Kriging Wind on Pressure Levels to enrich the Statistical Modelling of Aircraft Trajectories

Rémi Perrichon[1], Xavier Gendre[2], Thierry Klein[1,2]

[1] ENAC - École Nationale de l'Aviation Civile, Université de Toulouse, France.
[2] IMT - Institut de Mathématiques de Toulouse, UMR5219, Université de Toulouse, France.

E-mail for correspondence: `remi.perrichon@enac.fr`

Abstract: Additional to the usual dimensions of an aircraft trajectory (longitude, latitude, altitude), it is often valuable to consider weather dimensions when studying a flight. Geostatistics provides powerful methods to associate a weather value to a given point of a trajectory. Using kriging equations allows to predict weather values for any point of the flight and to take uncertainties into account. We present the steps to perform kriging of wind speed values on pressure levels with drift and anisotropy. Focus is made on the spatial dimension.

Keywords: Geostatistics; Kriging; Drift; Anisotropy; Trajectory.

1 Motivation and problem statement

For a given flight, the position of an aircraft is recorded for a finite set of observation times. This indexed set of positions is interesting but may be an incomplete summary of the flight. Indeed, knowing the experienced weather at each observation time may help to better understand the dynamics of fuel consumption or noise emission. The goal of this work is to associate each point of a trajectory with a weather value, so that experienced weather during the flight is a piece of information that can be used in further statistical analyses.

Past weather data are not available at any instant in time (if only for storage reasons). Rather, weather data are processed so that a three-dimensional weather grid is available every hour. Because most flights in Europe last more than an hour, the task of matching weather values typically involves several weather grids as schematized in Figure 1.

This paper was published as a part of the proceedings of the 37th International Workshop on Statistical Modelling (IWSM), Dortmund, Germany, 16–21 July 2023.

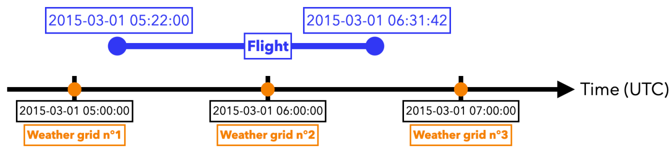

FIGURE 1. Adding weather data to a flight departing from Toulouse-Blagnac (LFBO) and landing at Paris-Orly (LFPO) in March 2015 may involve at least three weather grids.

Formally, this problem is often tackled as an *interpolation* or *spatio-temporal prediction* task. This task is common in environmental sciences as testifies the review of spatial interpolation methods written by Li and Heap (2014). In this work, a focus is made on the spatial aspect of the problem. In other words, a simple rule is adopted for the time dimension: for each point of a trajectory, the closest weather grid in time is used to perform the spatial interpolation. The interpolation problem boils down to a three-dimensional kriging problem involving an unknown drift and anisotropy. The solution is detailed in the sequel.

2 Raw data, scope of the study

Two data sources are used in the paper.
Trajectory data are taken from the R&D data archive that contains more than 18 million flights as of January 2023. The data are collected by Eurocontrol from all commercial flights operating in and over Europe. Data are available for 4 months each year: March, June, September and December.
Weather data are taken from ERA5 hourly data on pressure levels. ERA5 is the fifth generation European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis for the global climate and weather.
We focus on the interpolation of the three weather grids presented in Figure 1. The weather variable of interest is the horizontal wind speed (expressed in $m.s^{-1}$) for the flight departing from Toulouse-Blagnac (LFBO) and landing at Paris-Orly (LFPO) in March 2015. For 23 pressure levels, horizontal wind speed values are given on a $0.25° \times 0.25°$ longitude-latitude grid. The weather grid on which kriging is done is three-dimensional. For a single weather grid, there are 57 (longitude values) $\times$ 41 (latitude values) $\times$ 23 (pressure values) $= 53,751$ wind values.

3 A geostatistical framework

3.1 Dealing with projection and pressure levels

Raw weather data are given on a three-dimensional grid, often called a *region of interest*, commonly denoted D in geostatistics. Projecting is a safe

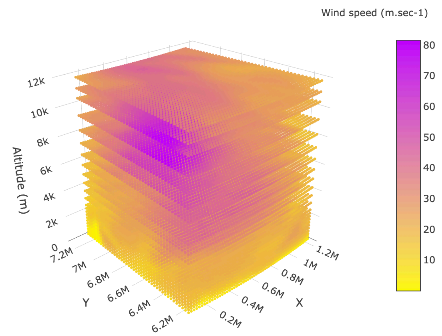

FIGURE 2. Weather grid n°1 giving wind speed values on 2015-03-01 05:00:00.

option when working with spatial data coming in longitude and latitude coordinates. It ensures that all statistical quantities based on the Euclidean distance are accurate. To safely use the Euclidean distance, pressure levels (in hectopascals) must be converted to altitude values in meters.
To go from a pressure level p to an altitude h in meters (m), the following formula is provided by the National Oceanic and Atmospheric Administration (NOAA):

$$h = \frac{145366.45\left[1 - \left(\frac{p}{1013.25}\right)^{0.190284}\right]}{3.281}.$$

It is based on the International Standard Atmosphere (ISA). The resulting grid once the two steps are performed (projection, conversion) is given in Figure 2. The Lambert 93 conformal conic projection is used as it is a very popular option for flights over Metropolitan France.

3.2 Mathematical framework

Every hour, raw data come as a collection of n *regionalized values* denoted $\{z(s_i), i = 1, ..., n\}$. Each location s on D is viewed as the realisation $z(s)$ of a random variable $Z(s)$. Values are said to be regionalized because they exhibit some spatial correlation. The family of real-valued random variables $\{Z(s), s \in D\}$ is traditionally called a *spatial random field*. In the sequel, we assume that the first moment as well as the usual second-order moments of the random field are well-defined. Contrary to usual multivariate statistics, there is only one realization of the random field making inference impossible without some assumptions. Geostatistics often relies on the *second-order stationary* hypothesis. The hypothesis is as follows:

1. The expectation exists and is constant, and therefore does not depend on the location s: $\mu(s) = \mu$.

2. The covariance exists for every pair of random variables, $Z(s)$ and $Z(s+h)$, and only depends on the vector h that joins the locations s and $(s+h)$, but not specifically on them: $C(Z(s), Z(s+h)) = C(h),\ \forall s \in D,\ \forall\ h \in \mathbb{R}^d$ such that $s + h \in D$.

3.3 A drift violates the second-order stationarity assumption

The second-order stationarity assumption doesn't hold for wind data as the mean of the random field depends on location. It is a *drift* problem. This smooth systematic non-random variation should be taken into account. To do so, the random field is broken down into the sum of two components, $Z(s) = \mu(s) + \varepsilon(s)$, where $\mu(s)$ denotes the unknown drift and $\varepsilon(s)$ the stochastic part that can be treated as second-order stationary.
Parametric models to the drift are often fit to detrend the data before attempting the analysis of the spatial correlation structure existing in the residuals. This approach is called *residual kriging* by Montero et al. (2015). This approach has been historically studied by Volpi and Gambolati (1978) through numerical simulations and applied to the mapping of an hydraulic head field of three major aquifers underlying the Venetian lagoon by Gambolati and Volpi (1979). Regarding our application, a quadratic trend has been found to be satisfactory to model the horizontal wind speed drift.
Characterization of the spatial dependence in the residuals relies on the empirical (or experimental) semivariogram. Note that the *variogram* of the random field is defined as the variance of the first differences of the random field:

$$2\gamma(s_i - s_j) = \mathbb{V}(Z(s_i) - Z(s_j)),\ \forall s_i, s_j \in D.$$

The function γ is called the *semivariogram*. In the case of second-order stationarity, the covariance function and the semivariogram are equivalent when it comes to defining the structure of spatial dependence displayed by the phenomenon. One reason for which the semivariogram is preferred to the covariogram is that it does not require the knowledge of the mean of the random field.

3.4 A key aspect: anisotropy

A given empirical semivariogram may not meet the theoretical properties of a valid semivariogram. These theoretical properties are given in most textbooks in geostatistics. The so-called *structural analysis* step is then concerned with the fitting a valid model to the empirical semivariogram. This step is necessary to make valid spatial predictions. Valid models are often *isotropic*. Isotropic covariance functions only depend on the distance between the locations s and $s + h$ as opposed to *anisotropic* ones. Regarding wind data, the dependence between $Z(s)$ and $Z(s + h)$ is obviously a function of both the magnitude and the direction of h. General anisotropy

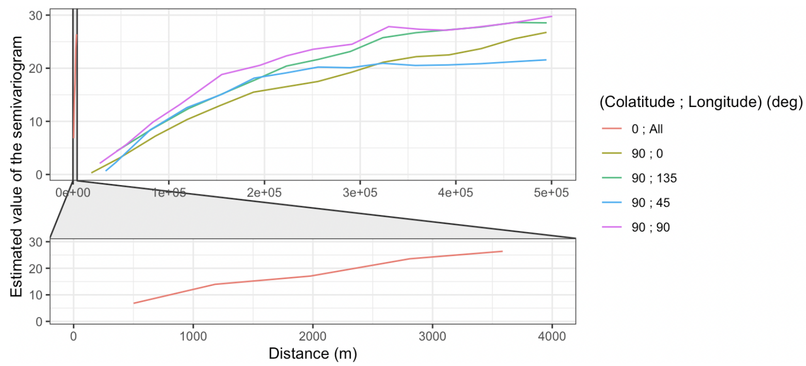

FIGURE 3. Estimated horizontal and vertical semivariograms on the residuals on 2015-03-01 05:00:00. A sample of 10,000 points (drawn at random out of 53,751 locations) is used in the estimation of the semivariograms to improve the computation time. The spatial dependence decreases rapidly in the vertical direction.

models have recently been studied by Allard et al. (2015). In practice, *geometric anisotropy* is the only one that can be corrected using a linear change of coordinates. Indeed, geometric anisotropy is obtained by some stretching of an isotropic model. Speaking in terms of semivariogram, geometric anisotropy is characterized by:

$$\gamma(h) = \gamma_{\text{iso}}(\|Ah\|_2)$$

where the matrix A defines the transformation from the initial space to the isotropic space. A linear transformation of the coordinates in enough to use an isotropic model. As put by Chilès and Delfiner (2012), the matrix A is usually written

$$A = \underbrace{\begin{pmatrix} b_1 & 0 & 0 \\ 0 & b_2 & 0 \\ 0 & 0 & b_3 \end{pmatrix}}_{T} \underbrace{\underbrace{\begin{pmatrix} \cos(\theta_3) & \sin(\theta_3) & 0 \\ -\sin(\theta_3) & \cos(\theta_3) & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{R_{\theta_3}} \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta_2) & \sin(\theta_2) \\ 0 & -\sin(\theta_2) & \cos(\theta_2) \end{pmatrix}}_{R_{\theta_2}} \underbrace{\begin{pmatrix} \cos(\theta_1) & \sin(\theta_1) & 0 \\ -\sin(\theta_1) & \cos(\theta_1) & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{R_{\theta_1}}}_{R}$$

where T is matrix of scaling factors and R a rotation matrix. Estimating anisotropy parameters is usually done with a *directional semivariogram.* In $\mathbb{R}^3$, taking anisotropy into account is key for good predictions because the vertical spatial dependence usually evolves very differently as compared to the horizontal one.

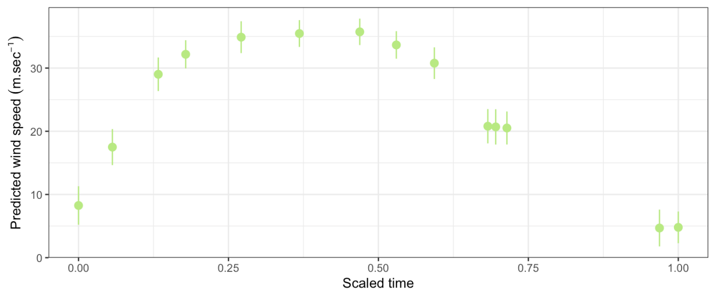

FIGURE 4. Geostatistical predictions of the wind speed values along the raw trajectory.

4 Results

For each grid, the trend is taken into account using Ordinary Least Squares (OLS). The horizontal and vertical semivariograms are then estimated on the residuals. As can be seen in Figure 3, a strong anisotropy should be taken into account, specifically in the vertical direction for which the spatial dependence is rapidly decreasing. Once corrected, predicted values are computed. Predicted wind values for the flight are shown in Figure 4. Note that the 95% confidence intervals only make sense if a Gaussian assumption holds for each weather grid. Confidence intervals are pointwise.

References

Allard, D., Senoussi R., Porcu E. (2015) Anisotropy models for spatial data. *Mathematical Geosciences*, **48** (3), 24 p.

Chilès, J.P. and Delfiner, P. (2012) *Geostatistics, modeling spatial uncertainty*. Wiley Series in Probability and Statistics.

Gambolati, G. and Volpi, G. (1979). Groundwater contour mapping in Venice by stochastic interpolators: 1. Theory. *Water Resources Research*, **15**, 281 – 290.

Li, J. and Heap, A.D. (2014). Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software*, **53**, 173 – 189.

Montero, J., Fernández-Avilés, G., Mateu, J. (2015). *Spatial and Spatio-Temporal Geostatistical Modeling and Kriging*. Wiley Series in Probability and Statistics.

Volpi, G. and Gambolati, G. (1978). On the use of a main trend for the kriging technique in hydrology. *Advances in Water Resources*, **1**, 345 – 349.