75 NRP Big Data
National Research Programme
FNSNF
Swiss National Science Foundation

UNIL | Université de Lausanne
Institut des dynamiques de la surface terrestre

Spatio-Temporal Modeling of Wind Speed Using EOF and Machine Learning

Fabian Guignard[1], Federico Amato[1], Sylvain Robert[2] and Mikhail Kanevski[1]

1. IDYST, Faculty of Geosciences and Environment, University of Lausanne
2. Swiss Re, Zurich
Contact: fabian.guignard@unil.ch

Spatio-Temporal Prediction of Irregularly Spaced Data

Background:

- The interpolation problem of **non-linear** continuous spatio-temporal fields measured on a set of **irregular** points in space, using Machine Learning (ML) is still under-investigated.

Aim:

- Introduce a framework for spatio-temporal prediction of climate and environmental data using ML.
- Show how spatio-temporal processes can be decomposed in terms of a sum of products of temporally referenced basis functions, and of stochastic spatial coefficients which can then be spatially modelled and mapped on a regular grid.
- Discuss an application on a real world dataset consisting of two years of wind speed measurements at hourly frequency collected in Switzerland with different ML models.

Methods

Basis function representation can describe the spatial, temporal and spatio-temporal dependencies in the data.

Here we use the reduced-rank basis obtained through a principal component analysis (PCA), also known as **Empirical Orthogonal Functions** (EOFs), to decompose the data into fixed temporal bases and their corresponding spatial coefficients.

Empirical Orthogonal Functions decomposition

Consider the spatio-temporal observations $\{Z(\mathbf{s}_i, t_j)\}$ at S spatial locations $\{\mathbf{s}_i : 1 \leq i \leq S\}$ and T time-indices $\{t_j : 1 \leq j \leq T\}$.
Let $\tilde{Z}(\mathbf{s}_i, t_j)$ be the spatially centered data,

$$\tilde{Z}(\mathbf{s}_i, t_j) := Z(\mathbf{s}_i, t_j) - \bar{\mu}(t_j), \quad \bar{\mu}(t_j) := \frac{1}{S}\sum_{i=1}^{S} Z(\mathbf{s}_i, t_j),$$

where $\bar{\mu}(t_j)$ is the global spatial mean at time t_j.
The centred data $\tilde{Z}(\mathbf{s}_i, t_j)$ can be represented as

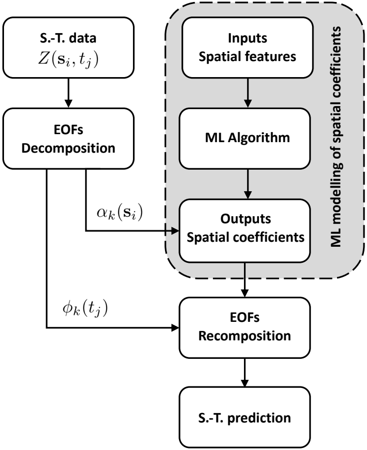
$$\tilde{Z}(\mathbf{s}_i, t_j) = \sum_{k=1}^{K} \alpha_k(\mathbf{s}_i)\phi_k(t_j), \tag{1}$$

where $\alpha_k(\mathbf{s}_i)$ is the coefficient with respect to the k-th basis function ϕ_k at spatial location $\mathbf{s}_i$, and $K = \min\{T, S-1\}$. The scalar coefficient $\alpha_k(\mathbf{s}_i)$ only depends on the location; the temporal basis function $\phi_k(t_j)$ doesn't depend of space.

Machine Learning modelling

The spatial coefficients can be considered as the target variable in a spatial regression problem, solved with ML techniques. In this work we will use :

- Extreme Learning Machine (ELM)
- K-Nearest Neighbours (KNN)
- Random Forest (RF)
- Support Vector Regression (SVR)
- General Regression Neural Networks (GRNN)



Case Study: Wind speed prediction in Switzerland

Here we will present the interpolation of wind speed data from 209 stations in Switzerland from 2017 to 2018 measured at 1 hour sampling period. Data were divided into train (166 stations) and test (43).

A 13-d input space has been designed including the geographical space (latitude, longitude and elevation) and features derived by applying filters and derivatives on the digital elevation model.

Wind phenomenon at such frequency and in complex mountainous region is known to be very difficult to model.

Case Study: wind speed prediction in Switzerland

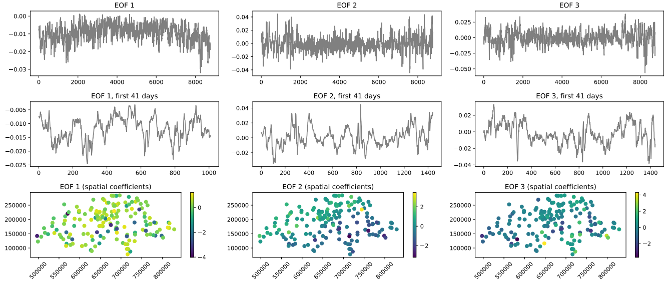


Figure 1: **EOF decomposition:** The first 3 components of the wind speed data EOFs decomposition. Top row : The temporally referenced basis functions. Center row : Temporal basis function for the first 42 days. Bottom row: The standardized spatial coefficients of the corresponding EOFs. Only the first 70 components will be kept for signal reconstruction, corresponding to 95% of the variability of the original data.

Example of prediction using Extreme Learning Machines

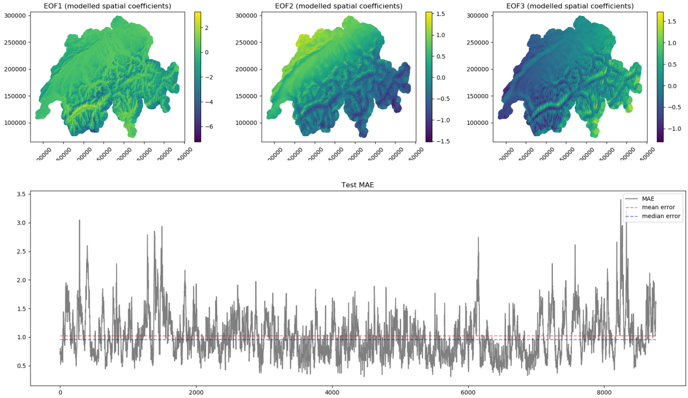


Figure 2: **Example of the output obtained with ELM modelling:** Top row : Predicted spatial coefficient maps for the first three components. Bottom: Test MAE over the two years investigated.

Example of prediction using Extreme Learning Machines

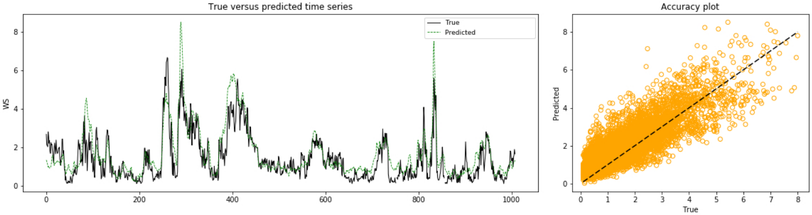


Figure 3: **Example of the output obtained with ELM modelling:** Left: The true time series (in black) at a testing station marked by a cross in the map below, and the predicted time series with the first 70 EOFs components (in green). For visualization purposes, only the first 42 days of the time series are shown. Right: Accuracy plot at the same testing station.

Comparison between different ML algorithms

Model	MAE (mean ± sd. dev.)
ELM	$1.010 \pm 9.5 \times 10^{-3}$
KNN	$0.968 \pm 2.2 \times 10^{-16}$
RF	$1.032 \pm 2.2 \times 10^{-3}$
SVR	$0.960 \pm 3.3 \times 10^{-16}$
GRNN	$1.013 \pm 2.2 \times 10^{-16}$

Table 1: **Comparison among different ML models:** The table shows the mean and standard deviation of the Mean Absolute Error averaged over the entire study period computed after 20 runs of each ML algorithm.

Discussion and conclusions

- We introduced a framework for spatio-temporal prediction of climate and environmental data using ML.
- The decomposition of the spatio-temporal signal into fixed temporal bases and stochastic spatial coefficients permits to fully reconstruct spatio-temporal fields starting from spatially irregularly distributed measurements.
- The spatial prediction of the stochastic coefficients can be performed using any ML algorithm.
- A promising direction for further research is to develop procedures to quantify the propagation of uncertainty through the diverse steps of the proposed framework.

75 NRP **Big Data**
National Research Programme

FNSNF
Swiss National Science Foundation

UNIL | Université de Lausanne
Institut des dynamiques de la surface terrestre