Journal Club
May 4, 2021

Jakob Schlör
Universität Tübingen

May 5, 2021
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
Motivation

**Ground truth** for the validation of process-based models

**Physical Equation-driven Earth and Climate Modelling**
- Main tool for quantifying the Earth's state under ongoing anthropogenic forcing
- **Persistent error sources**, e.g., non-explicit description of subgrid-scale processes and insufficient model domain coupling
- Combination of deep process understanding with recently revealed advantages of machine learning

**Process-based models and neural networks will be coupled** as actively learning and self-validating hybrid models

**Earth System Observation Data**

**Available data pool for extracting specific training environments** for neural networks

**Earth Data-driven Machine Learning**
- Highly specialized agents that uncover hidden patterns and relate geophysical quantities
- General lack of process knowledge leads to fundamental shortcomings, e.g., for predicting non-stationary climate processes
- Training based on model and observation data allows neural networks to accurately predict Earth system processes

**Hybrid models will start to outperform traditional models** in terms of physical consistency and predictive power

Successive research on explainable and interpretable AI will make hybrid models more physically tractable

Combining the advantages of process-based with machine learning models will drastically improve Earth system and climate projections

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

\[
\frac{\partial I}{\partial t} + (\mathbf{\omega} \cdot \nabla) I = D \nabla^2 I
\]

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

\[ \frac{\partial I}{\partial t} + (\omega \cdot \nabla) I = D \nabla^2 I \]

Global Solution:

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

RBF - kernel

\[ k(x, x') = \frac{1}{4\pi Dt}e^{-\frac{1}{4Dt}|x-x'|^2} \]

I(x,t) : sea surface temperature
ω ~ \frac{Δx}{Δt} : motion field
D: diffusion coefficient
Motion estimation

\[ \hat{I}_{t+1}(x) = \sum_{x'} k(x - \hat{\omega}(x), x') I_{t}(x') \]

CDNN → Motion Field → Warping Scheme → \( \hat{I}_{t+1} \) → Supervision
Covolution Deconvolution NN (CDNN)

Properties:
- Skip connections
- Batch normalization
- Leaky ReLU

Loss:

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

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

Observation

Proposed model

Motion field

Numerical model

ACNN

ConvLSTM
Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Average score (MSE)</th>
<th>Average time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numerical model (Béréziat and Herlin 2015)</td>
<td>1.99</td>
<td>4.8</td>
</tr>
<tr>
<td>ConvLSTM (Shi et al 2015)</td>
<td>5.76</td>
<td>0.018</td>
</tr>
<tr>
<td>ACNN</td>
<td>15.84</td>
<td>0.54</td>
</tr>
<tr>
<td>GAN video generation (Mathieu et al 2015)</td>
<td>4.73</td>
<td>0.096</td>
</tr>
<tr>
<td>Proposed model with regularization</td>
<td>1.42</td>
<td>0.040</td>
</tr>
<tr>
<td>Proposed model without regularization</td>
<td>2.01</td>
<td>0.040</td>
</tr>
</tbody>
</table>

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

Forecasting ability seems 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
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

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 $\rho_D$ and $\rho_G$, and weights $\lambda_{adv}, \lambda_{lp}$.

while not converged do

Update the discriminator $D$:
Get $M$ data samples $(X, Y) = (X^{(1)}, Y^{(1)}), \ldots, (X^{(M)}, Y^{(M)})$

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

Update the generator $G$:
Get $M$ new data samples $(X, Y) = (X^{(1)}, Y^{(1)}), \ldots, (X^{(M)}, Y^{(M)})$

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

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