DEEP LEARNING FOR WEATHER FORECASTING WITH VIDEO PREDICTION METHODS

BING GONG, MICHAEL LANGGUTH, YAN JI, MARTIN SCHULTZ

Contact: b.gong@fz-juelich.de

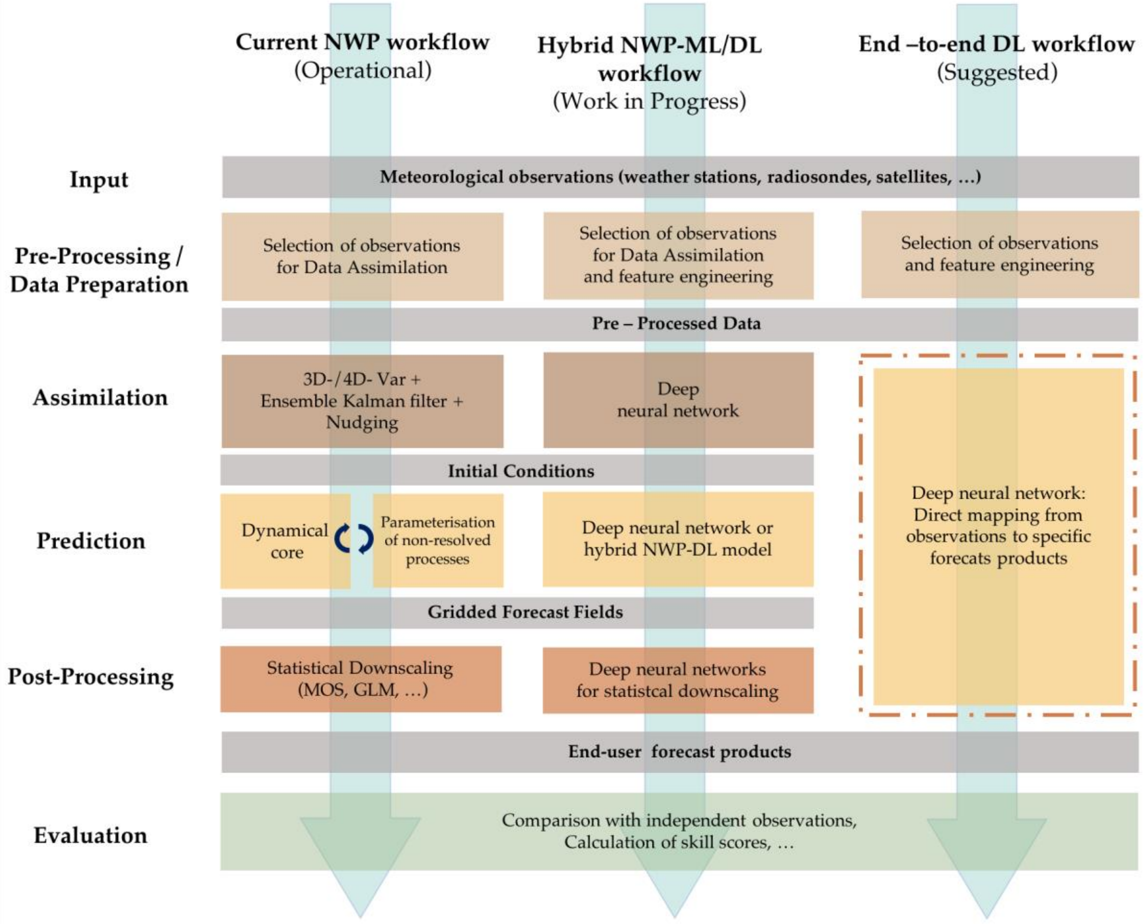
Member of the Helmholtz Association

JÜLICH Forschungszentrum

MOTIVATION

Modeling the atmosphere and the potential avenues for deep learning

- Numerical atmospheric models: backbone of operational weather prediction

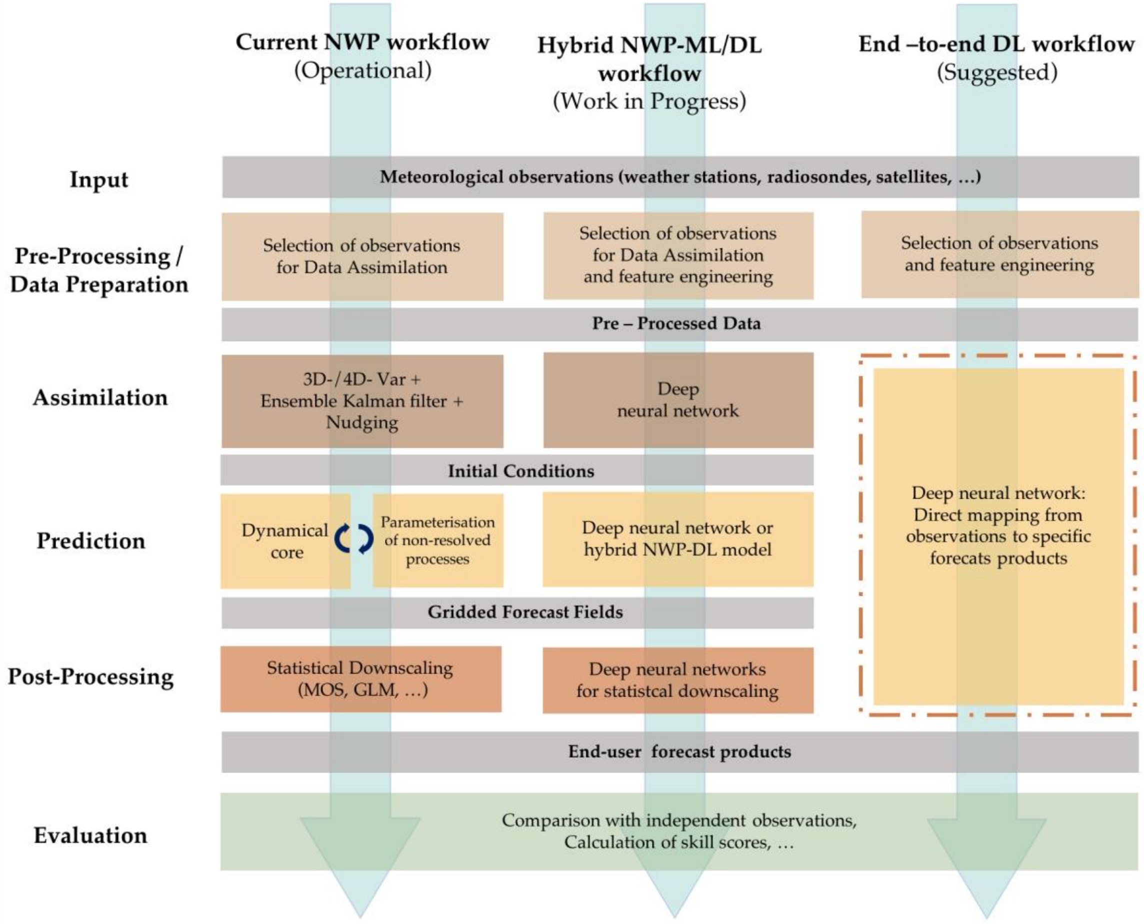


JÜLICH Forschungszentrum

MOTIVATION

Modeling the atmosphere and the potentical avenues for deep learning

- Numerical atmospheric models: backbone of operational weather prediction
- Increasing success of deep neural networks (DNNs) in various applications

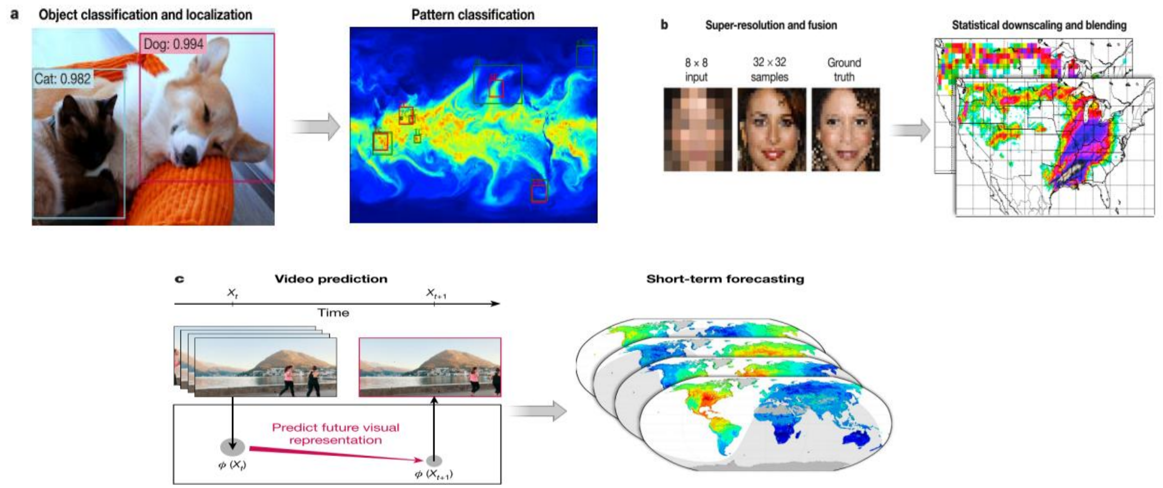


JÜLICH Forschungszentrum

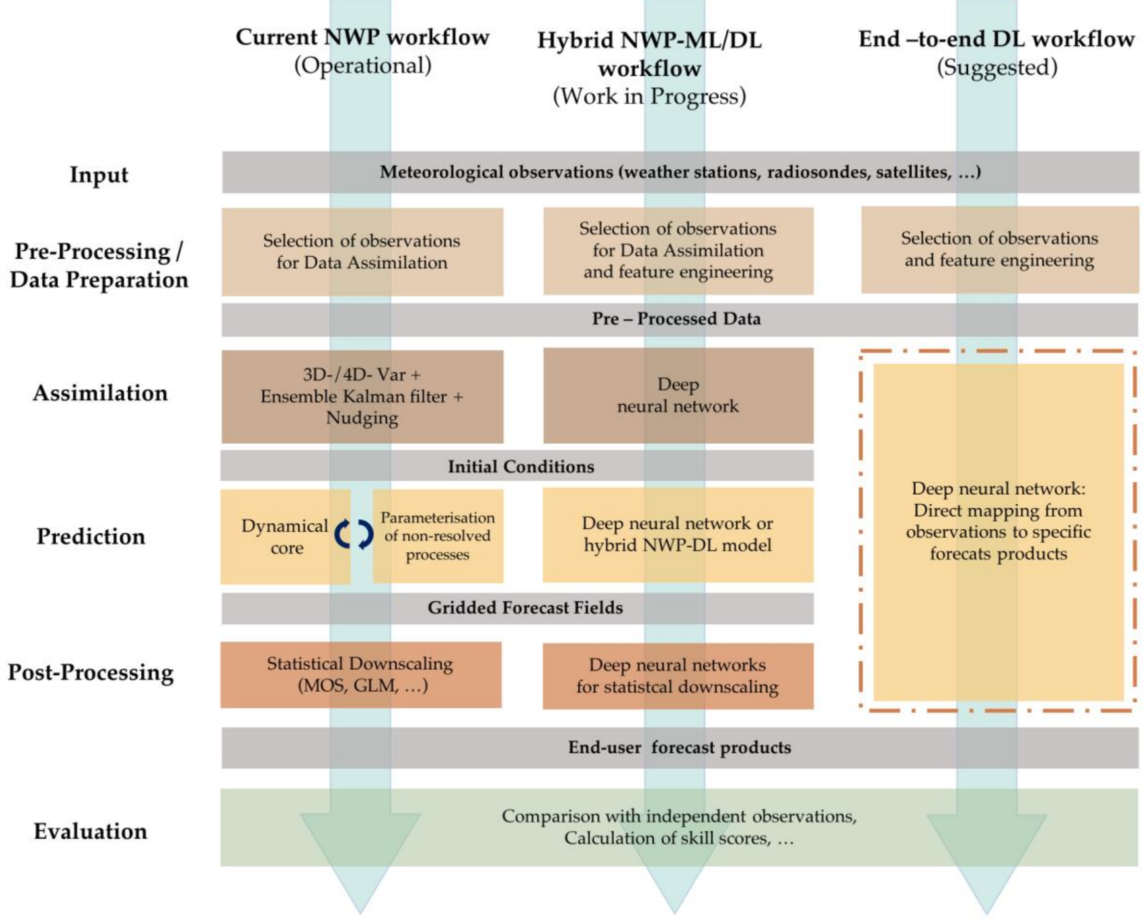
MOTIVATION

Modeling the atmosphere and the potential avenues for deep learning

- Numerical atmospheric models: backbone of operational weather prediction
- Increasing success of deep neural networks (DNNs) in various applications
- DNNs for new applications in the weather prediction workflow (see, e.g., *Schultz et al., 2021*)



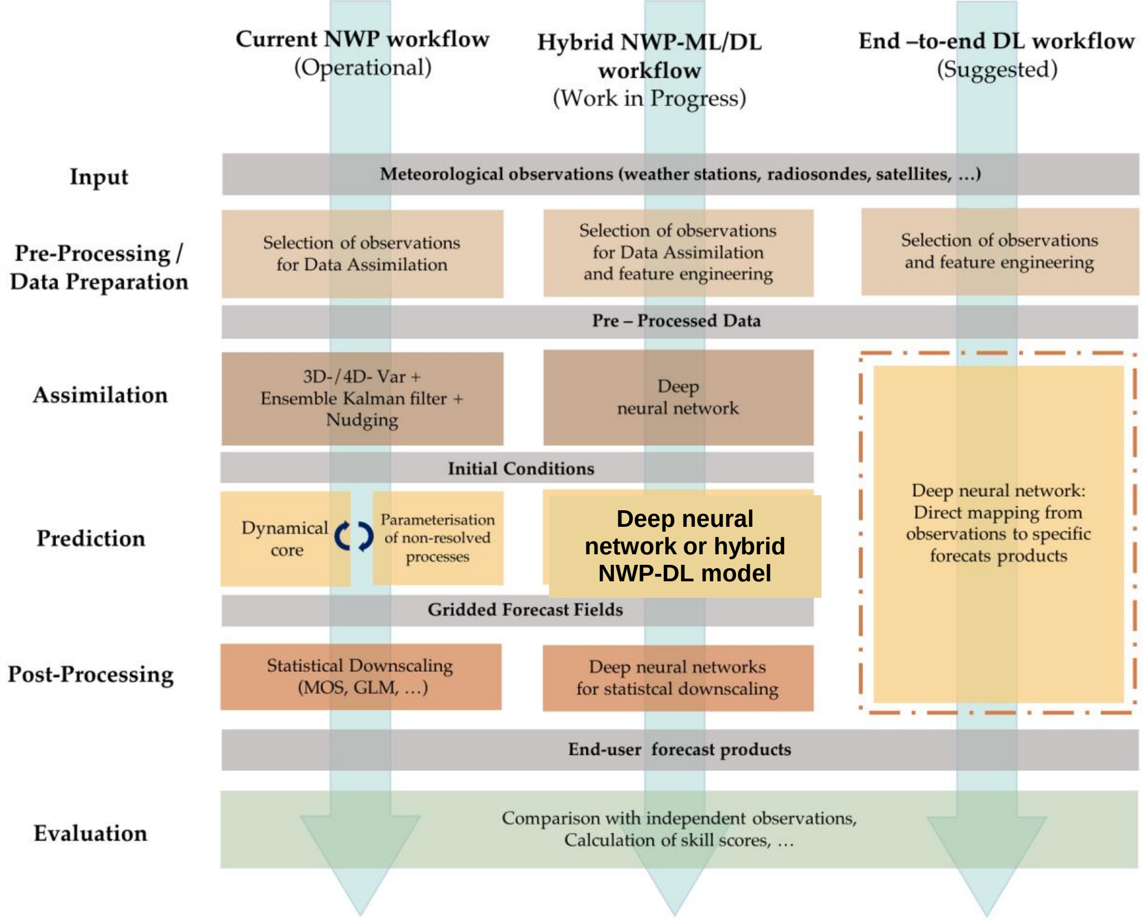
(Reichstein, Markus et al. 2019)



MOTIVATION

Modeling the atmosphere and the potentical avenues for deep learning

- Numerical atmospheric models: backbone of operational weather prediction
- Increasing success of deep neural networks (DNNs) in various applications
- DNNs for new applications in the weather prediction workflow (see, e.g., *Schultz et al., 2021*)
- DNNs for weather forecasting:
  - FourCastNet by Patha et al. on 8th August 2022
  - PanguWeather by Bi et al. on 3th November 2022
  - GraphCast by Lam et al. on 24th December 2022

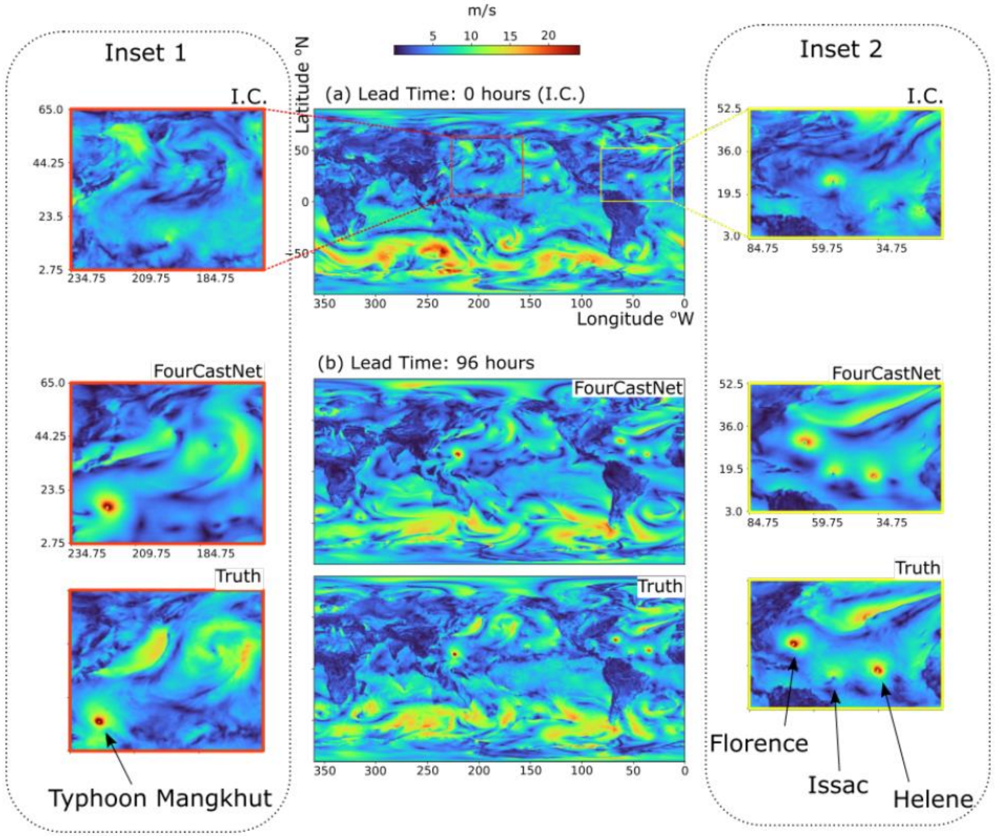


JÜLICH Forschungszentrum

WEATHER FORECAST WITH DEEP LEARNING

State-of-the-art

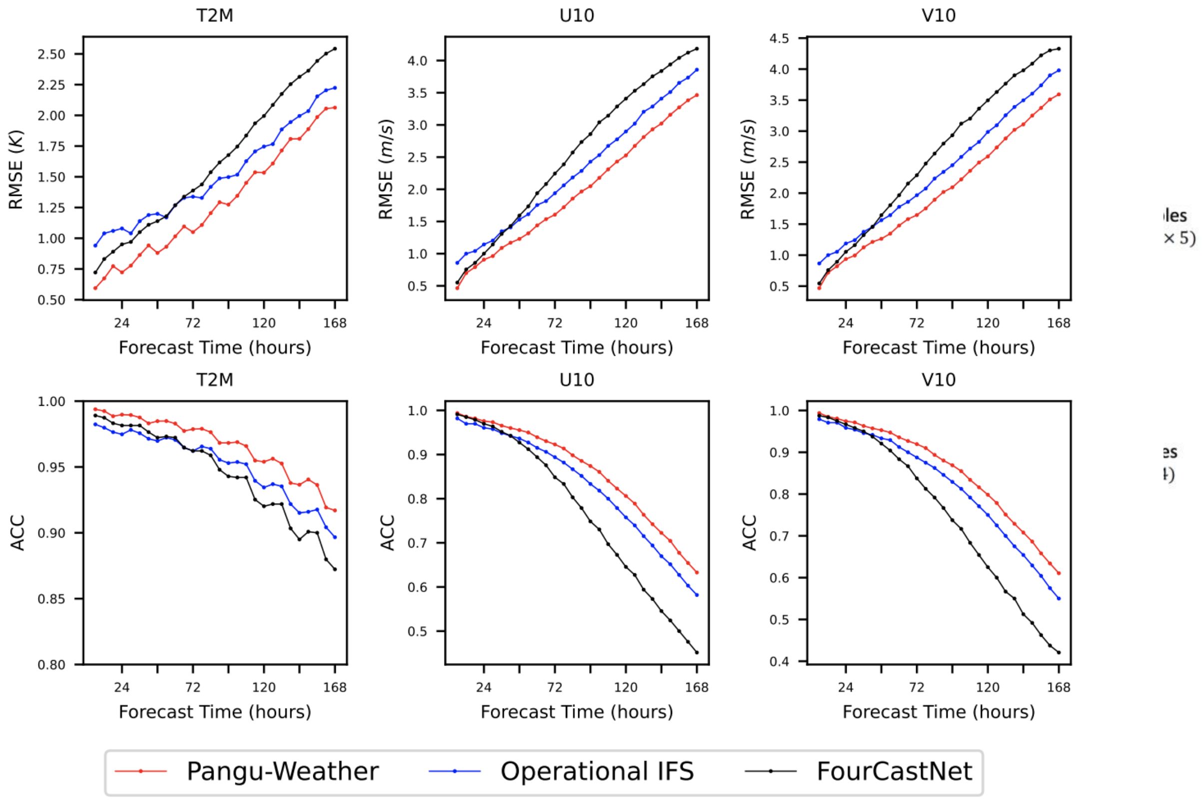
- FourCastNet by Patha et al. on 8th August 2022
- Vision Transformer (ViT)-based model
- FourCastNet is about 45,000 times faster than traditional NWP models on a node-hour basis.
- FourCastNet's predictions are comparable to the IFS model on metrics of Root Mean Squared Error (RMSE) and Anomaly Correlation Coefficient (ACC) at lead times of up to three days.



WEATHER FORECAST WITH DEEP LEARNING

State-of-the-art

- PanguWeather by Bi et al. on 3th November 2022
- … is a deep learning based system for fast and accurate global weather forecast
- … is a ViT-based model (256 million parameter)
- … shows good performance for short to medium-range forecast (i.e., forecast time ranges from one hour to one week)
- …outperforms state-of-the-art numerical weather prediction IFS model, especially for loner lead times

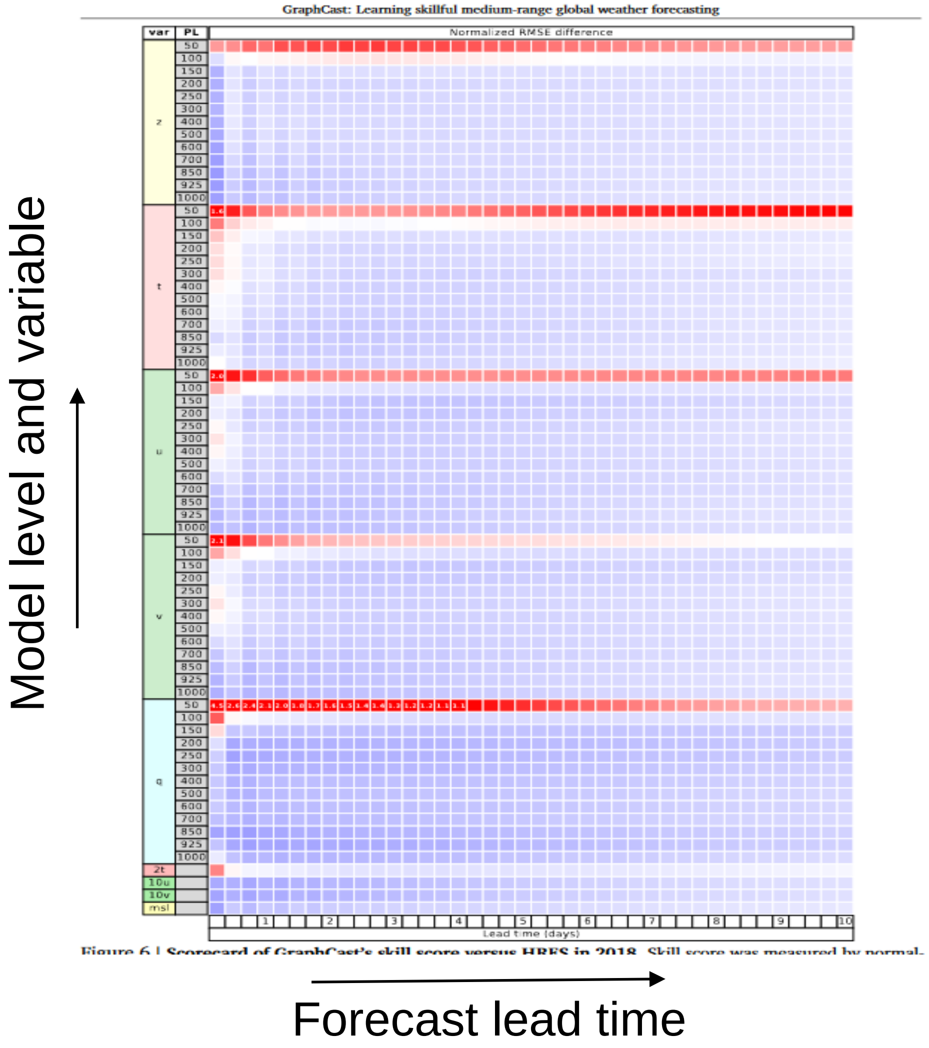


JÜLICH Forschungszentrum

WEATHER FORECAST WITH DEEP LEARNING

State-of-the-art

- GraphCast by Lam et al. on 24th December 2022
- …outperforms the deterministic IFS model, as well as all previous ML baselines
- …can make 10-day forecasts, at 6-hour time intervals, of five surface variables and six atmospheric variables, each at 37 vertical pressure levels
- …can generate a 10-day forecast (35 gigabytes of data) in under 60 seconds on Cloud TPU v4 hardware



JÜLICH Forschungszentrum

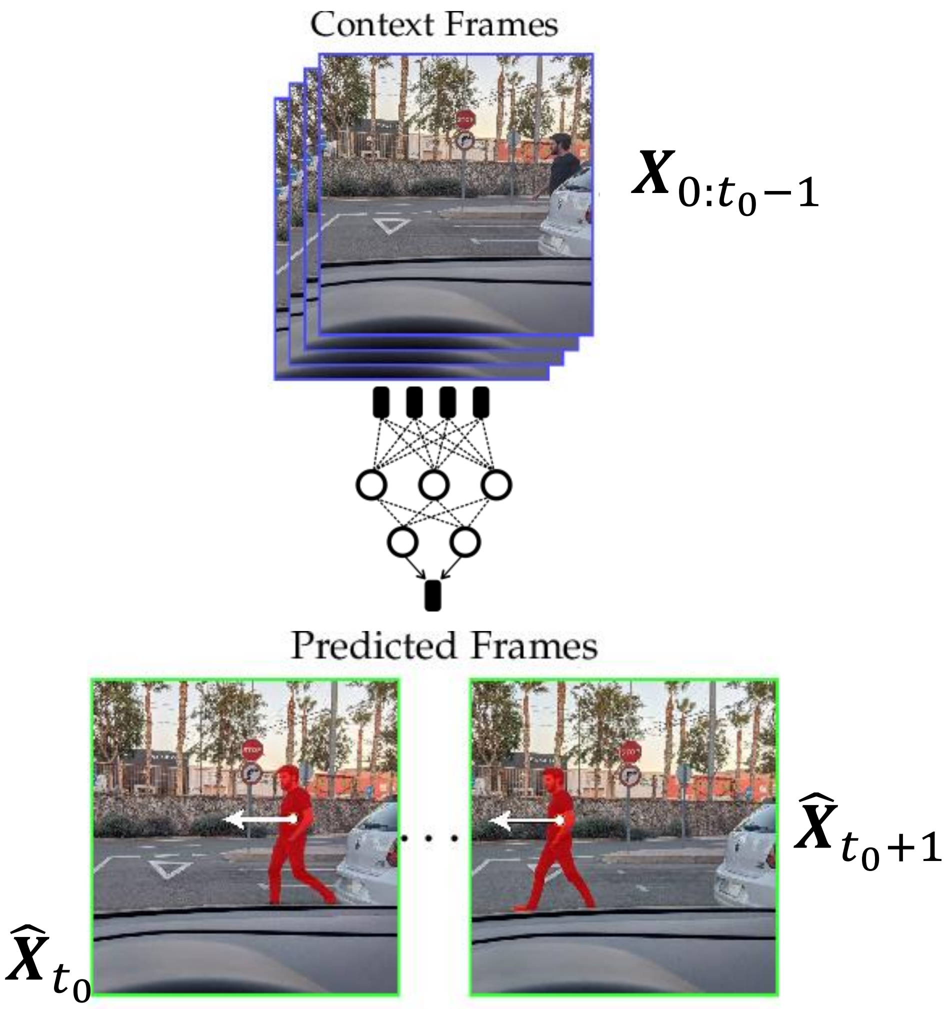
RESEARCH QUESTIONS

Video prediction for weather forecasting

Q1: Can we use video prediction approach to predict the diurnal cycle of 2m temperature?

Q2: Are advanced video prediction models beneficial for predicting the 2m temperature compare to shallow ones?

Q3: Can Generative Adversarial Networks (GANs) help to enhance the performance of 2m temperature forecasting?



From *Oprea et al. 2020*

JÜLICH Forschungszentrum

PRINCIPLES OF VIDEO PREDICTION WITH DL

Spatial-temporal learning via video prediction methods

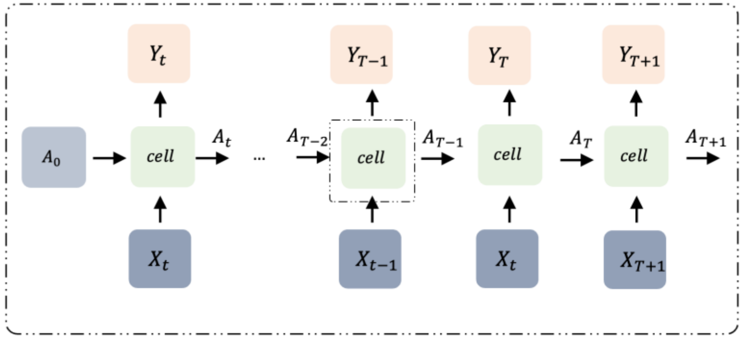
- ConvLSTM consists of two networks, an encoding network and a forecasting network (decoder)

- The decoder is conditioned on the last generated frame.

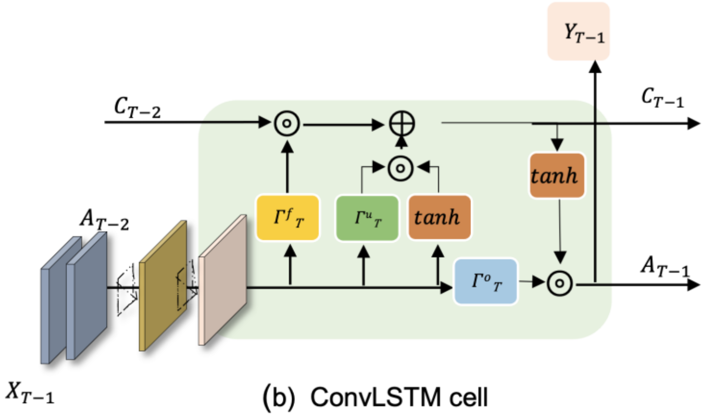
- A convolution operator for the state-to-state and input-to-state transitions

- Loss function (L1-loss):

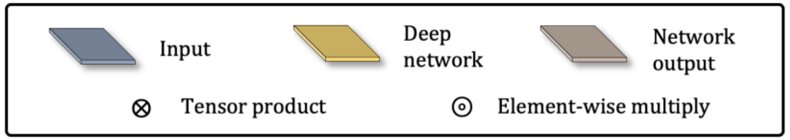
$$\mathcal{L}_1(\boldsymbol{X}_{t_0:T}, \widehat{\boldsymbol{X}}_{t_0:T}) = \sum_{t=t_0}^{T} |\boldsymbol{X}_t - \widehat{\boldsymbol{X}}_t|$$



(a) LSTM Future Predictor Model



(b) ConvLSTM cell



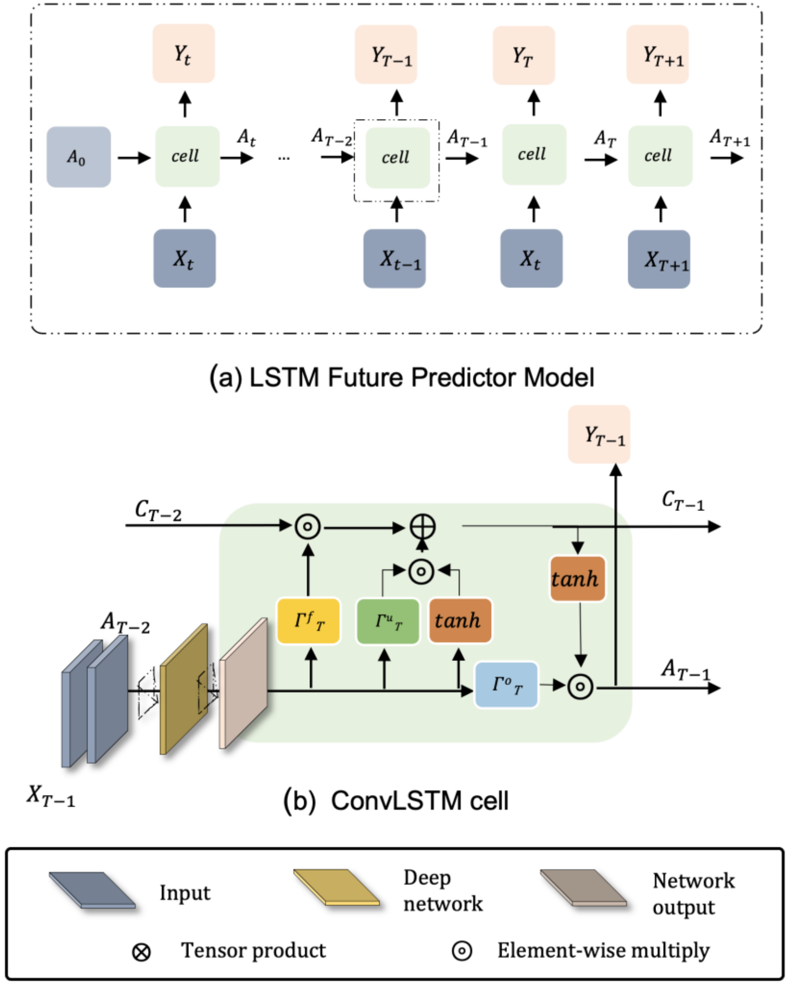
PRINCIPLES OF VIDEO PREDICTION WITH DL

429K

Spatial-temporal learning via video prediction methods

- ConvLSTM consists of two networks, an encoding network and a forecasting network (decoder)
- The decoder is conditioned on the last generated frame.
- A convolution operator for the state-to-state and input-to-state transitions
- Loss function (L1-loss):

$$\mathcal{L}_1(\boldsymbol{X}_{t_0:T}, \widehat{\boldsymbol{X}}_{t_0:T}) = \sum_{t=t_0}^{T} |\boldsymbol{X}_t - \widehat{\boldsymbol{X}}_t|$$



(a) LSTM Future Predictor Model

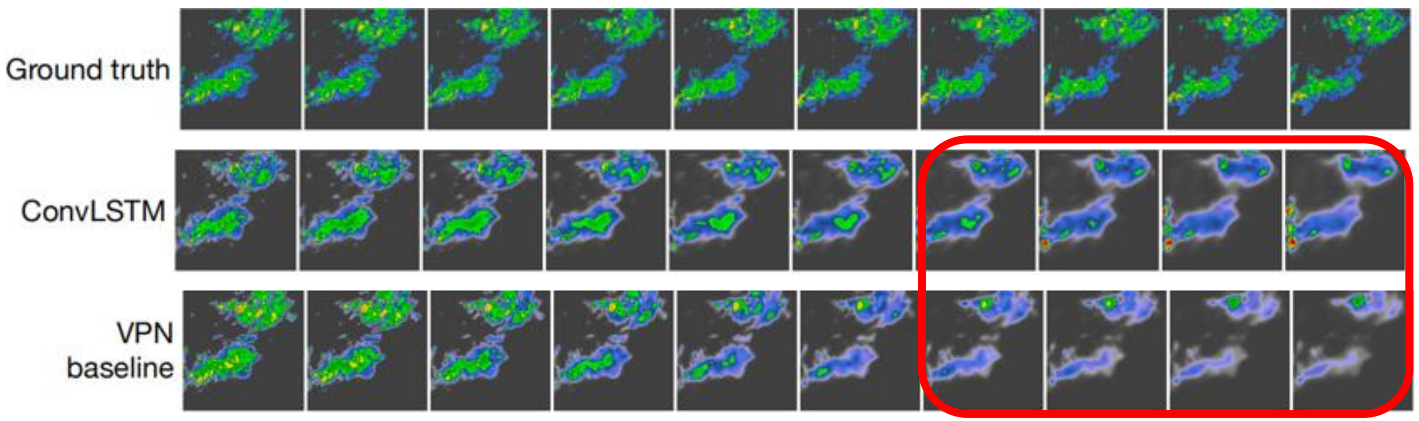
(b) ConvLSTM cell

CAVEATS OF SIMPLE VIDEO PREDICTION MODELS

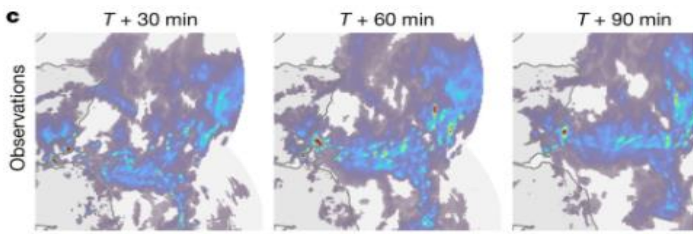
Limitation of applying pixel-wise loss

- Models based on point-to-point losses generate blurry images in autoregressive forecasting →
  - Decreased local spatial variability
  - Deteriorated capability for predicting extremes

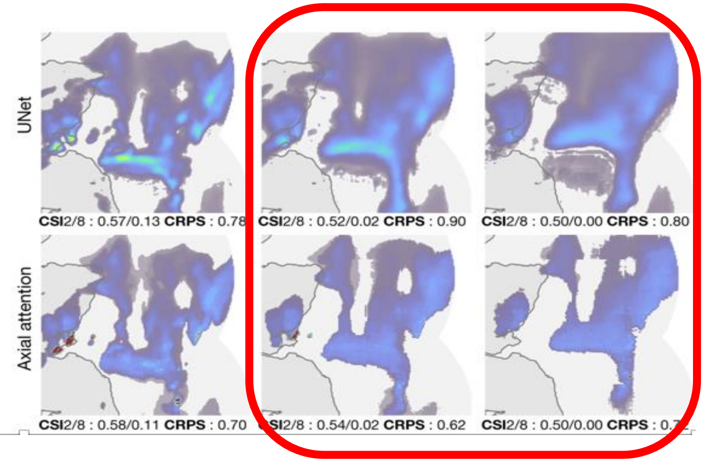
✓ Solution: Generative modelling



Adopted from Wang et al. (2017)



Adopted from Ravuri et al. (2021)



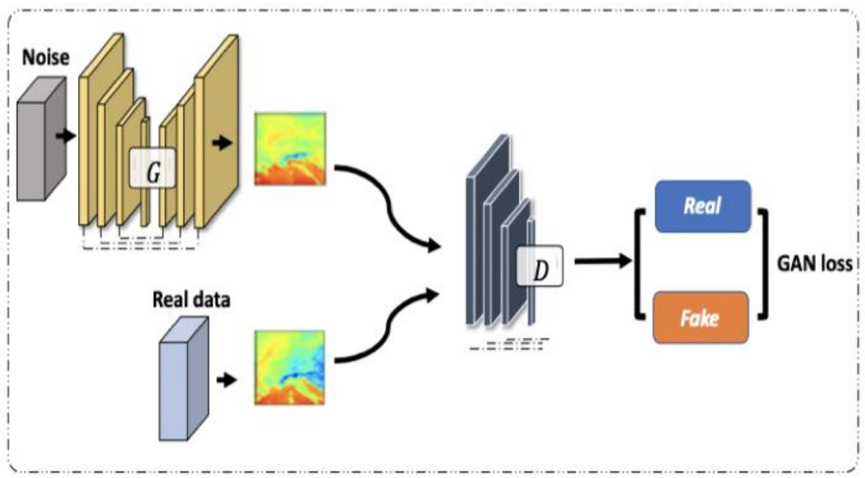
(Ravuri S et. al, 2021)

JÜLICH Forschungszentrum

GENERATIVE MODELS

Implicit Density Modeling

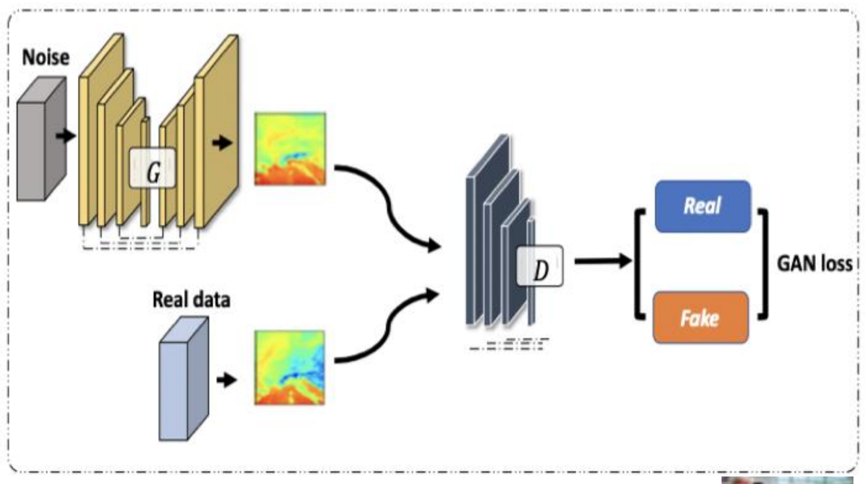
- Generator to reconstruct data at next time step
- Discriminator to distinguish between generated and real data sequences
- Generator and discriminator are trained adversarial in a minimax-optimization

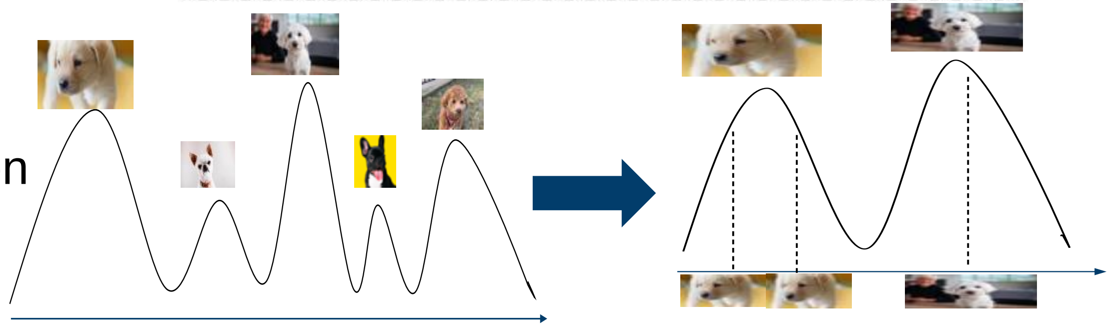


GENERATIVE MODELS

Implicit Density Modeling

- Generator to reconstruct data at next time step
- Discriminator to distinguish between generated and real data sequences
- Generator and discriminator are trained adversarial in a minimax-optimization
- **Problem:**
  - mode collapse (reduced diversity in prediction)
- **Remedy:**
  - Couple with VAE → Stochastic Adversarial Video prediction
  - Wasserstein GAN
  - Diffusion models





GENERATIVE MODELS

14M

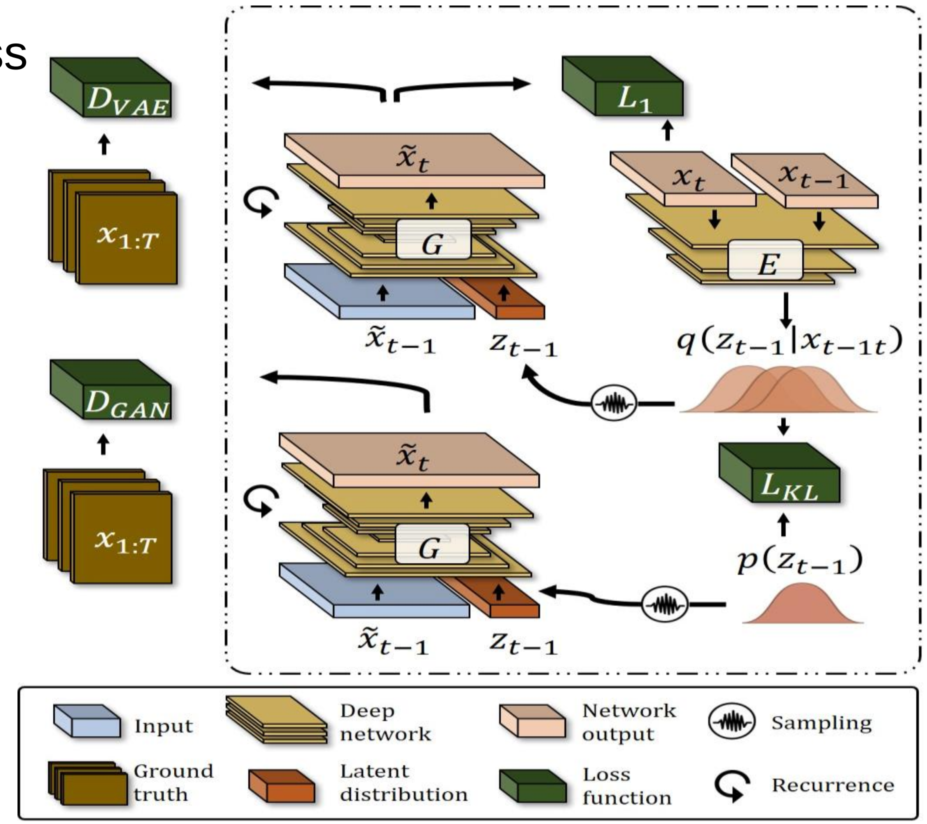
Stochastic adversarial video prediction (SAVP)

Composite model architecture: GAN (to overcome blurriness issue) + VAE (to overcome mode collapse issue)

- **SAVP** shares the same generator for VAE & GAN, but deploys two different discriminators

- **Loss function:**

$$G^* = \arg \min_{D} \max_{G} \lambda_1 \mathcal{L}_1(G,E) + \lambda_{kl}\ \mathcal{L}_{kl}(E) + \mathcal{L}_{GAN}(E,D) + \mathcal{L}_{GAN}^{VAE}(E,D)$$

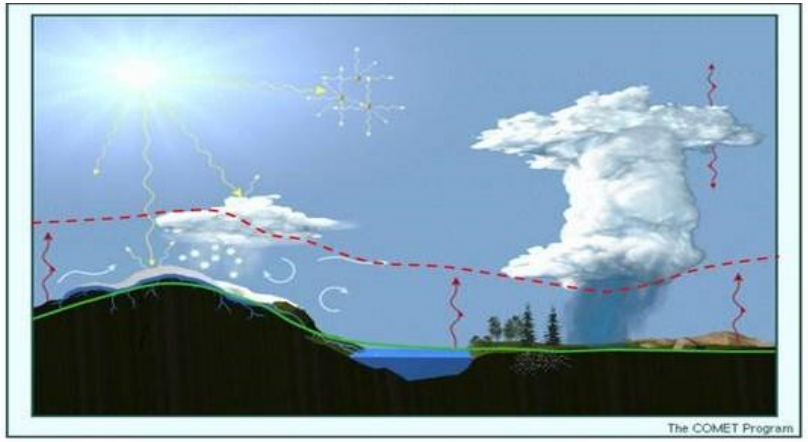


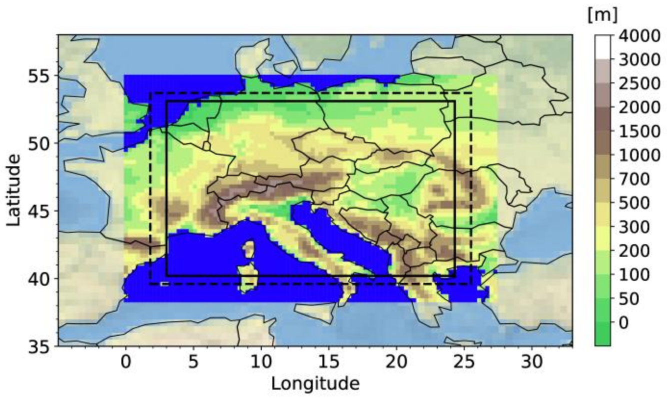
JÜLICH Forschungszentrum

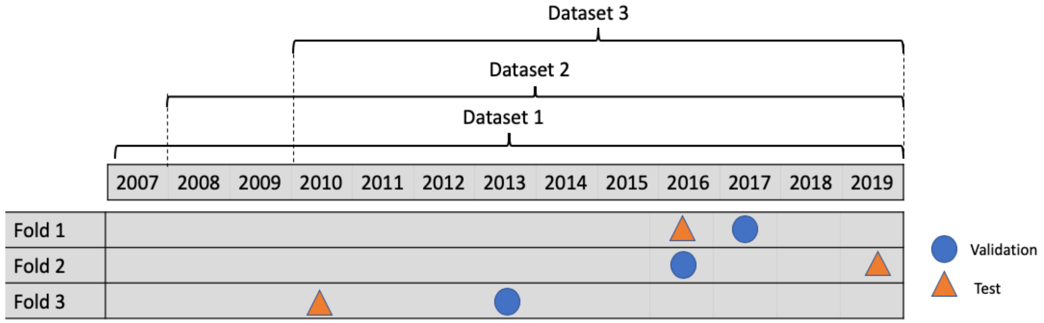
SHORT-TERM FORECASTING OF 2M TEMPERATURE

Experimental set-up

- Data source: ERA5 dataset from ECMWF
- Region: Crop hourly ERA5 reanalysis data to Central Europe: 92x56 grid points with Δx=0.3°
- Inputs: 2m temperature, 850 hPa temperature, Total cloud cover (hourly, preceding 12 hours)
- Outputs: 2m temperature (hourly, 12 hours lead time)
- Data period: 2007- 2019 (11 years for training)

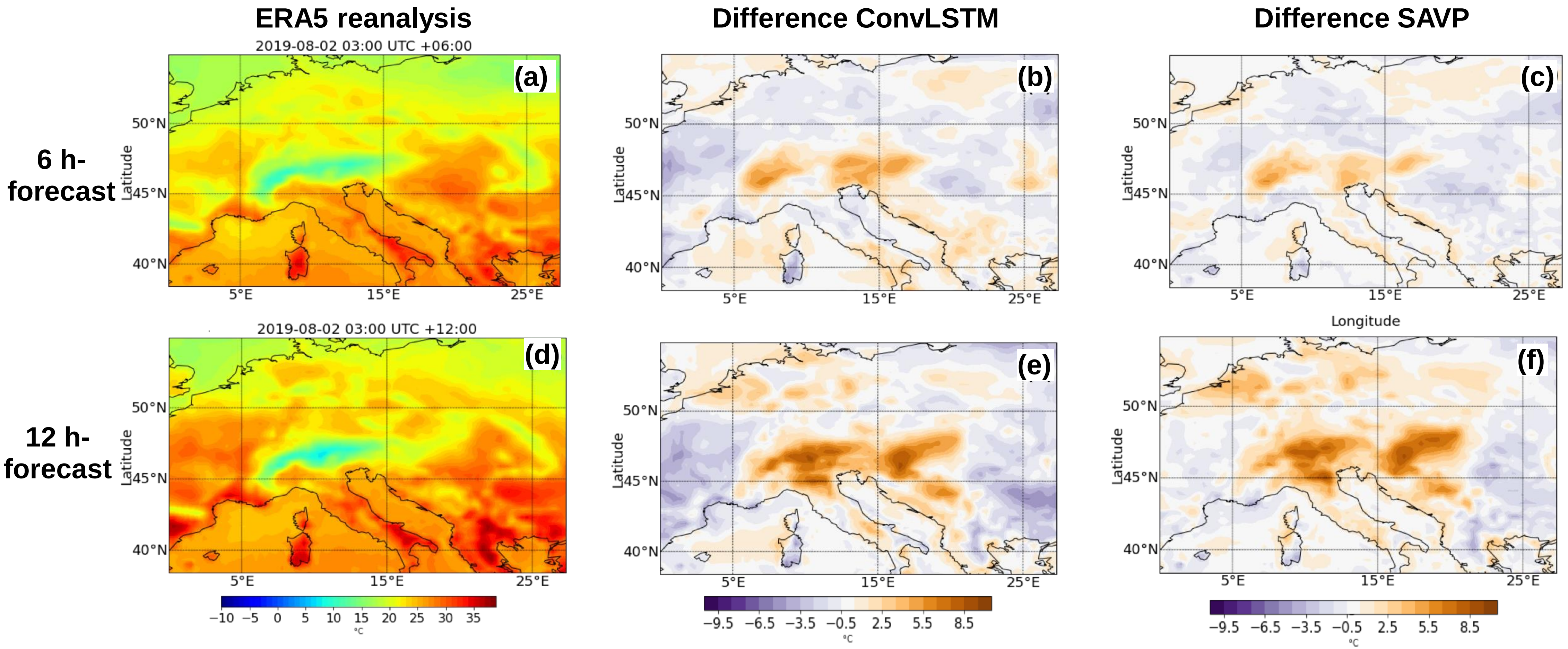






SHORT-TERM FORECASTING OF 2M TEMPERATURE

Results: An illustrative case study…



JÜLICH Forschungszentrum

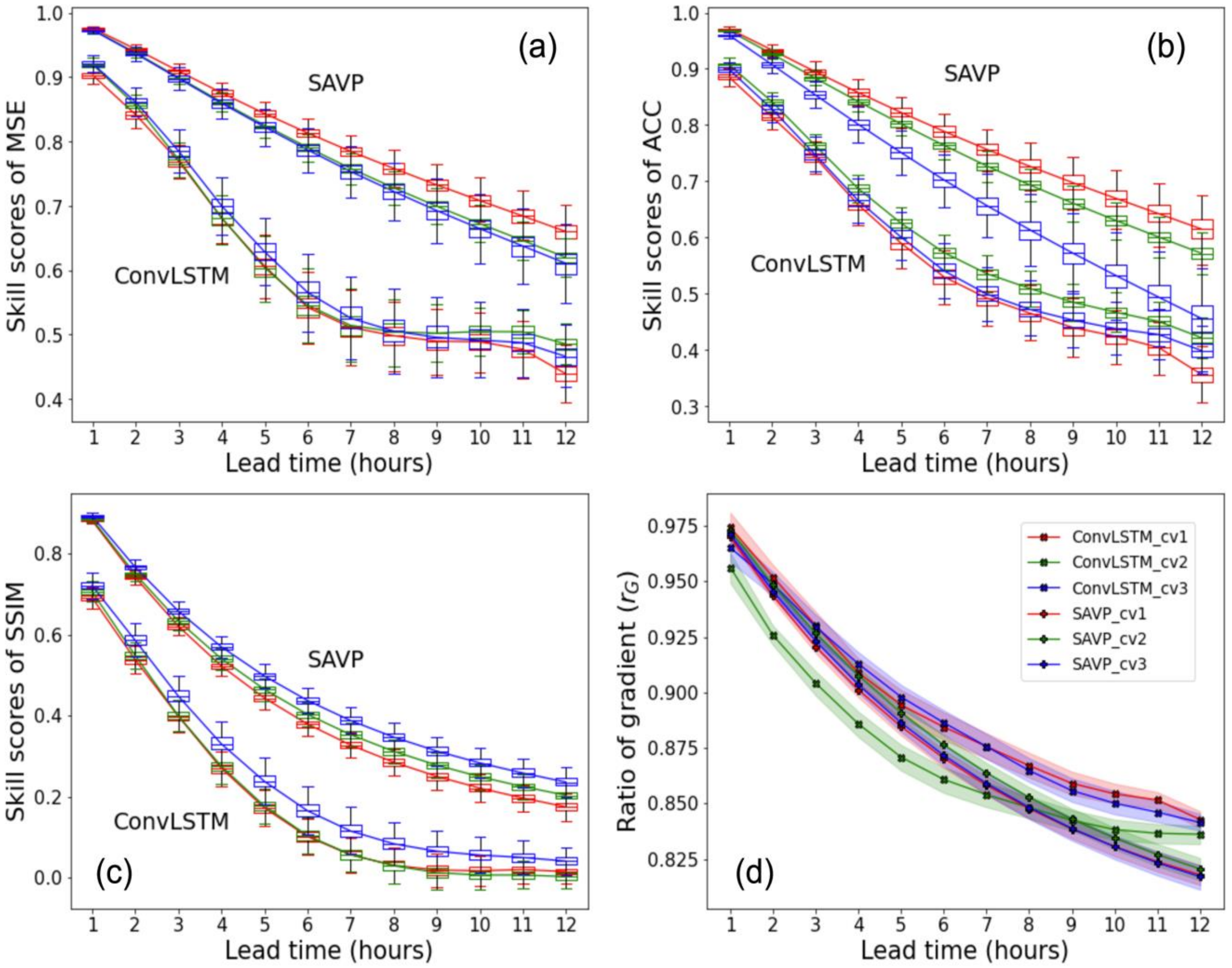
SHORT-TERM FORECASTING OF 2M TEMPERATURE

Comparison between deep learning models …

$$G^* = \arg\min\max \lambda_1 \mathcal{L}_1(G,E) + \lambda_{kl}\ \mathcal{L}_{kl}(E) \quad + \mathcal{L}_{GAN}(E,D) + \mathcal{L}_{GAN}^{VAE}(E,D)$$

Set-up : Strong scaling factor for L1-error in SAVP loss function ($\lambda_1 = 10^4$)

✓ Both models significantly outperform persistence forecasting (skill scores > 0)



JÜLICH Forschungszentrum

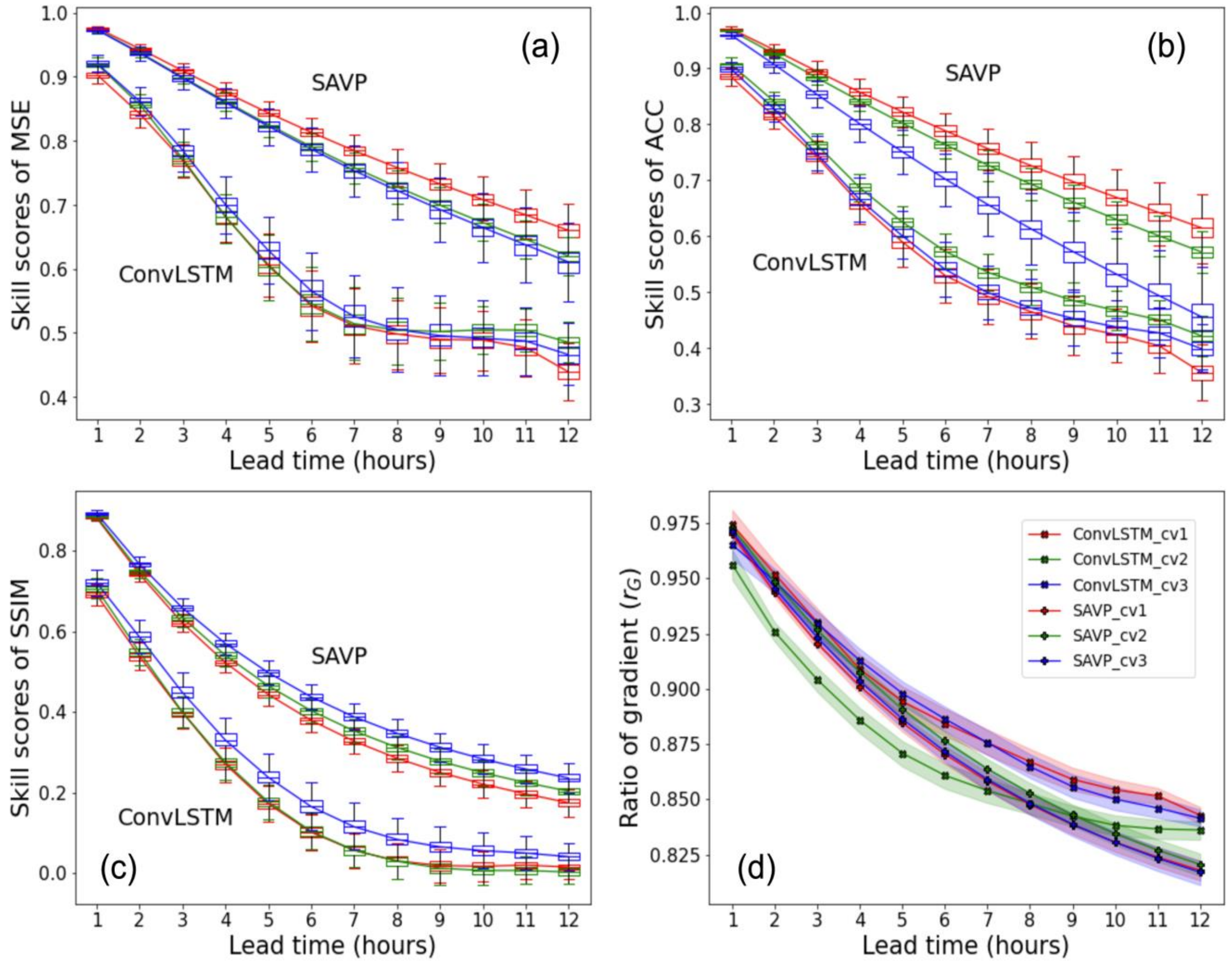
SHORT-TERM FORECASTING OF 2M TEMPERATURE

Comparison between deep learning models …

$$G^* = \arg \min \max \lambda_1 \mathcal{L}_1(G,E) + \lambda_{kl}\ \mathcal{L}_{kl}(E) \quad + \mathcal{L}_{GAN}(E,D) + \mathcal{L}_{GAN}^{VAE}(E,D)$$

Set up : Strong scaling factor for L1-error in SAVP loss function ($\lambda_1 = 10^4$)

- ✓ Both models significantly outperform persistence forecasting (skill scores > 0)
- ✓ SAVP is significantly superior to ConvLSTM



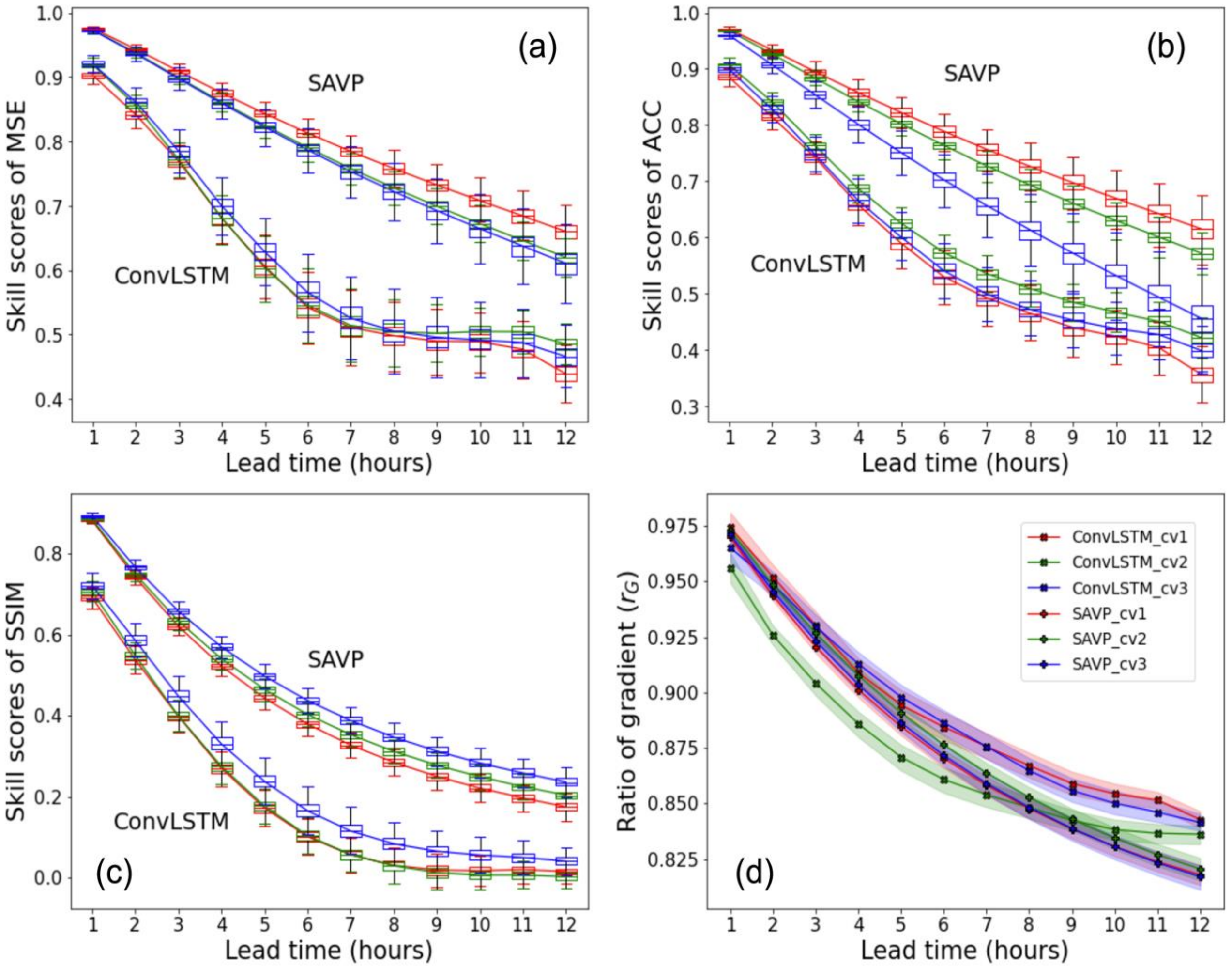
SHORT-TERM FORECASTING OF 2M TEMPERATURE

Comparison between deep learning models …

$$G^* = \arg\min\max \lambda_1 \mathcal{L}_1(G,E) + \lambda_{kl}\ \mathcal{L}_{kl}(E) \quad + \mathcal{L}_{GAN}(E,D) + \mathcal{L}_{GAN}^{VAE}(E,D)$$

Set up : Strong scaling factor for L1-error in SAVP loss function ($\lambda_1 = 10^4$)

- ✓ Both models significantly outperform persistence forecasting (skill scores > 0)
- ✓ SAVP is significantly superior to ConvLSTM
- ✓ ConvLSTM is notorious (notorious for what? -> missing adjective) for longer lead times



JÜLICH Forschungszentrum

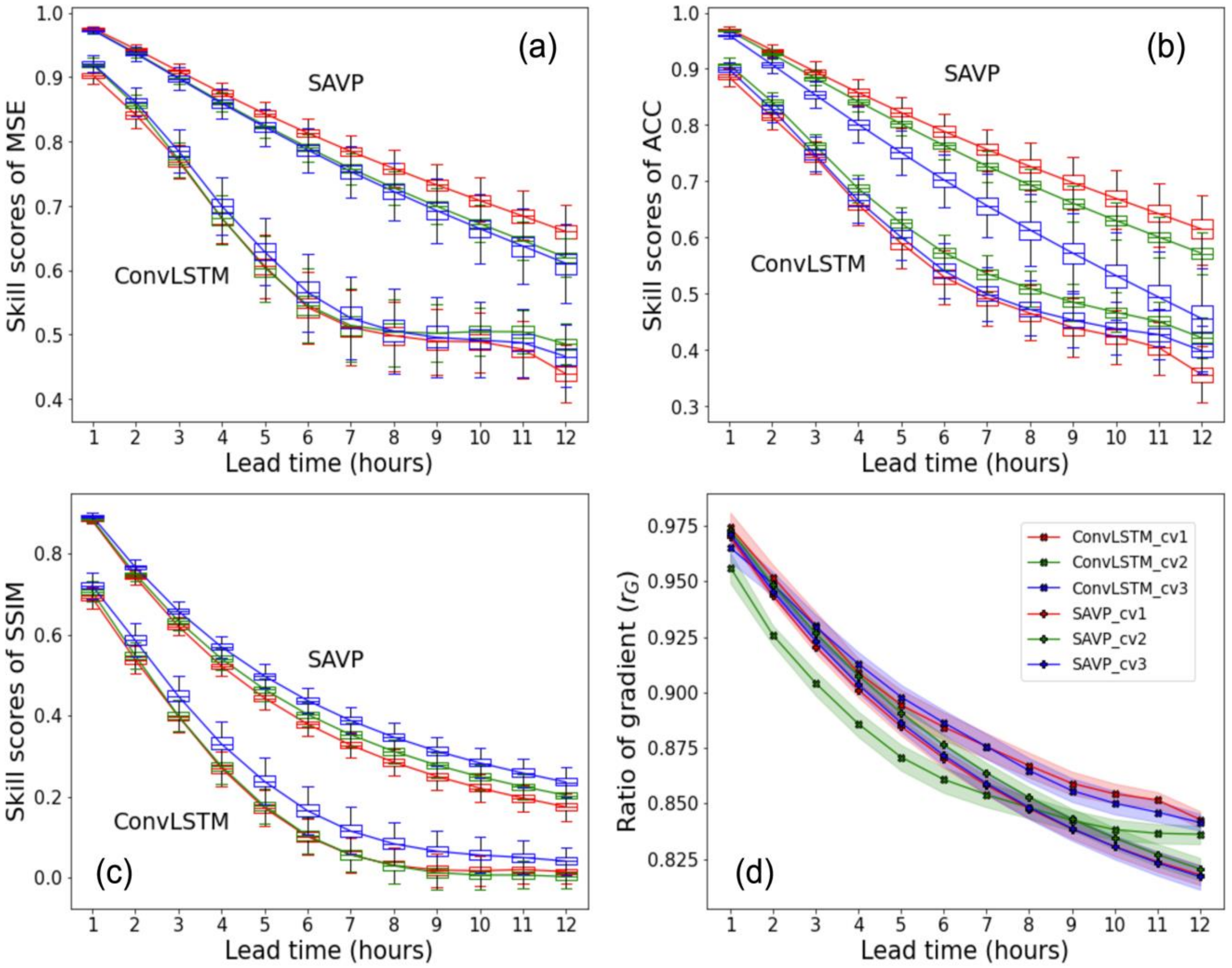
SHORT-TERM FORECASTING OF 2M TEMPERATURE

Comparison among deep learning models …

$$G^* = \arg\min\max \lambda_1 \mathcal{L}_1(G,E) + \lambda_{kl}\ \mathcal{L}_{kl}(E) \quad + \mathcal{L}_{GAN}(E,D) + \mathcal{L}_{GAN}^{VAE}(E,D)$$

Set up : Strong scaling factor for L1-error in SAVP loss function ($\lambda_1 = 10^4$)

- ✓ Both models significantly outperform persistence forecasting (Skill scores > 0)
- ✓ SAVP is significantly superior to ConvLSTM
- ✓ ConvLSTM is notorious (notorious for what? -> missing adjective) for longer lead times
- ✓ Strong generator enables performance improvement (MSE, ACC, SSIM)



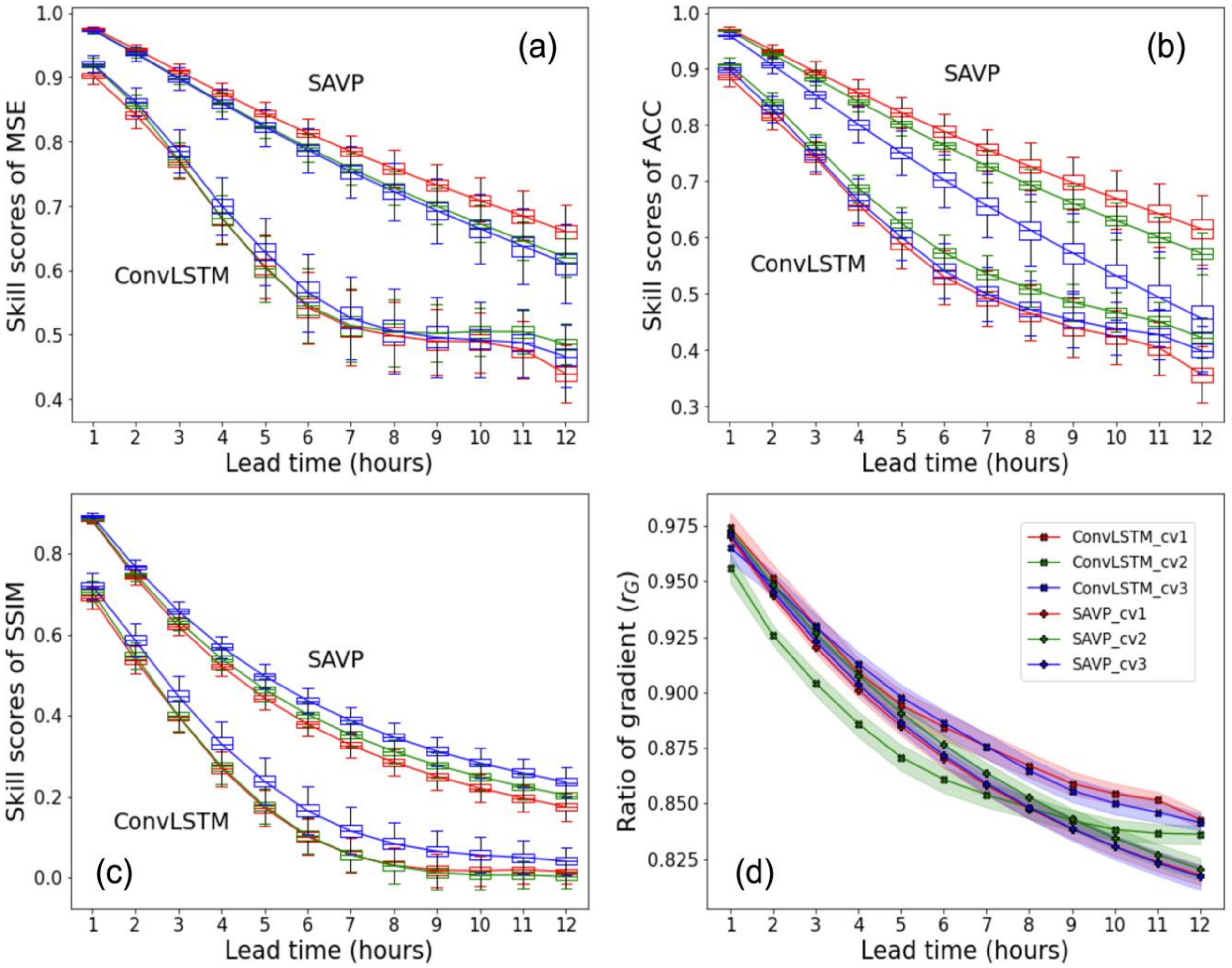
SHORT-TERM FORECASTING OF 2M TEMPERATURE

Comparison among deep learning models …

$$G^* = \arg \min \max \lambda_1 \mathcal{L}_1(G,E) + \lambda_{kl} \mathcal{L}_{kl}(E) + \mathcal{L}_{GAN}(E,D) + \mathcal{L}_{GAN}^{VAE}(E,D)$$

Set up : Strong scaling factor for L1-error in SAVP loss function ($\lambda_1 = 10^4$)

- ✓ Both models significantly outperform persistence forecasting (Skill scores > 0)
- ✓ SAVP is significantly superior to ConvLSTM
- ✓ ConvLSTM is notorious (notorious for what? -> missing adjective) for longer lead times
- ✓ Strong generator enables performance improvement (MSE, ACC, SSIM)
- ✓ No significant difference in terms of local spatial variability



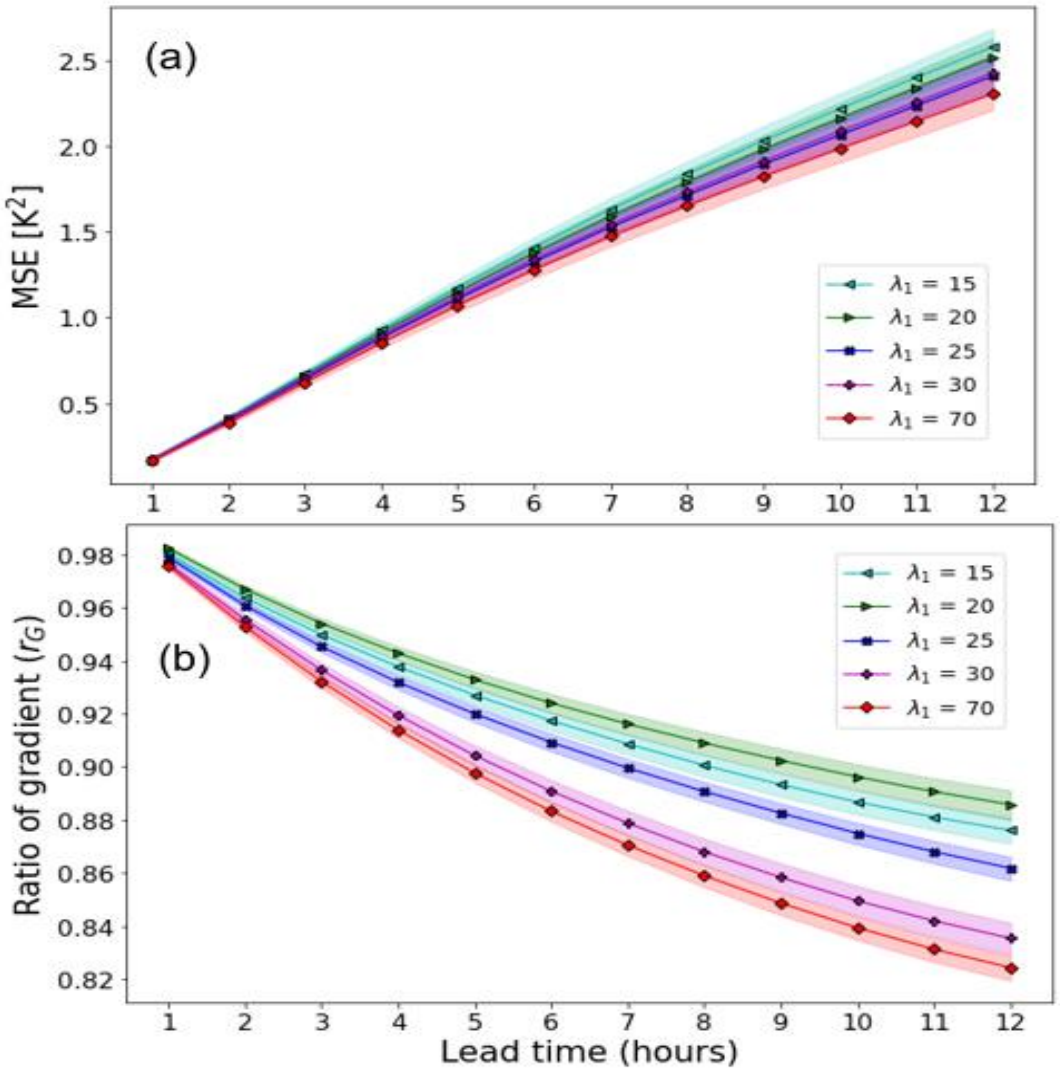
JÜLICH Forschungszentrum

SHORT-TERM FORECASTING OF 2M TEMPERATURE

Ablation study

$$G^* = \arg\min\max \lambda_1 \mathcal{L}_1(G,E) + \lambda_{kl}\ \mathcal{L}_{kl}(E) + \mathcal{L}_{GAN}(E,D) + \mathcal{L}_{GAN}^{VAE}(E,D)$$

✓ Small sensitivity for $\lambda_1 > 100$, but larger sensitivity for $\lambda_1 < 100$ (= Strong increase in importance of GAN-component)



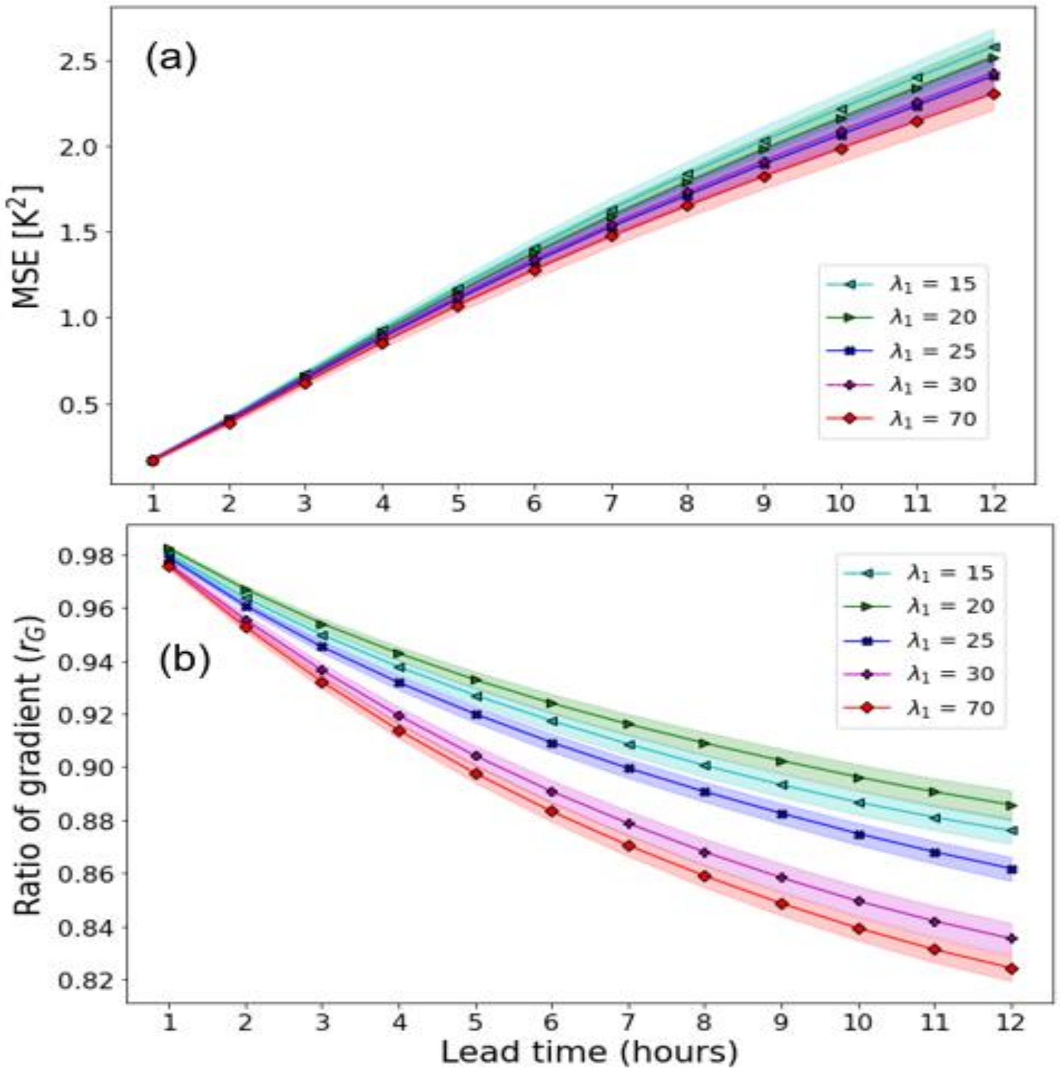
JÜLICH Forschungszentrum

SHORT-TERM FORECASTING OF 2M TEMPERATURE

Ablation study

$$G^* = \arg \min \max \lambda_1 \mathcal{L}_1(G,E) + \lambda_{kl} \mathcal{L}_{kl}(E) + \mathcal{L}_{GAN}(E,D) + \mathcal{L}_{GAN}^{VAE}(E,D)$$

- ✓ Small sensitivity for $\lambda_1 > 100$, but larger sensitivity for $\lambda_1 < 100$ (= Strong increase in importance of GAN-component)
- ✓ MSE slightly increases when λ_1 is decreased
- ✓ Gradient ratio (local variability) increases significantly



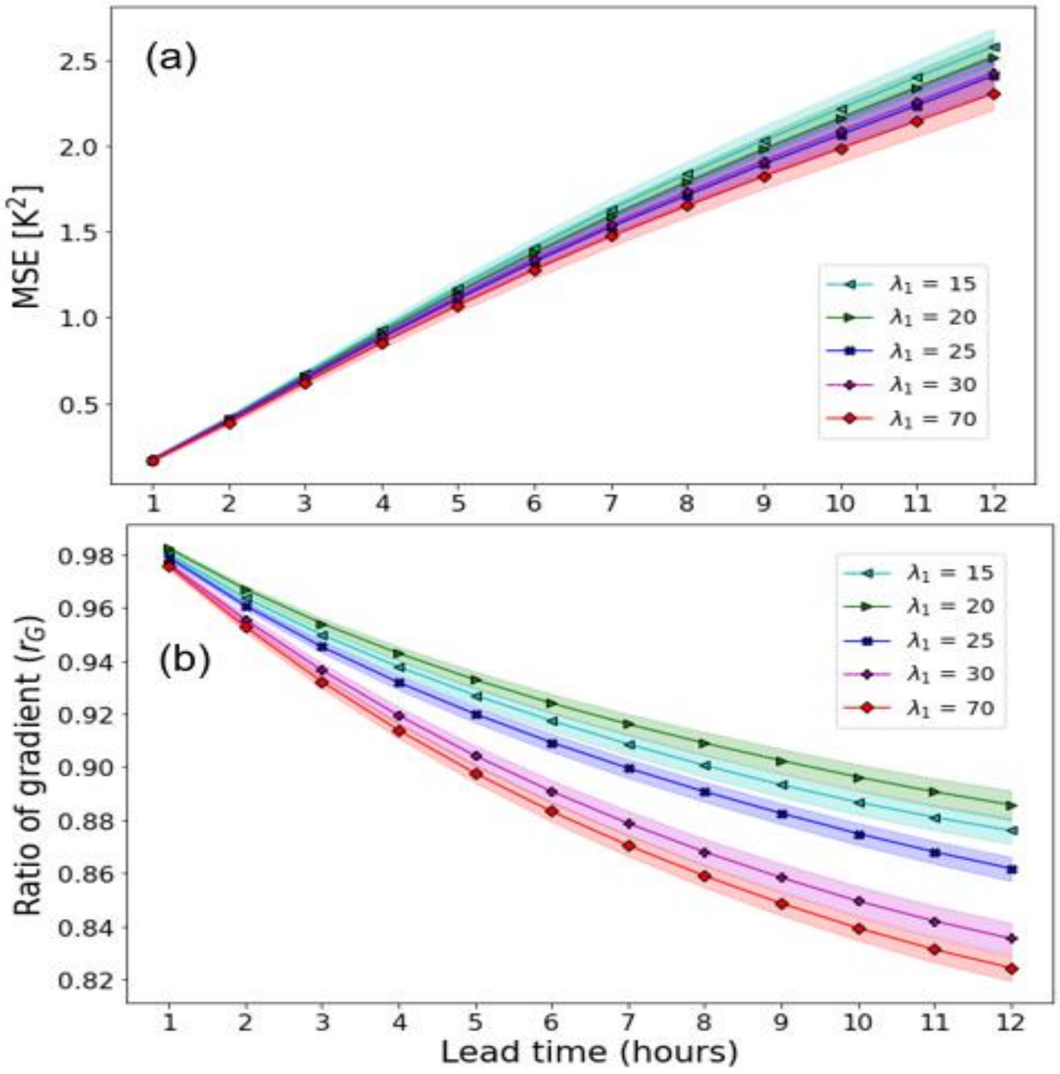
JÜLICH Forschungszentrum

SHORT-TERM FORECASTING OF 2M TEMPERATURE

Ablation study

$$G^* = \arg \min \max \lambda_1 \mathcal{L}_1(G,E) + \lambda_{kl} \mathcal{L}_{kl}(E) + \mathcal{L}_{GAN}(E,D) + \mathcal{L}_{GAN}^{VAE}(E,D)$$

- ✓ Small sensitivity for $\lambda_1 > 100$, but larger sensitivity for $\lambda_1 < 100$ (= Strong increase in importance of GAN-component)
- ✓ MSE slightly increases when λ_1 is decreased
- ✓ Gradient ratio (local variability) increases significantly
- ✓ **Trade-off** between MSE and Gradient ratio



CONCLUSION

Key messages:

1) **Can we use video prediction approach to predict the diurnal cycle of 2m temperature?**
   - **Yes,** the video prediction attain predictive skills, also for 2m temperature on sub-daily scale

JÜLICH Forschungszentrum

CONCLUSION

Key messages:

1) **Can we use video prediction approach to predict the diurnal cycle of 2m temperature?**
   - **Yes,** the video prediction attain predictive skills, also for 2m temperature on sub-daily scale.
2) **Are advanced video prediction models beneficial for predicting the 2m temperature compare to shallow ones?**
   - **Yes,** the state-of-the-art video prediction model can significantly improve the 2 m temperature accuracy. The predictors and the size of target region are also essential factors.

CONCLUSION

Key messages:

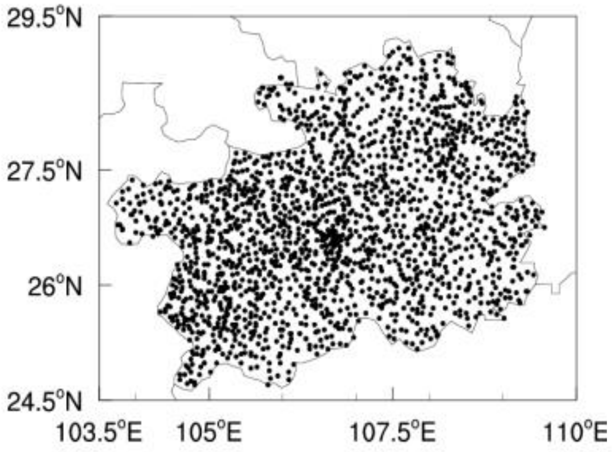
1) **Can we use video prediction approach to predict the diurnal cycle of 2m temperature?**
   - **Yes,** the video prediction attain predictive skills , also for 2m temperature on sub-daily scale
2) **Are advanced video prediction models beneficial for predicting the 2m temperature compare to shallow ones?**
   - **Yes,** the state-of-the-art video prediction model can significantly improve the 2 m temperature accuracy. The predictors and the size of target region are also essential factors.
3) **Can Generative Adversarial Networks (GANs) help to enhance the performance of 2m temperature forecasting?**
   - A strong generator is key to improve the performance in terms of point-to-point evaluation metrics (MSE and ACC) and global scale structure evaluation (SSIM). GAN-component is beneficial for the local scale variability (gradient amplitude ratio). A trade-off between MSE and local variability is observed (dependent on weight for the GAN-component)

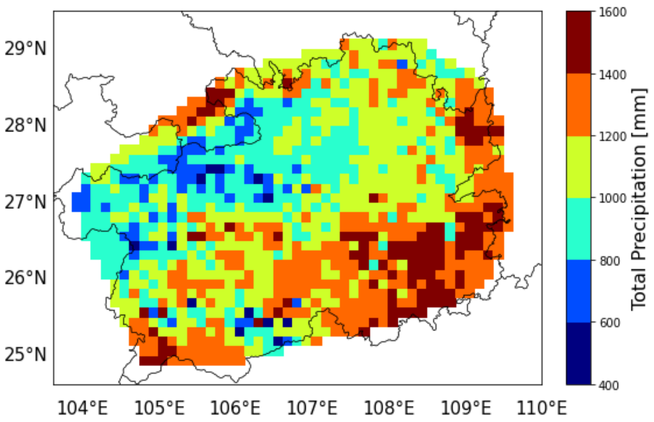
PRECIPITATION NOWCASTING

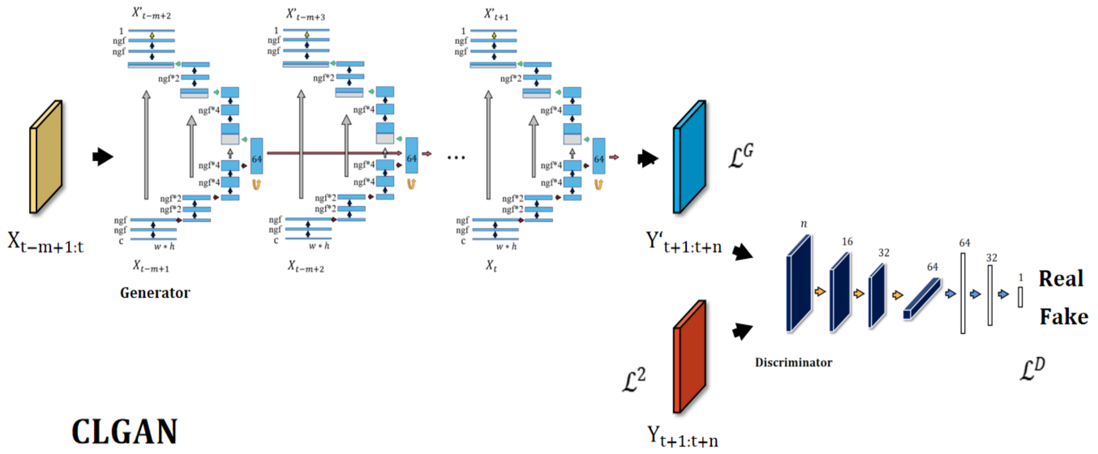
Experiment setting

- Data source: Guizhou_minute_AWS_data
- Time resolution: 10-minutes
- Variables: prcp
- Spatial resolution: 0.125 degrees
- Data period:
  - 2013- 2017 (training), 2018 (validation), 2019 (Test)
- Data preprocessing:
  - Bilinear interpolation
  - Rainy sequence selection
  - Log transformation
  - Min-Max normalization
- Model: CLGAN

$$\mathcal{L}^{*} = (1-\lambda)\mathcal{L}^{G} + \lambda\mathcal{L}^{2}, \lambda \in [0,1]$$







CLGAN

JÜLICH Forschungszentrum

PRECIPITATION NOWCASTING

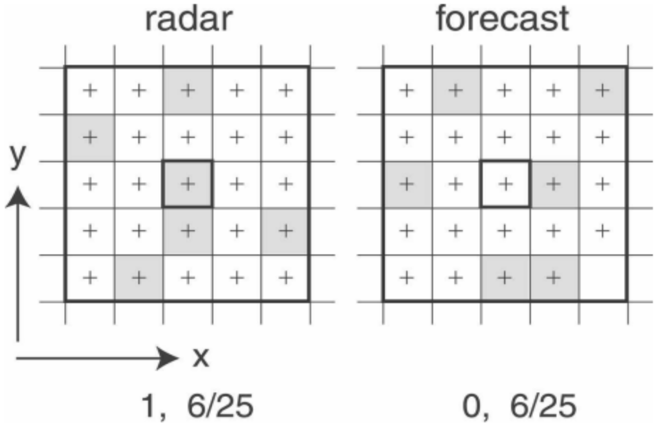
Verification methods

- Methods for forecasts of continuous variables: RMSE, Correlation Coefficient
- Methods for dichotomous (yes/no) forecasts: CSI
- Methods for spatial forecasts:
  - FSS (Fractions skill score)
  - MODE (the Method for Object-based Diagnostic Evaluation)

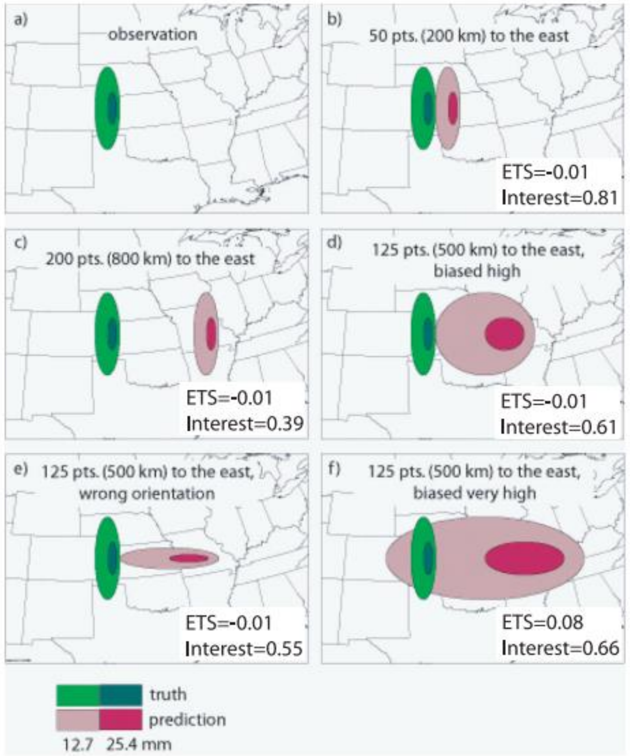
$$\boldsymbol{CSI} = \frac{hits}{hits + misses + false\ alarms}$$

Contingency Table

		Observed		
		yes	no	Total
Forecast	yes	***hits***	***false alarms***	***forecast yes***
	no	***misses***	***correct negatives***	***forecast no***
Total		***observed yes***	***observed no***	***total***



Adopted from Roberts and Lean (2008)

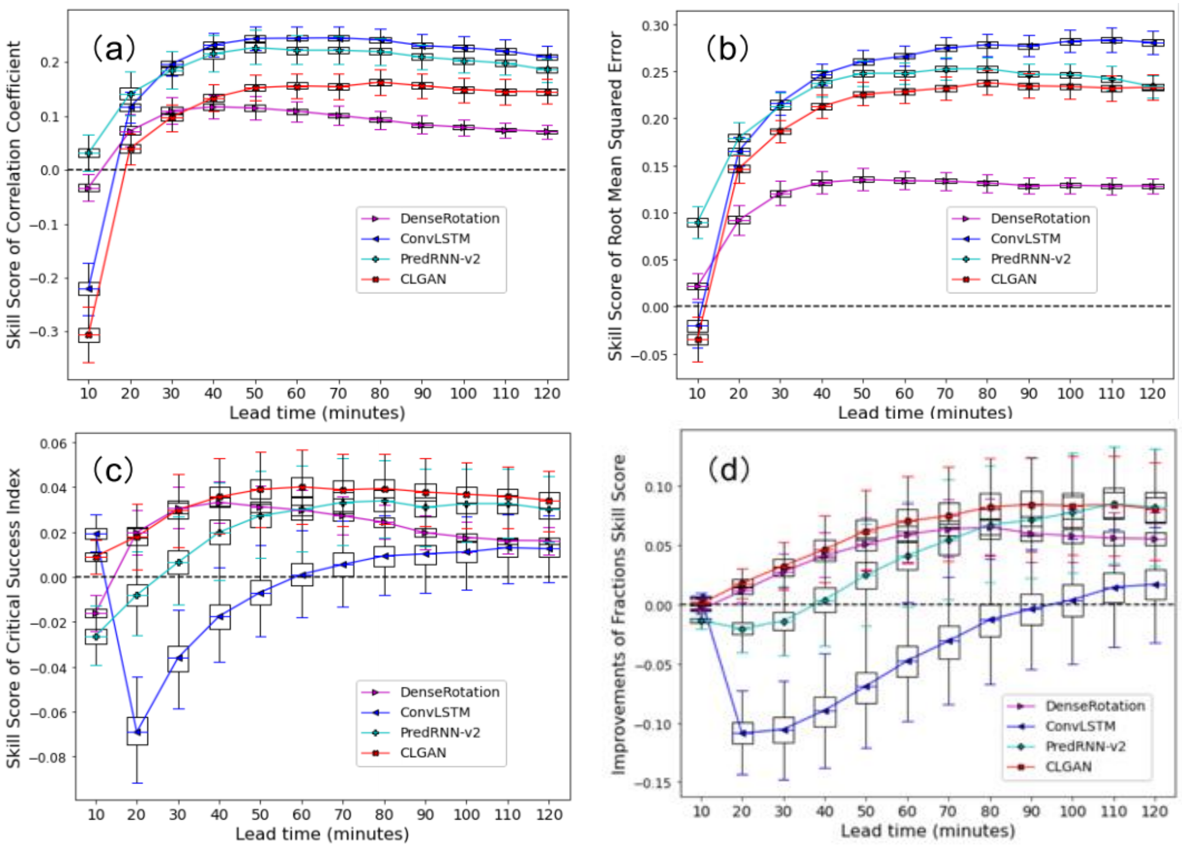


Adopted from Davis et al. (2009)

PRECIPITATION NOWCASTING

Results: Comparison among deep learning models …

- ✓ ConvLSTM and PredRNN-v2 are superior in terms of point-by-point scores (CC and RMSE)
- ✓ CLGAN performs best in terms of for dichotomous and spatial forecast scores (CSI and FSS)
  - ✓ More capability to forecast heavy precipitation events
  - ✓ More accurate prediction of the precipitation location
- ✓ CLGAN and PredRNN-v2 are able to capture precipitation area fairly well
- ✓ The location of precipitation centroids is generally well captured by all models
- X The orientation angle and the aspect ratio of the precipitation objects cannot be well simulated

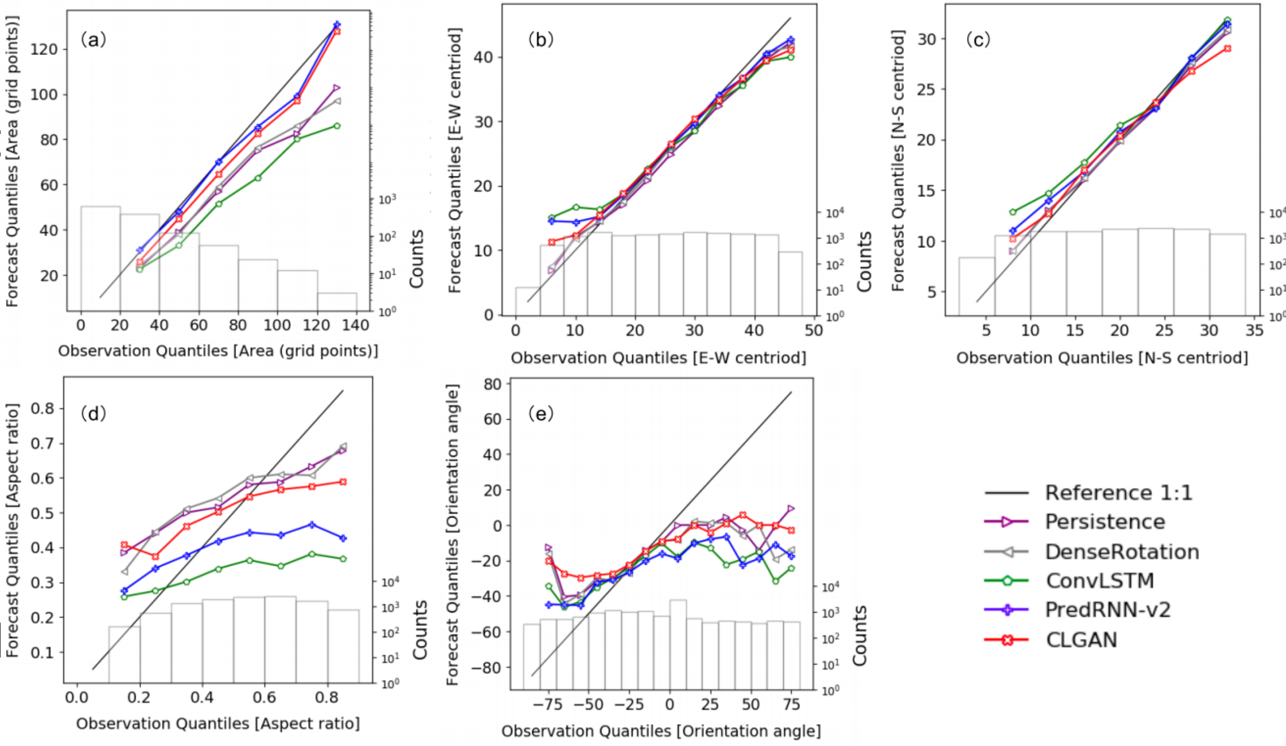


JÜLICH Forschungszentrum

PRECIPITATION NOWCASTING

Results: Comparison among deep learning models …

- ✓ ConvLSTM and PredRNN-v2 are superior in terms of point-by-point scores (CC and RMSE)
- ✓ CLGAN performs best in terms of for dichotomous and spatial forecast scores (CSI and FSS)
  - ✓ More capability to forecast heavy precipitation events
  - ✓ More accurate prediction of the precipitation location
- ✓ CLGAN and PredRNN-v2 are able to capture precipitation area fairly well
- ✓ The location of precipitation centroids is generally well captured by all models.
- X The orientation angle and the aspect ratio of the precipitation objects cannot be well simulated



PUBLICATIONS

https://doi.org/10.5194/gmd-15-8931-2022

Model D

Temperature forecasting by deep learning metho

Bing Gong, Michael Langguth, Yan Ji, Amirpasha Mozaffari, Scarlet Stadtler, Kari

Jülich Supercomputing Centre, Forschungszentrum Jülich, 52425 Jülich, Germany

Correspondence: Bing Gong (b.gong@fz-juelich.de)

Received: 22 December 2021 – Discussion started: 8 March 2022
Revised: 1 September 2022 – Accepted: 7 September 2022 – Published: 13 December 202

Abstract. Numerical weather prediction (NWP) models solve a system of partial differential equations based on physical laws to forecast the future state of the atmosphere. These models are deployed operationally, but they are computationally very expensive. Recently, the potential of deep neural networks to generate bespoke weather forecasts has been explored in a couple of scientific studies inspired by the success of video frame prediction models in computer vision. In this study, a simple recurrent neural network with convolutional filters, called ConvLSTM, and an advanced generative network, the Stochastic Adversarial Video Prediction (SAVP) model, are applied to create hourly forecasts of the

hances the forecast quali
a larger spatial domain.
cover as predictor or re
to 8 years has only smal
forecasts obtained in this
temporary NWP models,
ticated deep neural netw
cast quality beyond the
driven way.

https://doi.org/10.5194/egusphere-2022-859
Preprint. Discussion started: 14 November 2022

CLGAN: A GAN-based video prediction model for precipitation nowcasting

Yan Ji[1,2], Bing Gong[2], Michael Langguth[2], Amirpasha Mozaffari[2], and Xiefei Zhi[1]

[1]Nanjing University of Information Science and Technology, 210044 Nanjing, China
[2]Jülich Supercomputing Centre, Forschungszentrum Jülich, 52425 Jülich, Germany

Correspondence: b.gong@fz-juelich.de

Abstract. The prediction of precipitation patterns at high spatio-temporal resolution up to two hours ahead, also known as precipitation nowcasting, is of great relevance in weather-dependant decision-making and early warning systems. In this study, we are aiming to provide an efficient and easy-to-understand model - CLGAN, to improve the nowcasting skills of heavy precipitation events with deep neural networks for video prediction. The model constitutes a Generative Adversarial Network (GAN) architecture whose generator is built upon an u-shaped encoder-decoder network (U-Net) equipped with recurrent LSTM cells to capture spatio-temporal features. A comprehensive comparison among CLGAN, and baseline models optical flow model DenseRotation as well as the advanced video prediction model PredRNN-v2 is performed. We show that CLGAN outperforms in terms of scores for dichotomous events and object-based diagnostics. The ablation study indicates that the GAN-based architecture helps to capture heavy precipitation events. The results encourage future work based on the proposed CLGAN architecture to improve the precipitation nowcasting and early-warning systems.

ACKNOWLEDGEMENT

Funding is provided through ERC Advanced grant ERC-2017-ADG #787576 by Martin Schultz

Funding is provided by the European Union's Horizon 2020 research and innovation programme and the EuroHPC Joint Undertaking under Grant Agreement 955513 and co-funded by the German Ministry of Education and Research (BMBF) under funding reference 16HPC029.