

RESEARCH ARTICLE  

# Multiphase flow applications of nonintrusive reduced-order models with Gaussian process emulation

Themistoklis Botsas<sup>1</sup> , Indranil Pan<sup>1,2,3,\*</sup> , Lachlan R. Mason<sup>1,2</sup> and Omar K. Matar<sup>1,2</sup>

<sup>1</sup>Data Centric Engineering, The Alan Turing Institute, 96 Euston Rd, London NW1 2DB, United Kingdom

<sup>2</sup>Department of Chemical Engineering, Imperial College London, Exhibition Rd, South Kensington, London SW7 2BX, United Kingdom

<sup>3</sup>School of Mathematics, Statistics & Physics, Newcastle University, Newcastle upon Tyne NE1 7RU, United Kingdom

\*Corresponding author. E-mail: [i.pan11@imperial.ac.uk](mailto:i.pan11@imperial.ac.uk)

**Received:** 13 November 2021; **Revised:** 23 February 2022; **Accepted:** 26 March 2022

**Keywords:** Autoencoders; deep Gaussian process; Gaussian process; reduced-order models

## Abstract

Reduced-order models (ROMs) are computationally inexpensive simplifications of high-fidelity complex ones. Such models can be found in computational fluid dynamics where they can be used to predict the characteristics of multiphase flows. In previous work, we presented a ROM analysis framework that coupled compression techniques, such as autoencoders, with Gaussian process regression in the latent space. This pairing has significant advantages over the standard encoding–decoding routine, such as the ability to interpolate or extrapolate in the initial conditions’ space, which can provide predictions even when simulation data are not available. In this work, we focus on this major advantage and show its effectiveness by performing the pipeline on three multiphase flow applications. We also extend the methodology by using deep Gaussian processes as the interpolation algorithm and compare the performance of our two variations, as well as another variation from the literature that uses long short-term memory networks, for the interpolation.

## Impact Statement

Reduced-order models are popular in various engineering fields since they replicate the behavior of their complex counterparts using minimal computational resources. By combining machine learning (ML) algorithms we can not only construct these models but also extend them in such a way that they incorporate knowledge from physical parameters, among other advantages. One advantage is that we can use these hybrid models to provide predictions from physical parameters even where data are not available, bypassing the standard expensive procedure of running new (physical and/or numerical) experiments. In the present study, we use one such combination in order to illustrate how this framework can be used in this manner and compare it with variations of other ML algorithms.



This research article was awarded an Open Data and Open Materials badge for transparent practices. See the Data Availability Statement for details.

## 1. Introduction

Reduced-order models (ROMs) are widely applicable to various fields of science and engineering involving partial differential equations (PDEs), since they can speed up analyses and reduce computational requirements. ROMs are of particular interest to computational fluid dynamics (CFD), where they can be used to predict the characteristics of multiphase flows. For this work, we construct the main framework with CFD-related challenges in mind, such as computational complexity and need for physical parameter estimation. Machine learning and deep learning techniques are amongst the most popular choices employed in order to solve dimensionality reduction-related problems, including ROMs. Thus, algorithms such as autoencoders have been used in this manner (Kim et al., 2019) and are now considered an established and attractive choice. In our previous work (Maulik et al., 2021a), we introduced a hybrid version of ROMs where we coupled three dimensionality reduction techniques, namely principal component analysis (PCA), convolutional autoencoders (CAE), and variational convolutional autoencoders (VAEs) with interpolation in the latent space using Gaussian process (GP) regression and focused on the various advantages of our methodology; these include uncertainty quantification (due to the deployment of GPs), the derivation of a finer temporal resolution, and enhanced interpretability.

In the present study, we focus on another major, and unexplored until now, advantage of the methodology developed in our previous paper (Maulik et al., 2021a), which involves the ability to interpolate in parameter space. In practice, the nature of the interpolation, along with its attachment to physical parameters, gives us the ability to predict flow regimes even when data are not available. This is particularly significant for multiple engineering domains where small changes in parameter values can lead to significant differences in the temporal progression of the system, and running a simulator for all the required parameter values can be computationally intractable.

Similar methods that substitute GPs for long–short term memory recurrent neural networks (LSTMs) have recently emerged in Gonzalez and Balajewicz (2018), Mohan and Gaitonde (2018), Hu et al. (2019), and Maulik et al. (2021b). We focus on the latter, which follows same principles as our methodology, and compare the performance of the GP- and LSTM-based interpolation techniques on the same data-sets and comment on the reasons underlying the differences observed. Finally, we extend the GP-based methodology by replacing the GPs with deep Gaussian processes (DGPs), an ML method that has gained traction in recent years and can be perceived as an extension of standard GPs in the same manner that a neural network is an extension of the generalized linear model.

To summarize, the main novelties of this article are:

- examination of the performance of parameter interpolation in the latent space; something that has already been done for the LSTM variations (Maulik et al., 2021b), but not for the GP variations of the ROM-with-interpolation models;
- extension of the methodology by the introduction and incorporation of DGPs;
- comparison of the GP-based methods with each other and with the method that uses LSTMs for interpolation;
- application of all methodologies to three new multiphase flow data-sets of different nature and complexity.

The remainder of this paper is organized as follows. In Section 2, we present the general methodology and we briefly introduce all of the algorithms involved. In Section 3, we demonstrate and compare the different variations of the methodology applied to three multiphase flow data-sets with increasing complexity. Finally, in Section 4, we summarize the main takeaways from our work and discuss the focus of our future research.

## 2. Methodology

The main pipeline that we will use for the remainder of this article is similar to the one introduced in our previous paper (Maulik et al., 2021a). It combines two ML algorithms: a compression algorithm that takes a simulation in the form of time-related snapshots as input and outputs a latent space, and an interpolation algorithm, that is used as a regression model upon this space.

### 2.1. Compression algorithms

We focus on three different compression algorithms: PCA, CAE, and VAEs. In the context of CFD, a simulation involves the numerical solution of a set of differential equations that usually requires considerable computational resources (particularly if the simulations are spatio-temporal and solved in 3D space). A set of simulations is fed into one of the aforementioned ML algorithms in the form of images (frames that correspond to simulation timestamps). The information from the simulations is compressed and summarized in the form of the latent space, where it can be further manipulated and analyzed. Subsequently, the decompression portion of the algorithm can be used to reconstruct the original space. The whole process is shown in Figure 1.

#### 2.1.1. Principal component analysis

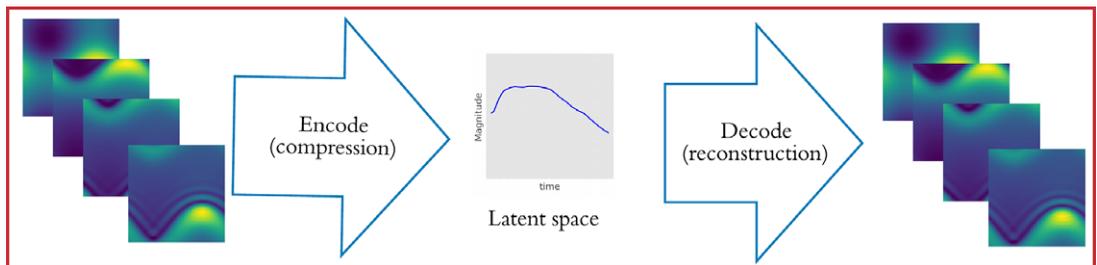
The PCA is a numerical method used to decompose a random vector field  $\mathbf{u}(\mathbf{d}, t)$  (in which  $\mathbf{d}$  and  $t$  denote space, which in turn can be represented by an appropriate coordinate system, and time, respectively). Following the decomposition step, a new basis is created where the new variables are linear combinations of the originals such that the explainable system variance is maximized. The  $n_r$ -dimensional latent space is created by selecting the  $n_r$  first components and discarding the rest.

To carry out a PCA, we first take  $p$  temporal snapshots of the field  $\mathbf{s}$ , which for  $n$  spatial elements yields the matrix  $\mathbf{S}$ :

$$\mathbf{S} = \begin{bmatrix} u(d_1, t_1) & \cdots & u(d_n, t_1) \\ \cdots & \cdots & \cdots \\ u(d_1, t_p) & \cdots & u(d_n, t_p) \end{bmatrix}.$$

Then, we compute the covariance matrix  $\mathbf{C}$  of  $\mathbf{S}$  as:

$$\mathbf{C} = \frac{1}{p-1} \mathbf{U}^T \mathbf{U},$$



**Figure 1.** Schematic of reduced-order modeling. Spatio-temporal outputs of a simulation (left) are being fed into an encoder and output a latent space (middle). The reconstruction of the original system (right) is the output of the decoder that uses as input the aforementioned latent space. Lower panel shows the interpolation and forecasting in the latent space for reconstruction.

where  $\mathbf{U}$  is an orthogonal matrix and consequently the eigenvalue diagonal matrix  $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_n\}$ , where  $\lambda_1 \leq \dots \leq \lambda_n$  and the corresponding eigenvector matrix  $\mathbf{W}$  are derived from:

$$\mathbf{C}\mathbf{W} = \mathbf{W}\Lambda.$$

Finally, the PCA basis is given by:

$$\theta = \mathbf{U}\mathbf{W},$$

and choosing the first  $n_r$  columns of  $\theta$  results in the creation of an  $n_r$ -dimensional latent space. The main advantage of the PCA is its simplicity and ease of computation, but the quality of the results is not necessarily equivalent to that of more sophisticated methods.

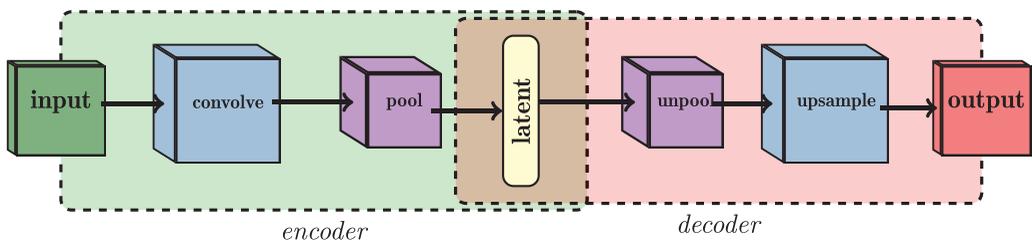
### 2.1.2. Convolutional autoencoders

Autoencoders are classes of neural network algorithms used primarily for unsupervised learning purposes, such as dimensionality reduction, data generation, and feature learning. Their structure is based on a bottleneck that combines two individual components: the *encoder*, which is a neural network that passes the input through layers that consist of a decreasing number of neurons, up until the bottleneck, where it outputs the *latent space*; and the *decoder*, which has the opposite architecture, uses the latent space as input and aims to reconstruct the original data by minimizing the reconstruction error. For this work, we use CAE, which is a class of autoencoders that includes layers with convolutions (LeCun et al., 1995), that is, a set of filters that extract specific features from images. The output  $y_{ijk}$  of a typical neuron in a convolutional layer has the form:

$$y_{ijk} = \varphi(\mathbf{f}_i \times \mathbf{p}_{jk} + b_i),$$

where  $\varphi$  is an activation function,  $\mathbf{f}_i$  is a single filter,  $\mathbf{p}_{jk}$  is a patch of data that shifts according to the dimensions  $j$  and  $k$ , and  $b_i$  is a bias term. In practice, the convolutional layers learn different features and patterns from the original data, particularly useful in image processing.

Another type of layer found in a convolutional neural network (CNN) is a pooling layer, which generally follows one (or more than one) convolutional layer with the aim of subsampling and summarizing the information from the filters. In a typical CNN, the continuous alternation of convolutional and pooling layers is how a neural network can extract high- and low-level features. In CAEs, the decoder comprises the opposite structure to that of the encoder. Instead of the convolutional and pooling layers, it consists of upsampling and unpooling layers respectively, where it produces an output of the same size as the input data using nearest-neighbor interpolation. A loss function such as the mean squared error (MSE) is then used during training in order to update the neural network's weights in a manner that minimizes the reconstruction error between the input and the output. The general structure of the CAE with all the components described above is shown in Figure 2.



**Figure 2.** Architecture of a convolutional autoencoder (CAE). The input frames are fed into the encoder, where convolutional (convolve) and pooling (pool) layers are used, in order to produce the latent space. The reverse scheme is used for the decoder, where unpooling (unpool) and upsampling (upsample) layers are used to produce an output as similar as possible to the original input.

### 2.1.3. Variational convolutional autoencoders

A similar class of algorithms is VAEs (Kingma and Welling, 2013). The main difference between these and their conventional convolutional counterparts is that the latent space is provided in the form of a probability distribution, usually a Gaussian. In practice this representation acts as *regularization* for the latent space, which can be particularly beneficial when quality of the data is poor. In order to reconstruct the input, a sample is drawn from the latent space distribution.

The encoder and decoder of a VAE can be described as the functions  $q(\mathbf{z}|\mathbf{x})$  and  $p(\mathbf{x}'|\mathbf{z})$  respectively, where  $\mathbf{x}$  is the input,  $\mathbf{z}$  is the latent space and  $\mathbf{x}'$  is the output. The latent space follows a Gaussian distribution  $\mathbf{z} \sim \mathcal{N}(\mu, \sigma^2)$  due to the Kullback–Leibler divergence (KL divergence)  $D_{\text{KL}}(q(\mathbf{z}|\mathbf{x}_i)||p(\mathbf{z}|\mathbf{x}_i))$ . The VAE loss has two components: the reconstruction loss (similar to the CAE) and the KL loss,

$$E_{q(\mathbf{z}|\mathbf{x}_i)}[\log p(\mathbf{x}_i|\mathbf{z})] - D_{\text{KL}}(q(\mathbf{z}|\mathbf{x}_i)||p(\mathbf{z}|\mathbf{x}_i)),$$

which ensure that the output is as similar as possible to the input and that the distribution of  $z$  is Gaussian, respectively. The practical difference between a CAE and a VAE is that, as shown in Figure 3, the output of the VAE encoder consists of two vectors  $\mu$  and  $\sigma$ . Given  $\varepsilon \sim N(0, I)$ , where  $I$  is the identity matrix we can use the *reparameterization trick* (Kingma and Welling, 2013) and write the latent space as  $z = \mu + \sigma\varepsilon$ , which is a form suitable for training. For the implementation of the autoencoders we use the TensorFlow package (Abadi et al., 2016).

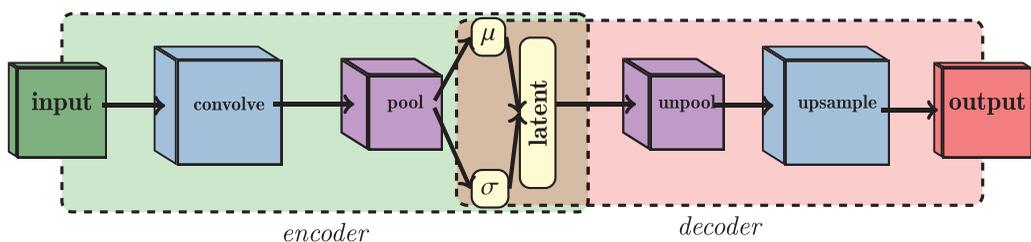
## 2.2. Interpolation algorithms

The enhanced ROM methodology that includes the interpolation step is shown in Figure 4. After we derive the latent space (in the same manner as in Section 2.1), we use it as input for an interpolation algorithm. The new model can be used for predictions with new sets of parameters, while the output can be further assessed and transformed back to the original space through the decoder.

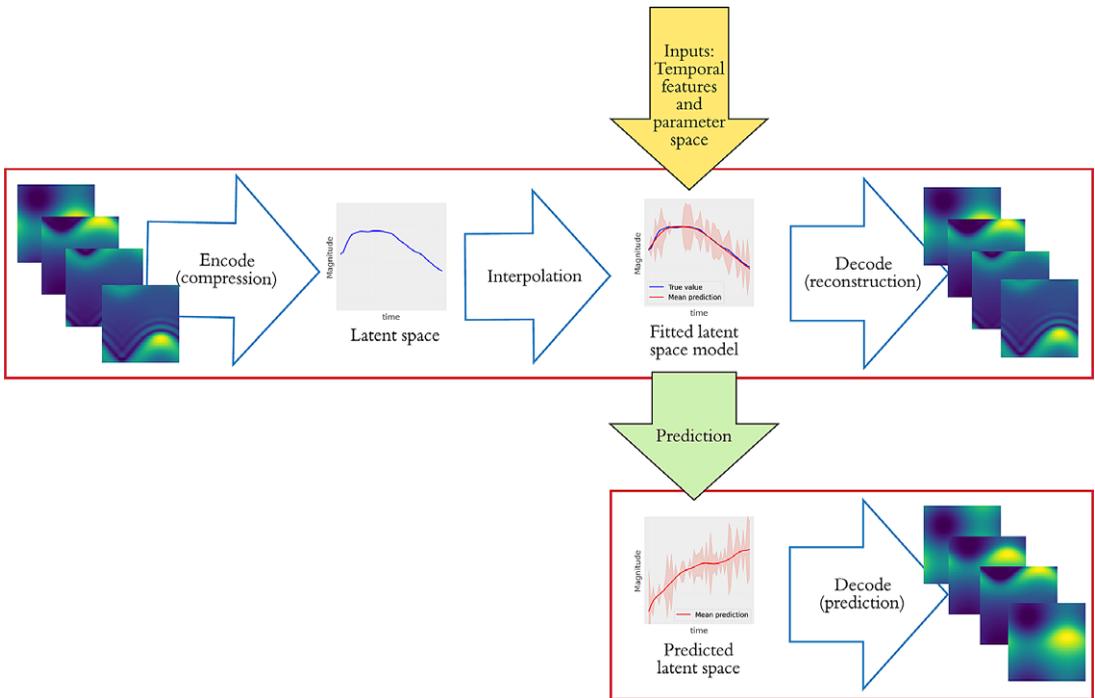
There are various advantages underlying latent space interpolation. Specifically, the one that we primarily focus on in this work is the ability to interpolate and extrapolate in the parameters of interest. This is an important advantage for multiphase flow applications since new simulations can be computationally costly, and, therefore, the ability to relocate this problem into the low-dimensional latent space, where predictions are easily performed, can allow for predictions, based on sets of parameters, even when data are not available. Other advantages that were explored by Maulik et al. (2021a) include interpolation in time, which can lead to finer temporal resolutions, increased interpretability since the interpolation algorithm provides valuable visualizations of the latent space that can show the quality of the compression; and, finally, in the case where the interpolation algorithm is a GP, uncertainty quantification is also possible.

### 2.2.1. Gaussian processes

GPs (Rasmussen and Williams, 2006) are generalizations of multivariate Gaussian distributions with infinite-dimensional space and a popular choice for regression (Williams and Rasmussen, 1996) due to



**Figure 3.** Architecture of a VAE. It is similar to the CAE equivalent, with the additional assumption that the latent space is a set of multivariate Gaussian distributions, and can be described as  $z \sim N(\mu, \sigma^2)$ .



**Figure 4.** Schematic of the enhanced reduced-order modeling which features additional steps to Figure 1 involving interpolation of the latent space and (if required) prediction for new parameters. The outputs are fed into the decoder for transformation back to the original space.

their versatility; the method is known as Gaussian process regression (GPR). We mainly focus on the mean prediction of a GP that corresponds to the maximum a posteriori estimate and use this as the decoder input.

Considering a mean function  $m(\mathbf{x})$  equal to 0, a GP can be completely specified by its second-order statistics; therefore, a positive definite covariance function (otherwise known as a kernel)  $k(\mathbf{x}, \mathbf{x}')$  is the only requirement. For a GPR model, we considered a GP  $f$  and noisy training observations  $\mathbf{y}$  of  $n$  datapoints  $\mathbf{x}$  derived from the true values  $f(\mathbf{x})$  with additive i.i.d. Gaussian noise  $\varepsilon$  with variance  $\sigma_n^2$ :

$$\begin{aligned} \mathbf{y} &= f(\mathbf{x}) + \varepsilon, \\ \varepsilon &\sim \mathcal{N}(0, \sigma_n^2), \\ f(\mathbf{x}) &\sim \text{GP}(0, k(\mathbf{x}, \mathbf{x}')), \end{aligned} \tag{1}$$

where  $k(\cdot, \cdot)$  is the kernel. We obtain the complete GP specification by maximizing the *marginal likelihood*, which we can acquire by integrating the product of the Gaussian likelihood and the GP prior over  $f$ :

$$p(\mathbf{y}|\mathbf{x}) = \int_f p(\mathbf{y}|f, \mathbf{x}) p(f|\mathbf{x}) df. \tag{2}$$

For testing input  $\mathbf{x}_*$  and output  $\mathbf{f}_*$ , we derive the joint marginal likelihood:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} k(\mathbf{x}, \mathbf{x}) + \sigma_n^2 \mathbf{I} & k(\mathbf{x}, \mathbf{x}_*) \\ k(\mathbf{x}_*, \mathbf{x}) & k(\mathbf{x}_*, \mathbf{x}_*) \end{bmatrix} \right),$$

where  $\mathbf{I}$  is the identity matrix. Finally, by conditioning the joint distribution on the training data and the testing inputs, we derive the predictive distribution

$$\mathbf{f}_* | \mathbf{x}, \mathbf{x}_*, \mathbf{y} \sim \mathcal{N}(\bar{\mathbf{f}}_*, \text{cov}(\mathbf{f}_*)), \tag{3}$$

where  $\bar{\mathbf{f}}_*$  and  $\text{cov}(\mathbf{f}_*)$  are given by

$$\begin{aligned} \bar{\mathbf{f}}_* &= k(\mathbf{x}_*, \mathbf{x}) [k(\mathbf{x}, \mathbf{x}) + \sigma_n^2 \mathbf{I}]^{-1} \mathbf{y} \\ \text{cov}(\mathbf{f}_*) &= k(\mathbf{x}_*, \mathbf{x}_*) - k(\mathbf{x}_*, \mathbf{x}) [k(\mathbf{x}, \mathbf{x}) + \sigma_n^2 \mathbf{I}]^{-1} k(\mathbf{x}, \mathbf{x}_*). \end{aligned} \tag{4}$$

We chose a single Matérn 3/2 kernel with lengthscale  $\mathbf{I}$  due to its versatility, flexibility, and smoothness. Specifically, we used the automatic relevance determination (ARD) extension Bishop (2006), which incorporates a separate parameter for each input variable:

$$k(\mathbf{x}, \mathbf{x}') = \left( 1 + \frac{\sqrt{3(\mathbf{x} - \mathbf{x}')^2}}{\mathbf{I}} \right) \exp \left( -\frac{\sqrt{3(\mathbf{x} - \mathbf{x}')^2}}{\mathbf{I}} \right). \tag{5}$$

Substitution of Equation (5) into Equation (4) yields:

$$\begin{aligned} \bar{\mathbf{f}}_* &= \left( 1 + \frac{\sqrt{3(\mathbf{x}_* - \mathbf{x})^2}}{\mathbf{I}} \right) \exp \left( -\frac{\sqrt{3(\mathbf{x}_* - \mathbf{x})^2}}{\mathbf{I}} \right) [(1 + \sigma_n^2) \mathbf{I}]^{-1} \mathbf{y}, \\ \text{cov}(\mathbf{f}_*) &= 1 - \left( 1 + \frac{\sqrt{3(\mathbf{x}_* - \mathbf{x})^2}}{\mathbf{I}} \right) \exp \left( -\frac{\sqrt{3(\mathbf{x}_* - \mathbf{x})^2}}{\mathbf{I}} \right) [(1 + \sigma_n^2) \mathbf{I}]^{-1} \\ &\quad \left( 1 + \frac{\sqrt{3(\mathbf{x} - \mathbf{x}_*)^2}}{\mathbf{I}} \right) \exp \left( -\frac{\sqrt{3(\mathbf{x} - \mathbf{x}_*)^2}}{\mathbf{I}} \right). \end{aligned} \tag{6}$$

During the reconstruction phase, we focus on the predictions that correspond to  $\bar{\mathbf{f}}_*$ .

### 2.2.2. Deep Gaussian processes

A DGP (Damianou and Lawrence, 2013) is a hierarchical composition of conventional GPs. In a DGP model, the data is modeled as the output of a multivariate GP, the inputs to that GP are governed by another GP and so on. In practice, DGPs are multilayer generalizations of GPs. For the purposes of this work, we will use the doubly stochastic variational inference variant of the DGP (Salimbeni and Deisenroth, 2017), according to which for  $L$  layers we derive the joint density:

$$p(\mathbf{y}, \{F^l, u^l\}_{l=1}^L) = \prod_{i=1}^N p(y_i | f_i^L) \prod_{l=1}^L p(F^l | u^l; F^{l-1}, z^{l-1}) p(u^l; z^{l-1}), \tag{7}$$

where  $F^0 = x$ . In Equation (7),  $x$  and  $y$  are the  $n$ -dimensional data (inputs and outputs respectively),  $F^l$  are stochastic functions with GPs as priors,  $f_i^L$  is the output of the last layer,  $z^i$  is a set of inducing points at layer  $i$ , and  $u^i$  the corresponding inducing function values. Note that in this parameterization each of the GPs has a zero mean and the Gaussian noise is absorbed into the kernel. By assuming that the posterior  $q$  of  $u^i$  is factorized between layers and  $q(u^i) \sim \mathcal{N}(m^i, S^i)$ , and after marginalizing the inducing variables of each layer, the marginal likelihood becomes:

$$q(\{F^l\}_{l=1}^L) = \prod_{i=1}^L q(F^l | m^l, S^l; F^{l-1}, Z^{l-1}) = \prod_{i=1}^L N(F^l | \bar{\mu}^l, \bar{\Sigma}^l). \tag{8}$$

For new predictions the following equation applies:

$$q(f_*^L) = \frac{1}{V} \sum_{v=1}^V q\left(f_*^L | m^L, S^L; f_*^{(v)(L-1)}, Z^{L-1}\right),$$

where  $f_*^{(v)(L-1)}$  are  $V$  samples from Equation (8). For the implementation of GPs and DGPs we used the GPyTorch library (Gardner et al., 2018).

### 2.2.3. Long short-term memory networks

The long short-term memory networks (LSTMs) (Hochreiter and Schmidhuber, 1997) are a special case of recurrent neural networks (RNNs), a class of neural networks that account for sequential data, thus being particularly useful for problems with temporal components. LSTMs, specifically, use gated cells that allow information transfer from both the recent past (short-term memory) and the distant past (long-term memory). This is an advantage of LSTMs compared to other RNN variations in terms of the results quality and also presents a solution to practical problems such as vanishing gradients (very small gradients during backpropagation that can render neurons inactive). A typical LSTM cell consists of:

$$\begin{aligned} \text{forget gate : } f_t &= \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \\ \text{input gate : } i_t &= \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \\ \text{output gate : } o_t &= \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \\ \text{cell input : } \bar{c}_t &= \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \\ \text{cell state : } c_t &= f_t \odot c_{t-1} + i_t \odot \bar{c}_t \\ \text{hidden state : } h_t &= o_t \odot c_t, \end{aligned}$$

where  $t$  denotes the time,  $x_t$  is the input,  $W_*$ ,  $U_*$ , and  $b_*$  are the weights of the input and recurrent connection matrices and bias terms of the quantity \* respectively,  $\odot$  is the element-wise product,  $\sigma_g$  is the sigmoid function and  $\sigma_c$  is the hyperbolic tangent function.

### 2.3. Methodology summary

The main steps of the methodology are:

1. Create realistic synthetic data-sets for different parameter values and time-steps.
2. Feed the data into the dimensionality reduction techniques (PCA, CAE, and VAE) in order to derive a latent space (reduced order).
3. Train an interpolation (regression) algorithm (GP, DGP, and LSTM) using the latent space, the associated parameters and time steps.
4. Given the trained interpolation algorithm, we make predictions for new sets of parameters, that were not included in the original data creation. This leads to new values of the latent space.
5. Feed the newly predicted latent space values into the inverse component of the dimensionality reduction algorithm (decoding part). The result gives an approximation of a simulation for the parameters that were used in step 4 (and were omitted from step 1).

Note that the whole procedure described here is presented visually in Figure 4.

## 3. Applications

We apply the ROM analysis pipeline to three fluid-dynamic simulation applications with increasing complexity: (a) an advection–diffusion equation, (b) a falling film flow, and (c) multicomponent polymer precipitation governed by a Cahn–Hilliard equation. We demonstrate how the methodology variants perform on simulation data-sets via appropriate visualizations and metrics.

### 3.1. Advection–diffusion

Advection–diffusion equations expressed as

$$\frac{\partial c}{\partial t} + \vec{v} \cdot \vec{\nabla} c = D \nabla^2 c \quad (9)$$

describe physical phenomena where mass, momentum, and energy are transported within a physical domain advectively by bulk motion and diffusively in response to differences in chemical potential and/or temperature. Here,  $c$  is a concentration,  $D$  is a constant diffusion coefficient,  $\vec{v} = (v_x(x, y), v_y(x, y))$  is the velocity field, and  $\vec{\nabla}$  denotes the gradient operator. Note that for a constant velocity,  $U_c$ , and an appropriate characteristic length scale,  $L_c$ , one can scale space and time on  $L_c$  and  $L_c/U_c$ , respectively, to arrive at

$$\frac{\partial c}{\partial t} + \frac{\partial c}{\partial x} + \frac{\partial c}{\partial y} = \frac{1}{Pe} \left( \frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2} \right), \quad (10)$$

where  $Pe = U_c L_c / D$  is a Péclet number.

To generate a characteristic advection–diffusion simulation data-set, we implement and solve Equation (10) using the framework of Bar-Sinai et al. (2019) and Zhuang et al. (2020). For training purposes, we generate 20 simulations for different values of  $Pe^{-1}$  ranging between 0.05 and 0.15, representing a blob of inert tracer placed in a constant-velocity fluid field. For each simulation, 50 evenly spaced time-snapshots with  $64 \times 64$  resolution are obtained to construct the full data-set. We split the simulations into 19 cases for training and 1 for testing, focusing on the *extrapolation* problem. The data are fed into the three algorithms presented in Section 2.1. We train the different compression algorithms for four degrees of freedom (DOF). For an extensive analysis on how different DOF affect the result, we refer the reader to the experiment section in Maulik et al. (2021a). For the encoders of the CAE and VAE we use five convolutional and pooling layers and an equal number of unpooling and upsampling layers for the decoders. We also train the data in mini-batches of four, and we use 1,000 epochs and early stopping based on a validation set that consists of 10% of the training data. We use TensorFlow (Abadi et al., 2016) for autoencoder implementations. For the GPs we use an ARD Matérn 3/2 kernel with two-dimensional lengthscale (one dimension corresponds to the diffusion coefficient and the other to the temporal parameter). For the DGPs we use 128 inducing points, two layers, each of which comprises of a single GP with a Matérn 3/2 kernel. Finally, for the LSTM architecture, we use three cells with 50 neurons in each cell, mini-batches of size 32, early stopping based on 10% of the training data that are set aside, and a time window of 10 points for the forecasts, following a similar approach to Maulik et al. (2021b). All the interpolation algorithms use the  $Pe$  value as an additional input to perform the latent space interpolation.

To assess the different methods, we use two metrics: the MSE and the mean absolute error (MAE). We remind the reader that output frames derived directly from the compression algorithms (in a typical ROM assessment fashion) are not possible to derive, since we make predictions for  $Pe$  values outside the training set. It is clear from Table 1 that PCA and DGP are consistently the worst performing of the compression and interpolation algorithms, respectively (with a few exceptions, such as DGP outperforming LSTM for the PCA case); the conclusion associated with PCA was one also reached by Maulik et al. (2021a). As for the DGP conclusion, it appears that the additional complexity of using the DGP is unwarranted when applied to relatively small and simple data-sets, and that the corresponding model may become over-parameterized. Instead, the standard GP is flexible enough to capture the variability of the data obviating the need for a DGP. It is also unclear based on the results presented which of the two autoencoders and remaining interpolation algorithms performs better. The MAE favors all the LSTM variations, and specifically the CAE–LSTM combination, while the MSE is lower for the GP variations and especially VAE–GP, indicating that the LSTM offers lower average error and the GP fewer error spikes. The error values, however, are very small signifying the overall effectiveness of the methodology.

In Figures 5 and 6, we show indicative input frames (from top to bottom: 11, 21, 31, 41, and 50), along with the different VAE-based reconstructions and the corresponding residual plots, respectively.

**Table 1.** Evaluation metrics for the PCA, CAE, and VAE models for the advection–diffusion problem.

Metric/Modelt	PCA		
	GP	DGP	LSTM
MAE	$2.51355 \times 10^{-5}$	$4.32963 \times 10^{-5}$	0.00203
MSE	0.00264	0.00292	0.01594
Metric/Modelt	CAE		
	GP	DGP	LSTM
MAE	$7.20011 \times 10^{-7}$	$1.28770 \times 10^{-6}$	<b><math>4.83279 \times 10^{-7}</math></b>
MSE	$4.48591 \times 10^{-5}$	$6.12765 \times 10^{-5}$	$4.98987 \times 10^{-5}$
Metric/Modelt	VAE		
	GP	DGP	LSTM
MAE	$6.31853 \times 10^{-7}$	$1.22549 \times 10^{-6}$	$4.87542 \times 10^{-7}$
MSE	<b><math>4.27802 \times 10^{-5}</math></b>	$6.30656 \times 10^{-5}$	$4.57708 \times 10^{-5}$

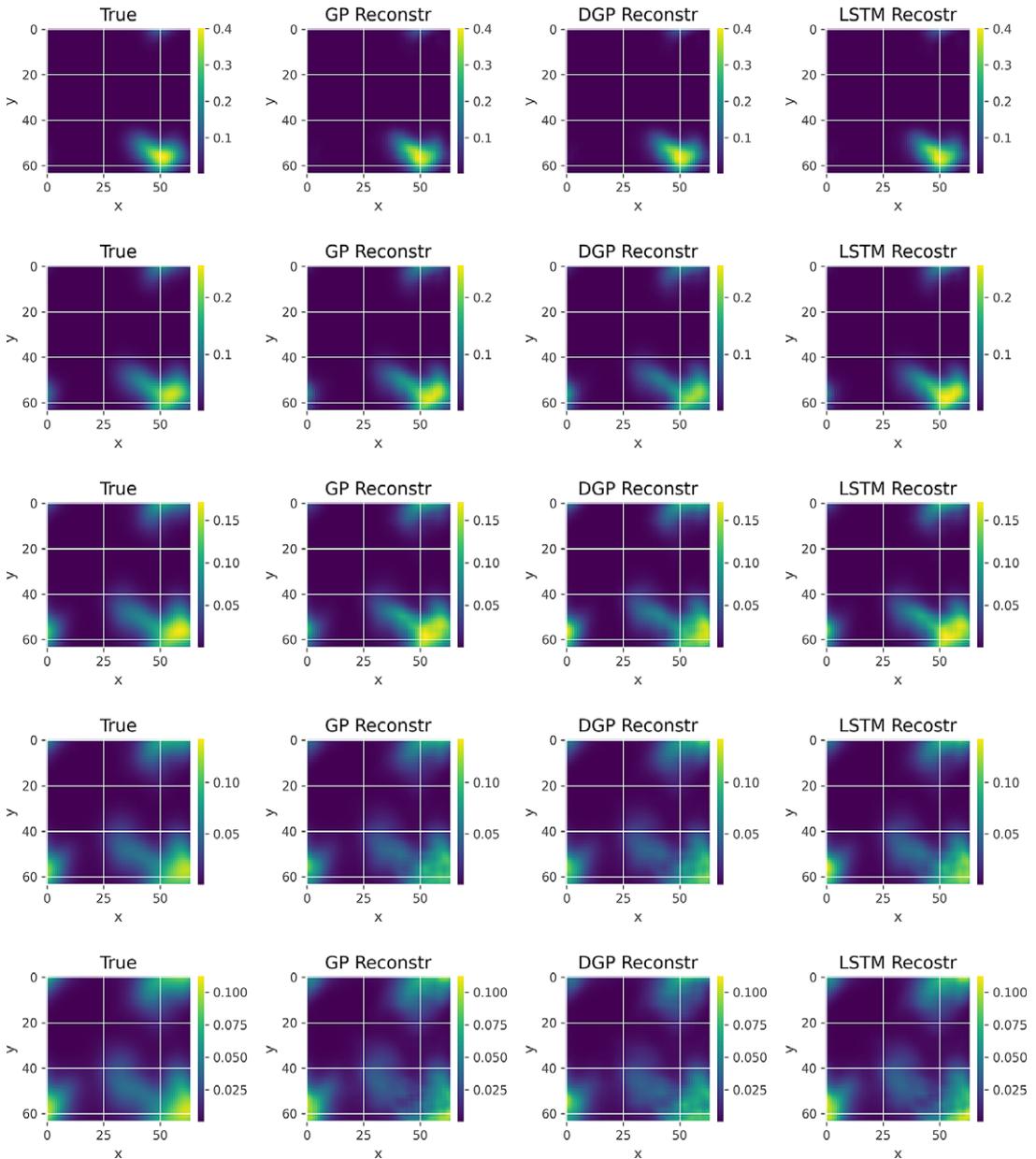
Note. The best performance for each metric (i.e., MAE and MSE) is highlighted in bold.

Abbreviations: CAE, convolutional autoencoder; DGP, deep Gaussian processes; GP, Gaussian process; LSTM, long short-term memory; MAE, mean absolute error; MSE, mean squared error; PCA, principal component analysis; VAE, variational convolutional autoencoders.

Specifically, the first column in Figure 5 corresponds to the true frames, the second to the GP reconstruction, the third to the DGP reconstruction, and the final to the LSTM reconstruction. Typically the true frames (left column) would not be available (nor required for the production of the two other columns), but here they are included for reference and model validation. The scaling is common for each row. Figure 6 uses the last three cases for the corresponding residuals. In order to show the differences among these plots we use flexible scaling. In terms of the actual frames we can see that the reconstruction is successful and it is hard to detect differences from the originals, even in the case of the DGP, which is the interpolation algorithm that performed more poorly, with some minor exceptions in the bottom frames. Regarding the residuals, we can see the similar error patterns generated, though, the scale makes it clear that the average error is larger for some of the frames in the GP and most in the DGP cases, when compared to the LSTM column.

In Figure 7, we gain insights into the practical differences between the three interpolation algorithms by showing the corresponding representations of the first dimension of the latent space. The blue line shows the actual first dimension from the VAE latent space, the red line the mean for the GP-based algorithms for the first two and the output of the LSTM for the third, and the light red shading the confidence intervals around 2 standard deviations from the mean. Starting from the GP (Figure 7a) we can see a very good fit and an inconsistent uncertainty level. Generally, the uncertainty is narrower close to the times where data are available and wider everywhere else. The DGP case (Figure 7b) shows a very similar mean result, but the over-parameterization resulted to a wide uniform credible interval along the time axis. Finally, the LSTM (Figure 7c) shows an almost perfect fit to the latent space. The missing line before  $t = 0.2$  of the plot corresponds to the timewindow of 10 points that is required for the forecast.

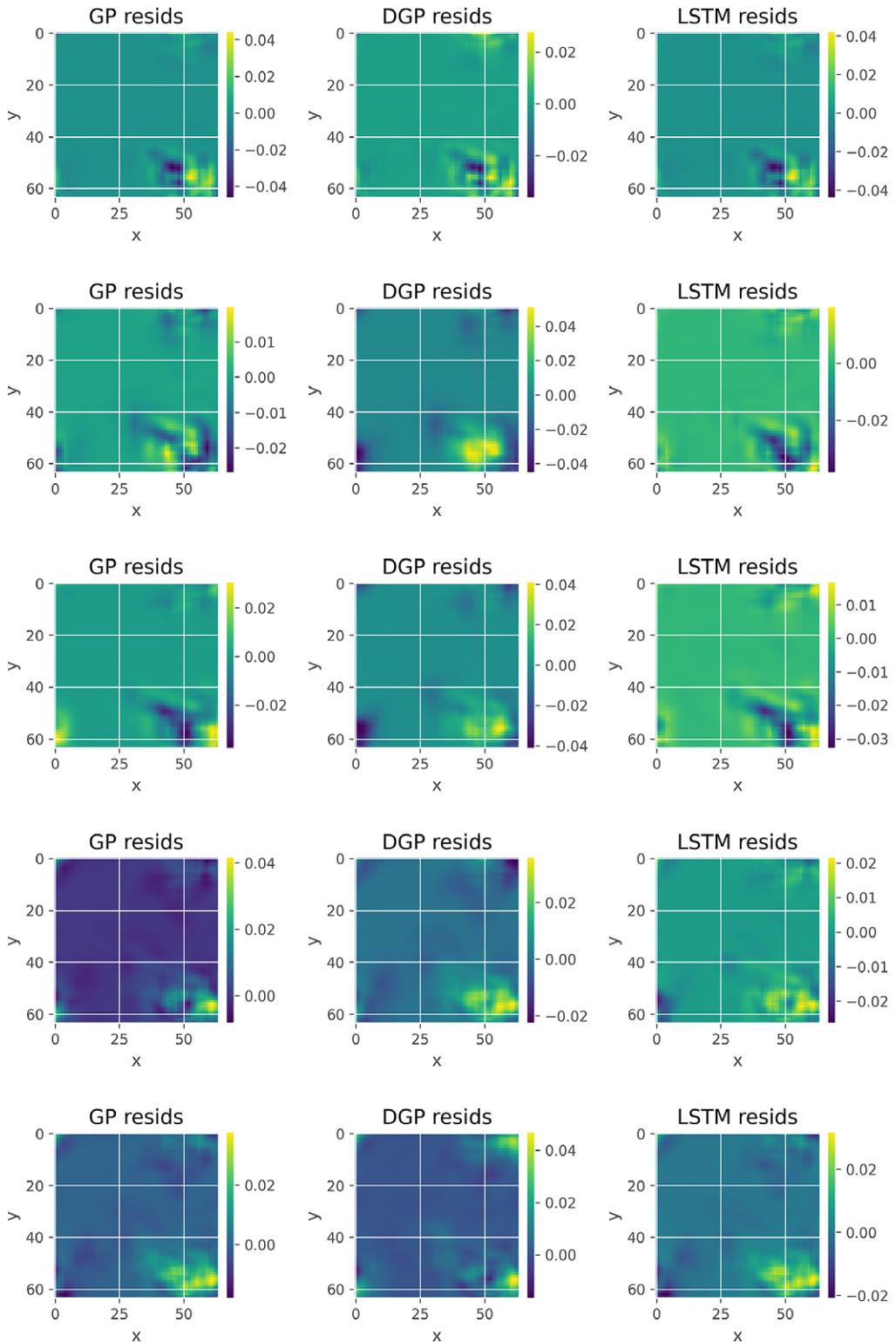
The computational cost for all the algorithms examined in this section is shown in Table 2. The compression algorithm is run first and its cost is common for all relative pipelines regardless of the choice of the interpolation. PCA is almost instantaneous, while the autoencoders require a similar amount of time with each other. The time to solution for the interpolation algorithms differ based on the compression algorithm they follow, but the scale remains approximately the same. Specifically, LSTMs are consistently the fastest, followed by GPs, while DGPs are the slowest. It is worth noting that different parameterizations of the algorithms can potentially lead to significant differences with respect to computational costs, within each class of compression and interpolation.



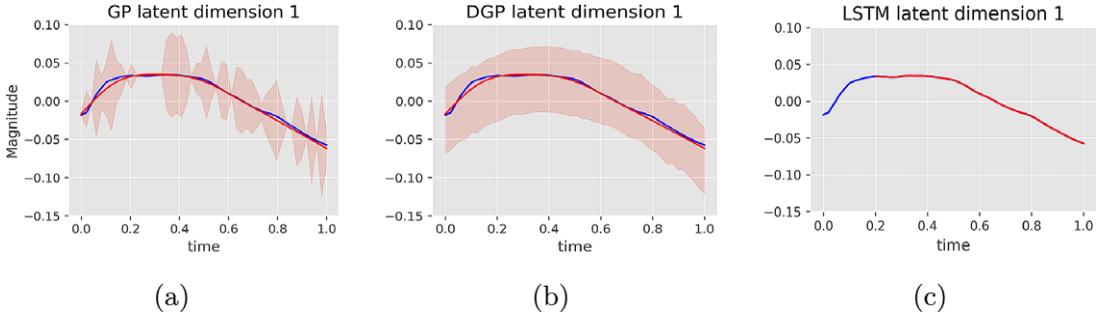
**Figure 5.** True and reconstructed frames for the concentration profiles of the VAE-related methods applied to the advection–diffusion problem. The rows correspond to time-steps 11, 21, 31, 41, and 50, respectively.

### 3.2. Falling films

We focus in this application on two-dimensional falling film flows, of importance to engineering applications including separation and heat removal units (Rohlfis et al., 2018). We apply the modeling framework of Scheid et al. (2006), which uses an averaged version of the Navier–Stokes equations together with Pade approximants; this yields a three-field equation system for the film thickness,  $h$ , and flow rates in the streamwise direction  $x$ ,  $q^x$ , and spanwise one  $y$ ,  $q^y$ , expressed by



**Figure 6.** Residual concentration profile plots of the VAE-related methods applied to the advection–diffusion problem. The rows correspond to time-steps 11, 21, 31, 41, and 50, respectively.



**Figure 7.** Comparison of the interpolation techniques; GP in (a), DGP in (b), and LSTM in (c), in the latent space for advection–diffusion.

**Table 2.** Computational cost (in seconds) for all compression and interpolation algorithms.

PCA		
GP	DGP	LSTM
$9.1 \times 10^1$	$5.96 \times 10^2$	$4.2 \times 10^1$
CAE		
GP	DGP	LSTM
$1.46 \times 10^2$	$6.19 \times 10^2$	$3.9 \times 10^1$
VAE		
GP	DGP	LSTM
$1.14 \times 10^2$	$4.3 \times 10^2$	$5.6 \times 10^1$

Abbreviations: CAE, convolutional autoencoder; DGP, deep Gaussian processes; GP, Gaussian process; LSTM, long short-term memory; PCA, principal component analysis; VAE, variational convolutional autoencoder.

$$\frac{\partial h}{\partial t} = -\frac{\partial q^x}{\partial x} - \frac{\partial q^y}{\partial y}, \tag{11}$$

$$\delta \frac{\partial q^x}{\partial t} = \delta \frac{\partial \mathcal{I}^{2D}}{\partial x} + \mathcal{G} \left( \frac{5}{6} h - \frac{5}{2h^2} \frac{\partial q^x}{\partial x} + \delta \frac{\partial \mathcal{I}^{3D}}{\partial x} + \eta \left[ \frac{\partial \mathcal{D}^{2D}}{\partial x} + \frac{\partial \mathcal{D}^{3D}}{\partial x} \right] + \frac{5h}{6} \frac{\partial \mathcal{P}}{\partial x} \right), \tag{12}$$

$$\delta \frac{\partial q^y}{\partial t} = \delta \frac{\partial \mathcal{I}^{2D}}{\partial y} - \frac{5}{2h^2} \frac{\partial q^y}{\partial y} + \delta \frac{\partial \mathcal{I}^{3D}}{\partial y} + \eta \left[ \frac{\partial \mathcal{D}^{2D}}{\partial y} + \frac{\partial \mathcal{D}^{3D}}{\partial y} \right] + \frac{5h}{6} \frac{\partial \mathcal{P}}{\partial y}, \tag{13}$$

where  $\mathcal{G} = [1 - (\delta/70)h^3 \frac{\partial h}{\partial x}]^{-1}$ ,  $\mathcal{I}$  and  $\mathcal{D}$  represent a collection of terms related to inertia and viscous dissipation, defined in equations (B1b) and (B1c) in Scheid et al. (2006), and  $\mathcal{P} = -\zeta h + h \left( \frac{\partial^2 h}{\partial x^2} + \frac{\partial^2 h}{\partial y^2} \right)$  represents contributions to the pressure due to gravity and capillarity. The parameters  $\delta$ ,  $\eta$ , and  $\zeta$  respectively represent the relative significance of inertia, gravity, and capillarity namely equations (2.9) and (2.10) in Scheid et al. (2006); the parameter  $\delta$  is referred to as a reduced Reynolds number.

Data were obtained from numerical solutions of Equations (11)–(13) simulated using a custom port of WaveMaker (Rohlfis et al., 2018) implemented in the Julia language (Bezanson et al., 2017). The general

structure for this application has a significant degree of similarity with the one applied in Section 3.1, but with a few key differences: first, the simulations are generated based on the reduced Reynolds number  $\delta$  that takes 20 equi-spaced values in the range of 1–69. This is also the parameter used for the interpolation algorithms. Second, in this application we focus on the *interpolation* problem. Specifically, we use the sixth simulation (that corresponds to  $\delta = 25.5$ ) for testing and the rest for training. Finally, we use five DOF.

The metrics for all the results are shown in Table 3. Similar to the first advection–diffusion application, the PCA and the DGP underperform when compared to the other methods. For this data-set the combination of CAE and LSTM was the best performing based on both the MAE and the MSE metrics. As with the advection–diffusion application, finding differences between the frames for the interpolation algorithms paired with the CAE compression, and the true frames, which are included for reference, on the leftmost column of Figure 8 is difficult, regardless of the significant differences shown by the metrics. The only notable exception is the DGP reconstruction that shows slight deviations from the true film thickness. It is interesting to note that for all methods in Figure 9, and particularly the LSTM, the error structure in the upper frames (early times) seems random, whereas in the bottom frames (late times) we can clearly see the general structure of the film thickness. This indicates that it is easier for the methodology to reconstruct the frames from the beginning of the simulations (where all training simulations are similar) rather than the ending (where deviations among simulations are visibly different from each other due to the effect of the different reduced Reynolds numbers).

### 3.3. Multicomponent polymer precipitation

We now apply the analysis pipeline to polymer precipitation dynamics, of importance to engineering design problems in high-performance plastics and membrane systems. The complex dynamics and rich pattern formation were considered recently by Inguva et al. (2020, 2021) (see also references therein) who used Cahn–Hilliard theory to model and simulate the spatio-temporal evolution of the emergent phase separation patterns; the relevant equations for a binary polymer blend are expressed by

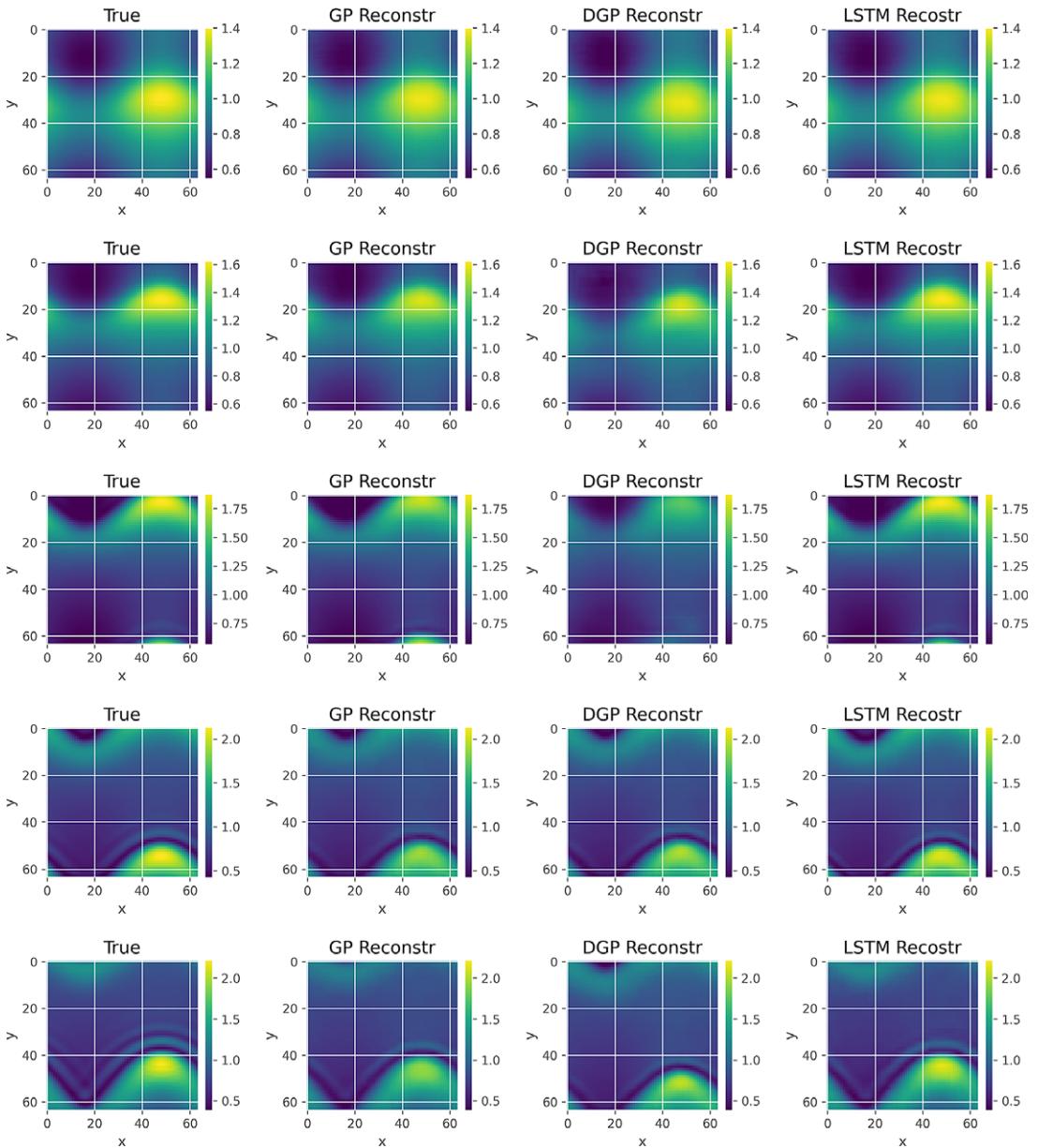
$$\frac{\partial \phi}{\partial t} - \nabla \cdot (M \nabla \mu) = 0, \quad (14)$$

**Table 3.** Evaluation metrics for the PCA, CAE, and VAE models.

Metric/Model	PCA		
	GP	DGP	LSTM
MAE	0.04865	0.04819	0.04055
MSE	0.17284	0.15188	0.16617
Metric/Model	CAE		
	GP	DGP	LSTM
MAE	$5.45812 \times 10^{-5}$	$3.28181 \times 10^{-4}$	<b><math>1.90501 \times 10^{-5}</math></b>
MSE	0.00056	0.00149	<b>0.00036</b>
Metric/Model	VAE		
	GP	DGP	LSTM
MAE	0.00013	0.00069	$5.46090 \times 10^{-5}$
MSE	0.00079	0.00229	0.00079

*Note.* The best performance for each metric (i.e., MAE, MSE) is highlighted in bold.

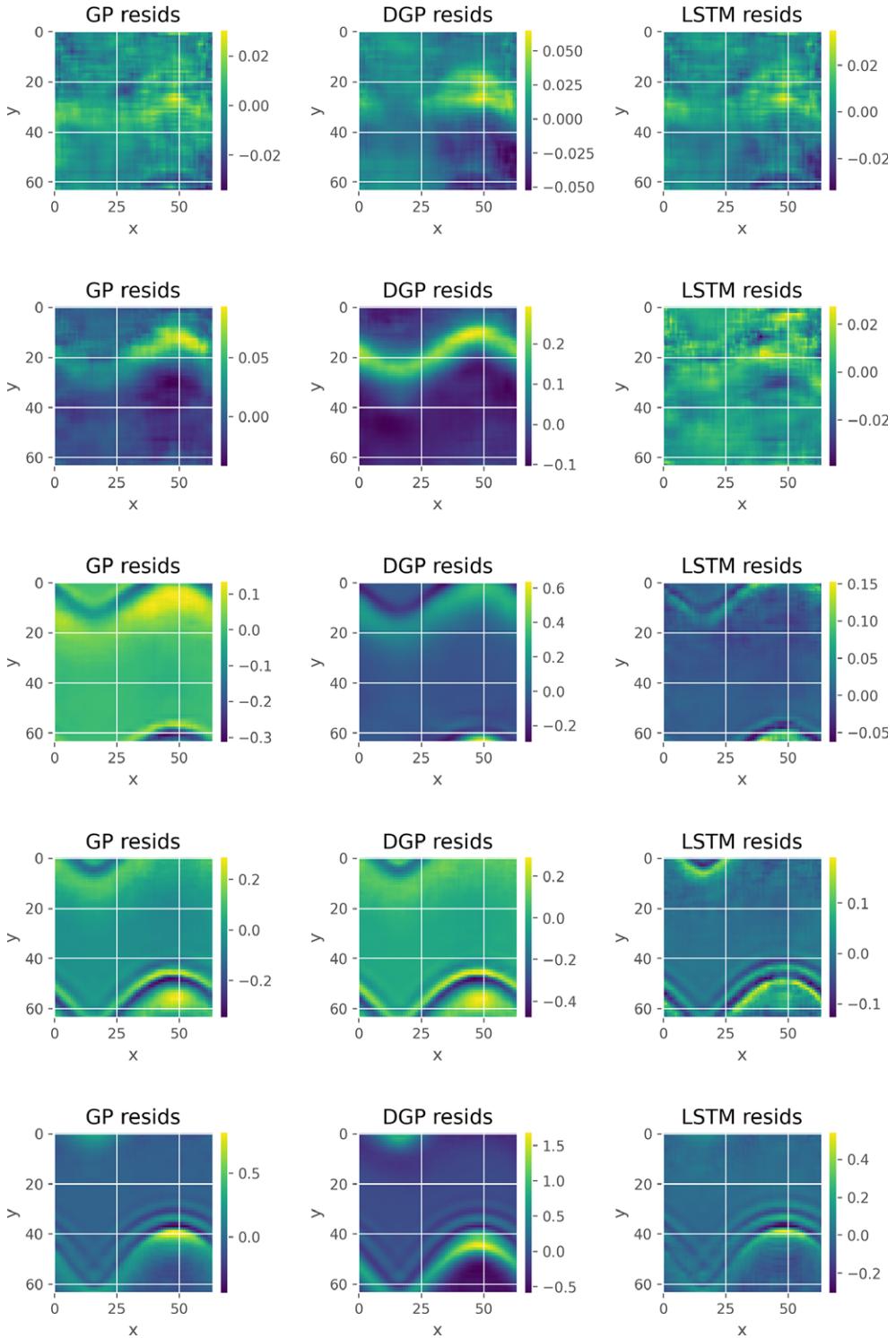
Abbreviations: CAE, convolutional autoencoder; DGP, deep Gaussian processes; GP, Gaussian process; LSTM, long short-term memory; MAE, mean absolute error; MSE, mean squared error; PCA, principal component analysis; VAE, variational convolutional autoencoders.



**Figure 8.** True and reconstructed frames of the film thickness for the CAE-related methods employed in the falling film problem. The rows correspond to time-steps 11, 21, 31, 41, and 50, respectively.

where  $\phi$  represents the volume fraction of one of the polymers in the blend,  $M$  is a constant mobility parameter, and  $\mu$  is a generalized chemical potential, which can be derived from the variational derivative of the Gibbs free energy functional:

$$\mu = \frac{df}{d\phi} - \lambda \nabla^2 \phi; \tag{15}$$



**Figure 9.** Residual plots of the film thickness associated with the CAE-related methods used in the falling film problem. The rows correspond to time-steps 11, 21, 31, 41, and 50, respectively.

**Table 4.** Evaluation metrics for the PCA, CAE, and VAE models.

Metric/Model	PCA		
	GP	DGP	LSTM
MAE	0.00756	0.00808	0.03973
MSE	0.06581	0.06553	0.15826
Metric/Model	CAE		
	GP	DGP	LSTM
MAE	<b><math>7.50197 \times 10^{-5}</math></b>	0.02313	0.00082
MSE	<b>0.00084</b>	0.01698	0.00277
Metric/Model	VAE		
	GP	DGP	LSTM
MAE	0.00081	0.02814	0.00111
MSE	0.00317	0.01871	0.00414

Note. The best performance for each metric (i.e., MAE and MSE) is highlighted in bold.

Abbreviations: CAE, convolutional autoencoder; DGP, deep Gaussian processes; GP, Gaussian process; LSTM, long short-term memory; MAE, mean absolute error; MSE, mean squared error; PCA, principal component analysis; VAE, variational convolutional autoencoders.

here,  $f$  denotes the homogeneous contribution to the Gibbs free energy per monomer, which is a nonconvex function of  $\phi$ , and  $\lambda$  is a gradient free energy parameter. Numerical solutions of the above equations are obtained subject to Neumann conditions:

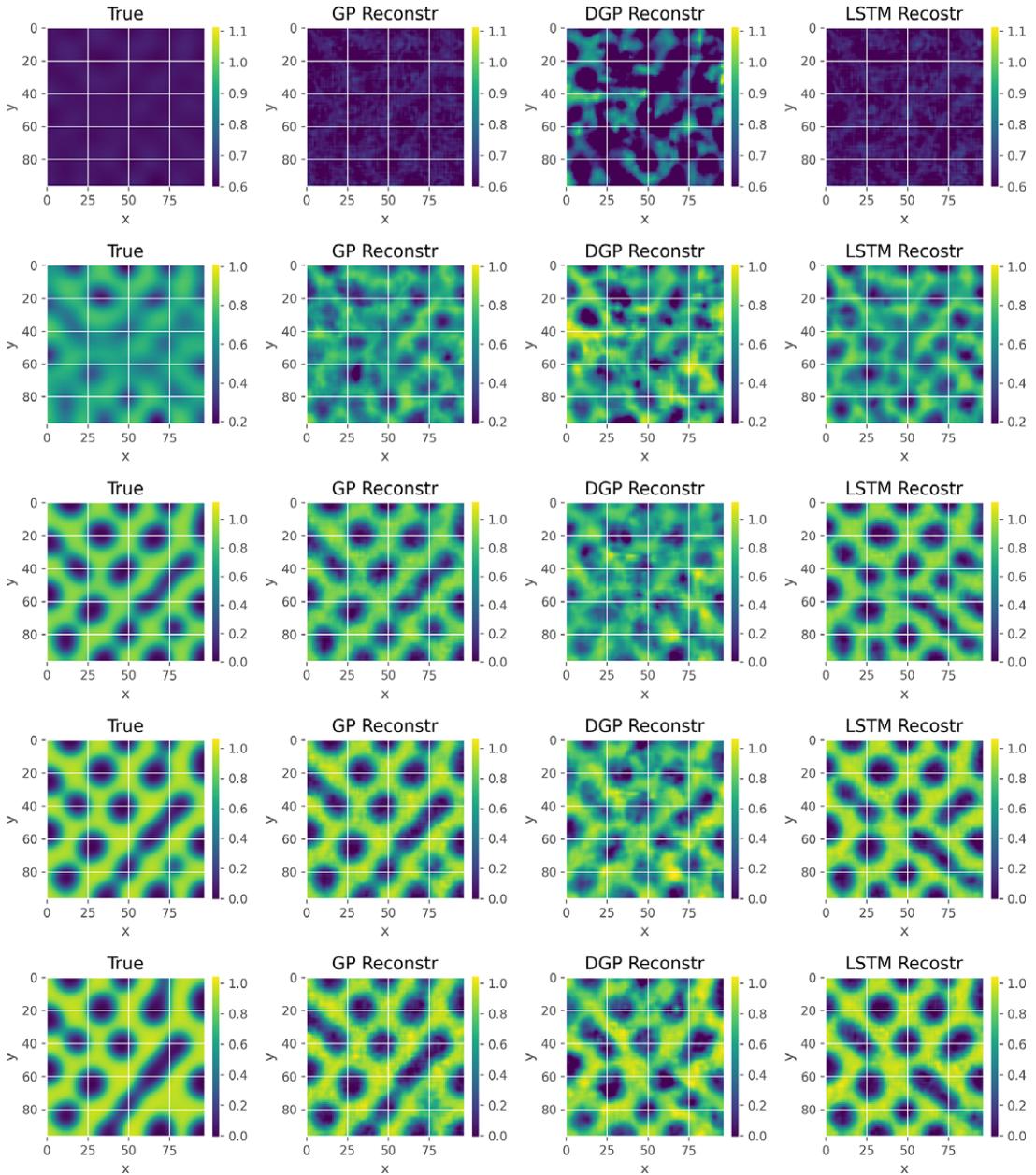
$$\nabla \mu \cdot \mathbf{n} = 0,$$

$$\nabla c \cdot \mathbf{n} = 0,$$

in which  $\mathbf{n}$  is the outward-directed boundary normal.

For training purposes, we generate 20 simulations based on values of  $\lambda$  within the range 0.01–0.0575, which is also the parameter used for the interpolation algorithms. We use the python API of the FEniCS project Alnæs et al. (2015) for the simulations. An interesting complication comes from the fact that the simulator provides results in the form of  $97 \times 97$  frames, which is not convenient for the standard structure of the autoencoders, since 97 is not a number that can be achieved with the up-sampling layers. This gives us the opportunity to examine how to address the issue of arbitrarily-shaped data. We tackle this with a zero padding that transforms the inputs to  $128 \times 128$  frames. Once again, we are concerned with *extrapolation*, attempting to reconstruct the frames that correspond to the last  $\lambda$  value, and use five DOF. All the hyper-parameters of the compression and interpolation algorithms remain the same.

According to the results presented in Table 4, the best performing combination is the CAE paired with the GP. Both the MAE and the MSE indicate significant improvement from the other compression and interpolation algorithms. The zero padding seems to affect the VAE negatively, which struggles to deal with the abrupt transitions between the zero-padded sections and the actual frames possibly due to the distributional aspect of the algorithm. The unusual structure of the data is not easy to capture as shown in Figure 10, where all the interpolation algorithms are shown for the CAE case. Note that the true volume fraction frames are included for reference. It is hard to distinguish anything in the earlier frames, mainly due to the common scaling and the fact that the DGP seems to be deviating significantly. In the frames that correspond to late times, we can observe the GP-reconstructed frames having some minor deviations from their true counterparts, but generally being able to capture the correct space of the circular features, while the other two methods deviate significantly.

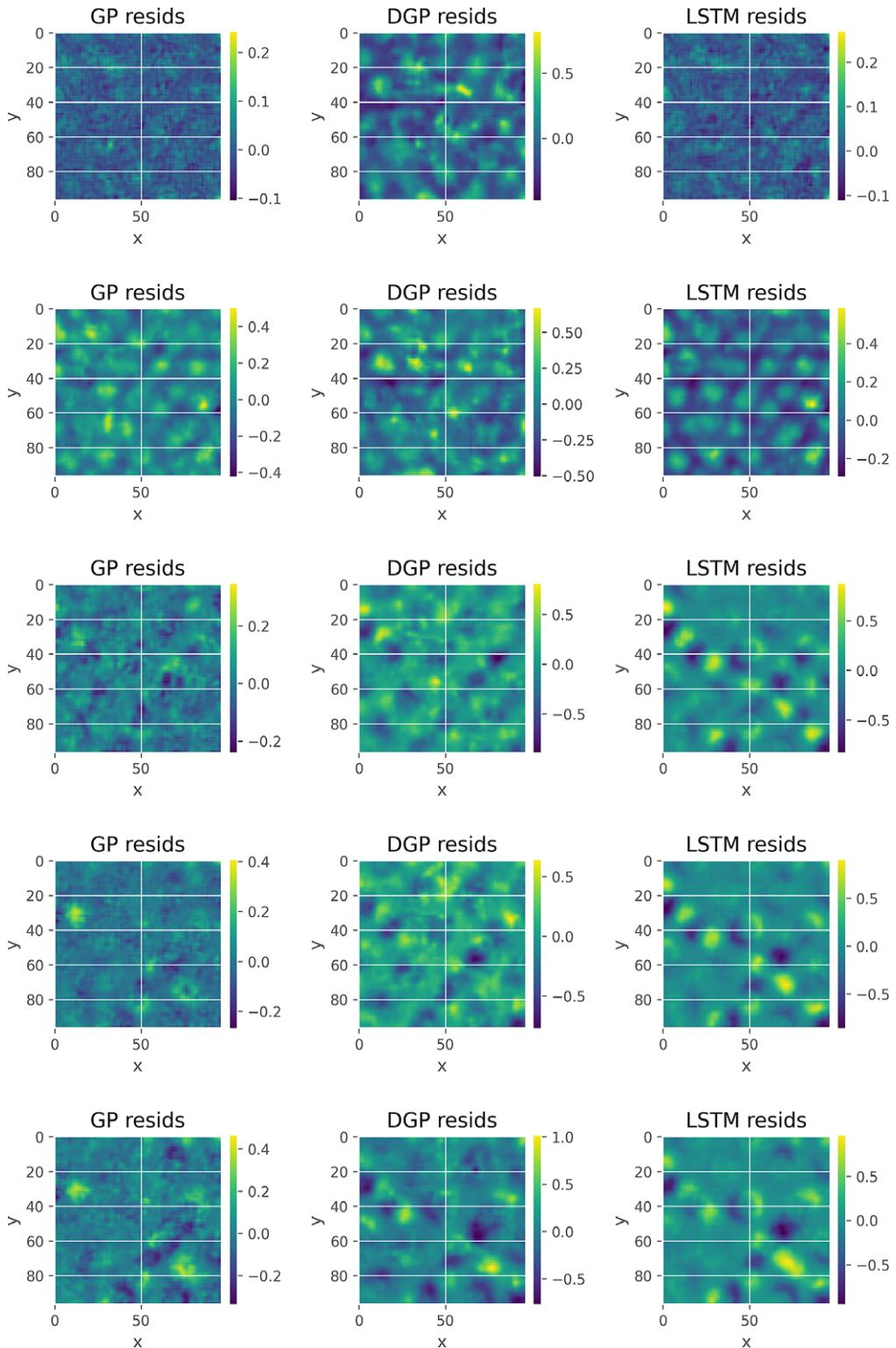


**Figure 10.** True and reconstructed frames for the volume fraction of the CAE-related methods used in the polymer precipitation problems. The rows correspond to time-steps 11, 21, 31, 41, and 50, respectively.

Finally, in Figure 11 we do not observe any specific pattern in the errors, which is expected from the structure of the data, but we can clearly see that the level of the errors is considerably lower for the late time frames in the GP case when compared to the other two.

**4. Conclusion**

In this study, we used the ROM interpolation framework from our previous paper (Maulik et al., 2021a) with the main purpose of showing a previously overlooked advantage, that is, the ability to emulate results



**Figure 11.** Residual volume fraction plots of the CAE-related methods in the polymer precipitation problems. The rows correspond to time-steps 11, 21, 31, 41, and 50, respectively.

for parameter values with the help of the GP interpolation, where data are not available. The effectiveness of this method is presented with three realistic multiphase flow applications with different levels of complexity and results are compared with two other interpolation techniques; the newly introduced DGPs, which are complex algorithms that use the GPs as building blocks, and the LSTMs, which have been used recently in literature (Maulik et al., 2021b) and have been found to work well for a class of flow problems. The results show that the choice of the best type of autoencoder is problem dependent, although CAE has a slight edge and is more versatile. In terms of the interpolation algorithms the GP provided good results that were on the same level as those of the LSTM method, which exists in the literature. The choice between the two is data-dependent. We believe that the main reason for the under-performance of the DGPs was over-parameterization, with respect to the complexity of the generated flow patterns. In future work, we plan on replicating the same comparison for larger and more complex data-sets, in order to address this issue, though we expect to be confronted with new challenges, such as the choice of hyper-parameters, especially for the LSTM and DGP.

**Acknowledgments.** We thank Romit Maulik and Nesar Ramachandra from Argonne National Laboratory for their collaboration in a previous project that was used as a basis for this article.

**Data Availability Statement.** Replication data and code can be found in the Github repository: [https://github.com/themisbo/ROM\\_applications.git](https://github.com/themisbo/ROM_applications.git).

**Author Contributions.** Conceptualization: T.B., I.P.; Data curation: L.R.M.; Formal analysis: T.B., I.P.; Funding acquisition: I.P., O.K.M.; Investigation: T.B.; Methodology: T.B., I.P., L.R.M.; Project administration: I.P., O.K.M.; Software: T.B., L.R.M.; Supervision: I.P., O.K.M.; Visualization: T.B.; Writing—original draft: T.B.; Writing—review and editing: I.P., L.R.M., O.K.M. All authors approved the final submitted draft.

**Funding Statement.** We acknowledge funding from the Engineering and Physical Sciences Research Council, UK, through the Programme Grant PREMIERE (EP/T000414/1), as well as funding through the Wave 1 of The UKRI Strategic Priorities Fund under the EPSRC Grant EP/T001569/1, particularly the *Digital Twins for Complex Engineering Systems* theme within that grant, and the Royal Academy of Engineering through their support of O.K.M.'s PETRONAS/RAEng Research Chair in Multiphase Fluid Dynamics. I.P. acknowledges funding from the NUAcT fellowship scheme at Newcastle University and Imperial College Research Fellowship scheme at Imperial College London.

**Competing Interests.** The authors declare no competing interests exist.

## References

- Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M (2016) Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv:1603.04467.
- Alnæs M, Blechta J, Hake J, Johansson A, Kehlet B, Logg A, Richardson C, Ring J, Rognes ME and Wells GN (2015) The Fenics project version 1.5. *Archive of Numerical Software* 3(100), 9–23.
- Bar-Sinai Y, Hoyer S, Hickey J and Brenner MP (2019) Learning data-driven discretizations for partial differential equations. *Proceedings of the National Academy of Sciences* 116(31), 15344–15349.
- Bezanson J, Edelman A, Karpinski S and Shah VB (2017) Julia: A fresh approach to numerical computing. *SIAM Review* 59(1), 65–98.
- Bishop CM (2006) *Pattern Recognition and Machine Learning*. New York, NY: Springer.
- Damianou A and Lawrence ND (2013) Deep gaussian processes. In *Artificial Intelligence and Statistics*. New York, NY: Proceedings of Machine Learning Research, pp. 207–215.
- Gardner JR, Pleiss G, Bindel D, Weinberger KQ and Wilson AG (2018) Gpytorch: Blackbox matrix-matrix Gaussian process inference with gpu acceleration. In *Advances in Neural Information Processing Systems*, 31.
- Gonzalez FJ and Balajewicz M (2018) Deep convolutional recurrent autoencoders for learning low-dimensional feature dynamics of fluid systems. arXiv:1808.01346.
- Hochreiter S and Schmidhuber J (1997) Long short-term memory. *Neural Computation* 9(8), 1735–1780.
- Hu R, Fang F, Pain C and Navon I (2019) Rapid spatio-temporal flood prediction and uncertainty quantification using a deep learning method. *Journal of Hydrology* 575, 911–920.
- Inguva PK, Mason LR, Pan I, Hengardi M and Matar OK (2020) Numerical simulation, clustering, and prediction of multicomponent polymer precipitation. *Data-Centric Engineering* 1, E13.
- Inguva PK, Walker PJ, Yew HW, Zhu K, Haslam AJ and Matar OK (2021) Continuum-scale modelling of polymer blends using the Cahn–Hilliard equation: Transport and thermodynamics. *Soft Matter* 17, 5645–5665.

- Kim B, Azevedo VC, Thuerey N, Kim T, Gross M and Solenthaler B** (2019) Deep fluids: A generative network for parameterized fluid simulations. In *Computer Graphics Forum*, Vol. 38. Hoboken, NJ: Wiley Online Library, pp. 59–70.
- Kingma DP and Welling M** (2013) Auto-encoding variational bayes. arXiv:1312.6114.
- LeCun Y, Bengio Y** (1995) Convolutional networks for images, speech, and time series. In *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA: MIT Press, p. 3361.
- Maulik R, Botsas T, Ramachandra N, Mason LR and Pan I** (2021a) Latent-space time evolution of non-intrusive reduced-order models using gaussian process emulation. *Physica D: Nonlinear Phenomena* 416, 132797.
- Maulik R, Lusch B and Balaprakash P** (2021b) Reduced-order modeling of advection-dominated systems with recurrent neural networks and convolutional autoencoders. *Physics of Fluids* 33(3), 037106.
- Mohan AT and Gaitonde DV** (2018) A deep learning based approach to reduced order modeling for turbulent flow control using lstm neural networks. arXiv:1804.09269.
- Rasmussen CE and Williams CKI** (2006) *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press.
- Rohlf W, Rietz M and Scheid B** (2018) Wavemaker: The three-dimensional wave simulation tool for falling liquid films. *SoftwareX* 7, 211–216.
- Salimbeni H and Deisenroth M** (2017) Doubly stochastic variational inference for deep gaussian processes. In *Advances in Neural Information Processing Systems*, 30.
- Scheid B, Ruyer-Quil C and Manneville P** (2006) Wave patterns in film flows: Modelling and three-dimensional waves. *Journal of Fluid Mechanics* 562, 183–222.
- Williams CK and Rasmussen CE** (1996) Gaussian processes for regression. In *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, pp. 514–520.
- Zhuang J, Kochkov D, Bar-Sinai Y, Brenner MP and Hoyer S** (2020) Learned discretizations for passive scalar advection in a 2-d turbulent flow. arXiv:2004.05477.