# Towards a Hybrid Approach to Physical Process Modeling

**Emmanuel de Bézenac [*], Arthur Pajot [*], Patrick Gallinari**
LIP6, UPMC
{name}@lip6.fr

## Abstract

We consider the use of Deep Learning methods for modeling complex phenomena like those occurring in natural physical processes. Using an example application, namely Sea Surface Temperature Prediction, we show how general background knowledge gained from the physics could be used as a guideline for designing efficient Deep Learning models. We demonstrate a formal link between the solution of a class of differential equations underlying a large family of physical phenomena and the proposed model. Experiments are then provided.

## 1 Introduction

The main scientific paradigm for modeling complex physical and natural processes consists in extracting knowledge from observations, formalizing this knowledge and validating the model experimentally. Conservation laws, physical principles or phenomenological behaviors are formalized e.g. using differential equations, while validation can take the form of forecasting the future states of the process. With the increased availability of large observation datasets, this physical modeling paradigm is being challenged by a statistical Machine Learning (ML) paradigm, which directly processes the data to infer prediction, bypassing the human formalization of process knowledge. However, despite impressive successes in a variety of domains as demonstrated by the deployment of Deep Learning methods in fields such as vision, language, speech, etc., the statistical approach is not ready to challenge the physical paradigm for modeling complex phenomena. We believe that both fields could benefit from closer interactions. Knowledge and techniques accumulated for modeling physical processes in fields such as scientific computing or physics are a source of prior information for designing efficient learning systems and conversely the ML paradigm could open new directions for modeling complex phenomena. This is the question we tackle: how could general knowledge gained from the physical modeling paradigm help in the design of efficient ML models? In the absence of a general approach to this issue, we choose to consider a specific physical modeling problem: forecasting sea surface temperature (SST). SST plays a significant role in analyzing and assessing the dynamics of weather and other biological systems. Weather satellites have made huge quantities of very high resolution SST data available [2]. Standard physical methods for forecasting SST use coupled ocean-atmosphere prediction systems, *e.g.* based on the Navier Stokes equations. Note that SST forecasting is used here as an illustrative and a representative example. We believe that the proposed general procedure could also be used for a more general class of transport problems.

We propose a Deep Neural Network (NN) model, inspired from general physical motivations which offers a new approach for solving a family of physical modeling problems. We first motivate our approach by introducing in section 2 the advection-diffusion equation, which is used in the modeling of a large range of transport and propagation phenomena in physics. Its solution is used as a guideline for introducing a Deep Learning architecture for SST prediction which is described in section 3. Experiments and comparison with a series of baselines are introduced in section 4.

[*]equal contribution

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

Our main contributions are: 1) an example showing how to incorporate general physical background for designing a NN aimed at modeling a relatively complex prediction task. We believe the approach to be general enough to be used for a family of transport problems obeying general advection-diffusion principles, 2) links between the mapping function implemented by the NN prediction model and the solution of a general advection diffusion PDE, 3) a proof of concept that data intensive approaches based on deep architectures can be competitive with state of the art dedicated numerical methods for modeling a family of physical processes when incorporating relevant physical prior knowledge.

## 2 Physical Motivation

Daily temperature acquisitions of the sea surface are captured via satellite imagery. Focusing on a specific area, forecasting SST can be seen as predicting future SST image frames given past ones. As in [1], classical approaches for forecasting SST introduce a numerical model representing prior knowledge on the conservation laws and physical principles taking place. Many fluid dynamic systems can be described using two equations: the *conservation equation* and the *momentum equation*. The conservation equation describes the *transport of some conserved quantity* $I$ (in our case, temperature), often relating the evolution of $I$ with a motion field $w$, while the momentum equation describes the evolution of the motion field $w$ (*e.g.* the Navier Stokes equation). By optimizing an energy functional enforcing the model's generated observations to be consistent with the acquired observations, the motion field is estimated, which is then used to forecast temperature.

Transport of a conserved quantity $I$ in fluid occurs through the combination of two principles: *advection* and *diffusion*. During advection, quantity $I$ is transported along $w$ via *bulk motion*. Diffusion corresponds to the movement which spreads out $I$ from areas of high concentration to areas of low concentration. The following conservation equation, known as the *advection-diffusion equation* describes this transport:

$$\frac{\partial I}{\partial t} + (w\,.\nabla)I = D\nabla^2 I \tag{1}$$

$\nabla$ and $\nabla^2$ denote the gradient and the Laplacian operator, respectively, while $D$ corresponds to the diffusion coefficient. This equation can be used to describe a large family of physical processes (e.g. fluid dynamics, heat conduction, wind dynamics, etc). Let us now state a result, characterizing the general solutions of equation 1.

**Theorem.** [2] *For any initial condition $I_0 \in L^1(\mathbb{R}^2)$ with $I_0(\pm\infty) = 0$, there exists a unique global solution $I(x,t)$ to the advection-diffusion equation 1, where pfs and $k(u,v) = \frac{1}{4\pi Dt}e^{-\frac{1}{4Dt}\|u-v\|^2}$ is a radial basis function kernel, or alternatively, a Gaussian probability density with mean $x - w$ and variance $2Dt$ in its second argument.*

Equation **??** provides a principled way to calculate $I$ for any time $t$ and position $x$, provided $I_0$, $w$, and $D$ are known. This result will be used in the following section to design a Deep Learning model for SST forecasting. This model will learn to predict a motion field analog to $w$ in equation **??**, which will be used in turn to forecast future images.

## 3 Model

The Deep Learning model consists of two main components, as illustrated in Figure 1. One predicts the motion field from a sequence of past input images using a convolutional-deconvolutional neural network (CDNN), and the other warps the last input image $I_t$ using the motion field estimate $\hat{w}$ from the first component, in order to produce an image forecast $\hat{I}_{t+1}$. The entire system is trained in an end-to-end fashion, using only the supervision from the target SST images $I_{t+1}$. By doing so, we are able to produce an interpretable latent state which corresponds in our problem to the velocity field advecting the temperatures.

**Motion field estimation.** As indicated in section 2, provided the underlying motion field is known, one can compute SST forecasts. Let us introduce how the motion field is estimated in our architecture.

[2] Starting from eq. 1 the proof follows from classical techniques from functional analysis

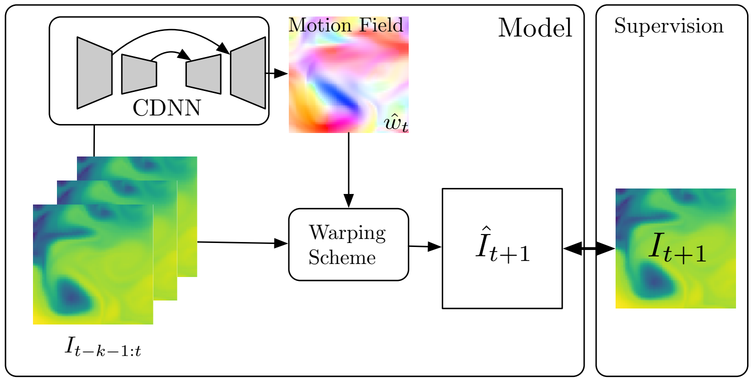


Figure 1: Motion field $\hat{w}$ is estimated with a convolutional-deconvolutional (CDNN) neural network taking as input a series of past images $I_{t-k-1}$ to $I_t$. A warping scheme then displaces the last input image $I_t$ along the motion estimate $\hat{w}$ to produce the future image forecast $\hat{I}_{t+1}$. The error signal is calculated using the target future image $I_{t+1}$, and is backpropagated through the warping scheme to correct the CDNN. To produce multiple time-step forecasts, the predicted image $\hat{I}_{t+1}$ is fed back in the CDNN in an autoregressive manner.

We are looking for a vector field $w$ which when applied to the geometric space $\Omega$ (the image manifold in $\mathbb{R}^2$) renders $I_t$ close to $I_{t+1}$, i.e. $I_{t+1}(x) \simeq I_t(x + w(x))$, $\forall x \in \Omega$. If $I_{t+1}$ were known, we could estimate $w$ via Optical Flow methods [4], but $I_{t+1}$ is precisely what we are looking for. Instead, we choose to use a CDNN architecture to predict a motion vector for each pixel. Since we usually do not have a direct supervision on the motion vector field, we will not be able to learn to predict motion by regressing to the target motion. Using the warping scheme introduced below, we will nonetheless be able to (weakly) supervise estimate $\hat{w}$, based on the discrepancy of the warped version $\hat{I}_{t+1}$ of image $I_t$ and the target image $I_{t+1}$.

**Warping scheme.** Discretizing the solution of the advection-diffusion equation in equation 2 and setting image $I_t$ as the initial condition, we obtain a method to calculate the future image, based on the motion field estimate $\hat{w}$. The latter is used in a warping scheme:

$$\hat{I}_{t+1}(x) = \sum_{y \in \Omega} k(\, x - \hat{w}(x),\, y)\; I_t(y) \tag{2}$$

where $k(x - \hat{w}, y) = \frac{1}{4\pi D \Delta t} e^{-\frac{1}{4D\Delta t}\|x - \hat{w} - y\|^2}$ is a radial basis function kernel, as in equation **??**, parameterized by the diffusion coefficient $D$ and the time step value $\Delta t$ between $t$ and $t+1$. Informally, to calculate the pixel value for time $t+1$ at position $x$, we first compute its previous position at time $t$, *i.e.* $x - \hat{w}$. We then center a Gaussian in that position in order to obtain a weight value for each pixel in $I_t$ based on its similarity with $x - \hat{w}$, and compute a weighted average of the pixel values of $I_t$. This weighted average will correspond to the new pixel value at $x$ in $\hat{I}_{t+1}$.

This warping mechanism has been inspired by the Spatial Transformer Network (STN) [5], originally designed to be incorporated as a layer in a convolutional neural network architecture in order to gain invariance under geometric transformations. Using the notations in [5], when the inverse geometric transformation $\mathcal{T}_\theta$ of the grid generator step is set to $\mathcal{T}_\theta(x) = x - \hat{w}(x)$, and the kernels $k(\,.\,;\Phi_x)$ and $k(\,.\,;\Phi_y)$ in the sampling step are rbf kernels, we recover our warping scheme. The latter can be seen as a specific case of the STN, without the localization step. This result theoretically grounds the use of the STN for Optical Flow in many recent articles [11] [10] [8] [3], as in equation 1 when $D \to 0$, we recover the brightness constancy constraint equation, ubiquitous in Optical Flow.

**Loss function.** Equation 1 is under constrained: for each equation we have 2 variables corresponding to the 2D motion field. Clearly, if $w$ is a solution of equation 1, for any $\eta$ for which $\eta \perp \nabla I$, $w + \eta$ is also solution. Additional physical prior knowledge on $w$ can be easily incorporated in our model, by adding penalty terms in the loss function. In our experiments, we have tested the influence of: divergence

$\nabla . \hat{w}_t(x)^2$, magnitude $\|\hat{w}_t(x)\|^2$ and smoothness $\|\nabla \hat{w}_t(x)\|^2$. We evaluate the influence of these terms in the experiments section. The complete loss function we used to train the model can then be written as:

$$L_t = \sum_{x\in\Omega} \left\|\hat{I}_{t+1}(x) - I_{t+1}(x)\right\|^2 + \lambda_{\text{div}}(\nabla . \hat{w}_t(x))^2 + \lambda_{\text{magn}} \|\hat{w}_t(x)\|^2 + \lambda_{\text{grad}} \|\nabla \hat{w}_t(x)\|^2 \quad (3)$$

# 4 Experiments

We used synthetic SST data of the Atlantic Ocean generated by NEMO (Nucleus for European Modeling of the Ocean)[3] [6], a state-of-the-art modeling framework of ocean-related engines. The resulting dataset consists of daily temperature acquisitions of 481 by 781 pixels, from 2006-12-28 to 2017-04-05 (3734 acquisitions). We extract 64 by 64 pixels sized subregions. We use data from years 2006 to 2015 for training and validation (94743 training $64 \times 64$ subregions examples), and years 2016 to 2017 for testing (1716 subregions). We withhold 20% of the training data for validation, selected uniformly at random at the beginning of each experiment. (1716 for test). We compare our model with several baselines. Each model is evaluated with a mean square error metric, forecasting images on a horizon of 6 (we forecast from $I_{t+1}$ to $I_{t+6}$ and then average the MSE). The hyperparameters are tuned using the validation set. Concerning the constraints on the vector field $w$ (equation 3). the regularization coefficients selected via validation are $\lambda_{\text{div}} = 1$, $\lambda_{\text{magn}} = -0.03$ and $\lambda_{\text{grad}} = 0.4$. We also compare the results with the model without any regularization. Our reference model for forecasting is [1], a state of the art numerical assimilation model for predicting ocean dynamics, here SST. The other baselines are 1) an autoregressive convolutional-deconvolutional NN (ACNN) with an architecture similar to our CDNN module, but trained to predict the future image directly. Each past observation is used as an input channel, and the output is used as new input for multi-step forecasting. 2) a ConvLSTM (Convolutional Long Short Term Memory) [9], which uses convolutional transitions in the inner LSTM module, and 3) the model in [7], a multi-scale ACNN trained as a Generative Adversial Network (GAN). Quantitatively, our model

| Model | Average Score (MSE) | Average Time |
|---|---|---|
| Numerical model [1] | 1.99 | 4.8 s |
| ConvLSTM [9] | 5.76 | 0.018 s |
| ACNN | 15.84 | 0.54 s |
| GAN Video Generation ([7]) | 4.73 | 0.096 s |
| Proposed model with regularization | **1.42** | 0.040 s |
| Proposed model without regularization | 2.01 | 0.040 s |

Table 1: Average score and average time on test data. Average score is calculated using the *mean square error* metric (MSE), time is in seconds.

performs well (table 1. The MSE score is better than any of the baselines. The closest NN baseline is [7] which regularizes a regression CDNN model with a GAN. The performance is however clearly below our proposed model and it does not allow to easily incorporate prior constraints inspired from the physics of the phenomenon. ACNN is a direct predictor of the image sequence, implemented via a CDNN module identical to the one used in our model. Its performance is limited. ConvLSTM performs better: as opposed to the ACNN, it seems to be able to capture a dynamic, although not very accurately. The state of the art numerical model [1], performs well but has a slightly lower performance than our regularized model, although it incorporates more prior constraints. This shows that pure ML models, when conceived adequately by incorporating general prior physical knowledge and when trained with enough data, can be competitive with state of the art dedicated models. We also provide in table 1 the average inference time of the models. All the experiments with NN have been performed og a Titan Xp GPU. The numerical model [1] has been run on a classical CPU (no GPU code) so that its elapsed time is not comparable.

[3]NEMO data are available at `http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&view=details&product_id=GLOBAL_ANALYSIS_FORECAST_PHY_001_024`

## 5 Conclusion

The data intensive paradigm offers alternative directions to the classical physical approaches for modeling complex natural processes. Cross fertilization of both paradigms is essential for pushing further the frontier of complex data modeling. By using as an example application a problem of intermediate complexity concerning ocean dynamics, we proposed a principled way to design Deep Learning models taking our inspiration from physics. This approach can be generalized to a class of problems for which the underlying dynamics follow advection-diffusion principles. It is able to reach performance comparable to a state of the art numerical model and clearly outperforms alternative NN models used as baselines.

**Acknowledgments**

This work was partially funded by ANR project LOCUST - ANR-15-CE23-0027 and by CLEAR Lab.

## References

[1] Dominique Béréziat and Isabelle Herlin. *Coupling Dynamic Equations and Satellite Images for Modelling Ocean Surface Circulation*, pages 191–205. Springer International Publishing, Cham, 2015.

[2] R. L. Bernstein. Sea surface temperature estimation using the noaa 6 satellite advanced very high resolution radiometer. *Journal of Geophysical Research: Oceans*, 87(C12):9455–9465, 1982.

[3] Chelsea Finn, Ian J. Goodfellow, and Sergey Levine. Unsupervised learning for physical interaction through video prediction. *CoRR*, abs/1605.07157, 2016.

[4] Denis Fortun, Patrick Bouthemy, and Charles Kervrann. Optical flow modeling and computation: a survey. *Computer Vision and Image Understanding*, 134:21, May 2015.

[5] Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu. Spatial transformer networks. *CoRR*, abs/1506.02025, 2015.

[6] G. Madec. *NEMO ocean engine*. Note du Pôle de modélisation, Institut Pierre-Simon Laplace (IPSL), France, No 27, ISSN No 1288-1619, 2008.

[7] Michaël Mathieu, Camille Couprie, and Yann LeCun. Deep multi-scale video prediction beyond mean square error. *CoRR*, abs/1511.05440, 2015.

[8] Viorica Patraucean, Ankur Handa, and Roberto Cipolla. Spatio-temporal video autoencoder with differentiable memory. *CoRR*, abs/1511.06309, 2015.

[9] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *CoRR*, abs/1506.04214, 2015.

[10] Jason J. Yu, Adam W. Harley, and Konstantinos G. Derpanis. *Back to Basics: Unsupervised Learning of Optical Flow via Brightness Constancy and Motion Smoothness*, pages 3–10. Springer International Publishing, Cham, 2016.

[11] Yi Zhu, Zhen-Zhong Lan, Shawn D. Newsam, and Alexander G. Hauptmann. Guided optical flow learning. *CoRR*, abs/1702.02295, 2017.