

4

- Ham, Kim & Luo, 2019
- ~~Mahesh et al., 2019~~
- ~~Cachay et al., 2021~~

LETTER

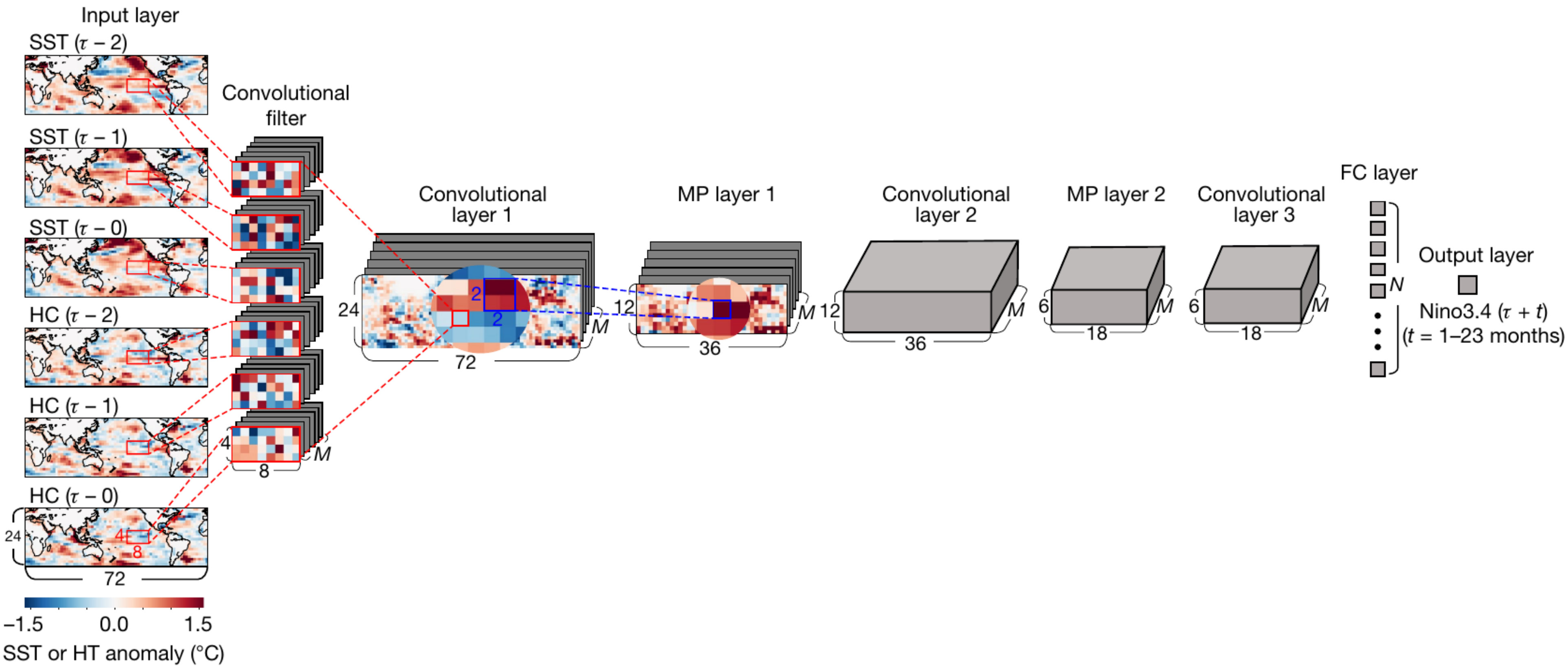
https://doi.org/10.1038/s41586-019-1559-7

Deep learning for multi-year ENSO forecasts

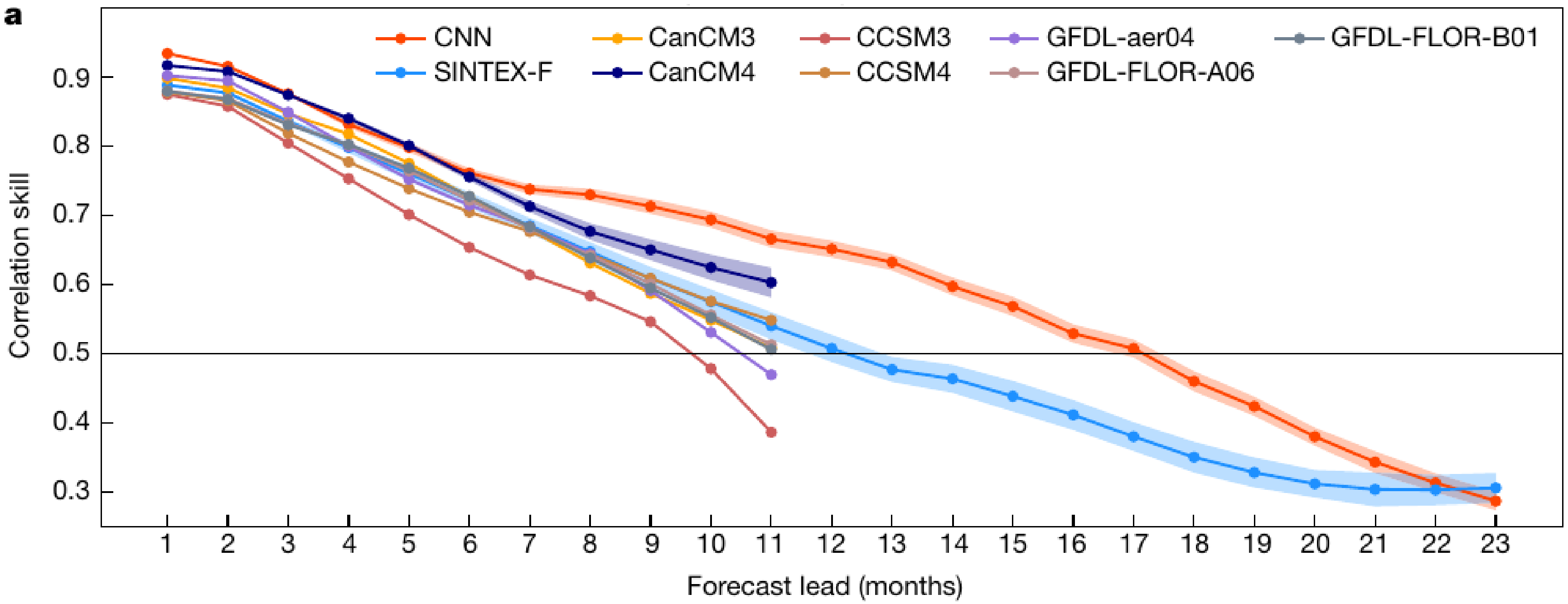
Yoo-Geun Ham[1]*, Jeong-Hwan Kim[1] & Jing-Jia Luo[2,3]

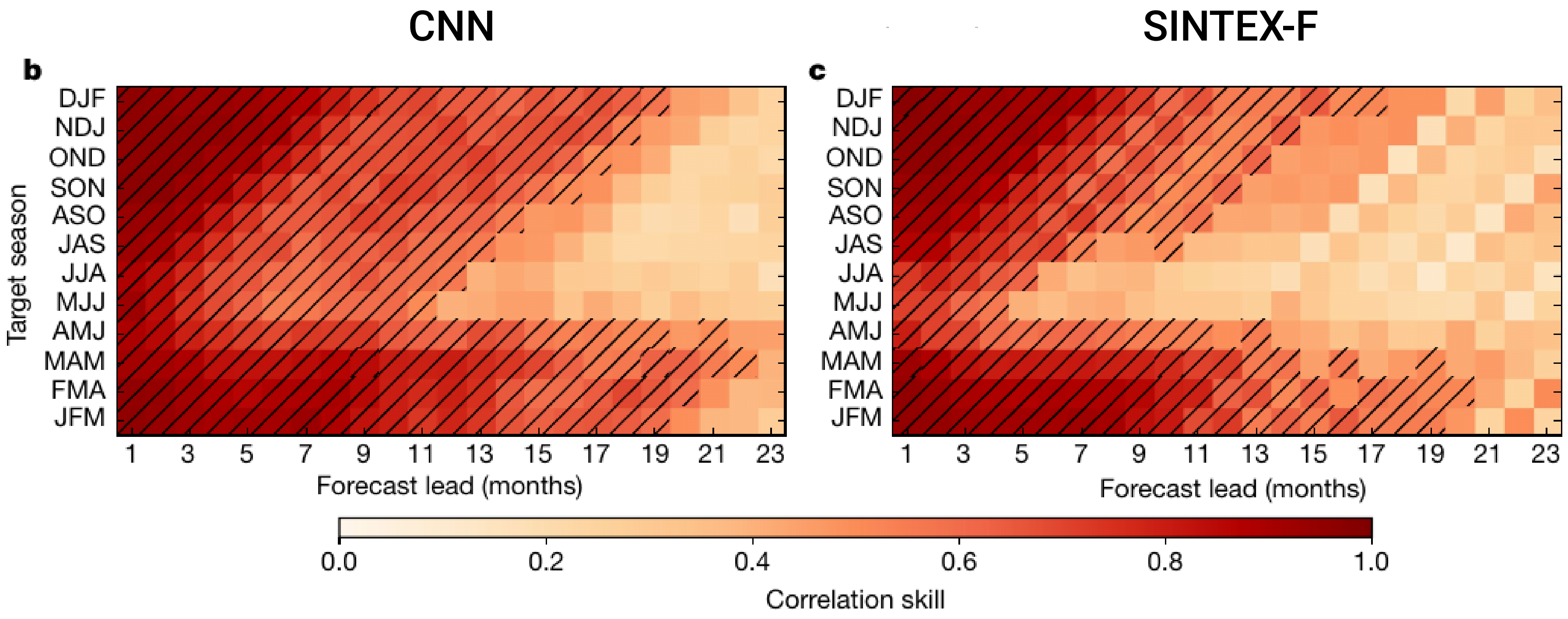
Main idea ...

- Use CNNs to predict ENSO index values up to one and half years in advance
- Overcome limited amount of observartions (in terms of El Niños and La Niñas) by
  - Training the CNN
    - On historical simulations
    - On reanalysis between 1871-1973
- Implicitly assumed
  - Statistical emulators can help predict ENSO behaiviour better than dynamical models
    - All dynamics are not knowable

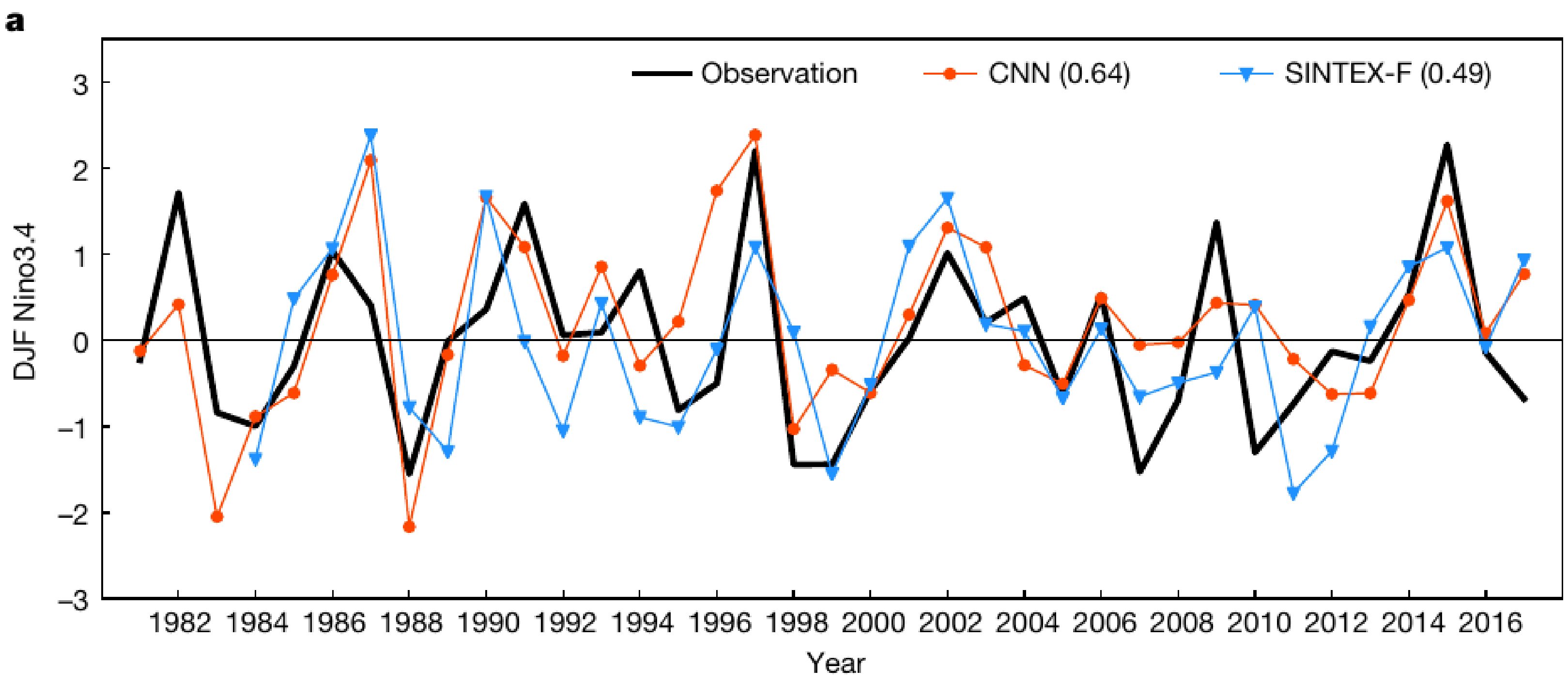


4. ENSO Forecasting → Ham, Kim & Luo, 2019

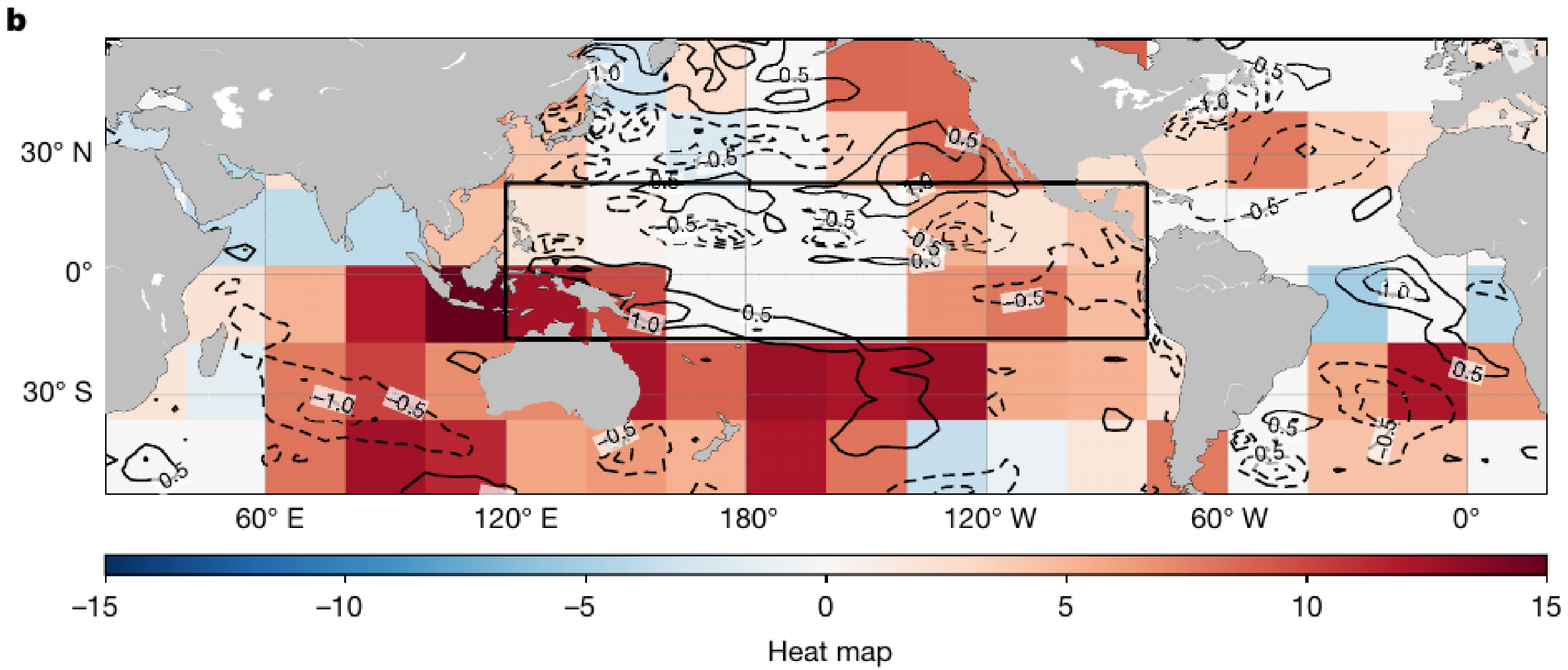

a
CNN
CanCM3
CCSM3
GFDL-aer04
GFDL-FLOR-B01
SINTEX-F
CanCM4
CCSM4
GFDL-FLOR-A06
Correlation skill
0.9
0.8
0.7
0.6
0.5
0.4
0.3
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
Forecast lead (months)



4. ENSO Forecasting → Ham, Kim & Luo, 2019



4. ENSO Forecasting → Ham, Kim & Luo, 2019



4. ENSO Forecasting → Ham, Kim & Luo, 2019

To summarise ...

- CNNs were used to forecast ENSO index values
  - Input data were sea surface temperatures and oceanic heat content
- CNNs were trained on both climate model output and reanalysis data
- The neural network approach method outpferformed all other methods
- Forecast skill was above 0.5 till around 16 months
- Heatmap analysis revealed extratropical souther Pacific and Indian oceans as important predictors of ENSO

Forecasting El Niño with Convolutional and Recurrent Neural Networks

Ankur Mahesh* ClimateAi
Maximilian Evans ClimateAi
Garima Jain ClimateAi
Mattias Castillo ClimateAi
Aranildo Lima ClimateAi
Brent Lunghino ClimateAi
Himanshu Gupta ClimateAi
Carlos Gaitan ClimateAi
Jarrett K. Hunt ClimateAi
Omeed Tavasoli ClimateAi
Patrick T. Brown ClimateAi, San Jose State University
V. Balaji Geophysical Fluid Dynamics Laboratory

The World as a Graph: Improving El Niño Forecasts with Graph Neural Networks

Salva Rühling Cachay[1], Emma Erickson[*2], Arthur Fender C. Bucker[*3, 4], Ernest Pokropek[*5], Willa Potosnak[*6], Suyash Bire[8], Salomey Osei[7], and Björn Lütjens[8]

[1]Technical University of Darmstadt, [2]University of Illinois at Urbana-Champaign, [3]University of São Paulo, [4] Technical University of Munich, [5]Warsaw University of Technology, [6]Duquesne University, [7]African Institute for Mathematical Sciences, [8]Massachusetts Institute of Technology

Weather Forecast Postprocessing

1

- Rasp & Lerch, 2018
- ~~Scher & Messori, 2018~~
- ~~Grönquist et al., 2020~~

Climate Model Parametrisation

2

- Krasnopolsky, Fox-Rabinovitz & Belochitski, 2013
- ~~Rasp, Pritchard & Gentine, 2018~~
- ~~Yuval, O'Gorman & Hill, 2021~~

Data-Driven Climate Modeling

3

- Scher, 2018
- ~~Scher & Messori, 2020~~
- ~~Rasp & Thuerey, 2021~~

ENSO Forecasting

4

- Ham, Kim & Luo, 2019
- ~~Mahesh et al., 2019~~
- ~~Cachay et al., 2021~~

Q
A