# Forecasting global climate drivers using Gaussian processes and convolutional autoencoders

James Donnelly [a,b,*], Alireza Daneshkhah [a], Soroush Abolfathi [b]

[a] *Centre for Computational Science & Mathematical Modelling, Coventry University, UK*
[b] *School of Engineering, University of Warwick, UK*

## ARTICLE INFO

## ABSTRACT

Machine learning (ML) methods have become an important tool for modelling and forecasting complex high-dimensional spatiotemporal datasets such as those found in environmental and climate modelling applications. ML approaches can offer a fast, low-cost alternative to short-term forecasting than expensive numerical simulation while addressing a significant outstanding limitation of numerical modelling by being able to robustly and dynamically quantify predictive uncertainty. Low-cost and near-instantaneous forecasting of high-level climate variables has clear applications in early warning systems, nowcasting, and parameterising small-scale locally relevant simulations. This paper presents a novel approach for multi-task spatiotemporal regression by combining data-driven autoencoders with Gaussian Processes (GP) to produce a probabilistic tensor-based regression model. The proposed method is demonstrated for forecasting one-step-ahead temperature and pressure on a global scale simultaneously. By conducting probabilistic regression in the learned latent space, samples can be propagated back to the original feature space to produce uncertainty estimates at a vastly reduced computational cost. The composite GP-autoencoder model was able to simultaneously forecast global temperature and pressure values with average errors of 3.82 °C and 638 hPa, respectively. Further, on average the true values were within the proposed posterior distribution 95.6% of the time illustrating that the model produces a well-calibrated predictive posterior distribution.

## 1. Introduction

Many scientific and engineering problems focus on dynamical systems, complex phenomena evolving in time and space, often characterised by non-linear interactions of underlying mechanisms on varying scales. Forecasting the behaviour of these complex and multi-faceted systems is challenging yet critical and relied upon every day, i.e., such as weather forecasts. In order to forecast these complex dynamical systems, sophisticated computer models have been developed that aim to simulate the underlying physics and processes governing the behaviour of the system. These models typically represent the system in terms of a set of partial differential equations (PDEs). For instance, numerical weather prediction (NWP) models may rely on finite difference implementations of the Navier–Stokes equations for incompressible fluid flow (Bauer et al., 2015; Giraldo and Restelli, 2008).

In recent decades, there has been significant advancement in computer models designed to solve systems of PDEs, aiming to predict the behaviour of such systems. This progress has been driven largely by breakthroughs in both software engineering and hardware development, enabling parallel simulations that leverage the immense computational power of hundreds of thousands of CPU and GPU cores (Worley et al., 2011). However, as these computer models continue to scale up in complexity, the computational resources they demand increase proportionally. This escalation often renders many simulation tasks, such as those integral to NWP, beyond the reach of much of the scientific and engineering community. Additionally, there is a need to better consider the economic and environmental costs associated with the operation of supercomputers, aspects that are often overshadowed or neglected (Portegies Zwart, 2020; Bates et al., 2015). Moreover, even when utilising the most advanced supercomputer and algorithms available, certain forecasting tasks, such as those involving real-time prediction or a large number of evaluations (for uncertainty quantification and sensitivity analysis) remain infeasible due to the computational time required.

Despite significant advances in robustness and accuracy of numerical algorithms for modelling complex systems, the resulting models are generally deterministic in nature, and fail to account for inherent uncertainty and stochasticity within a given system. In the context of

---

* Corresponding author at: Centre for Computational Science & Mathematical Modelling, Coventry University, UK.
*E-mail addresses:* jamie.donnelly@warwick.ac.uk (J. Donnelly), Alireza.Daneshkhah@Coventry.ac.uk (A. Daneshkhah), Soroush.Abolfathi@warwick.ac.uk (S. Abolfathi).

forecasting dynamical systems, uncertainty can represent variability in data, incomplete knowledge of the underlying physics, errors as a result of discretisations and simplifications, and general measurement errors (Xiu, 2009). Therefore, in order to generate robust forecasts for the behaviour of complex dynamical systems, it is imperative to incorporate a systematic approach for quantifying and propagating uncertainty in the forecasts. Monte Carlo sampling (Eigel et al., 2019; Lozanovski et al., 2020) and Bayesian methods (Haukaas and Gardoni, 2011; Behmanesh and Moaveni, 2015; Cockayne et al., 2019; Esmaeil-beigi et al., 2023) have been implemented into numerical computer models to address the issue of uncertainty. However, robust incorporation of uncertainty and probabilistic methods into numerical algorithms remains an active area of research, with ongoing debates regarding the reliable incorporation of various sources of uncertainty, such as uncertain inputs, into complex numerical models (Chakraborty and Chowdhury, 2016).

The integration of ML approaches to forecasting dynamical systems serves several benefits. Primarily, it presents an alternative to the computationally intensive process of simulation when making new inferences about a system. ML models generally operate in a "slow to train, fast to evaluate" paradigm. Once a model is trained, making new inferences is generally very low-cost and occurs in real-time. Further, ML models also offer the benefit of being able to straightforwardly leverage the growing abundance of empirical data in scientific and engineering domains, such as sensor and satellite data, resolving some of the questions surrounding data assimilation in numerical modelling applications such as ocean modelling (Wilkin et al., 2017; Fringer et al., 2019; Fanous et al., 2023a,b).

ML-based approaches offer significant advantages in terms of reducing computational demands for inference and optimising data utilisation. However, some algorithms also have the capability to facilitate probabilistic inference, producing distributions of outputs and allowing for the quantification of uncertainty associated with a forecast. Despite these capabilities, many existing applications of ML to forecasting tasks still rely on deterministic algorithms, especially when dealing with complex high-dimensional and multi-task data like the spatiotemporal data characteristic of dynamical systems. This is because there is generally a trade-off between scalability and having the ability to quantify uncertainty or introduce probabilistic aspects to a model. This study explores efficient and scalable approaches to forecasting high-dimensional multi-task data in a robust and probabilistic approach. The capacity to manage uncertainty is of paramount importance as it enables a more nuanced comprehension of potential system outcomes and variabilities.

To demonstrate the appropriateness and robustness of the proposed methodology, it is applied to a case study of forecasting global climate data. Global and local climatic patterns are examples of a well-studied dynamical system, where forecasts can have significant, far-reaching implications, influencing global economic policy and politics which ultimately impacts the lives of the majority of the population. In this study, a data-driven surrogate model (Donnelly et al., 2023) is developed for probabilistic modelling of global climate data from the ERA5 global reanalysis datasets (Hersbach et al., 2020). The dataset consists of accurate hindcast data for a wide collection of meteorological and environmental measurements across the planet dating back to 1940. Despite not being strictly empirical data, the ERA5 dataset underscores the need and significance of ML methods in facilitating easy scalability and leveraging of vast datasets. The ERA5 dataset, which is already several petabytes in size and continuously updated, serves as a prime example of the importance of efficient handling and analysis of voluminous data.

This study leverages the abundant empirical and quasi-empirical (simulation of historical data) climatic data to develop a data-driven reduced order model (ROM). This ROM is constructed using deep convolutional neural network (CNN)-based autoencoders and GP regression models to simultaneously forecast temperature and pressure observations on a global scale. By integrating these methodologies, this study outlines an advanced and novel approach to quantify predictive uncertainty in high-dimensional feature spaces parameterised by multi-channel tensor data. This is achieved via a latent-space sampling approach whereby uncertainty in the latent space can be robustly decoded back to the original feature space to provide dynamic estimates of uncertainty for multi-task regression problems.

The proposed method is demonstrated to be highly flexible, scalable, and applicable to general dynamical systems based on spatiotemporal data and can provide accurate point-wise one-step-ahead forecasts as well as robust and dynamic estimates of predictive uncertainty. This ROM is demonstrated as a viable alternative to resource-intensive numerical simulations, facilitating rapid inference capabilities essential for downstream applications such as nowcasting, early warning systems (EWS), and producing realistic input for locally relevant lower-cost simulations. The advantages of a model such as this are clear when considering both the costs of constant re-evaluation using expensive multi-core supercomputers and the existing limitations with respect to uncertainty quantification.

The proposed model extends the current state-of-the-art (SOTA) with respect to probabilistic regression in high-dimensional multi-task settings and demonstrates how advances in computer vision can be effectively applied to address problems in dynamical systems and climate forecasting. Furthermore, this study outlines the significance of employing a probabilistic strategy in dynamical system regression, as opposed to traditional point-wise forecasting methods, and its potential applicability to real-world challenges like global climate modelling.

## 2. Related work

The intractability of many approaches to ML on very high-dimensional datasets has led to dimensionality reduction often being an integral part of any ML modelling framework (Ayesha et al., 2020). The recent advances in deep learning have led to the development and adoption of autoencoders as a flexible, non-parametric approach to dimensionality reduction (Tschannen et al., 2018; Sewak et al., 2020). Autoencoders offer increased scalability and complexity over well-known parametric approaches to dimensionality reduction such as principal component analysis (PCA).

Within the dynamical systems literature, the ROM is often employed similarly, aiming to provide low-dimensional descriptions of the underlying behaviour of a complex system (Loiseau et al., 2018). Historically, approaches to ROM were parametric and involved projecting data onto lower-dimensional linear subspaces, such as proper orthogonal decomposition (Lee and Carlberg, 2020). In recent years, data-driven non-parametric approaches to model reduction have been adopted within the dynamical systems literature, leveraging models such as convolutional autoencoders for a variety of applications such as fluid dynamics (Agostini, 2020; Lee and Carlberg, 2020; Otto and Rowley, 2019; Xu and Duraisamy, 2020; Daneshkhah et al., 2020).

Learning low-dimensional latent representations can facilitate fast regression, as an alternative to time-consuming numerical simulations. This type of ROM-enabled regression has been applied to a variety of problems such as energy forecasting (Mudunuru et al., 2017), fluid flow predictions (Wiewel et al., 2019; Daneshkhah et al., 2020) and wildfire forecasting (Cheng et al., 2022). However, the approaches outlined are flexible and generally not context-dependent, advanced regression models such as long short-term memory (LSTM) networks have been shown to effectively learn the latent features of general dynamical systems (Maulik et al., 2020).

Maulik et al. (2021) used convolutional autoencoders to produce latent features of a dynamical system governed by the Navier–Stokes equations. A GP-based regression model was then developed to learn the temporal evolution of the latent space features. The data was applied to a single task, i.e., a single-channel ($1{\times}64{\times}64$) tensor, regression problem, and found that the model was able to accurately reproduce

the underlying data. However, this is a toy example and does not reflect representative real-world problems and does not demonstrate the scalability of the proposed method. Furthermore, the proposed methodology is not fully utilised by quantifying original feature space after reconstructing the latent representations. Similarly, Guo and Hesthaven (2018) uses GPs for latent space modelling in structural analysis, however, the authors relied on existing parametric model-reduction techniques rather than more flexible data-driven approaches.

Quilodrán-Casas and Arcucci (2023) propose a sophisticated data-driven approach to the development of a surrogate model to forecast latent space dynamics for 3D unstructured computational fluid dynamics (CFD) simulations. The authors construct a ROM by utilising both PCA and a data-driven autoencoder to produce latent features from their high-dimensional simulation data before training an LSTM one-step-ahead regression model on their latent features. The application to unstructured 3D domains is novel and the results highlighted the robustness of the proposed approach.

PCA involves a linear transformation of the original data onto a new basis. A fully-connected dense autoencoder with a linear activation function is capable of retrieving principal components when trained to minimise the squared distance between the original data and its reconstruction. Therefore, the two-step encoding process can be reinterpreted as including an additional pre-trained fully-connected layer, parameterised by a weight vector $W_0^*$, located at the beginning of the autoencoder. Although PCA is a widely used technique for dimensionality reduction, it does not easily extend to the case of multi-channel data and can suffer from computational costs in very high-dimensional spaces. Additionally, compared to a GP model, an LSTM does not easily extend to probabilistic inference, which can lead to a lack of consideration for uncertainty in the predictions.

There are many examples of attempts to forecast environmental and climatic variables using ML and neural network-based approaches. Tran et al. (2021) outline a review of the existing literature for using neural networks to forecast air temperature. The studies outlined utilise a variety of neural network architectures, mostly feed-forward networks, and LSTM models, however, in more recent years CNNs have become increasingly adopted for spatiotemporal forecasts. Wang et al. (2021) similarly outlines a review of neural network-based approaches to forecasting wind speed, and consequently wind power, and found growing adoption of CNNs and autoencoders for feature extraction.

Zhang and Dong (2020) explore a convolutional recurrent neural network (CRNN) model for a temperature forecasting case study in China. By discretising the domain into a $32 \times 62$ matrix the authors produce a neural network operating on $4 \times 32 \times 64$ tensors by aggregating historical data to produce the next temperature map in the sequence. The authors manage to obtain a Root Mean Squared Error (RMSE) of $1.7°C$ for their proposed model.

Similarly, Kreuzer et al. (2020) propose convolutional LSTM for temperature forecasting in Germany and obtain an RMSE of $2.1°C$, however, this study only forecasted measurements at five weather stations rather than forecasting a full grid of values across a spatial domain. However, these studies explore deterministic models producing point-wise estimates and forecasts making no considerations for predictive uncertainty. Moreover, these studies focus on the forecasting of single-level measurements and the approaches outlined do not attempt multi-task regression with robust quantification of uncertainty.

Despite some successful applications thus far of data-driven approaches to ROM used in conjunction with advanced regression models like GPs, the existing literature fails to tackle a truly high-dimensional and multi-task (i.e., predicting multiple spatially varying measurements) problem, generally modelling relatively small-scale data with shallow autoencoders. Existing neural network-based approaches to forecasting environmental and climate data generally make no consideration for uncertainty and aim to forecast single levels over a 1D grid-based output or for discrete locations aggregated into vectors.

The existing literature does not tackle scenarios of high-dimensional multi-task data with robust consideration of predictive uncertainty. Furthermore, while GP models have been adopted, the existing literature fails to focus on the probabilistic aspects of GP regression. The existing literature primarily emphasises GP as a flexible non-parametric regression model, but it fails to acknowledge the benefits of generating full predictive posterior distributions or quantifying uncertainty in complex regression problems.

## 3. Methods

### 3.1. Probabilistic tensor regression

Tensor data is ubiquitous in scientific and engineering applications, e.g., a single-channel 2D tensor describing a scalar field for some spatially varying measurement, or a three-channel tensor representing RGB images. This study is concerned with time-indexed tensors, $\mathbf{R}^{(t)} \in \mathbb{R}^{C \times H \times W}$, describing the state of some dynamical system at time $t$, i.e. at each point in time there are $C$ variables being measured, each varying over a rectangular $H \times W$ grid. Given a collection of time-indexed measurements, $\{\mathbf{R}^{(t)}\}_{t=1}^T$, the aim is to construct a probabilistic regression model,

$$\hat{\mathbf{R}}^{(t+1)} = f(\mathbf{R}^{(t)}) \tag{1}$$

such that given the state of the system at time $t$, a probabilistic forecast for the state of the system at time $t+1$ can be made using some model, $f$. The regression model will be constructed in order to provide both point-wise estimates of the system at time $t+1$, $\mathbb{E}[\hat{\mathbf{R}}^{(t+1)}]$, as well as a measure of predictive uncertainty, $\mathbb{V}[\hat{\mathbf{R}}^{(t+1)}]$.

To construct the probabilistic regression model, $f$, a GP is utilised. A GP is a stochastic process defined as a collection of random variables, any finite number of which have a joint Gaussian distribution (Williams and Rasmussen, 2006). A GP can be considered as a generalisation of the multivariate Gaussian distribution, retaining many convenient mathematical properties of the multivariate Gaussian distribution (Donnelly et al., 2022). However, instead of being defined over finite-length vectors and parameterised by a mean vector and covariance matrix, a GP is defined over functions and parameterised by mean and covariance functions. A GP is therefore a (possibly infinite) distribution that is fully specified by its mean function $m(\mathbf{x}^{(i)})$ and its kernel-covariance function $k(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$.

GPs are a robust Bayesian regression model, allowing for straightforward inferences about predictive uncertainty without expensive sampling techniques. However, GPs are limited in their scalability and flexibility. For standard regression problems, GPs can only learn functions of the form $f : \mathbb{R}^p \rightarrow \mathbb{R}$. i.e., a scalar output, $y^{(t+1)}$, is regressed onto a vector valued input, $\mathbf{x}^{(t)}$,

$$y^{(t+1)} = f(\mathbf{x}^{(t)}) + \epsilon^{(t)}, \tag{2}$$

where, $f$ is a GP and $\epsilon^{(t)}$ is additive Gaussian noise. However, in GP regression, instead of producing a deterministic prediction, $f(\mathbf{x}^{(t)})$, a full predictive posterior distribution over the output is constructed,

$$p(y^{(t+1)}|\mathbf{x}^{(t)}) = N(\mu_*^{(t+1)}, \sigma_*^{(t+1)}), \tag{3}$$

where $\mu_*^{(t)}$ and $\sigma_*^{(t)}$ are the predictive mean and standard deviation, respectively. Samples can be drawn from this posterior. However, the expected value and predictive uncertainty are determined analytically without sampling where, $\mathbb{E}[y^{(t+1)}] = \mu_*^{(t+1)}$ and $\mathbb{V}[y^{(t+1)}] = \sigma_*^{(t+1)}$.

The regression problem outlined in Eq. (1) cannot be solved using standard GPs as these are three-dimensional tensors, not scalars. However, GPs can be extended to handle vector-to-vector regression problems. This can be achieved by modelling each output independently using a standard GP or through multi-task models using approaches such as the Linear Model of Coregionalization (Alvarez et al., 2012). Other methods (Liu et al., 2018) also exist, which generally model the
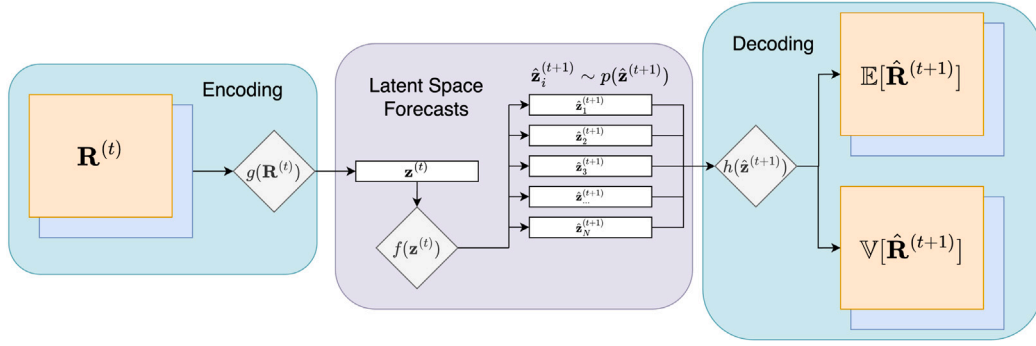
**Fig. 1.** Illustration of the GP-autoencoder, showing how sampling in the latent space using a GP propagates uncertainty through to the original feature space.

problem as a standard GP but use structured kernel matrices to handle multiple tasks and specify inter-task dependencies.

GPs suffer from high computational complexity costs, scaling cubically, $O(N^3)$, with the number of training instances, $N$, due to the inversion of a $N \times N$ matrix during training (Williams and Rasmussen, 2006; Chatrabgoun et al., 2022). Modelling $d$ tasks with an exact multitask GP similarly scale with $O(d^3 N^3)$ complexity. While approximate GPs using a subset of inducing points, thereby reducing the size of the matrix being inverted, have been demonstrated effectively (Liu et al., 2020; Foreman-Mackey et al., 2017), a robust probabilistic model that demonstrates scalability like neural networks does not exist.

To overcome these limitations, the underlying data, $\{\mathbf{R}^{(t)}\}_{t=1}^{T}$, can be modified to make it amenable for GP regression. By producing a sophisticated encoder, $g$, and corresponding decoder, $h$, to transform the original tensors, $\mathbf{R}^{(t)}$, into latent vectors, $\mathbf{z}^{(t)}$, with minimal reconstruction error, $L(\hat{\mathbf{R}}^{(t)}, \mathbf{R}^{(t)})$,

$$\mathbf{z}^{(t)} = g(\mathbf{R}^{(t)})$$
$$\hat{\mathbf{R}}^{(t)} = h(\mathbf{z}^{(t)}) \tag{4}$$

the regression problem can be reinterpreted as,

$$\mathbf{z}^{(t+1)} = f(\mathbf{z}^{(t)}) + \epsilon^{(t)}. \tag{5}$$

where $f$ is a multi-output GP model, and $d$, where $\mathbf{z}^{(t)} \in \mathbb{R}^d$, is small enough to build such a model. Using a GP, a predictive posterior can be constructed over the latent features, $p(\mathbf{z}^{(t+1)}|\mathbf{z}^{(t)})$, as shown in Eq. (3).

By sampling in the latent space, producing a collection of realisations of the latent target, $\mathbf{z}_i^{(t+1)}$, uncertainty can be propagated back into the original tensor space,

$$\mathbb{E}[\hat{\mathbf{R}}^{(t+1)}] = \mathbb{E}[h(\mathbf{z}_i^{(t+1)})] \tag{6}$$

$$\mathbb{V}[\hat{\mathbf{R}}^{(t+1)}] = \mathbb{V}[h(\mathbf{z}_i^{(t+1)})] \tag{7}$$

producing robust point-wise forecasts (Eq. (6)) and estimates of predictive uncertainty (Eq. (7)) for each element in the tensor simultaneously.

A robust and accurate encoder–decoder scheme that generates non-discontinuous features in the latent $z$-space with minimal reconstruction error could enable a GP regression model to accurately reproduce the system's behaviour in the encoded space. For an encoder–decoder to have sufficient capability to produce meaningful representations in the latent space from large tensors, $g$ and $h$ are likely to be complex, non-linear functions. The overall methodology for the study is outlined in Fig. 1, by sampling in the latent space and propagating samples through the decoder, point-wise forecasts and estimates of uncertainty can be propagated through to the original feature space.

### 3.2. Autoencoders

Autoencoders are a class of neural network consisting of an encoder sub-network, $g$, and a decoder sub-network, $h$, which take an arbitrary input $\mathbf{X}$ and maps back to an output $\hat{\mathbf{X}}$ with the same shape

as the input. Autoencoders generally contain an under-parameterised bottleneck such that the encoder sub-network produces an intermediate output, $\mathbf{z}$, smaller than the original input. Using self-supervised training, where the input and the output are the same, and a distance-based loss function, $L$, the network optimises its weights to minimise the reconstruction error between the input and the reconstructed output, $L(\hat{\mathbf{X}}, \mathbf{X})$. Since the bottleneck of the network is under-parameterised, the network simultaneously learns an efficient encoding scheme for the original data, $\mathbf{z} = h(\mathbf{X})$, and decoder for the latent features, $\hat{\mathbf{X}} = g(\mathbf{z})$, as previously expressed in Eq. (4).

Neural networks can learn non-parametric, non-linear functions, which means that both the encoder and decoder sub-networks can be infinitely expressive and capable of learning complex patterns. By constructing highly complex networks and leveraging large datasets to optimise the weights, autoencoders can be a more powerful dimensionality reduction tool than parametric approaches such as PCA. Furthermore, neural network architectures, such as CNN are well-suited for to handling non-linear data with multiple channels in the input and output making them ideal for flexible tensor-based compression. An example of two-channel convolutional autoencoder is illustrated in Fig. 2.

To construct highly flexible and powerful autoencoders, existing state-of-the-art CNN classification architectures were used as the basis of the network. The ResNet50 (illustrated in Fig. 3 model outlined in He et al. (2016) is a 150-layer CNN designed to classify RGB images into distinct categories. The architecture obtained SOTA performance on ImageNet (Krizhevsky et al., 2017) and CIFAR-10 (Krizhevsky et al., 2009). In this architecture, multi-channel tensors are down-sampled through convolutions and pooling layers to finally produce classification probabilities in a vector-based output. The network can be modified such that the classification layer becomes the bottleneck of the autoencoder with a fixed size. In addition to the original ResNet50 architecture, smaller versions of ResNet, such as ResNet18 and ResNet34, which were inspired by the original architecture, were also tested to compare their performance.

### 3.3. Data

The proposed predictive model is developed for the case study of predicting temperature and pressure data recorded at 3-hour intervals on a global scale. The dataset consists of the full record of simulated hindcast data from the ERA5 reanalysis (Hersbach et al., 2020) for the year 2022. This meant the proposed regression model involved simultaneously forecasting global temperature and pressure values for 3-hours ahead based on the current values. The spatial resolution of the data is $1° \times 1°$, therefore for each time-step, the data is described by tensors $\mathbf{R}^{(t)} \in \mathbb{R}^{2 \times 180 \times 360}$ and the full dataset consisted of approximately 2900 samples. This resulted in an aggregated dataset $\mathbf{X} \in \mathbb{R}^{2900 \times 2 \times 180 \times 360}$. Examples of the original data are illustrated in Fig. 4, showing the transition of both temperature and pressure across 3 time steps. This data also highlights the vastly different scales of the two measurements, varying in units of measurement and the magnitude of the average values.
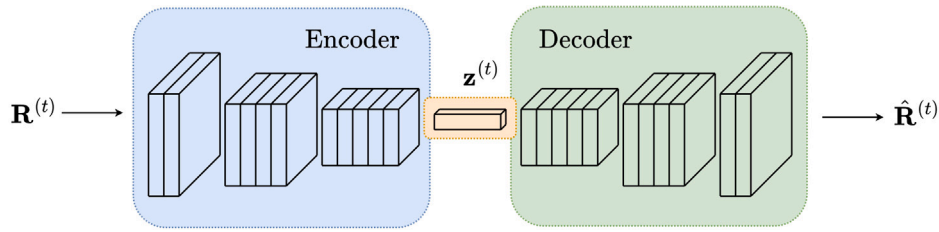
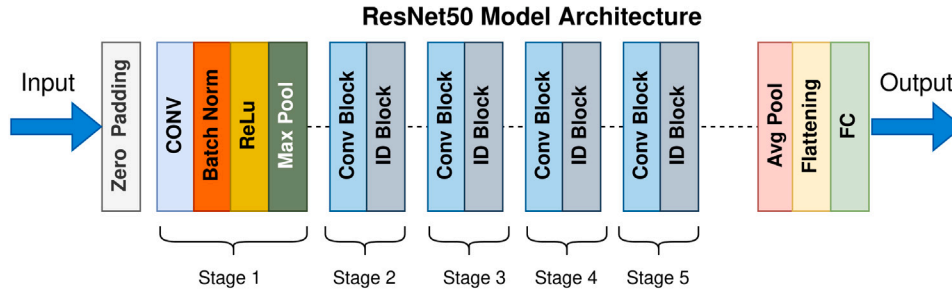**Fig. 2.** Example of an under-parameterised convolutional autoencoder architecture.



**Fig. 3.** ResNet50 architecture
*Source:* Figure adopted from Gorlapraveen123: Wikimedia Commons (2021).
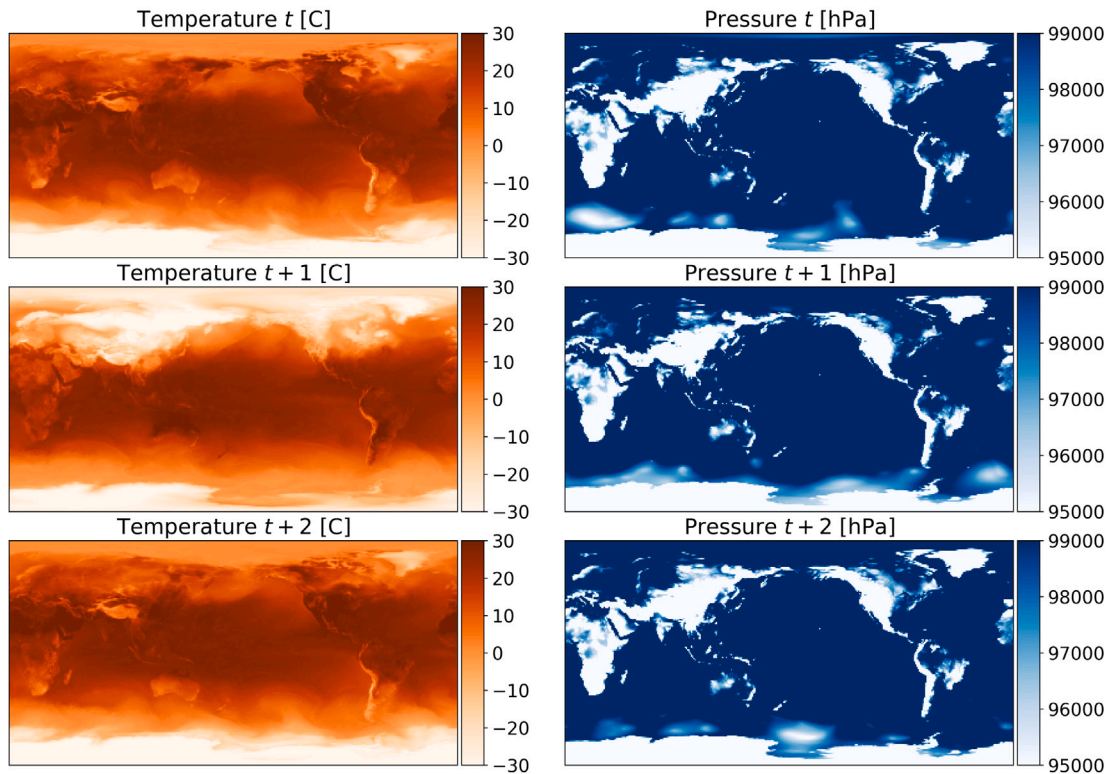


**Fig. 4.** Un-transformed global temperature and pressure data for 2 temporal transitions.

## 4. Results

### 4.1. Experimental setup

In this study three autoencoders were trained and assessed, with the encoder's sub-network based on the existing ResNet18, ResNet34, and ResNet50 object-classification architectures with the output layers modified to produce continuous latent vectors rather than discrete classifications. In order to determine the dimensionality of the bottleneck and latent vectors, a trade-off between the expressiveness of the latent representations and the complexity of the GP model needs to be balanced. Balancing this trade-off along with preliminary testing determined a latent dimension size of 12 features such that $\mathbf{z}^t \in \mathbb{R}^{12}$. The training cost and computational resources required to train the autoencoder are not sensitive to the size of $\mathbf{z}^t$, the GP is highly sensitive to this choice. The computational complexity of the multi-task GP model is $O(d^3 N^3)$ where $d$ is the number of features and $N$ is the
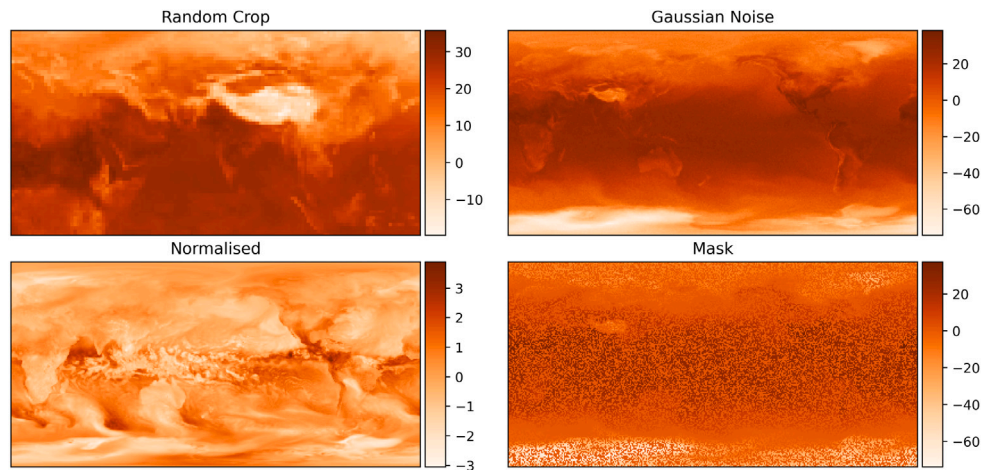
**Fig. 5.** Examples of augmented temperature data.

number of instances, so as the latent dimension $d$ grows, the GP model can quickly become prohibitively expensive.

In order to construct robust autoencoders that can effectively capture the intricate features of the underlying data and generalise well, it is necessary for the network to possess adequate flexibility. Additionally, the application of data augmentation and pre-processing strategies during training has been shown to be crucial for achieving good performance. The data were normalised to have zero mean and unit variance across each feature, then a series of augmentations were applied randomly during the training process. These augmentations consisted of crops and resize (i.e., randomly cropping then re-scaling back to the original shape), alternating the order of the channels (i.e., swapping the order of temperature and pressure data), applying a mask (i.e., randomly zeroing out cells in the data) and randomly applying Gaussian noise to cells. Examples of some of these augmentation strategies applied independently can be seen in Fig. 5.

The training of the autoencoders was parallelised across 6 Nvidia Tesla K80 GPUs using a batch size of 32, Adam optimiser and initial learning rate of $1e^{-4}$. Training was conducted for 100 epochs, or until convergence was reached. After producing the latent feature dataset from the trained autoencoder, a multi-task GP parameterised by a multi-task Matern$_{3/2}$ kernel function and a constant mean function was trained. The GP training, i.e., tuning of the hyperparameters of the kernel function and additional noise parameters, was performed through minimising the negative marginal log likelihood (Williams and Rasmussen, 2006).

Additionally, a second phase of GP training was initiated whereby the noise hyperparameter of the GP likelihood, i.e., the parameter determining how much the true target is corrupted by noise as illustrated in Eq. (1), is modified to fine-tune the level of predictive uncertainty. Purely data-driven optimisation of these parameters can overfit the data and produce values of $\mathbb{V}[\mathbf{z}^{(t+1)}]$ that are far too low in practice. However, the influence of the noise parameter on the value of $\mathbb{E}[\mathbf{z}^{(t+1)}]$ is minimal and therefore does not significantly change the performance of point-wise predictive performance.

### 4.2. Autoencoder

The performance of the ResNet-based autoencoders can be seen in Table 1. Increasing the network's capacity resulted in a consistent decrease in the RMSE for temperature predictions, but it did not necessarily lead to an improvement in predictive performance for pressure measurements. ResNet-34 is likely to be the right balance between good predictive performance, while not being too highly over-parameterised. Training and inference costs also scale with size of the model and therefore again, ResNet-34 seemed the best architecture
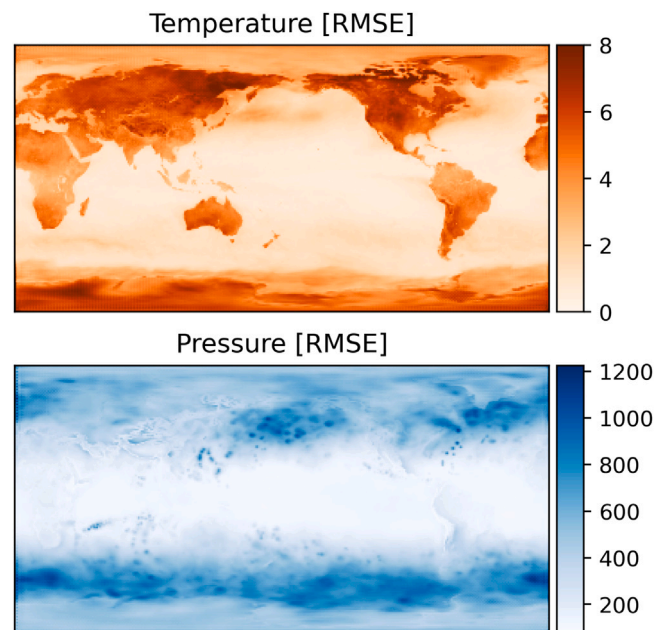


**Fig. 6.** Reconstruction RMSE for temperature and pressure on a cell-by-cell basis across the domain using the ResNet-34 based autoencoder.

overall out of those tested (ResNet-18, ResNet-34 and ResNet-50) in this study because of this trade-off between accuracy and computational resources. Furthermore, ResNet-34 obtained the best predictive score for the pressure measurements out of all the models. An RMSE of 2.79 for temperature predictions on the test set means that on average, the predicted value in each cell is incorrect by 2.79°C, and likewise, for pressure, each cell is incorrect by, on average, 434 hPa.

In Fig. 6, the reconstruction error on the test set can be seen for the ResNet-34 autoencoder. The plot shows the RMSE for each cell in the domain across all time-steps in the test set. In spite of the aggregate RMSE value being 2.79°C, it is observed that much of the domain, especially those locations on land, have a higher average loss. Similarly, the values over the oceans tend to have a lower average loss. This outcome can be explained by Fig. 7, which shows the standard deviation of the test set values, highlighting that the areas of the domain with the most variability are the same areas where the prediction error is highest.

In Fig. 8, the predictive performance for a collection of test samples is illustrated by showing the $L_1$ loss on the temperature and pressure data. As can be seen across both variables, the reconstruction error
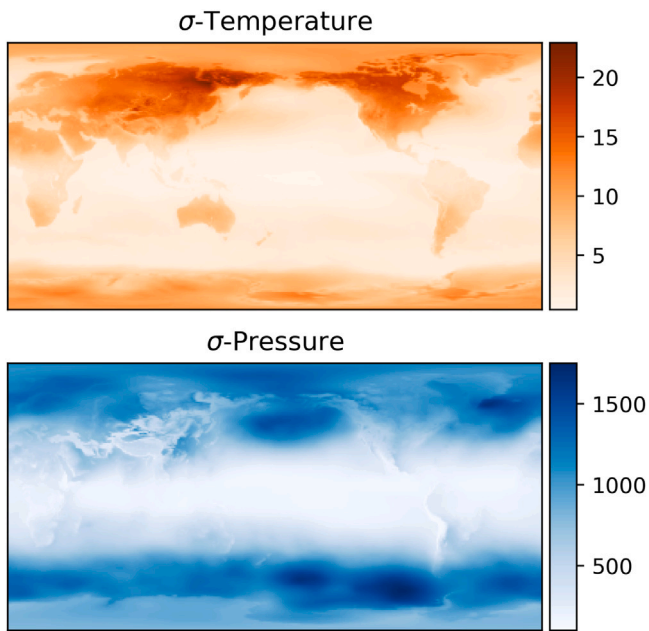
**Fig. 7.** Cell-by-cell standard deviation over the test set samples.

**Table 1**
Validation performance of different popular CNN-architecture autoencoder backbones with ResNet-18 having fewest parameters and ResNet-50 the most.

| Architecture | Temp. (RMSE) | Pressure (RMSE) |
|---|---|---|
| ResNet-18 | 3.04 | 482 |
| ResNet-34 | 2.79 | 434 |
| ResNet-50 | 2.66 | 533 |

### 4.3. GP-autoencoder

After having trained the autoencoder, the latent feature data was used to train the GP regression model. Regression performance of the GP model in the latent space is illustrated in Fig. 9 along with the 2 standard deviation predictive intervals. An accurate model would expect to see the true value within these intervals for 95% of the observations. The autoencoder resulted in non-discontinuous and smooth latent features facilitating very high latent space regression accuracy. Having trained both the autoencoder and GP regression model, the composite GP-autoencoder model can be used for inference by producing point-wise forecasts for the true tensor, $\mathbb{E}[\hat{\mathbf{R}}^{(t+1)}]$, and a tensor of equal shape, $\mathbb{V}[\hat{\mathbf{R}}^{(t+1)}]$, outlining predictive uncertainty associated with each element where $\mathbb{V}[\hat{\mathbf{R}}^{(t+1)}]_{ijk} = \sigma_{ijk}^{(t+1)}$.

The results of the GP-autoencoder performance on the test set can be seen in Table 2. The full GP-autoencoder model increases the average loss by $1.03°C$ for the temperature predictions, and by 234 hPa for the pressure predictions compared to the baseline reconstruction error for the autoencoder alone. When compared to the RMSE values of $1.7°C$ and $2.1°C$ for temperature predictions in the studies by Zhang and Dong (2020) and Kreuzer et al. (2020), respectively, the RMSE of $3.82°C$ achieved is lower in point-wise accuracy. However, it is important to note that these are not like-for-like comparisons. The proposed model performed in significantly higher dimensions, conducts simultaneous multi-task forecasting, and provides predictive uncertainty estimates.

is low generally, with small areas where the reconstruction error is significantly higher. For instance, in sample 1, while the majority of the domain has a reconstruction error below 2, certain areas in North America exhibit a significantly higher reconstruction error, ranging from 8 and above. This is a common result of data-driven modelling in which predictive performance is lower when results deviate most strongly from their training set means.
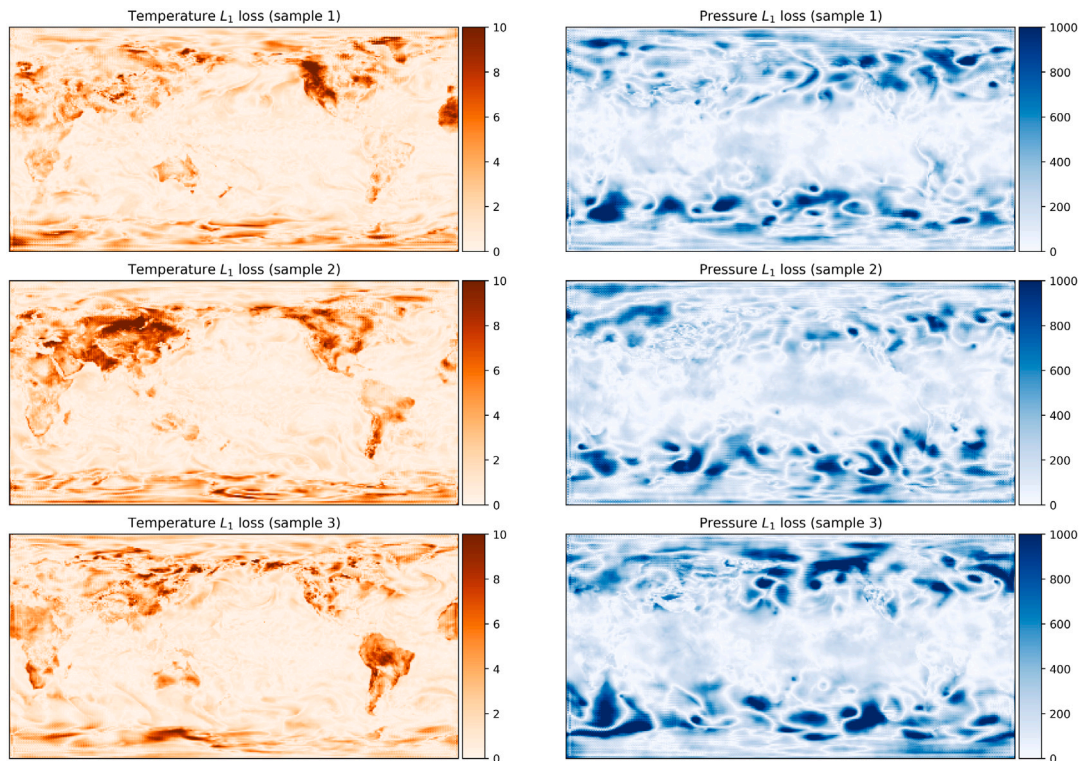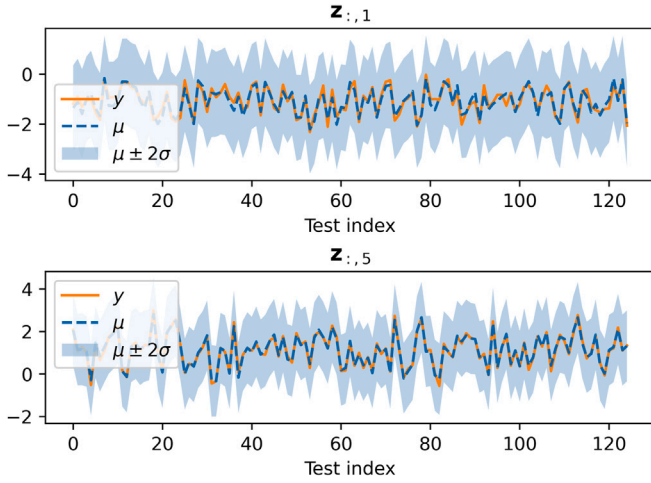


**Fig. 8.** $L_1$ loss for samples in the test set.

**Table 2**

Validation performance of the composite GP-autoencoder regression model.

| Temp. (RMSE) | Pressure (RMSE) |
|---|---|
| 3.82 | 638 |



**Fig. 9.** Test performance in the latent space for the first and fifth component of the set of vectors $\{\mathbf{z}^{(t+1)}\}$ showing the mean prediction and 2 s.d. predictive uncertainty.

Considering these caveats, our model illustrates notable performance, achieving a high level of accuracy.

Furthermore, in order to assess the robustness of the estimated predictive posterior distribution a custom probabilistic validation metric was proposed. This metric involved, for each sample in the test data, counting how many of the true individual tensor elements, $\mathbf{R}_{ijk}^{(t+1)}$, were within the plausible range of the posterior distribution. This was done by counting the number of tensor elements within 3 standard deviations of the mean of the predictive posterior,

$$\sum_i \sum_j \sum_k \mathbb{I}_{\{3\sigma_{ijk}^{(t+1)}\}} \tag{8}$$

where,

$$\mathbb{I}_{\{3\sigma_{ijk}^{(t+1)}\}} = \begin{cases} 1, & \text{if } \mathbf{R}_{ijk}^{(t+1)} \in \left[ \mathbb{E}[\hat{\mathbf{R}}^{(t+1)}]_{ijk} - 3\sigma_{ijk}^{(t+1)}, \ \mathbb{E}[\hat{\mathbf{R}}^{(t+1)}]_{ijk} + 3\sigma_{ijk}^{(t+1)} \right] \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

On average it was found that 95.6% of the true values were within this 3 s.d. range. For a fully accurate Gaussian probabilistic model it would be expected that approximately 99% of the values would fall within the 3 s.d. range. However, this importantly highlights that even when the point-wise estimate is not close to the target value, the true target value has a 95% chance of being within the range of the predictive posterior. This underscores the significance and practical value of incorporating uncertainty quantification in complex regression models.

Examples of some of the best and worst performance with respect to this metric can be seen in Figs. 10–11, illustrating predictions where the true value falls within the predictive posterior distribution (3 s.d.). The cells marked in red represent cells where the respective forecast did not
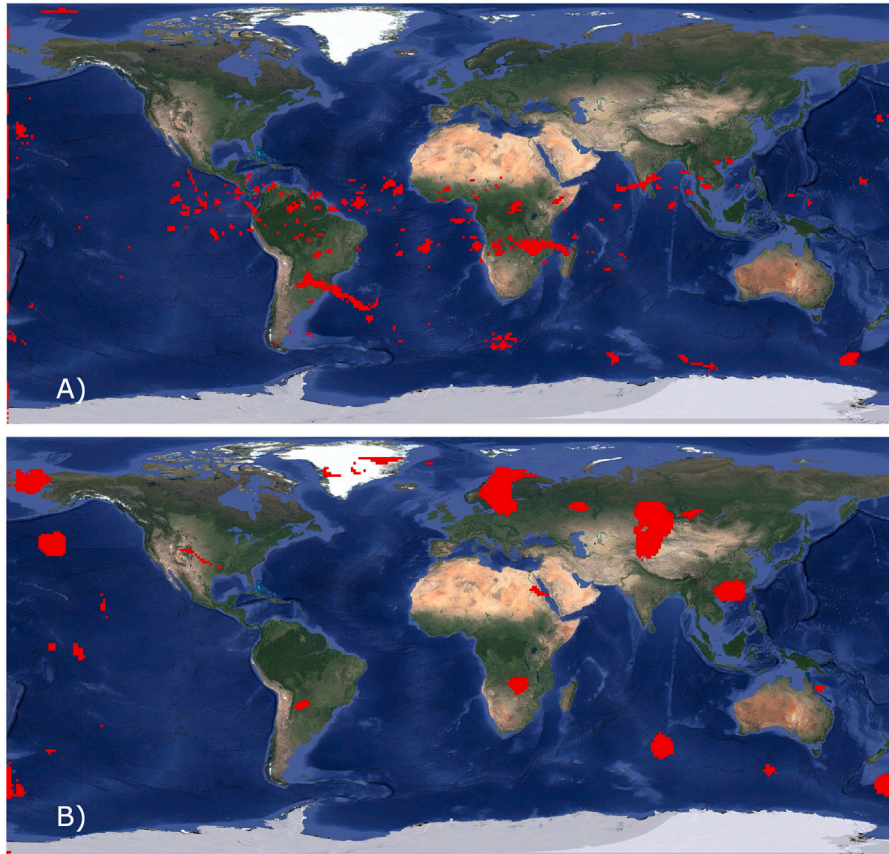


**Fig. 10.** Test predictions where (A) 99% of the cells fall within the 3 s.d. range for the temperature forecast and (B) 98% of the cells fall within the 3 s.d. range for pressure. The red cells represent those not within the range.

**Fig. 11.** Test predictions where (A) 82% of the cells fall within the 3 s.d. range for the temperature forecast and (B) 93% of the cells fall within the 3 s.d. range for pressure. The red cells represent those not within the range.
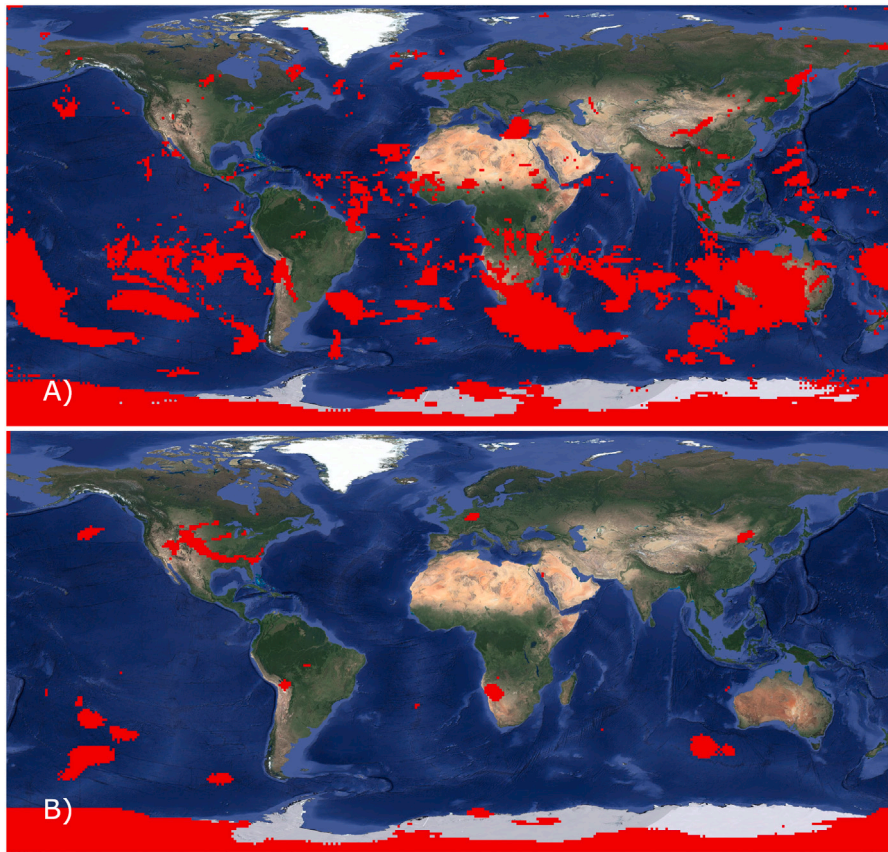
fall in the range for A) temperature and (B) pressure. Poor performance again results from true values that deviate mostly strong from their training set values. The average rate of 95.6% of cells falling within the reasonable range of the predictive posterior suggests that the model generally performs very well with respect to this metric. However, to handle more extreme cases additional probabilistic components could be considered in the model, such as a probabilistic encoding scheme through the use of variational autoencoders.

In Fig. 12, the average RMSE for both temperature and pressure is illustrated alongside the average predictive uncertainty (one standard deviation) across the test data. This figure shows that the model generates mostly accurate point-wise estimates but also high-resolution dynamic estimates of predictive uncertainty. The figure only outlines the predictive uncertainty averaged across test samples. However, these estimates of predictive uncertainty vary from sample to sample are not constant. Fig. 7 showed that there is low variation in ocean conditions around the equator and as such the model produces very low estimates of predictive uncertainty, meaning point-wise forecasts can be interpreted with a high degree of confidence. Similarly, in areas where there is significant variation, a robust approach to forecasting would be to observe the full range of possible values and give less credibility to a single point-wise estimate. In the forecasting of complex systems, it is better to be imprecisely right than precisely wrong. Furthermore, by using a more diverse sample of training data the model could easily be improved with respect to predictive performance. However, for demonstrative purposes, only data from 2022 was used to train and validate the model.

## 5. Limitations

The proposed methodology has been demonstrated highly effective for the case study of forecasting temperature and pressure on a global scale. This is a highly versatile approach to forecasting problems involving high-dimensional tensor-based data with clear applications beyond the case study outlined here. This methodology relies on autoencoders for data compression and GPs for learning the latent-space dynamics. Both of these models are highly flexible; autoencoders can be made with arbitrary encoder/decoder sub-networks and GPs can be parameterised with a variety of kernel and mean functions.

A limitation of the proposed method, the individual autoencoder, and GP models is that the data should exhibit some measure of smoothness, or facilitate a smooth representation through pre-processing. In the case study outlined here, this would mean that it should be expected that values within a pixel neighbourhood in the tensors should be similar and that the changes in the value of a pixel between time steps should not be very large and discontinuous. Dealing with highly nonlinear and discontinuous data has always been known as a limitation of GP models which later influenced the development of Deep Gaussian Process models (Damianou and Lawrence, 2013).

Autoencoders are sensitive to the size of the latent dimension and as the size of the latent dimension converges to the size of the input, the reconstruction error will converge to zero. Conversely, the greater the compression the more information can be expected to be lost. Similarly, depending on the choice of kernel function, GPs will result in a smooth interpolation between points in the output space which can result in poor performance on outlying and extreme data points. Given these limitations, their composition into a single model could result in a prohibitively high amount of information loss from the original data if poorly constructed.

An existing limitation of GP regression models is the high computational complexity during training, with multi-task GP models scaling with $O(d^3 N^3)$ computational complexity for $d$ tasks/output features and $N$ instances. While autoencoders are highly scalable and can scale
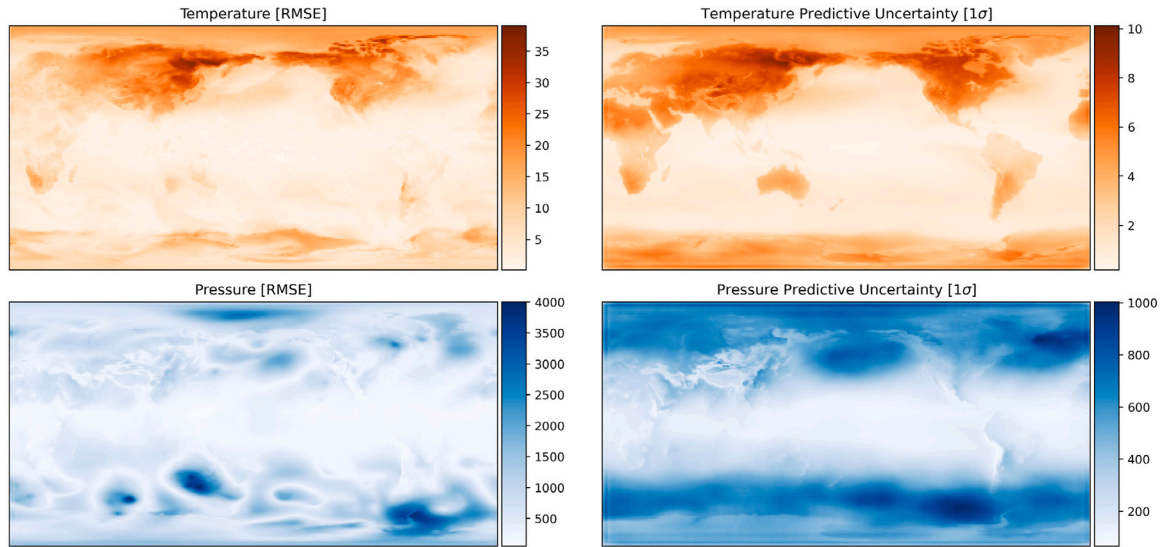
**Fig. 12.** Average RMSE for temperature and pressure predictions alongside average predictive uncertainty.

to arbitrarily large $N$, the model will be bottlenecked by the complexity of the latent-space GP model. Therefore, careful consideration of training data is required to obtain an optimal ratio of information-data through the selection of the $N$ training instances.

## 6. Conclusions

Multi-task spatially-varying regression is a challenging task in data-driven modelling, with models needing to be sufficiently flexible to capture the underlying patterns while also not suffering from high computational costs and the curse of dimensionality. However, many natural and artificial dynamical systems generate this data in abundance through improvements in numerical simulation, sensor technology or the proliferation of satellite data. Forecasting the behaviour of these systems is important in many scientific and engineering disciplines, as well as crucial to modern life generally.

To make robust forecasts and sophisticated inferences about the future behaviour of these systems, an appropriate consideration must be made for the uncertainty that can arise from the inherently chaotic nature of many of these systems. Therefore, to produce effective regression models for the behaviour of these systems, the models need to generalise well to high-dimensional spatiotemporal data and quantify the uncertainty associated with the forecasts. The existing literature and ML models fail to achieve this for representative problems in the real-world. Developing models with these capabilities is still an ongoing challenge in the existing literature due to clear challenges of tractability and computational complexity associated with expensive sampling procedures or Bayesian approaches.

By utilising SOTA approaches to deep learning and probabilistic inference, a flexible and robust non-parametric approach to probabilistic tensor-based regression has been outlined. Using advanced CNN-based architectures from computer vision applications as the basis of autoencoder networks, it is shown possible to accurately reconstruct complex multi-channel tensor data from very small latent vectors. Further, the latent vectors produced were smooth, non-discontinuous, and amenable to exact multi-task GP regression.

The proposed GP-autoencoder model has been applied to the case study of simultaneously forecasting one-step-ahead global temperature and pressure observations, significantly important high-level underlying drivers of global climate behaviour. Developing an ML-based model that facilitates near-instantaneous is important to reduce the reliance on expensive numerical simulation and consequently allows for real-time forecasting where simulation would not. This would be crucial to the development of early warning systems, nowcasting, and parameterise the inputs of smaller regional simulation models. Furthermore, by including dynamic estimates of predictive uncertainty the forecasts would allow for a more sophisticated and nuanced interpretation than a deterministic point-wise estimate. This can be particularly valuable in complex regression problems, where uncertainty quantification is crucial for informed decision making.

The model was demonstrated to have good accuracy, obtaining an average error of $3.82°C$ and 638 hPa for the temperature and pressure observations, respectively. Additionally, the model produced well-calibrated predictive posterior distributions, with the posterior distribution on average containing the true observation 95% of the time. This aspect of the model demonstrates that even when the point-wise error was above average there was a strong chance that the true observation falls within the predicted distribution. Contrasting these results with the existing SOTA outlined in Section 2 showed the proposed model to have a slightly higher error for temperature forecasting. However, the problem tackled in this study uses data order of magnitudes larger, forecasts multiple variables simultaneously, and provides probabilistic predictions. These contributions while still obtaining accurate and comparable forecasts, demonstrate the contributions outlined in this study.

A commonality of all data-driven forecasting is that performance was poorest when the data being predicted was more anomalous and less similar to the examples within the training data. For demonstrative purposes in this study, the data used was limited only to 2022, which inevitably leads to lower performance in some of the test cases as only a fraction of the true range of outputs is observed. The proposed model's predictive power could be readily enhanced through the utilisation of a more extensive and sophisticated training dataset. However, these limitations are intrinsic to data-driven modelling, and are to some extent irreducible since finite training data is unlikely to ever fully capture the behaviour of a dynamic system. Probabilistic modelling can effectively mitigate many of these problems by offering dynamic and robust estimates of predictive uncertainty, thereby eliminating the issues associated with producing point-wise estimates on previously unseen data.

The novel methodological approach outlined in this study can readily be repurposed to a variety of complex modelling problems involving

time-indexed tensor data. Based on the results outlined, the authors believe that using a highly-flexible autoencoder and a GP regression model is the best existing approach to forecasting high-dimensional multi-channel data in a robust and probabilistic manner.

## CRediT authorship contribution statement

**James Donnelly:** Conceptualization, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. **Alireza Daneshkhah:** Supervision, Writing – review & editing. **Soroush Abolfathi:** Supervision, Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The data used is from publicly available sources cited in the manuscript.

## Acknowledgement

## References

Agostini, L., 2020. Exploration and prediction of fluid dynamical systems using auto-encoder technology. Phys. Fluids 32 (6), 067103.

Alvarez, M.A., Rosasco, L., Lawrence, N.D., et al., 2012. Kernels for vector-valued functions: A review. Found. Trends® Mach. Learn. 4 (3), 195–266.

Ayesha, S., Hanif, M.K., Talib, R., 2020. Overview and comparative study of dimensionality reduction techniques for high dimensional data. Inf. Fusion 59, 44–58.

Bates, N., Ghatikar, G., Abdulla, G., Koenig, G.A., Bhalachandra, S., Sheikhalishahi, M., Patki, T., Rountree, B., Poole, S., 2015. Electrical grid and supercomputing centers: An investigative analysis of emerging opportunities and challenges. Informatik-Spektrum 38, 111–127.

Bauer, P., Thorpe, A., Brunet, G., 2015. The quiet revolution of numerical weather prediction. Nature 525 (7567), 47–55.

Behmanesh, I., Moaveni, B., 2015. Probabilistic identification of simulated damage on the dowling hall footbridge through Bayesian finite element model updating. Struct. Control Health Monit. 22 (3), 463–483.

Chakraborty, S., Chowdhury, R., 2016. Modelling uncertainty in incompressible flow simulation using Galerkin based generalized ANOVA. Comput. Phys. Comm. 208, 73–91.

Chatrabgoun, O., Esmaeilbeigi, M., Cheraghi, M., Daneshkhah, A., 2022. Stable likelihood computation for machine learning of linear differential operators with Gaussian processes. Int. J. Uncertain. Quantif. 12 (3).

Cheng, S., Prentice, I.C., Huang, Y., Jin, Y., Guo, Y.-K., Arcucci, R., 2022. Data-driven surrogate model with latent data assimilation: Application to wildfire forecasting. J. Comput. Phys. 464, 111302.

Cockayne, J., Oates, C.J., Sullivan, T.J., Girolami, M., 2019. Bayesian probabilistic numerical methods. SIAM Rev. 61 (4), 756–789.

Damianou, A., Lawrence, N.D., 2013. Deep gaussian processes. In: Artificial Intelligence and Statistics. PMLR, pp. 207–215.

Daneshkhah, A., Chatrabgoun, O., Esmaeilbeigi, M., Sedighi, T., Abolfathi, S., 2020. Uncertainty quantification of Darcy flow through porous media using deep Gaussian process. arXiv preprint arXiv:2011.01647.

Donnelly, J., Abolfathi, S., Daneshkhah, A., 2023. A physics-informed neural network surrogate model for tidal simulations. ECCOMAS Proceedia 836–844.

Donnelly, J., Abolfathi, S., Pearson, J., Chatrabgoun, O., Daneshkhah, A., 2022. Gaussian process emulation of spatio-temporal outputs of a 2D inland flood model. Water Res. 225, 119100.

Eigel, M., Schneider, R., Trunschke, P., Wolf, S., 2019. Variational Monte Carlo—bridging concepts of machine learning and high-dimensional partial differential equations. Adv. Comput. Math. 45, 2503–2532.

Esmaeilbeigi, M., Chatrabgoun, O., Daneshkhah, A., Shafa, M., 2023. On the impact of prior distributions on efficiency of sparse Gaussian process regression. Eng. Comput. 39 (4), 2905–2925.

Fanous, M., Daneshkhah, A., Eden, J.M., Remesan, R., Palade, V., 2023a. Hydromorphodynamic modelling of mangroves imposed by tidal waves using finite element discontinuous Galerkin method. Coast. Eng. 182, 104303.

Fanous, M., Eden, J.M., Remesan, R., Daneshkhah, A., 2023b. Challenges and prospects of climate change impact assessment on mangrove environments through mathematical models. Environ. Model. Softw. 105658.

Foreman-Mackey, D., Agol, E., Ambikasaran, S., Angus, R., 2017. Fast and scalable Gaussian process modeling with applications to astronomical time series. Astron. J. 154 (6), 220.

Fringer, O.B., Dawson, C.N., He, R., Ralston, D.K., Zhang, Y.J., 2019. The future of coastal and estuarine modeling: Findings from a workshop. Ocean Model. 143, 101458.

Giraldo, F.X., Restelli, M., 2008. A study of spectral element and discontinuous Galerkin methods for the Navier–Stokes equations in nonhydrostatic mesoscale atmospheric modeling: Equation sets and test cases. J. Comput. Phys. 227 (8), 3849–3877.

Gorlapraveen123: Wikimedia Commons, 2021. ResNet50 - Wikimedia commons, the free media repository.

Guo, M., Hesthaven, J.S., 2018. Reduced order modeling for nonlinear structural analysis using Gaussian process regression. Comput. Methods Appl. Mech. Engrg. 341, 807–826.

Haukaas, T., Gardoni, P., 2011. Model uncertainty in finite-element analysis: Bayesian finite elements. J. Eng. Mech. 137 (8), 519–526.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146 (730), 1999–2049.

Kreuzer, D., Munz, M., Schlüter, S., 2020. Short-term temperature forecasts using a convolutional neural network—An application to different weather stations in Germany. Mach. Learn. Appl. 2, 100007.

Krizhevsky, A., Hinton, G., et al., 2009. Learning Multiple Layers of Features from Tiny Images. Toronto, ON, Canada.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. Imagenet classification with deep convolutional neural networks. Commun. ACM 60 (6), 84–90.

Lee, K., Carlberg, K.T., 2020. Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders. J. Comput. Phys. 404, 108973.

Liu, H., Cai, J., Ong, Y.-S., 2018. Remarks on multi-output Gaussian process regression. Knowl.-Based Syst. 144, 102–121.

Liu, H., Ong, Y.-S., Shen, X., Cai, J., 2020. When Gaussian process meets big data: A review of scalable GPs. IEEE Trans. Neural Netw. Learn. Syst. 31 (11), 4405–4423.

Loiseau, J.-C., Noack, B.R., Brunton, S.L., 2018. Sparse reduced-order modelling: Sensor-based dynamics to full-state estimation. J. Fluid Mech. 844, 459–490.

Lozanovski, B., Downing, D., Tran, P., Shidid, D., Qian, M., Choong, P., Brandt, M., Leary, M., 2020. A Monte Carlo simulation-based approach to realistic modelling of additively manufactured lattice structures. Addit. Manuf. 32, 101092.

Maulik, R., Botsas, T., Ramachandra, N., Mason, L.R., Pan, I., 2021. Latent-space time evolution of non-intrusive reduced-order models using Gaussian process emulation. Physica D 416, 132797.

Maulik, R., Mohan, A., Lusch, B., Madireddy, S., Balaprakash, P., Livescu, D., 2020. Time-series learning of latent-space dynamics for reduced-order model closure. Physica D 405, 132368.

Mudunuru, M.K., Karra, S., Harp, D.R., Guthrie, G., Viswanathan, H.S., 2017. Regression-based reduced-order models to predict transient thermal output for enhanced geothermal systems. Geothermics 70, 192–205.

Otto, S.E., Rowley, C.W., 2019. Linearly recurrent autoencoder networks for learning dynamics. SIAM J. Appl. Dyn. Syst. 18 (1), 558–593.

Portegies Zwart, S., 2020. The ecological impact of high-performance computing in astrophysics. Nat. Astron. 4 (9), 819–822.

Quilodrán-Casas, C., Arcucci, R., 2023. A data-driven adversarial machine learning for 3D surrogates of unstructured computational fluid dynamic simulations. Physica A 615, 128564.

Sewak, M., Sahay, S.K., Rathore, H., 2020. An overview of deep learning architecture of deep neural networks and autoencoders. J. Comput. Theor. Nanosci. 17 (1), 182–188.

Tran, T.T.K., Bateni, S.M., Ki, S.J., Vosoughifar, H., 2021. A review of neural networks for air temperature forecasting. Water 13 (9), 1294.

Tschannen, M., Bachem, O., Lucic, M., 2018. Recent advances in autoencoder-based representation learning. arXiv preprint arXiv:1812.05069.

Wang, Y., Zou, R., Liu, F., Zhang, L., Liu, Q., 2021. A review of wind speed and wind power forecasting with deep neural networks. Appl. Energy 304, 117766.

Wiewel, S., Becher, M., Thuerey, N., 2019. Latent space physics: Towards learning the temporal evolution of fluid flow. In: Computer Graphics Forum, Vol. 38. Wiley Online Library, pp. 71–82.

Wilkin, J., Rosenfeld, L., Allen, A., Baltes, R., Baptista, A., He, R., Hogan, P., Kurapov, A., Mehra, A., Quintrell, J., et al., 2017. Advancing coastal ocean modelling, analysis, and prediction for the US integrated ocean observing system. J. Oper. Oceanogr. 10 (2), 115–126.

Williams, C.K., Rasmussen, C.E., 2006. Gaussian Processes for Machine Learning, Vol. 2. MIT press Cambridge, MA.

Worley, P.H., Mirin, A.A., Craig, A.P., Taylor, M.A., Dennis, J.M., Vertenstein, M., 2011. Performance of the community earth system model. In: Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis. pp. 1–11.

Xiu, D., 2009. Fast numerical methods for stochastic computations: A review. Commun. Comput. Phys. 5 (2–4), 242–272.

Xu, J., Duraisamy, K., 2020. Multi-level convolutional autoencoder networks for parametric prediction of spatio-temporal dynamics. Comput. Methods Appl. Mech. Engrg. 372, 113379.

Zhang, Z., Dong, Y., 2020. Temperature forecasting via convolutional recurrent neural networks based on time-series data. Complexity 2020, 1–8.