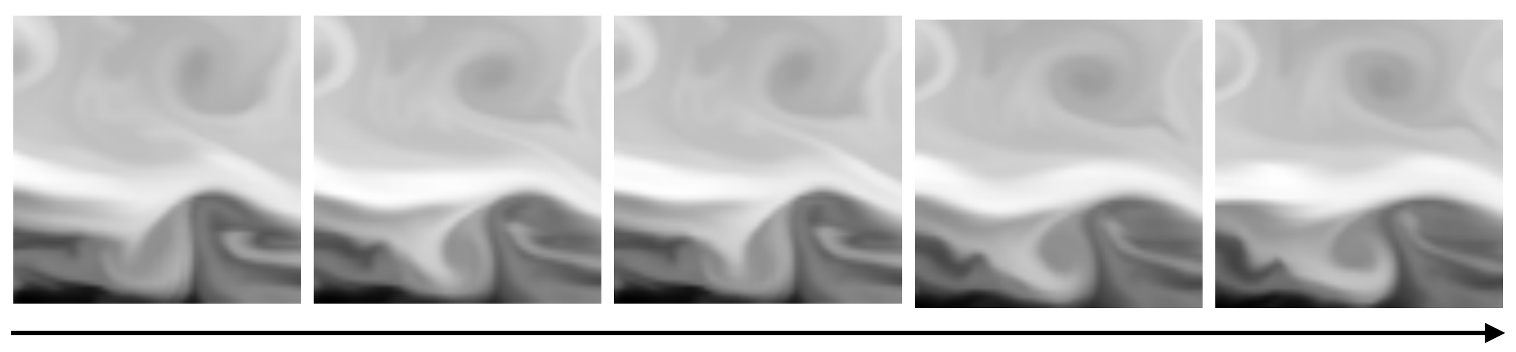
Model order reduction using a deep orthogonal decomposition

Daniel J. Tait, `dtait@turing.ac.uk`

WARWICK THE UNIVERSITY OF WARWICK · The Alan Turing Institute

Introduction

The complex, dynamical systems describing our natural world can be complete to forward simulate, never-mind parameterise and calibrate. At the the same time these processes exhibit rich, informative features which persist over time.



Temporally persistent informative features

This naturally suggests the use of the power feature extraction of modern machine learning. In this work we aim to combine this feature extraction process with the generalisability of a well specified physical model. We aim to perform model order reduction while pushing the dynamics to a latent space enabling a tight link between DL-based encoding and the dynamics

Ritz-Galerkin projection

We utilise the Ritz-Galerkin projection, a classical method of solving a PDE like the convection-diffusion PDE

$$\frac{\partial u}{\partial t} = -\nabla \cdot (\alpha \nabla u) + \mathbf{w} \cdot \nabla u$$

by projection onto a finite dimensional subspace spanned by a set of orthogonal function. This gives a new dynamical system

$$\frac{d}{dt}\mathbf{z} = \mathbf{L}_{\varphi}, \qquad (\mathbf{L}_{\varphi})_{ij} = \int_{\Omega} \alpha \nabla \varphi_j \cdot \nabla \varphi_i dx + \int \mathbf{w} \cdot \nabla \varphi_j \varphi_i dx$$

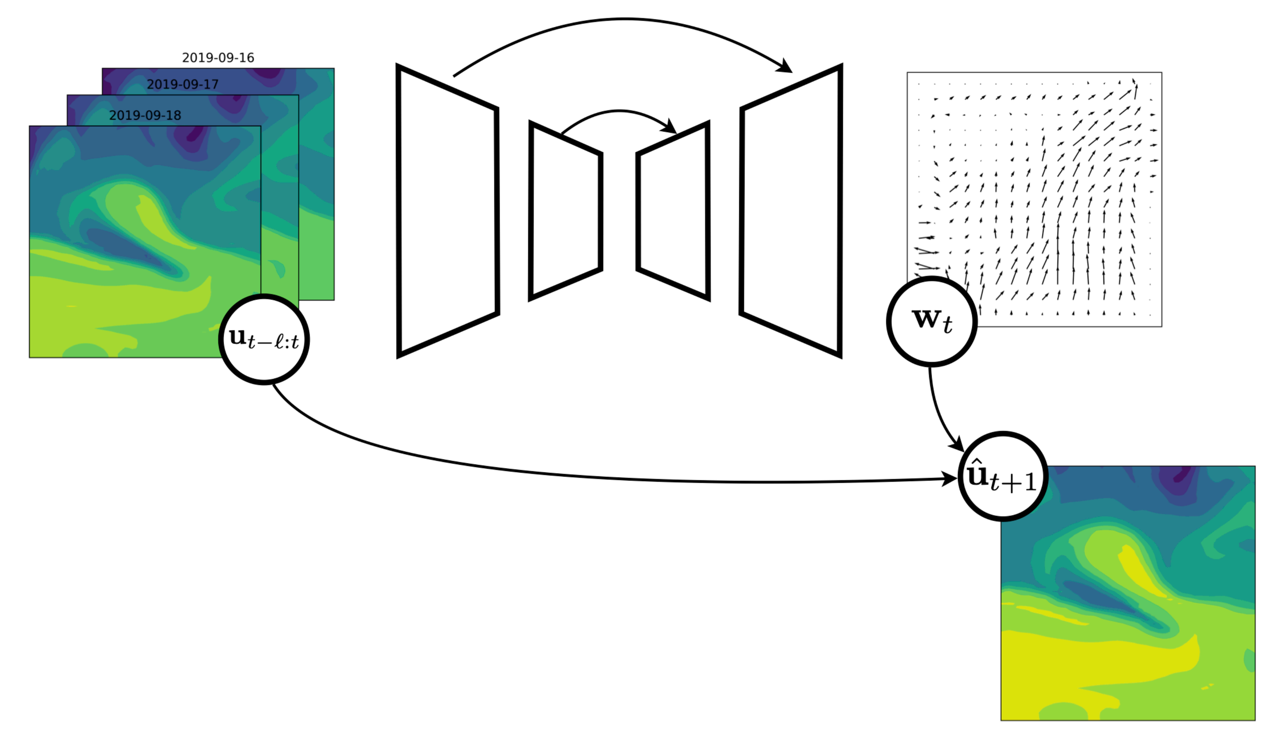
in a finite dimensional state vector z. The resulting operator will depend on the basis functions used to create the finite dimensional subspace. The *Proper orthogonal decomposition* takes the basis function as the principle components of a PCA analysis applied to snapshots of the system.

Our approach will use a deep network, referred to as the `LocalGalerkinProjection`, to parameterise these functions. Rather than the global information extraction of POD we shall use only the most recent inputs, relying on DL to extract sufficient information from this reduced set — similar to a localised nonlinear kernel PCA [Lawrence, 2005].

The Multitask Model

Bezenac et al., 2018

Previous work by Bezenac et al. has considered the use of the convection-diffusion PDE to predict the solution. A deep encoder-decoder network is used to parameterise the vector field of the transport equation

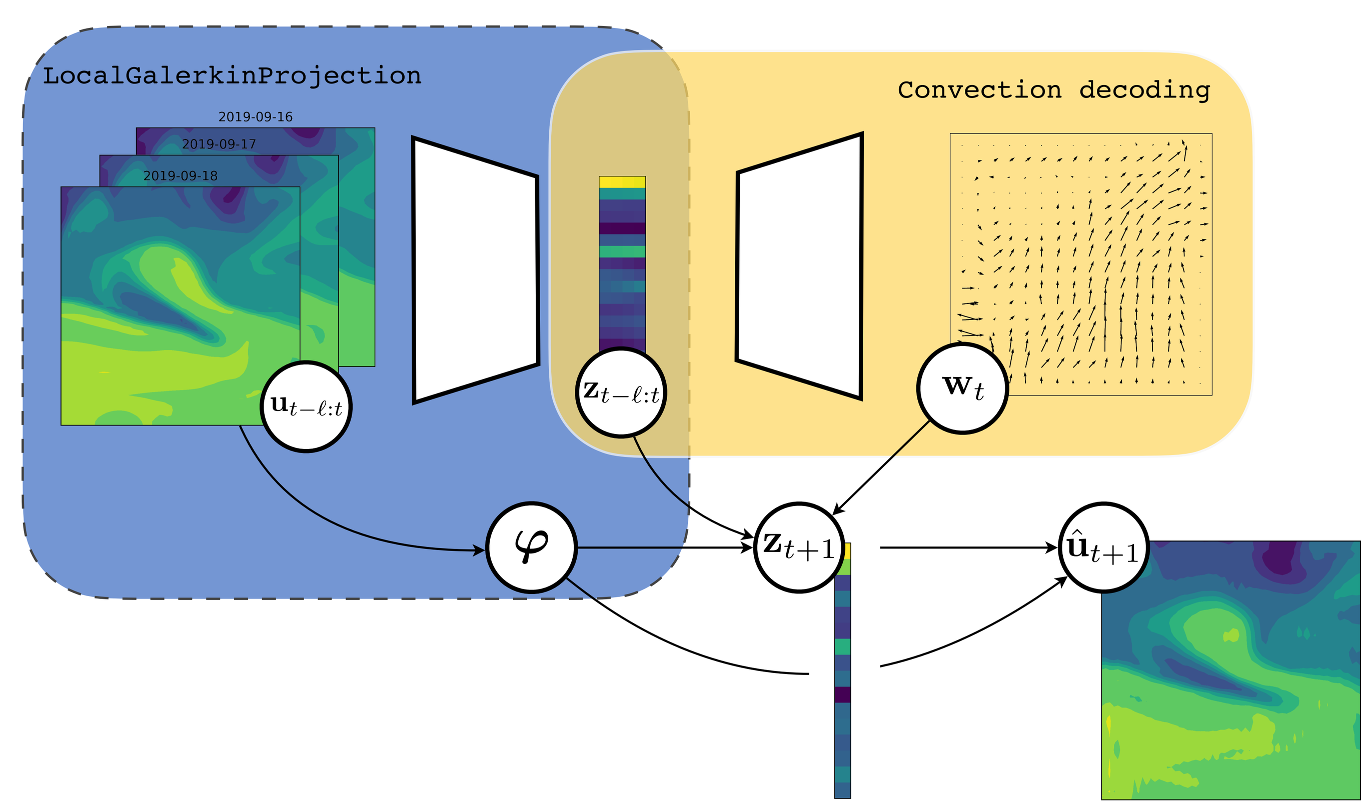


The method of [Bezenac et al., 2018] uses DL to estimate a motion field, and then transports the solution in the original data-space. No further use is made of the latent representations.

This is a powerful approach to parameterising the transport vector field, but ultimately still linear dynamics in the original space.

Our approach

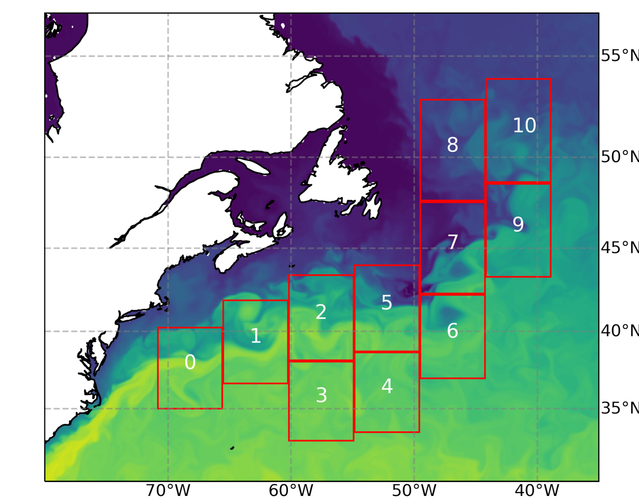
We use the Galerkin projection to first transform to the latent space. Then decode the projected variable advance the linear model in the latent space and then finally decode. The additional encoding step is then able to simultaneously learn a projection which makes the linear approximation more appropriate.

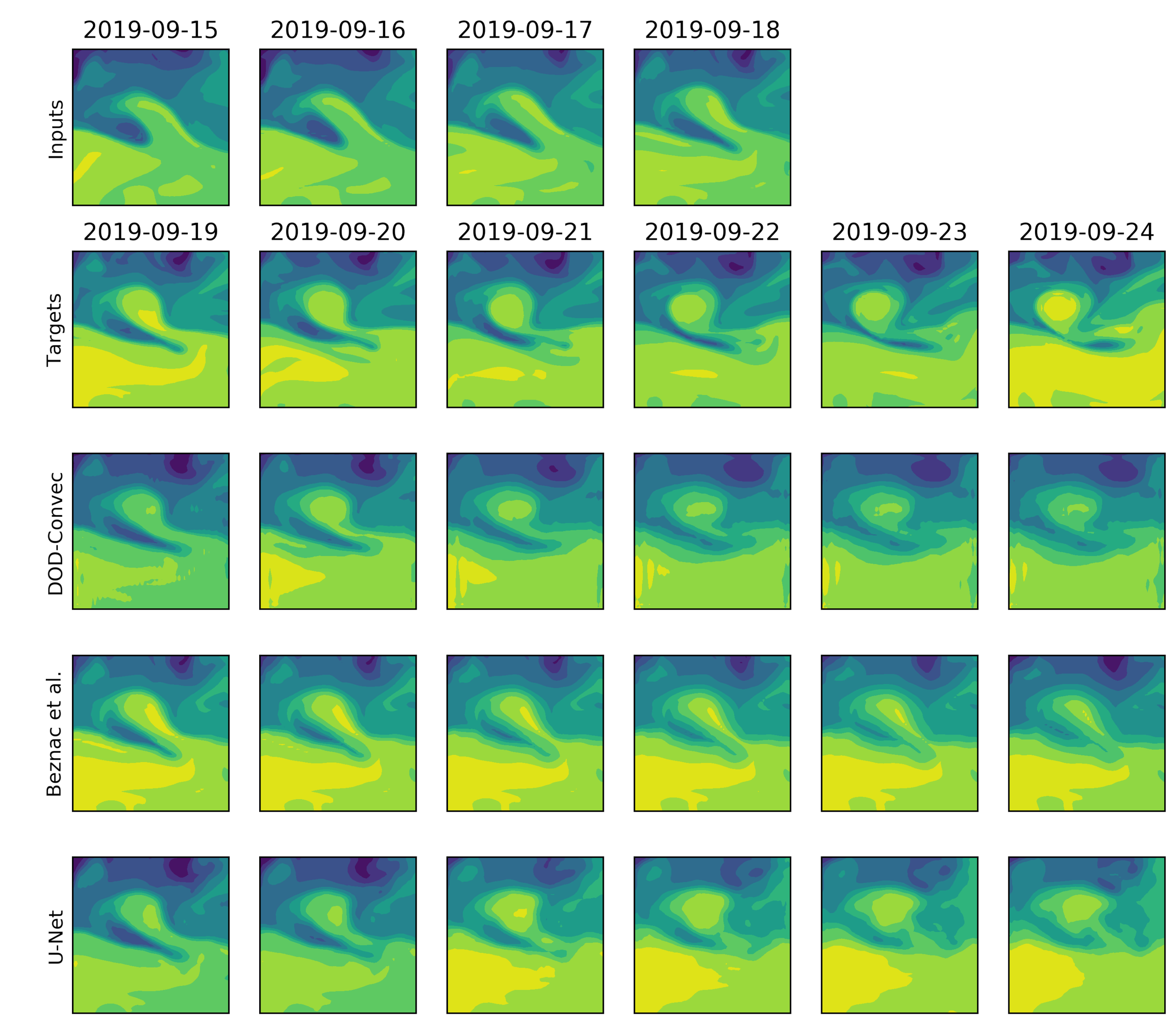


Our method uses the `LocalGalerkinProjection` and then advances the solution in the latent space using information from the encoded variables to parameterise the convection model, before reconstructing in the original space

Experiments

Sea surface temperature from the Nucleus for European Modelling of the Ocean (NEMO) engine. Which uses assimilation and so provides accurate records of historical sea surface temperature data. Train on 4 input images and predict the next 6





Predicted surfaces from an input of length four, top row, compared to a target output sequence of length six, second row. We compare the physically informed models (ours and the model of [Bezenac et al., 2018] and ML sequence-to-sequence approaches, the U-Net model.

	Average Score (MSE)	No. of Parameters	Run-Time [s]
Conv-LSTM	0.2073	1,779,073	0.43
U-Net	0.1473	31,032,446	0.79
Flow [Bezenac et al.]	0.1304	22,197,736	0.60
DOD-Convec [Our method]	**0.1132**	10,106,339	0.48

Discussion

An important feature of our method is the use of deep learning and Galerkin projection to jointly push the observations and dynamics to a latent space. Future work will investigate how to maintain fidelity of these simplified representative models to more complex descriptions of the problem which are typically given in the original data space.

References

Lawrence (2005). “Probabilistic Non-linear Principal Component Analysis with Gaussian Process Latent Variable Models”. In: JMLR 6, pp. 1783 –1816

Bezenac et al (2018). “Deep learning for physical processes. Incorporating prior scientific knowledge.” In: ICLR