A JOINT SPACE-TIME ENCODER FOR GEOGRAPHIC TIME-SERIES DATA

David Mickisch[1], Konstantin Klemmer[2], Mélisande Teng[1, 3], and David Rolnick[1, 4]

[1]Mila - Quebec Artificial Intelligence Institute
[2]Microsoft Research
[3]Université de Montréal
[4]McGill University

ABSTRACT

Many real-world processes are characterized by complex spatio-temporal dependencies, from climate dynamics to disease spread. Here, we introduce a new neural network architecture to model such dynamics at scale: the *Space-Time Encoder*. Building on recent advances in *location encoders*, models that take as inputs geographic coordinates, we develop a method that takes in geographic and temporal information simultaneously and learns smooth, continuous functions in both space and time. The inputs are first transformed using positional encoding functions and then fed into neural networks that allow the learning of complex functions. We implement a prototype of the *Space-Time Encoder*, discuss the design choices of the novel temporal encoding, and demonstrate its utility in climate model emulation. We discuss the potential of the method across use cases, as well as promising avenues for further methodological innovation.

1 INTRODUCTION

Spatio-temporal dynamics occur frequently in tasks within climate science and adaptation that involve geographic information – data that can be mapped onto the sphere of our planet. A common problem in domains such as atmospheric science, ecology, and agriculture is to build dense maps from observations that are sparse in space and time. Traditionally, Gaussian Processes (GPs) have been used for such interpolation tasks, however, GPs struggle to adapt to the ever-growing volume of geographic data available. There is a clear need for scalable neural network methods for processing geographic time series and for learning complex spatio-temporal functions (Reichstein et al., 2019). In this work, we address this research gap by introducing the *Space-Time Encoder*, a neural network architecture aimed at learning complex spatio-temporal patterns at scale.

While previous work has investigated spatial-only geographic location encoders for supervised learning and pretraining tasks (Mac Aodha et al., 2019; Mai et al., 2023), geospatial data often contains additional temporal components that affects the observed process, and may be expected to change the optimal spatial encoding. Such spatio-temporal dynamics occur in diverse problems, such as climate model emulation (Watt-Meyer et al., 2023), species distribution modeling of migratory birds (Zuckerberg et al., 2016), and crop yield forecasting (Cai et al., 2018). It is thus natural to consider whether location encoding methods may be expanded to account for temporal information.

Research on geospatial embeddings indicates that choosing orthonormal function families for encoding geocoordinates boosts prediction performance (Rußwurm et al., 2023). Therefore, it is of interest to construct orthonormal function families for the spatio-temporal domain and compare the prediction performance of such families with the performance of non-orthonormal encodings.

In this work, we introduce the *Space-Time Encoder*, including a novel temporal encoding system and proposing several different, modular design options. We show that our approach improves performance on a sparse, multi-variate climate model emulation task on the ACE dataset (Watt-Meyer et al., 2023), representing an important problem in monitoring long-term atmospheric dynamics under climate change. We report results for different configurations of the temporal encoding component

and consider the effect of orthonormal regularization. We believe our work represents a meaningful step towards spatio-temporal encoding systems that may be useful across many tasks relevant to climate science and adaptation.

2 METHOD

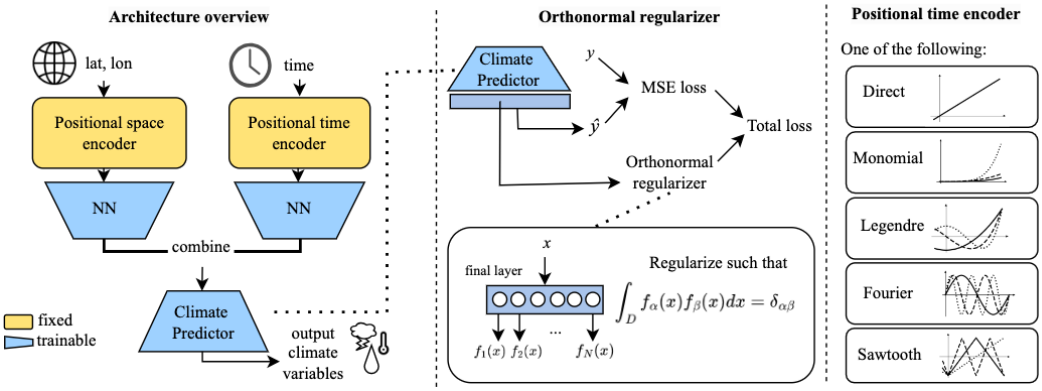

Figure 1: Overview of our proposed *Space-Time Encoder* (left panel), with a focus on the orthonormal regularizer (center panel) and different time encoding configurations (right panel)

Our method addresses the direct prediction problem of learning functions mapping spatio-temporal coordinates – expressed in longitude, latitude, and time – to environmental variables such as air temperatures at different altitudes. The proposed space-time encoder first maps the spatial and temporal coordinates to two separate embedding vectors. For the space encoder, we achieve this by applying a set of spherical harmonics functions to the spatial coordinates, which has proved effective in previous work. The resulting vector can then be passed through a neural network module to obtain the final space embedding vector. For the time encoder, we proceed similarly by choosing an encoding function, which we then apply to the time coordinate to obtain an embedding vector. A neural network can then transform this embedding vector further to obtain the final time embedding vector. The two embedding vectors are combined and passed through a neural network trained to output the prediction vector.

The focus of our investigation is the choice of function family for the time encoder. We investigate different choices of families in order to improve on the direct embedding of the time coordinate. We further investigate a novel regularizer whose design is guided by insights from previous investigations of geo-spatial encoders. An overview of our proposed method is given in Figure 1.

2.1 DEFINITION OF ORTHONORMALITY

Let $\mathcal{F} = \{f_1, \dots f_N\}$ denote a set of functions mapping a common domain, D, to the real numbers, that is, $f_i : D \to \mathbb{R}$. For our method, this domain will usually be the sphere S^2, the interval I, or their Cartesian product $S^2 \times I$. Recall that the scalar product of two functions $f, g : D \to \mathbb{R}$ can be defined as the integral of their point-wise product over the domain D: $\langle f, g \rangle = \int_D f(x)g(x)dx$. The set of functions $\mathcal{F}$ is an *orthonormal* set of functions if the scalar product of any two of its elements equals the Kronecker delta: $\langle f, g \rangle = \delta_{f,g}$, which is 1 if $f = g$ and 0 otherwise.

We now consider the neurons of a neural network layer as functions, $n_i : D \to \mathbb{R}$ from the domain of the neural network to the real numbers. We define a neural network layer to be orthonormal if its neurons are representing a set of orthonormal functions. Note that an orthonormal layer consisting of a weight matrix A, bias vector b, and activation function σ, need not have an orthonormal weight matrix A and that A being an orthonormal matrix need not imply that the layer is orthonormal.

2.2 EMBEDDING THE SPATIO-TEMPORAL DOMAIN

Previous work has shown the effectiveness of using sets of spherical harmonics function to create geo-spatial coordinate embeddings. Those sets of functions are orthonormal and we are therefore constructing sets of orthonormal functions for spatio-temporal coordinates from $S^2 \times I$. There are several examples of orthonormal function sets on the interval such as Fourier bases, Legendre polynomials - the analogous concept to spherical harmonics on the interval - and sawtooth bases. Orthonormal functions on the sphere and on the interval can be combined to orthonormal functions on cross product of sphere and interval as follows: Consider an orthonormal set of functions $\mathcal{S} = \{f_1, \dots f_N\}$, which are defined on the sphere, S^2, and an orthonormal set of functions which are defined on the interval, $I, \mathcal{I} = \{g_1, \dots, g_M\}$. Then the set $\mathcal{S} \otimes \mathcal{I} = \{f_1 \otimes g_1, \dots, f_i \otimes g_j, \dots, f_N \otimes g_M\}$ is orthonormal on the Cartesian product $S^2 \times I$. Here, the tensor product of two functions $f : S^2 \to \mathbb{R}$ and $g : I \to \mathbb{R}$ is just defined as $f \otimes g : S^2 \times I \to \mathbb{R}, (x, y) \mapsto f(x) \cdot g(y)$. The proof that $\mathcal{S} \otimes \mathcal{I}$ is indeed orthonormal, can be done via Fubini's theorem.

2.3 ORTHONORMAL REGULARIZATION

Given a uniform sample from the data manifold $\mathcal{X} = \{x_1, \dots, x_N\}$, one can approximately evaluate the integrals that are necessary for testing orthonormality: $\langle f, g \rangle = \int_D f(x)g(x)dx \approx \frac{1}{N} \sum_i f(x_i)g(x_i)$. We define a data-dependent scalar product using this Monte Carlo approximation: $\langle f, g \rangle_{\mathcal{X}} := \frac{1}{N} \sum_i f(x_i)g(x_i)$. Using this new scalar product, for a layer $\mathcal{L} = \{n_1, \dots n_K\}$ we can express an approximate version of the orthonormality condition as: $\langle n_j, n_k \rangle_{\mathcal{X}} - \delta_{j,k} = 0$. By squaring each such equation and summing over the equations for all $j, k \in [K]$, we obtain an orthonormality regularizer $\mathcal{N}(\mathcal{L}|\mathcal{X}) := \sum_j \sum_k (\langle n_j, n_k \rangle_{\mathcal{X}} - \delta_{j,k})^2$. Notice that computing this regularizer needs $\mathcal{O}(NK^2)$ additions and $\mathcal{O}(K^2)$ squaring operations.

3 EXPERIMENTS

Our experiments are based on the dataset of the AI2 Climate Emulator (Watt-Meyer et al., 2023). The original dataset is composed of 11 climate model simulations each over a period of 10 years. The simulations associate values for 55 climate variables to spatio-temporal location coordinates. The temporal resolution is 6h and the spatial resolution is 100km.

We select 1 year of data from 1 simulation and further select 8 climate variables representing temperatures at different altitudes. We consider the task of interpolation given a sparse set of spatio-temporal coordinates. We used 3% of the available grid-points for training set, validation set and test set in equal proportion, i.e. 1% of data, or around 1 million space-time coordinates sampled randomly from a uniform distribution, for each set.

We first consider a direct time encoding baseline in which the time is input directly to the model without further encoding. We then explore two methods for improving further on this baseline. First, we vary the type of the positional time encoding considering five encoding families as shown in 1. We further investigate adding an orthonormal regularization term as described in 2.3 for last layer during training.

We use the FCNet architecture which has been proved effective for geospatial interpolation (Rußwurm et al., 2023) for the positional encoding networks and the climate predictor modules. The model has 4 Residual Blocks with 1024 hidden neurons in each layer with a total of 10.2M trainable parameters. We fix the positional space encoding to a set of 1600 spherical harmonics basis functions also following the approach in Rußwurm et al. (2023). The size of the positional time encoding is fixed to 120 dimensions across all experiments. Both positional encodings are concatenated to form a combined spatio-temporal positional encoding. All the models are trained for 5 epochs with the Adam optimizer and a Mean Square Error loss function. The training data are mean-centered and scaled by their standard deviation. The models are evaluated with Root Mean Square Error (RMSE) averaged over the 8 considered temperature variables.

Table 1: Average Test RMSE for 8 temperature variables from ACE dataset. Orthonormal regularization improves the test performance of our space-time encoder for most types of temporal encodings

Time Embedding Type	***Without* Regularization**	***With* Regularization**
Drop Time Coordinate	5.691	5.737
Direct	3.147	*3.144*
Monomial	3.222	*3.190*
Legendre	2.971	*2.953*
Fourier	2.731	***2.704***
Sawtooth	3.456	*3.44*

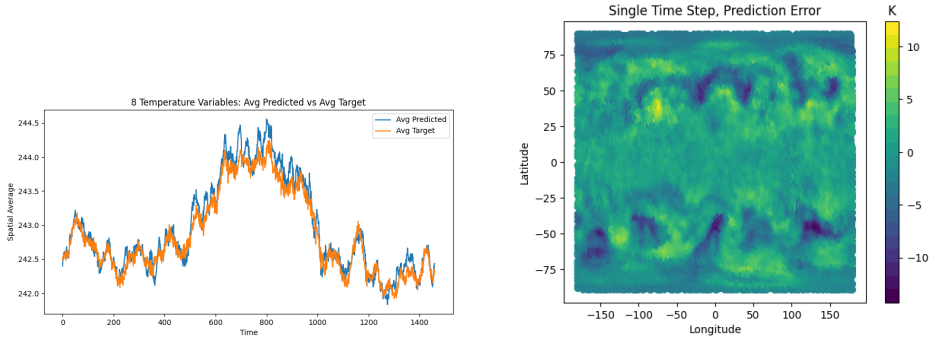

Figure 2: *Left part* shows spatial averages over target and predictions from Space-Time Encoder with Fourier embeddings and orthonormal regularization for points from a test set with 30% of points from 12 months of data. *Right part* shows example predictions for a randomly sampled single time step and temperature variable 4.

4 Results & Discussion

We report the RMSE of the different models, averaged over the 8 temperature variables in Table 1. A map of the prediction error for an single time step is given in Figure 2 As anticipated, we observe that adding the time coordinate as an input to the model improves on using spatial coordinates alone for our task, demonstrating the relevance of including the temporal domain.

The fact that Legendre embeddings improve on the direct time embedding while monomial embeddings actually result in worse performance indicate the importance of using orthonormal encodings since one can obtain Legendre functions by orthonormalizing sets of monomials. The fact that Fourier embeddings perform best also points to the importance of orthonormality at the encoding level. Additionally, we find that adding orthonormal regularization improves performance, for all considered time embeddings with the best performance being achieved by the model with Fourier time embedding and orthonormal regularization. While these experiments are still a proof of concept on a small subset of the ACE dataset, they point to the potential of the Space-Time encoder for the task of climate prediction and of the use of orthonormal regularization in this context. Qualitatively, we observe that the spatial averages align well, as shown in Fig. 2 on the left. Immediate next steps include extending experiments to use a larger subset of the ACE dataset, considering multiple years

and more climate variables, as well as tuning the models more rigorously, and comparing different architectures for the positional encodings networks and the climate predictor module.

5 CONCLUSION

We introduced the Space-Time Encoder, aimed at capturing spatio-temporal dynamics, which are crucial in real-world processes such as climate prediction. We implement a prototype that we test on the task of climate dynamics prediction on a subset of the ACE dataset. We explore different possibilities for the encoding the time coordinate, and show that adding orthonormal regularization is a promising avenue to improve predictions. Our experiments demonstrate the potential of the Space-Time Encoder, and we plan on comparing it to other methods that can take as input continuous space-time coordinates such as vanilla neural networks or Gaussian Processes. Future work includes applying this approach to other climate-relevant tasks, including crop yield prediction and species range map estimation. We plan on using the Space-Time encoder as a prior to condition models, building on the work of Mac Aodha et al. (2019). We will also investigate pre-training general spatio-temporal embeddings which could then serve as space-time prior, using unlabeled Earth observation data, in a similar way to SatCLIP location embeddings (Klemmer et al., 2023).

REFERENCES

Yaping Cai, Kaiyu Guan, Jian Peng, Shaowen Wang, Christopher Seifert, Brian Wardlow, and Zhan Li. A high-performance and in-season classification system of field-level crop types using time-series landsat data and a machine learning approach. *Remote sensing of environment*, 210:35–47, 2018.

Konstantin Klemmer, Esther Rolf, Caleb Robinson, Lester Mackey, and Marc Rußwurm. Satclip: Global, general-purpose location embeddings with satellite imagery. *arXiv preprint arXiv: 2311.17179*, 2023.

Oisin Mac Aodha, Elijah Cole, and Pietro Perona. Presence-only geographical priors for fine-grained image classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

Gengchen Mai, Ni Lao, Yutong He, Jiaming Song, and Stefano Ermon. Csp: Self-supervised contrastive spatial pre-training for geospatial-visual representations. *International Conference on Machine Learning*, 2023. doi: 10.48550/arXiv.2305.01118.

Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and F Prabhat. Deep learning and process understanding for data-driven earth system science. *Nature*, 566(7743):195–204, 2019.

Marc Rußwurm, Konstantin Klemmer, Esther Rolf, Robin Zbinden, and Devis Tuia. Geographic location encoding with spherical harmonics and sinusoidal representation networks. *arXiv preprint arXiv:2310.06743*, 2023.

Oliver Watt-Meyer, Gideon Dresdner, Jeremy McGibbon, Spencer K. Clark, Brian Henn, James Duncan, Noah D. Brenowitz, Karthik Kashinath, Michael S. Pritchard, Boris Bonev, Matthew E. Peters, and Christopher S. Bretherton. Ace: A fast, skillful learned global atmospheric model for climate prediction. *arXiv preprint arXiv: 2310.02074*, 2023.

Benjamin Zuckerberg, Daniel Fink, Frank A. La Sorte, Wesley M. Hochachka, and Steve Kelling. Novel seasonal land cover associations for eastern north american forest birds identified through dynamic species distribution modelling. *Diversity and Distributions*, 22(6):717–730, 2016. doi: https://doi.org/10.1111/ddi.12428. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/ddi.12428`.