

Analysis of Sea Surface Temperature Variability Using Machine Learning



Said Ouala, Bertrand Chapron, Fabrice Collard, Lucile Gaultier,
and Ronan Fablet

Abstract Sea surface temperature (SST) is a critical factor in the global climate system and plays a key role in many marine processes. Understanding the variability of SST is therefore important for a range of applications, including weather and climate prediction, ocean circulation modeling, and marine resource management. In this study, we use machine learning techniques to analyze SST anomaly (SSTA) data from the Mediterranean Sea over a period of 33 years. The objective is to best explain the temporal variability of the SSTA extremes. These extremes are revealed to be well explained through a non-linear interaction between multi-scale processes. The results contribute to better unveil factors influencing SSTA extremes, and the development of more accurate prediction models.

Keywords Sea surface temperature · Machine learning · Stochastic models · Extremes

1 Introduction

Sea surface temperature (SST) is a critical parameter in the global climate system [1, 2] and plays a vital role in many marine processes, including ocean circulation, evaporation, and the exchange of heat and moisture between the ocean and atmosphere [3, 4, 5].

In recent years, particular attention has been attracted by marine heat waves, when SST largely exceeds the local expected average values [6, 7]. Extreme SST

S. Ouala (✉) · R. Fablet
IMT Atlantique, Lab-STICC, Brest, France
e-mail: said.ouala@imt-atlantique.fr

B. Chapron
Ifremer, LOPS, Plouzané, France

F. Collard · L. Gaultier
ODL, Locmaria-Plouzané, France

© The Author(s) 2024

B. Chapron et al. (eds.), *Stochastic Transport in Upper Ocean Dynamics II*,
Mathematics of Planet Earth 11, https://doi.org/10.1007/978-3-031-40094-0_11

247

can cause coral bleaching [8, 9], with cascading effects on the entire ecosystem. Additionally, localized events affect the amount of atmospheric moisture available, to impact precipitation patterns and the likelihood of drought or flooding in certain regions [10]. Better uncovering factors contributing to these extreme events is therefore of great importance to help predicting and mitigating their impacts.

The SST dynamics compound many processes that interact across a continuum of spatio-temporal scales. A first-order approximation of such a system was initially introduced by [11, 12]. Hasselmann pioneered a two-scale stochastic decomposition, to represent the interactions between slow and fast variables. In this study, we focus on SSTA data collected in the Mediterranean Sea, and examine the potential of machine learning techniques to derive relevant dynamical models. Focus is given on the seasonal modulation of the SSTA and we wish to unveil factors influencing the temporal variability of SSTA extremes. The proposed analysis builds on Hasselmann’s assumption that the variability of the SSTA can be decomposed into slow and fast components. The slow variables mostly follows the seasonal cycle, while the fast variables are linked to rapid processes, e.g. the wind variability. We thus approximate the probability density function of the SSTA data, using a stochastic differential equation in which the drift function represents the seasonal cycle and the diffusion function represents the envelope of the fast SSTA response.

The paper is organized as follows. We start by introducing the general underlying state space model of the SST anomaly. Rather than directly presenting the stochastic model, we first assume that an underlying deterministic ordinary differential equation (ODE) can represent the non-periodic variability of the SSTA. Considering a phase space reconstruction setting, we use the neural embedding of dynamical systems (NbedDyn) framework [13, 14] for this task. We then discuss the limitations of such a representation, and present the stochastic model. We conclude by summarizing our findings and potential future directions.

2 Method

Let us assume the following state-space model

$$\dot{\mathbf{z}}_t = f(\mathbf{z}_t) \tag{1}$$

$$\mathbf{x}_t = \mathcal{H}(\mathbf{z}_t) \tag{2}$$

where $t \in [0, +\infty]$ is time. The variables $\mathbf{z}_t \in \mathbb{R}^s$ and $\mathbf{x}_t \in \mathbb{R}^n$ represent the state variables and the SST anomaly observations respectively. f and \mathcal{H}_t are the dynamical and observation operators. The impact of noise on the dynamics and observation models is omitted for simplicity of the presentation.

2.1 Deterministic Model Hypothesis

The NbedDyn Framework If we assume that \mathbf{z}_t is asymptotic to a limit-set $L \subset \mathbb{R}^s$ and that the observations model is not an embedding [15], The NbedDyn model allows one to jointly derive a geometric reconstruction of the unseen phase space from partial observations and a corresponding dynamical model. For any given operator \mathcal{H} of a deterministic dynamical system, Takens theorem [16] guarantees that such an augmented space exists. However, instead of using a delay embedding, NbedDyn defines a d_E -dimensional augmented space with states $\mathbf{u}_t \in \mathbb{R}^{d_E}$ as follows:

$$\mathbf{z}_t^T = [\mathbf{x}_t^T, \mathbf{y}_t^T] \quad (3)$$

where $\mathbf{y}_t \in \mathbb{R}^{d_E-r}$ are stated as the latent states and T represents the matrix transpose. They account for the unobserved components of the true state \mathbf{z}_t .

The augmented state \mathbf{u}_t is assumed to satisfy the following state space model:

$$\dot{\mathbf{u}}_t = f_{\theta_1}(\mathbf{u}_t) \quad (4)$$

$$\mathbf{x}_t = (\mathbf{G}\mathbf{u}_t) \quad (5)$$

where \mathbf{G} is a projection matrix that satisfies $\mathbf{x}_t = \mathbf{G}\mathbf{u}_t$. The dynamical operator f_{θ_1} belongs to a given family of neural network operators parameterized by a parameter vector θ_1 . In this work, we follow [14] and use a linear quadratic parameterization of f_{θ_1} . This particular parameterization allows us to guarantee boundedness of the ODE (4) using the Schlegel boundedness theorem [17]. A linear quadratic ODE model can be written as follows:

$$\dot{\mathbf{u}}_t = f_{\theta_1}(\mathbf{u}_t) = \mathbf{c} + \mathbf{L}\mathbf{u}_t + [\mathbf{u}_t^T \mathbf{Q}^{(1)} \mathbf{u}_t, \dots, \mathbf{u}_t^T \mathbf{Q}^{(d_E)} \mathbf{u}_t]^T \quad (6)$$

where $\mathbf{c} \in \mathbb{R}^{d_E}$, $\mathbf{L} \in \mathbb{R}^{d_E \times d_E}$ and $\mathbf{Q}^{(i)} = [q_{i,j,k}]_{j,k=1}^{d_E}$, $i = 1, \dots, d_E$. The above approximate model is shifted according to $\tilde{\mathbf{u}}_t = \mathbf{u}_t - \mathbf{m}$ with $\mathbf{m} \in \mathbb{R}^{d_E}$. The approximate dynamical equation of the shifted state can be written as:

$$\dot{\tilde{\mathbf{u}}}_t = \mathbf{d} + \mathbf{A}\tilde{\mathbf{u}}_t + [\tilde{\mathbf{u}}_t^T \mathbf{Q}^{(1)} \tilde{\mathbf{u}}_t, \dots, \tilde{\mathbf{u}}_t^T \mathbf{Q}^{(d_E)} \tilde{\mathbf{u}}_t]^T \quad (7)$$

with

$$\mathbf{d} = \mathbf{c} + \mathbf{L}\mathbf{m} + [\mathbf{m}^T \mathbf{Q}^{(1)} \mathbf{m}, \dots, \mathbf{m}^T \mathbf{Q}^{(s)} \mathbf{m}]^T \quad (8)$$

and

$$\mathbf{A} = \left(a_{ij} \right) = \left(l_{ij} + \sum_{k=1}^s (q_{i,j,k} + q_{i,k,j}) m_k \right) \quad (9)$$

Given an observation time series of size $N + 1$ $\{\mathbf{x}_{t_0}, \dots, \mathbf{x}_{t_N}\}$, the training setting comes to jointly learning the model parameters $\theta_1 = \{\mathbf{c}, \mathbf{L}, \mathbf{Q}^1, \mathbf{Q}^2, \dots, \mathbf{Q}^{d_E}, \mathbf{m}\}$ and the latent states \mathbf{y}_t according to the following constrained optimization problem

$$\begin{aligned} \hat{\theta}_1, \{\hat{\mathbf{y}}_t\}_{t=0}^{i=N-1} = \arg \min_{\theta_1, \{\mathbf{y}_t\}} & \sum_{i=1}^N \|\mathbf{x}_{t_i} - \mathbf{G}\Phi_{\theta_1, t_i}(\mathbf{u}_{t_{i-1}})\|^2 \\ & + \lambda_1 \|\mathbf{u}_{t_i} - \Phi_{\theta_1, t_i}(\mathbf{u}_{t_{i-1}})\|^2 \\ & + \lambda_2 \mathcal{C}_1 \\ & + \lambda_3 \mathcal{C}_2 \end{aligned} \quad (10)$$

with $\Phi_{\theta_1, t}(\mathbf{u}_{t-1}) = \mathbf{u}_{t-1} + \int_{t-1}^t f_{\theta_1}(\mathbf{u}_w) dw$ is the flow of the ODE (6) (in our work, this flow is approximated using a Runge Kutta 4 scheme) and $\mathcal{C}_1 = \sum_{i,j,k=1}^s \|q_{i,j,k} + q_{i,k,j} + b_{j,i,k} + b_{j,k,i} + b_{k,i,j} + b_{k,j,i}\|^2$ and $\mathcal{C}_2 = \sum_{i=1}^s \text{Max}(\alpha_i, 0) / \text{Max}(\alpha_i + 1, 0)$ where $\alpha_i, i = 1, \dots, d_E$ the eigenvalues of the matrix $\mathbf{A}_s = \frac{1}{2}(\mathbf{A} + \mathbf{A}^T)$. The variables $\lambda_{1,2,3}$ are constant weighting parameters. The first constraint \mathcal{C}_1 steams from the energy-preserving condition of the quadratic non-linearity. It forces the contribution of the quadratic terms of f_{θ_1} to the fluctuation energy to sum up to zero. The second constraint, \mathcal{C}_2 , ensures that the eigenvalues of \mathbf{A}_s are negative. Satisfying these constraints guarantees that the model f_{θ_1} is bounded through the existence of a monotonically attracting trapping region that includes the limit-set revealed by the minimization of the forecasting loss. Similarly to the Takens delay embedding technique, the sequence:

$$R_{t_0, t_N} = \{\hat{\mathbf{u}}_{t_i}^T = [\mathbf{x}_{t_i}^T, \hat{\mathbf{y}}_t^T]\} \text{ with } t_i = t_0, \dots, t_N \quad (11)$$

represents a geometric reconstruction of the phase space. In addition to this reconstruction, the NbedDyn model can be used to forecast new observations by determining an initial condition of the unobserved component \mathbf{y}_t and performing a numerical integration of the ODE model (6). We infer the initial condition using a minimization of an objective function similar to (10), but only with respect to the latent states \mathbf{y}_t . This minimization can be seen as a variational data assimilation problem, with partial observations of the state-space variables and known dynamical and observation models [18].

Related Works Related state-of-the-art techniques mainly rely on the reconstruction of a phase space using delay embedding [16]. This includes both traditional parametric and non-parametric modeling techniques [19, 20] as well as recurrent neural networks (RNNs). The latter family of methods includes both simple RNN

parameterizations of dynamical systems, as well as latent space inference techniques that are built on an approximation of a posterior distribution that requires the parameterization of a delay embedding [21, 22, 23].

The interest of the NbedDyn framework in contrast to delay embedding based approaches resides in the fact that we do not exploit either a delay embedding or an explicit modeling of the inference model (i.e., the reconstruction of the latent states given the observed time series). As such, our scheme only involves the selection of the class of ODEs of interest. This model reduces the complexity of the overall scheme to the complexity of the ODE representation and guarantees the consistency of the reconstructed latent states w.r.t. the learnt ODE.

2.2 Stochastic Model Hypothesis: The Stochastic NbedDyn

When using phase space reconstruction techniques, one should not forget about the assumptions that this theory is built on. For any embedding to work, we are assuming that the dynamical model in (1) exists and can be represented by an ordinary differential equation [15]. For several realistic applications, this ODE may not exist or can have an extremely large dimension. In geoscience, for instance, the dimension of a state space variable can reach $s \approx O(10^9)$. In these situations, reconstructing such a high-dimensional phase space becomes significantly more challenging. In practice, the model returned by any embedding technique can be complemented by an appropriate closure. The form of this closure term can be deterministic using for example the framework of [24] or stochastic through an appropriate calibration of a noise forcing.

When considering SST anomaly data, after calibration of the neural embedding model, an unpredictable, high frequency residual remains. Based on Hasselmann's idea, we assume this residual component represents the effect of fast-scale processes, e.g. passages of atmospheric and oceanic eddies. To first order, it can be represented as a modulated white noise. Indeed, this residual, shown in Fig. 3, exhibits correlations with the slow-scale SST anomaly data.

To model stochastic SST anomalies, the deterministic NbedDyn model described above is first optimized, and further complemented (6) with a stochastic forcing as follows:

$$\begin{cases} \dot{\mathbf{u}}_t = f_{\theta_1}(\mathbf{u}_t) + g_{\theta_2}(\mathbf{u}_t)\xi_t \\ \mathbf{x}_t = \mathbf{G}\mathbf{u}_t \end{cases} \quad (12)$$

with ξ_t is a white noise. We derive the parameters of the model (12), as follows. Given an observation time series of size $N + 1$ $\{\mathbf{x}_{t_0}, \dots, \mathbf{x}_{t_N}\}$, similarly to the deterministic case, we optimize the diffusion parameters θ_2 to minimize the forecast of the observations. In addition to the diffusion parameters, we also reconstruct a noise realization ξ^{rec} that generates the observations process under (12). Overall,

the optimization problem can be written as follows:

$$\hat{\theta}_2, \{\hat{\xi}_{t_i}^{rec}\}_{i=0}^{N-1} = \arg \min_{\theta_2, \xi^{rec}} \sum_{t=1}^T \|\mathbf{x}_t - \mathbf{G}\Phi_{\theta,t}(\mathbf{u}_{t-1}, \xi_{t-1}^{rec})\|^2$$

$$\text{Subject to } \begin{cases} \mathbf{u}_t & = \Phi_{\theta,t}(\mathbf{u}_{t-1}, \xi^{rec}) \\ \mathbf{G}\mathbf{u}_t & = \mathbf{x}_t \\ \mathbf{R}_{\xi^{rec}}\xi^{rec}(\tau) = 0 & \text{for all } \tau \neq 0 \end{cases} \tag{13}$$

with $\{\hat{\xi}_{t_i}^{rec}\}_{i=0}^{N-1}$ is the noise realization that minimizes the objective function in (13) and $\Phi_{\theta,t}$:

$$\Phi_{\theta,t}(\mathbf{u}_{t-1}, \xi^{rec}) = \mathbf{u}_{t-1} + \int_{t-1}^t f_{\theta}(\mathbf{u}_w)dw + \int_{t-1}^t g_{\theta}(\mathbf{u}_w)\xi_w^{rec}dw$$

the solution of the stochastic model. This solution is approximated in this work using an Euler-Maruyama scheme, which makes the model converge to an Ito SDE.

In practice, we use the following regularized optimization problem:

$$\hat{\theta}_2, \hat{\xi}^{rec} = \arg \min_{\theta_2, \xi^{rec}} \sum_{t=1}^T \|\mathbf{x}_t - \mathbf{G}\Phi_{\theta,t}(\mathbf{u}_{t-1}, \xi^{rec})\|^2$$

$$+ \lambda_4\mathcal{C}_3$$

$$+ \lambda_5\mathcal{C}_4$$

$$+ \lambda_6\mathcal{C}_5 \tag{14}$$

with $\mathcal{C}_3 = \|\mathbf{R}_{\xi^{rec}}\xi^{rec}(\tau)\|^2$, $\mathcal{C}_4 = \mathbf{Var}(\Phi_{\theta,t}(\mathbf{u}_{t-1}, \xi^{samp}))$, $\mathcal{C}_5 = \|\Phi_{\theta,t}(\mathbf{u}_{t-1}, \xi^{rec}) - \mathbf{E}[\Phi_{\theta,t}(\mathbf{u}_{t-1}, \xi^{samp})]\|^2$ and ξ^{samp} is a sampled Gaussian white noise. The variables $\lambda_{4,5,6}$ are constant weighting parameters. The first constraint \mathcal{C}_3 makes the reconstructed noise path white. The second and third constraints, $\mathcal{C}_{4,5}$, ensure that the SDE generalizes to sampled white noises. Specifically \mathcal{C}_3 makes an ensemble of trajectories generated from sampled white noise close to the trajectory generated from the reconstructed noise and \mathcal{C}_2 reduces the spread of the ensemble around the trajectory simulated from the reconstructed noise.

After optimization, we can couple the optimization problems (10) and (14) and calibrate jointly all the model parameters $\theta_1, \theta_2, \mathbf{y}_t, \xi^{rec}$. This fine tuning step is not essential but allows both the drift and diffusion parameters of the model to adapt to each other.

3 Numerical Experiments

3.1 Data

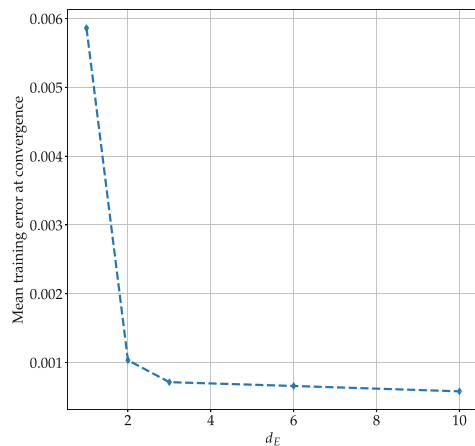
Sea Surface Temperature Anomalies (SSTA) in the Mediterranean Sea correspond to the Ligurian Sea at 8.6°E , 43.8°N . The anomalies are computed based on a yearly average of the annual 99th percentile of the SST reanalysis [25, 26]. The time series is made up of daily SST anomaly measurements from 1987 to 2019. We use the daily data from 1987 to 2014 as training data. Figure 3e illustrates the time series. These time series include a seasonal cycle and non-periodic high temperature extremes in the summer.

3.2 Analysis of the Deterministic Model

In this first experiment, we investigate if the deterministic neural embedding model is able to model the non-periodic variability of the SSTA extremes. For this purpose, we test 3 models with dimensions of the embedding ranging from 1 to 10.

Analysis of the Embeddings The choice of the dimension d_E is linked to the number of independent variables that can be used to model the dynamics using, in our context, a bounded autonomous linear quadratic ODE. We start by studying the direct impact of d_E on the performance of the NbedDyn model. Figure 1 shows the impact of d_E on the training error between the observations and the model simulation. Other criteria could be used (please refer to [13, 14] for a more in depth analysis of this parameters on other case studies), but overall, the training error provides a direct measure of the effectiveness of the embedding dimension in the training phase. The first evaluation of the training error reported in Fig. 1

Fig. 1 Mean training error at convergence. We report the mean training error at convergence of the deterministic NbedDyn model for different dimensions d_E of the embedding. This error is averaged over the training time series, and we highlight here both the mean and standard deviations



corresponds to d_E equal to the dimension of the measurements, i.e. $d_E = 1$. In this experiment, no latent states \mathbf{y}_t are used and the embedding $\mathbf{u}_t = \mathbf{x}_t$. In such situations, the ODE model can not perfectly fit the data. Furthermore, at this particular value of d_E , the models are more likely to display a bad asymptotic behavior. As the dimension increases, this training error decreases which confirms better modeling abilities using the NbedDyn model. In the following, we study the models with $d_E = 3, 6$ and 10 .

Asymptotic Properties of the Models We evaluate the asymptotic behavior of the deterministic models for $d_E = 3, 6$ and 10 . For this purpose, we run the nbedDyn models for a period of 27 years. The resulting simulation is visualized with respect to the reconstructed phase space (11) of the training data in Fig. 2. Overall, the

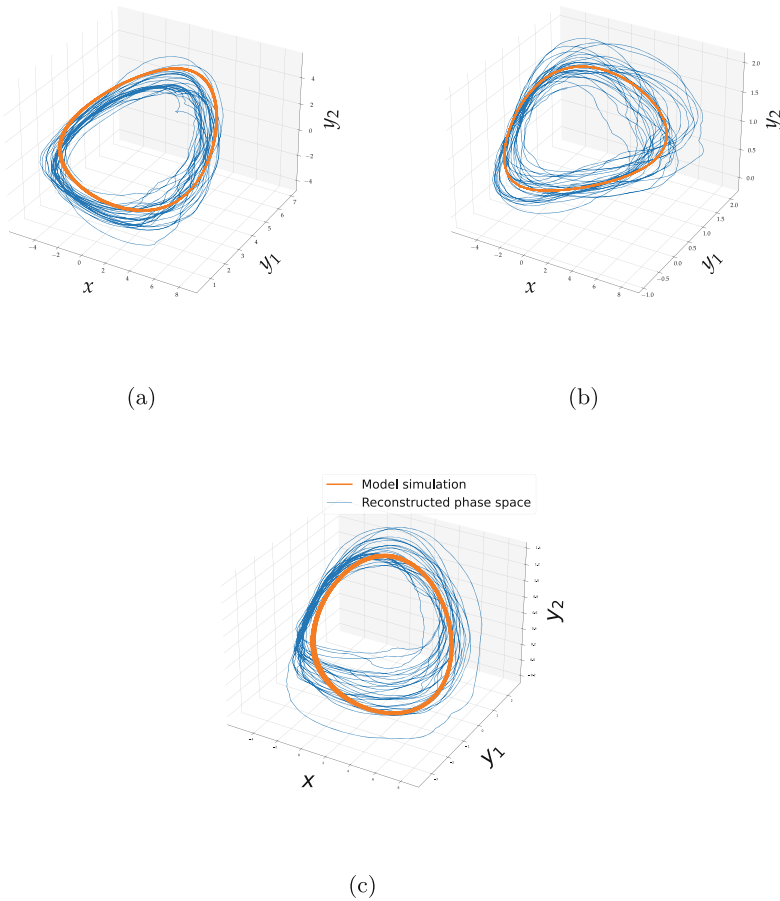


Fig. 2 Asymptotic solution of the deterministic models. We visualize the simulation of the deterministic models with respect to the reconstructed phase space. The models with $d_E = 3, 6$ and 10 are given in figure (a), (b) and (c) respectively. For $d_E > 3$, we project the simulation and the reconstructed phase space into \mathbb{R}^3

models are only able to reproduce the seasonal cycle of the SST anomaly data. Other experiments (not shown here) suggest that even a farther increase of the dimension of the embedding does not allow the model to capture the non-periodic behavior of the SST anomaly extremes.

Analysis of the Training Residuals To further investigate the asymptotic behavior of the deterministic models, we visualize in Fig. 3 the training residual $\{\mathbf{x}_{t_i} - \mathbf{G}\Phi_{\theta,t_i}(\mathbf{u}_{t_{i-1}})\}$ with $t_i = t_0, \dots, t_N$. When the dimension of the embedding increases above $d_E = 1$, a qualitative and quantitative change in the residual error is present. This is due to the fact that a two-dimensional ODE (in \mathbb{R}) is needed to capture the oscillations of the seasonal cycle of the SSTA. However, when the dimension of the embedding increases above 2, no clear qualitative or quantitative change is present. Furthermore, the residual is much more high frequency than the training SSTA data, which suggests that the errors are due to a missing high frequency scales that can not be modeled using the standard deterministic model.

Based on these considerations, and motivated by Hasselmann's works on stochastic climate models with applications on SST anomaly data, we proposed the stochastic NbedDyn model. In this framework, the SST residual of Fig. 3 is modeled as a stochastic forcing.

3.3 Analysis of the Stochastic Model

We focus our analysis on the model with $d_E = 6$. We add a stochastic forcing to the neural embedding model (the parameters of the diffusion function g_{θ_2} are optimized according to Appendix 1). Figure 4 shows the reconstructed phase space under this new model, as well as a model simulation of 27 years. When compared to the simulations of the deterministic model in Fig. 2, the stochastic model is able to cover the whole reconstructed phase space, including the regions with high temperature extremes. This shows that including the high frequency forcing is crucial for the model to capture the non-periodic behavior of the extremes.

These observations are further illustrated in the simulation example given in Fig. 5. The stochastic model is able to produce an ensemble of SST anomaly trajectories that reproduce the non-periodic variability of the extremes. Furthermore, the trajectories generated from a sampled white noise match the one of the reconstructed noise, which validates the proposed training procedure.

We can also discuss the marginal PDF of the stochastic model and compare it to the one computed from the data in Fig. 6. The PDF of the model is computed over a simulation of 109 years. Overall, the model is able to correctly model the high SST anomalies (in the summer), including the non-periodic extremes that form the tail of the distribution. The negative SST anomalies (in the winter) are not approximated as good as in the summer case. This is due to the fact that the model flattens the PDF in the winter by generating trajectories that have more spread (as highlighted in the ensemble prediction experiment in Fig. 5). We did not investigate this problem

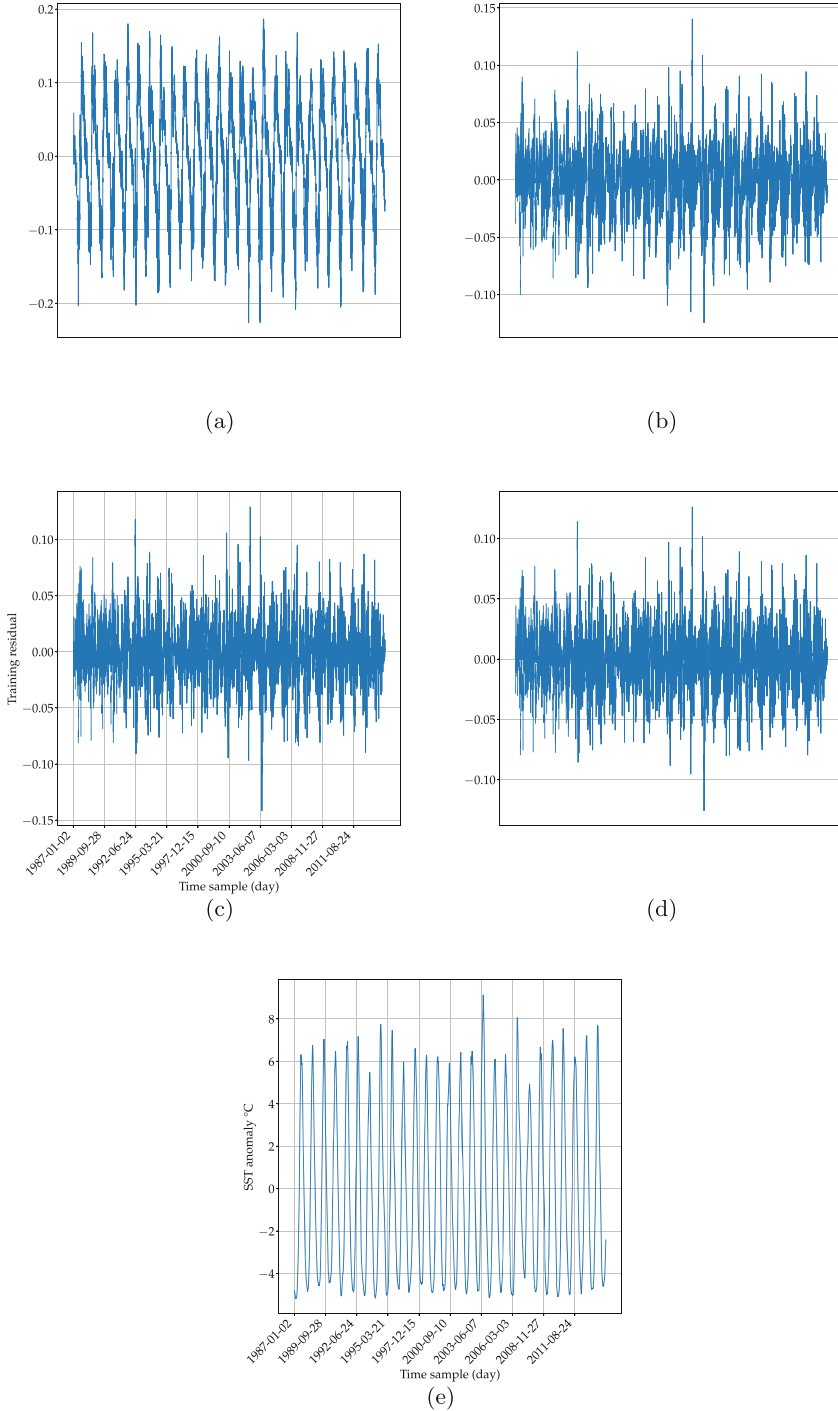
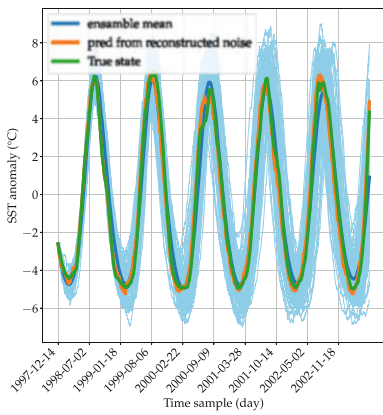
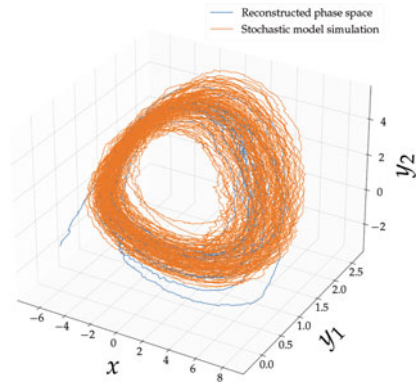
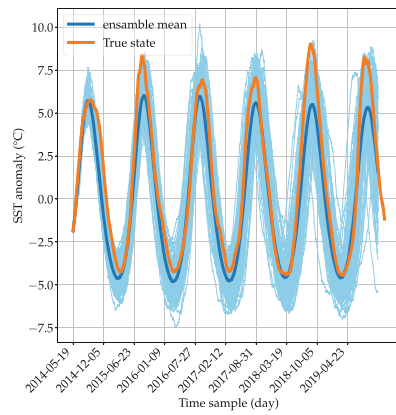


Fig. 3 Training residual and corresponding training data. We visualize the training residual for $d_E = 1$ in (a), $d_E = 3$ in (b), $d_E = 6$ in (c) and $d_E = 10$ in (d). Corresponding training data are given in (e)

Fig. 4 Simulation of the stochastic model. We visualize a stochastic model simulation and compare it to the reconstructed phase space. Both the model simulation and reconstructions are projected into \mathbb{R}^3



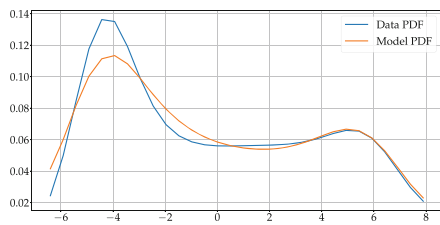
(a)



(b)

Fig. 5 Ensemble simulation of the stochastic model. We visualize an ensemble simulation of the stochastic model, both in the training (a) and test (b) sets. The simulation in the training set is carried to compare trajectories computed from a sampled noise to the one issued from the reconstructed noise ξ^{rec}

Fig. 6 PDF of the data and the stochastic model. We compare the marginal PDF of the stochastic model with the one of the data



within the present study. However, we can make the PDF sharper in the winter by forcing the diffusion of the model to be closer to zero during this season.

4 Conclusion

In this work, we examined the potential of machine learning techniques to derive relevant dynamical models of sea surface temperature anomaly data in the Mediterranean Sea. We focused on the seasonal modulation of SSTA extremes and used a neural embedding model to reconstruct the phase space of SSTA data. We then added a stochastic forcing term to account for the missing high frequency variability. Our results contribute to the understanding of the factors influencing SSTA extremes and the development of more accurate prediction models. In particular, the analysis highlights the importance of including these fast high-frequency scales in the modeling of SSTA data.

One potential avenue for future work is to investigate the white noise hypothesis in comparison to other types of stochastic models, such as those based on colored noise or fractional Brownian motion. Furthermore, it would be interesting to apply the methodology to other regions and compare the results to evaluate the local impacts of the fast-scales on the slower ones. The ability of using this model as an emulator and studying its predictive skills with respect to standard ocean data assimilation based systems is also a promising perspective.

Finally, and from a methodological point of view, this work highlights the importance of complementing models that are returned by an embedding methodology. Specifically, and as discussed in Sect. 2.2, in complex applications such as the ones in geosciences, the dimension of the underlying state variables is likely to be huge and defining ways of complementing reduced order models through appropriate closure terms is mandatory in order to capture the variability of the data. Analysing the residual of the model fitting procedure is a natural way to define and optimize this closure terms.

Appendix 1: Training

The trainable parameters of the deterministic NbedDyn models i.e. the linear quadratic ODE and initial conditions of the latent states are initially sampled from a uniform distribution. The training of all models is carried using the Adam optimizer. We use a varying learning rate (from 0.1 to 0.001) in all the experiments to speed up the training. Regarding the weighting parameters $\{\lambda_i\}_{i=1}^3$, we set $\lambda_1 = 1$ during all the training. The weights responsible for the boundedness constraints were set at higher values in the beginning of the training i.e. $\lambda_2 = 100$ and $\lambda_3 = 1000$ and then reduced to $\lambda_2 = 1$ when $\lambda_3 C_2 = 0$. The training is stopped using cross-validation. Regarding the stochastic forcing, the parameters of the diffusion are initially sampled from a uniform distribution and the noise path ξ^{rec} is initialized from a standard normal distribution. We use a learning rate of 0.001 and the weighting parameters $\{\lambda_i\}_{i=4}^6$ to one during all the training. We finished the training

with a fine tuning step, in which all the parameters of the model are optimized jointly with a learning rate of 0.0001.

Appendix 2: Parameterization of the Diffusion Function

The diffusion function $g_{\theta_2} : \mathbb{R}^{d_E} \longrightarrow \mathbb{R}^{d_E \times d_N}$ where d_N is the dimension of the noise. In our experiment, we parameterized this function using a fully connected neural networks with 2 hidden layers with a sigmoid activation and 400 neurones per hidden layer. The dimension of the noise d_N is set to the dimension of the state \mathbf{u}_t .

References

- [1] Grant R Bigg, TD Jickells, PS Liss, and TJ Osborn. The role of the oceans in climate. *International Journal of Climatology: A journal of the Royal Meteorological Society*, 23(10):1127–1159, 2003.
- [2] Song Yang, Zhenning Li, Jin-Yi Yu, Xiaoming Hu, Wenjie Dong, and Shan He. El niño–southern oscillation and its impact in the changing climate. *National Science Review*, 5(6):840–857, 2018.
- [3] Gerold Siedler, John Gould, and John Church. *Ocean circulation and climate: observing and modelling the global ocean*. Elsevier, 2001.
- [4] Alexander Otto, Friederike EL Otto, Olivier Boucher, John Church, Gabi Hegerl, Piers M Forster, Nathan P Gillett, Jonathan Gregory, Gregory C Johnson, Reto Knutti, et al. Energy budget constraints on climate response. *Nature Geoscience*, 6(6):415–416, 2013.
- [5] Kevin E Trenberth and John T Fasullo. An apparent hiatus in global warming? *Earth's Future*, 1(1):19–32, 2013.
- [6] Eric CJ Oliver, Simon J Wotherspoon, Matthew A Chamberlain, and Neil J Holbrook. Projected Tasman Sea extremes in sea surface temperature through the twenty-first century. *Journal of Climate*, 27(5):1980–1998, 2014.
- [7] Deborah Pardo, Stéphanie Jenouvrier, Henri Weimerskirch, and Christophe Barbraud. Effect of extreme sea surface temperature events on the demography of an age-structured albatross population. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1723):20160143, 2017.
- [8] Terry P Hughes, James T Kerry, Mariana Álvarez-Noriega, Jorge G Álvarez-Romero, Kristen D Anderson, Andrew H Baird, Russell C Babcock, Maria Beger, David R Bellwood, Ray Berkelmans, et al. Global warming and recurrent mass bleaching of corals. *Nature*, 543(7645):373–377, 2017.
- [9] PC Mohanty, A Kushabaha, RS Mahendra, RK Nayak, BK Sahu, E Rao, and T Sinivasa Kumar. Persistence of marine heat waves for coral bleaching and their spectral characteristics around andaman coral reef. *Environmental Monitoring and Assessment*, 193(8):1–9, 2021.
- [10] Wenju Cai, Simon Borlace, Matthieu Lengaigne, Peter Van Rensch, Mat Collins, Gabriel Vecchi, Axel Timmermann, Agus Santoso, Michael J McPhaden, Lixin Wu, et al. Increasing frequency of extreme el niño events due to greenhouse warming. *Nature climate change*, 4(2):111–116, 2014.
- [11] Klaus Hasselmann. Stochastic climate models part I. Theory. *Tellus*, 28(6):473–485, 1976.

- [12] Claude Frankignoul and Klaus Hasselmann. Stochastic climate models, part II application to sea-surface temperature anomalies and thermocline variability. *Tellus*, 29(4):289–305, 1977.
- [13] S Ouala, D Nguyen, L Drumetz, B Chapron, A Pascual, F Collard, L Gaultier, and R Fablet. Learning latent dynamics for partially observed chaotic systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 30(10):103121, 2020.
- [14] Said Ouala, Steven L Brunton, Bertrand Chapron, Ananda Pascual, Fabrice Collard, Lucile Gaultier, and Ronan Fablet. Bounded nonlinear forecasts of partially observed geophysical systems with physics-constrained deep learning. *Physica D: Nonlinear Phenomena*, page 133630, 2023.
- [15] Tim Sauer, James A. Yorke, and Martin Casdagli. Embedology. *Journal of Statistical Physics*, 65(3):579–616, Nov 1991.
- [16] Floris Takens. Detecting strange attractors in turbulence. In David Rand and Lai-Sang Young, editors, *Dynamical Systems and Turbulence, Warwick 1980*, pages 366–381, Berlin, Heidelberg, 1981. Springer Berlin Heidelberg.
- [17] Michael Schlegel and Bernd R. Noack. On long-term boundedness of Galerkin models. *Journal of Fluid Mechanics*, 765:325–352, 2015.
- [18] Peter Lynch and Xiang-Yu Huang. *Initialization*, pages 241–260. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.
- [19] Henry D. I. Abarbanel. *Modeling Chaos*, pages 95–114. Springer New York, New York, NY, 1996.
- [20] Johan Paduart, Lieve Lauwers, Jan Swevers, Kris Smolders, Johan Schoukens, and Rik Pintelon. Identification of nonlinear systems using Polynomial Nonlinear State Space models. *Automatica*, 46(4):647–656, April 2010.
- [21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. *arXiv:1512.03385 [cs]*, December 2015. arXiv: 1512.03385.
- [22] Rahul G. Krishnan, Uri Shalit, and David Sontag. Structured Inference Networks for Nonlinear State Space Models. *arXiv:1609.09869 [cs, stat]*, September 2016. arXiv: 1609.09869.
- [23] Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, pages 6571–6583, 2018.
- [24] Matthew Levine and Andrew Stuart. A framework for machine learning of model error in dynamical systems. *Communications of the American Mathematical Society*, 2(07):283–344, 2022.
- [25] S Simoncelli, C Fratianni, N Pinardi, A Grandi, M Drudi, P Oddo, and S Dobricic. Mediterranean sea physical reanalysis (cmems med-physics)[data set]. *Copernicus Monitoring Environment Marine Service (CMEMS)*, 2019.
- [26] Karina von Schuckmann et al. Copernicus marine service ocean state report, issue 3. *Journal of Operational Oceanography*, 12(sup1):S1–S123, 2019.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

