Journal Club
May 4, 2021

Jakob Schlör

Universität Tübingen

machine learning in climate science

May 5, 2021

Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

Journal Club

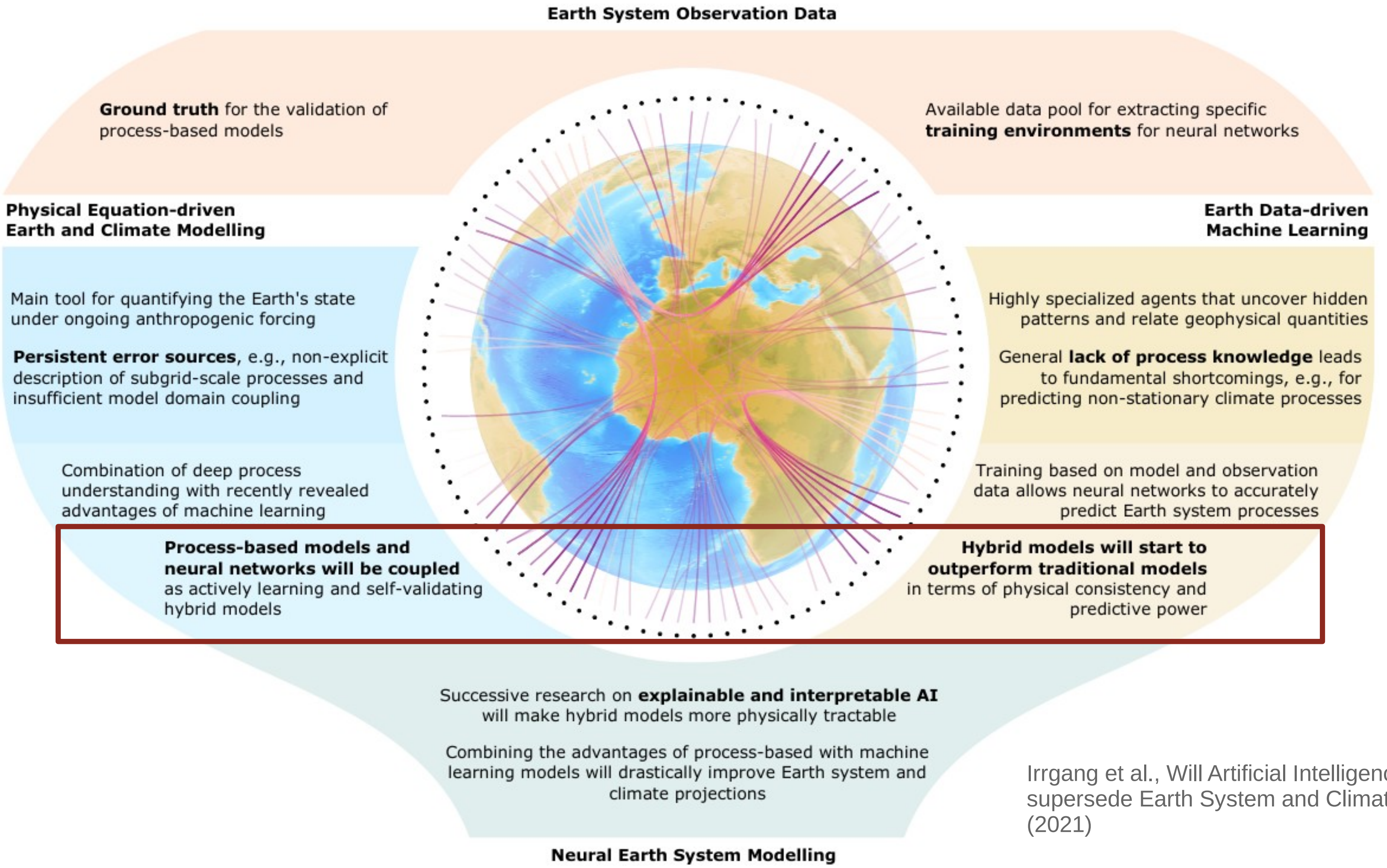
Deep learning for physical processes: incorporating prior scientific knowledge*

Emmanuel de Bézenac[1,3], Arthur Pajot[1,3] and Patrick Gallinari[2]

[1] Sorbonne Universites, UMR 7606, LIP6, F-75005 Paris, France
[2] Sorbonne Université, UMR 7606, LIP6, Paris and Criteo AI Lab, Paris, France

Jakob Schlör

Motivation



Irrgang et al., Will Artificial Intelligence supersede Earth System and ClimateModels? (2021)

In short

Question:

How to incorporate physical knowledge for designing a NN aimed at forecasting sea surface temperatures?

Results:

Improve SST forecasting (6 days) by combining NN with the advection-diffusion equation.

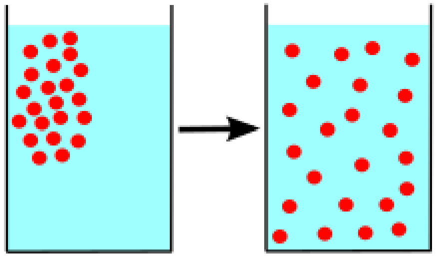
Impact:

Proposed hybrid model generalizes to a class of problems for forecasting spatio-temporal data.

Advection-diffusion equation

Advection: transport of substance or quantity by motion of a fluid

Diffusion: Movement of substance or quantity from regions of higher to regions of lower concentration



$$\underbrace{\frac{\partial I}{\partial t} + (\omega \cdot \nabla) I}_{\text{advection}} = \underbrace{D \nabla^2 I}_{\text{diffusion}}$$

$\mathrm{I(x,t)}$: sea surface temperature

$\omega \sim \frac{\Delta x}{\Delta t}$: motion field

D: diffusion coefficient

Advection-diffusion equation

$$\underbrace{\frac{\partial I}{\partial t} + (\omega \cdot \nabla) I}_{\text{advection}} = \underbrace{D\nabla^2 I}_{\text{diffusion}}$$

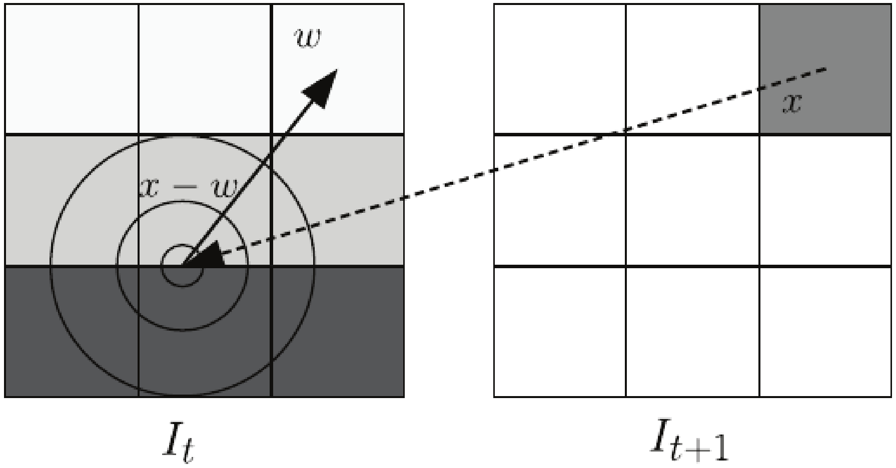
I(x,t) : sea surface temperature
$\omega \sim \frac{\Delta x}{\Delta t}$: motion field
D: diffusion coefficient

Global Solution:

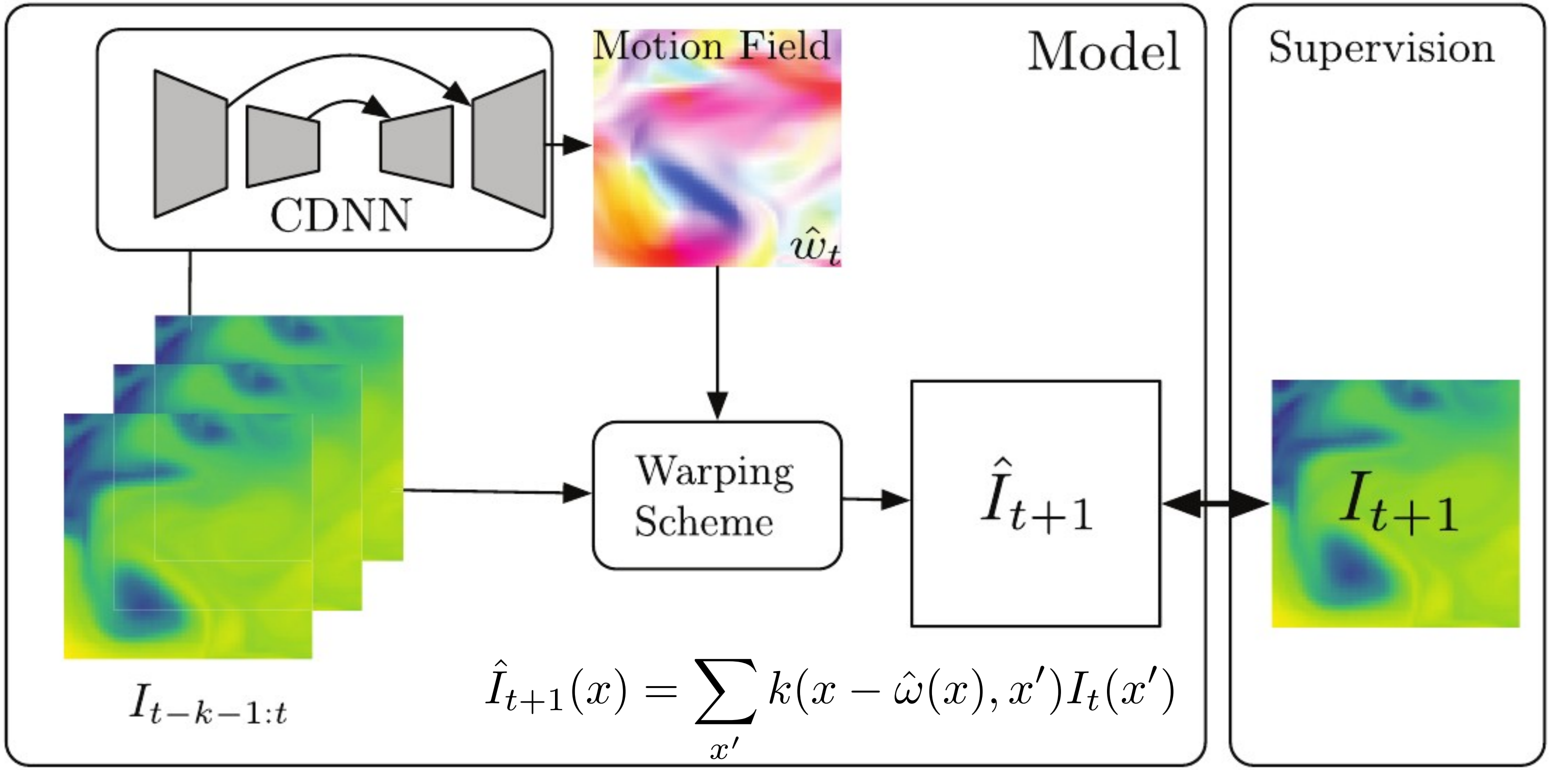
$$I(x,t) = \int_{\mathbb{R}^2} k(x - t\omega, x') I_0(x') dx'$$

RBF - kernel

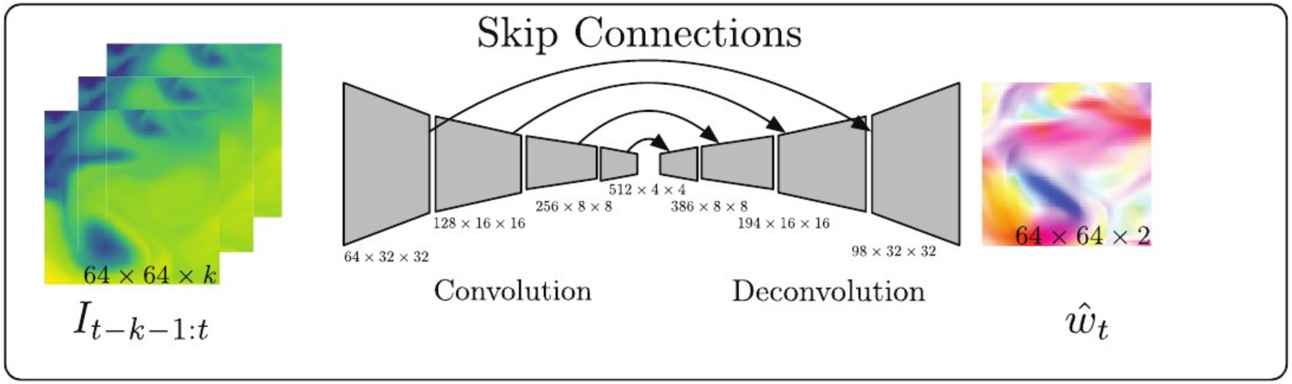
$$k(x,x') = \frac{1}{4\pi D t} e^{-\frac{1}{4Dt}|x-x'|^2}$$



Motion estimation



Covolution Deconvolution NN (CDNN)



Properties:

- Skip connections
- Batch normalization
- Leaky ReLU

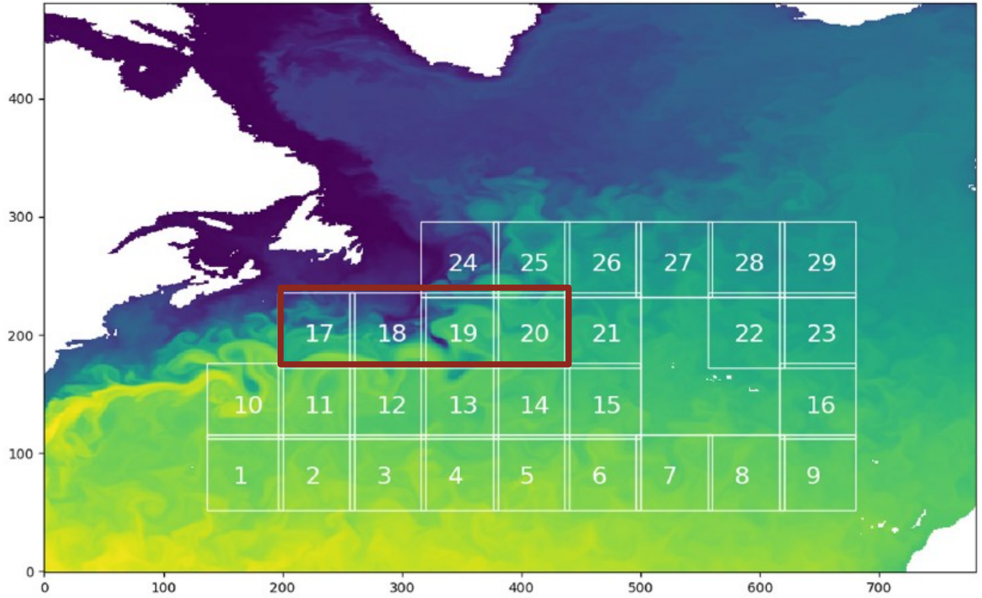
Loss:

$$L_t = \sum_{x \in \Omega} \rho\left(\hat{I}_{t+1}(x) - I_{t+1}(x)\right) + \underbrace{\lambda_{\text{div}} \left(\nabla \cdot w_t(x)\right)^2}_{\text{divergence}} + \underbrace{\lambda_{\text{magn}} \left\|w_t(x)\right\|^2}_{\text{magnitude}} + \underbrace{\lambda_{\text{grad}} \left\|\nabla w_t(x)\right\|^2}_{\text{smoothness}}$$

Charbonnier penalty function: $\rho(x) = (x + \epsilon)^{\frac{1}{\alpha}}$

Dataset

- Normalized sea surface temperature anomalies
- Daily temperature
- NOAA 6 satellite (with NEMO assimilation)
- Training/Validation data: 2006-2015
- Test data: 2016-2017



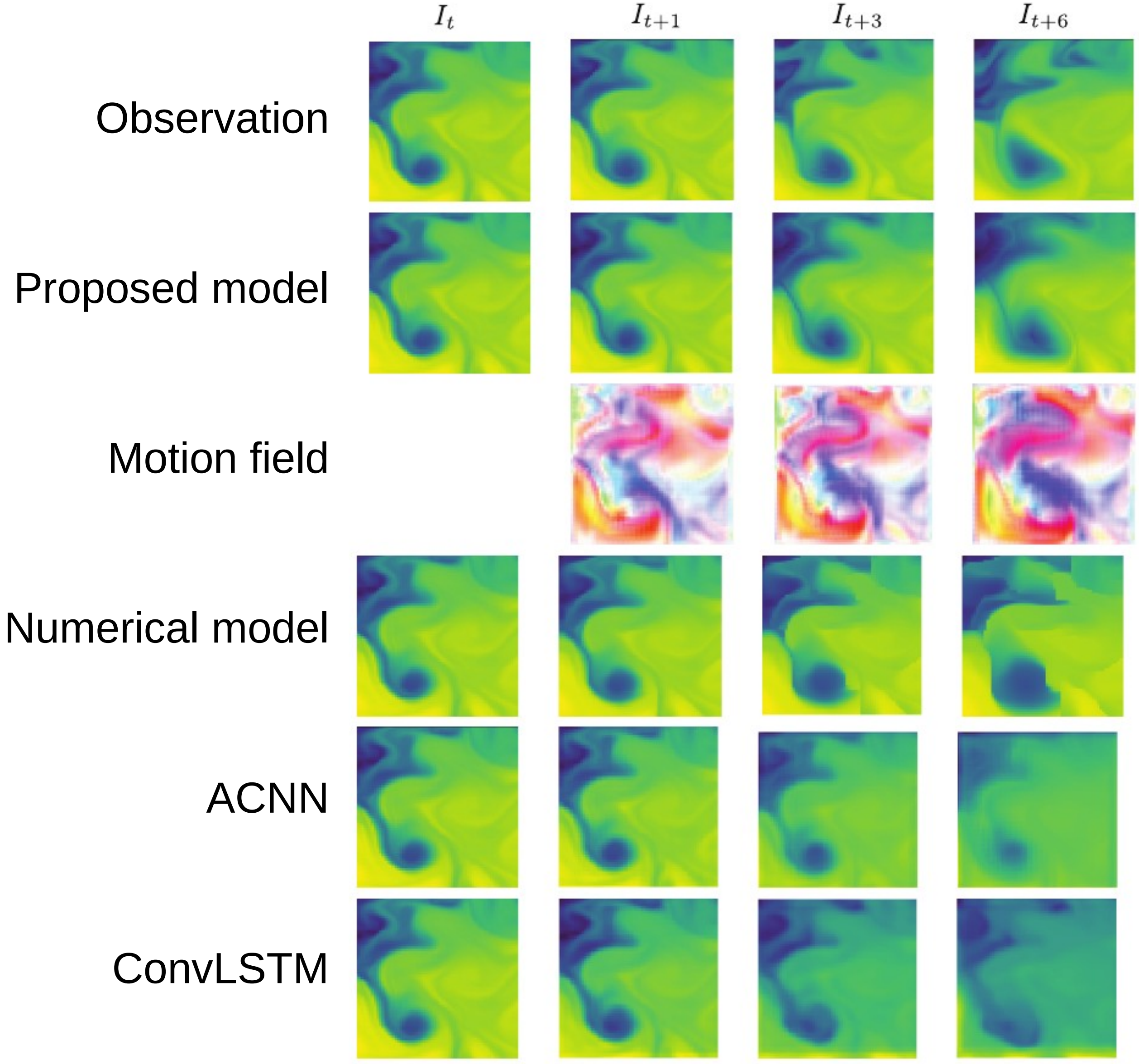
Assumption: sub-region contains enough information for forecasting

Results

- 6 day forecasts: $I_t \rightarrow I_{t+6}$

- Comparison to:
  - Numerical model based on shallow-water equation
  - Autoregressive CDNN directly on SST prediction
  - ConvLSTM
  - Autoregressive CNN trained as a GAN

Results

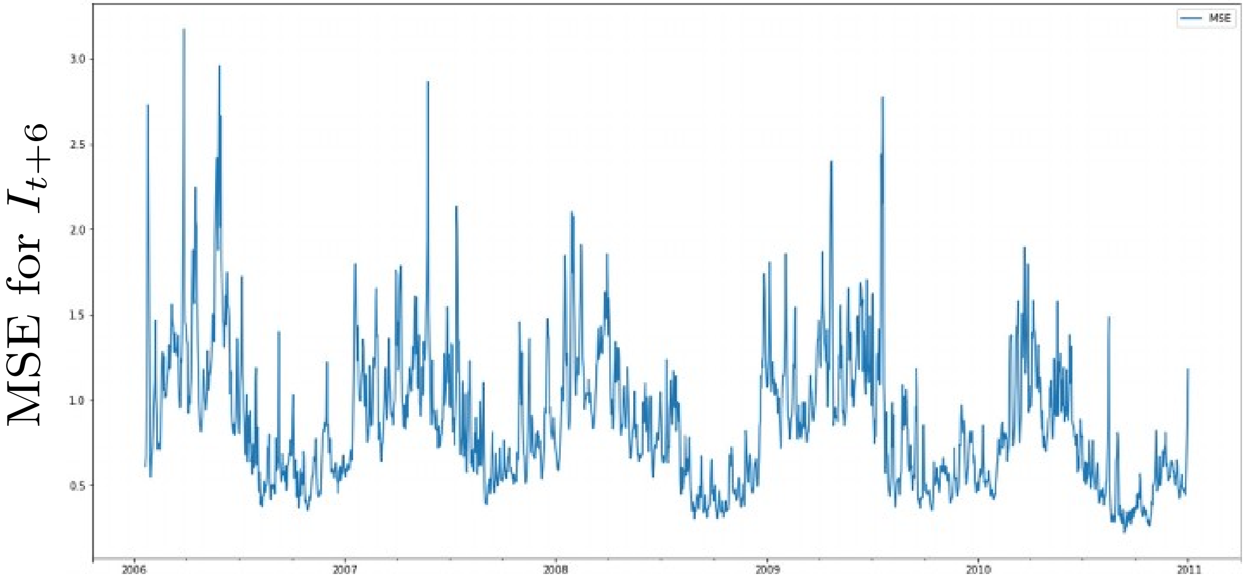

I_t
I_{t+1}
I_{t+3}
I_{t+6}
Observation
Proposed model
Motion field
Numerical model
ACNN
ConvLSTM

Results

Model	Average score (MSE)	Average time (s)
Numerical model (Béréziat and Herlin 2015)	1.99	4.8
ConvLSTM (Shi *et al* 2015)	5.76	0.018
ACNN	15.84	0.54
GAN video generation (Mathieu *et al* 2015)	4.73	0.096
Proposed model with regularization	**1.42**	0.040
Proposed model without regularization	2.01	0.040

- Proposed model performs similarly to the numerical model
- Computational time is strongly decreased in comparison to the numerical model

Results



Forecasting ability seem to be seasonal dependent

Summary

- Combining physical knowledge and CDNN outperforms purely data-driven NN models

- Proposed approach reaches comparable performance than numerical model

- Generalizes to problems which follow advection-diffusion principles

Shortcomings and improvements

- Uncertainty prediction
- Validating motion field
- Incorporating additional terms not captured by advection-diffusion equation
- Other examples to show generalizability

Take home message

- Read model papers before using a data-driven approach
- Incorporating known equations or principles from physics to a NN
  - Model architecture
  - Loss function

Kashinath et al., Physics-informed machine learning: case studies for weather and climate modelling (2021)

Numerical Model

Dynamics are based on the shallow water equations:

- Derived from depth-integrated Navier-Stokes equation (animation)
- conservation of mass and momentum
- Group all terms not related to advection into one Lagrangian variable
- Initial conditions derived from data assimilation

Convolutional LSTM

- Convolution operator in the state-to-state and input-to-state transitions
- Used for precipitation nowcasting

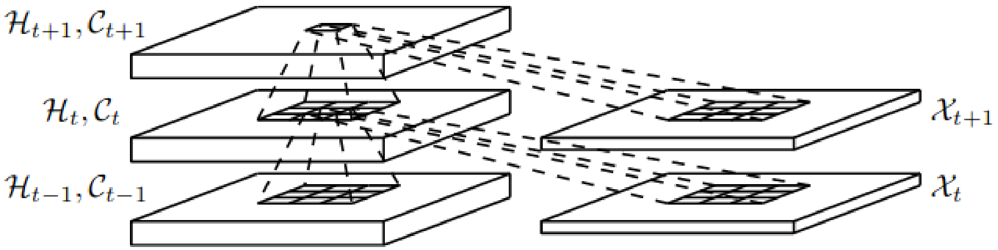


Figure 2: Inner structure of ConvLSTM

Journal Club

Jakob Schlör

Shi et al. (2015) Convolutional LSTM network: a machine learning approach for precipitation nowcasting

GAN video generation

- Autoregressive CDNN as a generative model
- Joined training of generative model and discriminative model

Algorithm 1: Training adversarial networks for next frame generation

Set the learning rates ρ_D and ρ_G, and weights $\lambda_{adv}, \lambda_{\ell_p}$.

while *not converged* **do**

Update the discriminator D**:**

Get M data samples $(X, Y) = (X^{(1)}, Y^{(1)}), \dots, (X^{(M)}, Y^{(M)})$

$$W_D = W_D - \rho_D \sum_{i=1}^{M} \frac{\partial \mathcal{L}_{adv}^{D}(X^{(i)}, Y^{(i)})}{\partial W_D}$$

Update the generator G**:**

Get M new data samples $(X, Y) = (X^{(1)}, Y^{(1)}), \dots, (X^{(M)}, Y^{(M)})$

$$W_G = W_G - \rho_G \sum_{i=1}^{M} \left(\lambda_{adv} \frac{\partial \mathcal{L}_{adv}^{G}(X^{(i)}, Y^{(i)})}{\partial W_G} + \lambda_{\ell_p} \frac{\partial \mathcal{L}_{\ell_p}(X^{(i)}, Y^{(i)})}{\partial W_G} \right)$$

Mathieu et al. (2015) Deep multi-scale video prediction beyond mean square error