Ordinary kriging for on-demand average wind interpolation of in-situ wind sensor data

Zlatko Zlatev, Stuart E. Middleton, and Galina Veres
IT Innovation Centre, University of Southampton, Southampton, SO16 7NP, UK
{zdz, sem, gvv}@it-innovation.soton.ac.uk

Abstract

We have developed a domain agnostic ordinary kriging algorithm accessible via a standards-based service-oriented architecture for sensor networks. We exploit the Open Geospatial Consortium (OGC) Sensor Web Enablement (SWE) standards. We need on-demand interpolation maps so runtime performance is a major priority.

Our sensor data comes from wind in-situ observation stations in an area approximately 200km by 125km. We provide on-demand average wind interpolation maps. These spatial estimates can then be compared with the results of other estimation models in order to identify spurious results that sometimes occur in wind estimation.

Our processing is based on ordinary kriging with automated variogram model selection (AVMS). This procedure can smooth time point wind measurements to obtain average wind by using a variogram model that reflects the wind phenomenon characteristics. Kriging is enabled for wind direction estimation by a simple but effective solution to the problem of estimating periodic variables, based on vector rotation and stochastic simulation.

In cases where for the region of interest all wind directions span 180 degrees, we rotate them so they lie between 90 and 270 degrees and apply ordinary kriging with AVMS directly to the meteorological angle. Else, we transform the meteorological angle to Cartesian space, apply ordinary kriging with AVMS and use simulation to transform the kriging estimates back to meteorological angle.

Tests run on a 50 by 50 grid using standard hardware takes about 5 minutes to execute backward transformation with a sample size of 100,000. This is acceptable for our on-demand processing service requirements.

Keywords

Ordinary kriging, variogram, spatial estimation, spatial interpolation, spatial smoothing, wind velocity, wind direction, periodic variable, in-situ, sensor data.

1. Introduction

In the SANY project [5] we have developed reference implementations of decision support and generalized data fusion services based on an environmental service architecture that is compliant with the Open Geospatial Consortium (OGC) Sensor Web Enablement (SWE) standards [6]. Data from existing sensor networks, emerging ad-hoc sensor networks, virtual sensors and so on are collected in sensors observations services (SOS) and can be pulled on demand and directed towards a web processing services (WPS).

We have been developing generic processing services to provide SANY with a fusion capability that is domain agnostic, not restricted to a particular phenomenon or set of dataset characteristics. As a case study we have looked at providing services to support risk assessment and decision making when tracking bathing water pollution at beaches. For these risk assessment and decision making tasks proprietary physical models are fused with average wind velocity estimates on a spatial grid.

Also, output from complex non-parametric models may be compared against spatial estimates from in-situ sensors in order to identify spurious results, which may sometimes occur in those more complex models. For example in [1] a number of reasons are stated that cause spurious results when estimating average wind speed and direction with Volume Imaging Lidar (VIL). Suspected VIL estimates may be compared against interpolated values from in-situ sensors of an existing sensor network covering the target area.

We haven’t been able to find readily available generic work and procedures to address our requirements for obtaining average wind estimates on a spatial grid on-demand from time point wind measurements. We developed ordinary kriging with automated variogram model selection (AVMS) to provide on-demand maps of interpolated values of average wind velocity for any given moment in time.

Our case study involves the study of wind velocity measurements from observation stations covering an area of approximately 200km by 125km. The sensor data is located in a database and accessed via a SOS, making use of the SWE metadata to read in self-described sensor datasets for the requested time period. We provide kriging via a WPS and our result sets are also formatted according to the SWE standards, making them also self-described. We have an initial, manual configuration for a new dataset by a domain expert based on an analysis of the dataset characteristics. After that our service uses the self-described nature of the input data to automatically configure itself to the observed phenomenon required and associated characteristics.

Section 2 outlines our sensor data characteristics, the challenges faced when handling periodicity and the solutions we adopted. Section 3 describes in detail our algorithm for ordinary kriging with AVMS.

2. Interpolating periodic data

Wind direction is represented by its meteorological angle, which is periodic. Most statistical estimation techniques are not directly fit for estimating periodic variables. For periodic variables, values at the period beginning and end are contextually close but numerically distant. An approach to resolve this disagreement could be to transform the periodic variable to Cartesian space, obtain estimates of the Cartesian components and then perform backward transformation to estimates in the original periodic space. The backward transformation shall be obtained as the arctangent of the ratio of the Cartesian components’ estimates, which are generally assumed to be Gaussians. The Gaussian ratio distribution has no simple analytical solution [3]. The resulting arctangent distribution can be multimodal and with undefined mean and variance making a simple backward transformation impossible.

A method based on an artificial neural network (ANN) with a mixture of periodic or Euclidean Gaussian kernel functions has been proposed by Bishop and Legleye [4] for the estimation of densities of periodic variables. However, applying estimation methods based on ANN on ad-hoc bases and on a spatial grid, e.g. 50 by 50, can be very time consuming which compromises the on-demand interpolation service requirement. In our ordinary kriging with AVMS implementation for wind direction spatial interpolation we tackle the problem of estimating a periodic random variable with a simple but effective solution based on vector rotation and Cartesian transformation with stochastic simulation.

Ordinary kriging estimates at unobserved locations are obtained from measurements at observed locations and a variogram, which models spatial dependences. When creating the variogram numerically close measurements point to high spatial dependence and vice versa. Wind directions' meteorological angles at the period's beginning and end are close in terms of direction but are distant numerically, which makes the variogram creation direct from the observed wind direction angle inappropriate.

We observed in our test datasets, containing observations for 17 locations over an area of 200km by 125 km for 5 years, about 80% of the wind direction angles spanned less then 180^{o}. In cases like these, before creating a variogram we will only need to rotate the wind directions in a way that the period onset doesn't intersect the wind directions span. This will eradicate the periodicity problem as the numerical distance between the wind direction angles will correspond to the distance between the wind directions and will enables variogram building and kriging. Figure 1 depicts this idea: the original wind directions (solid line) in quadrants 1 and 4, are rotated (dashed line) to lie in quadrants 2 and 3. As kriging smoothes extreme values the interpolated values are expected to be within the extreme wind direction angle values. With this transformed data a variogram is created and ordinary kriging used to obtain estimates. The kriging estimates are rotated backwards to obtain the final results, as depicted in Figure 2.

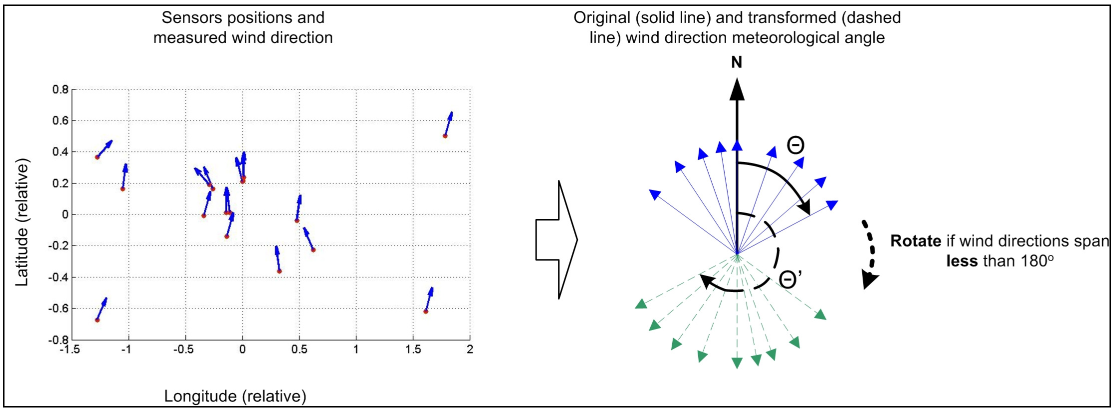

Figure 1. Transforming observed wind direction meteorological angles when observed wind directions span less then 180^{o}.

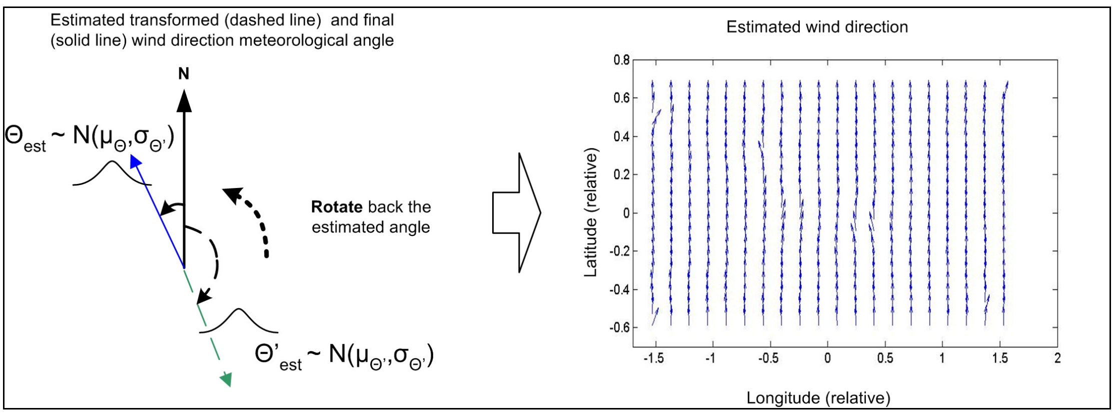

Figure 2. Transforming estimated wind direction meteorological angles when observed wind directions span less then 180°.

For cases where the observed wind directions span more then 180°, we transform the meteorological angle to Cartesian space and perform kriging on the transformed data. For the backward transformation, from Cartesian components estimates to meteorological angle, we use stochastic simulations and calculate the mode and 0.025 & 0.925 percentiles from the simulated sample. Figures 3 and 4 show respectively the forward (before variogram building and kriging) and backward (after variogram building and kriging) data transformation.

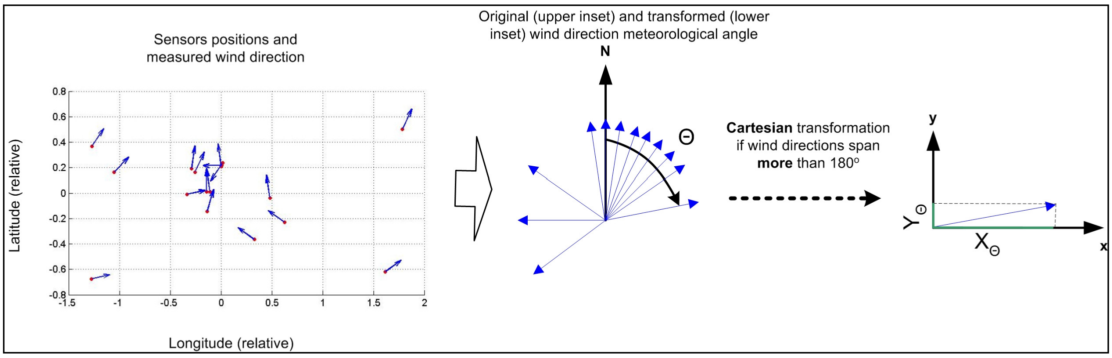

Figure 3. Transforming observed wind direction meteorological angles when observed wind directions span more then 180°.

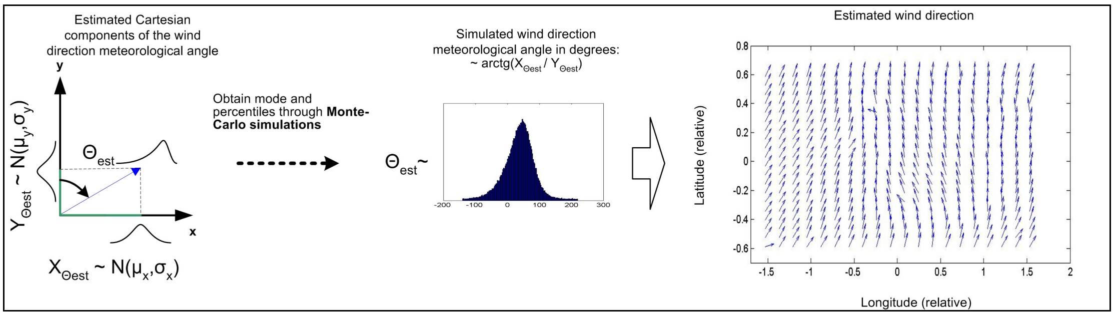

Figure 4. Obtaining estimated wind direction meteorological angles when observed wind directions span more then 180°.

We expect that for the majority of the cases only direction rotation will be needed. The simulation approach was tested on a 50 by 50 grid using standard hardware. Backward transformation with sample size of 100,000, allowing results to approach direct angle kriging results, takes on average about 5 minutes. Thus, the service on-demand requirements for the WPS interpolation service are not compromised.

3. Ordinary kriging with automated variogram model selection

Traditional in-situ wind speed and direction measuring instruments provide point measurements at certain locations in space and time points. For environment monitoring applications we are interested in area or volume average wind speed and direction values. Generally, point measurements in the convective boundary layer do not provide representative average wind estimates. This is because point measurements often include contaminations from large eddies, roll vortices, and topographic influences, which causes them to differ from the average values [1]. Time averaging often cannot compensate temporal fluctuations without losing some information about the changing value of the average wind [1]. Therefore, for wind interpolation from point measurements we need a method that accounts for its phenomenology and smoothes the fluctuations accordingly.

The wind velocity (speed and direction) power-spectrum will depend on the geographical area and the surface topology but in general wind has a slowly varying component and a rapidly fluctuating component. The slowly varying component can be considered as current average wind velocity level, relatively constant over a period of time of several hours. The rapidly fluctuating component is with a period of fluctuations from seconds to minutes. To extract the average wind velocity vector a time series power-spectrum analysis can be performed. However, for the purposes of the SANY project it is required that the wind spatial interpolation service is to work on ad-hoc bases (i.e. dataset agnostic processing) on a given time-slice data. Consequently, the power spectrum analysis from temporal data may not be performed but instead we needed to use a smoothing procedure that can utilise background information about the phenomenon.

For spatial interpolation our WPS provides ordinary kriging with automated variogram model selection (AVMS). Background information describing the phenomenon characteristics can be reflected by the variogram model used for kriging. For the ordinary kriging with AVMS, along the sensor data, metadata is supplied that impose constraints to the variogram model to be selected in a way that reflects the phenomenology of the interpolated phenomenon. Figure 5 depicts the different stages of the ordinary kriging with AVMS interpolation procedure.

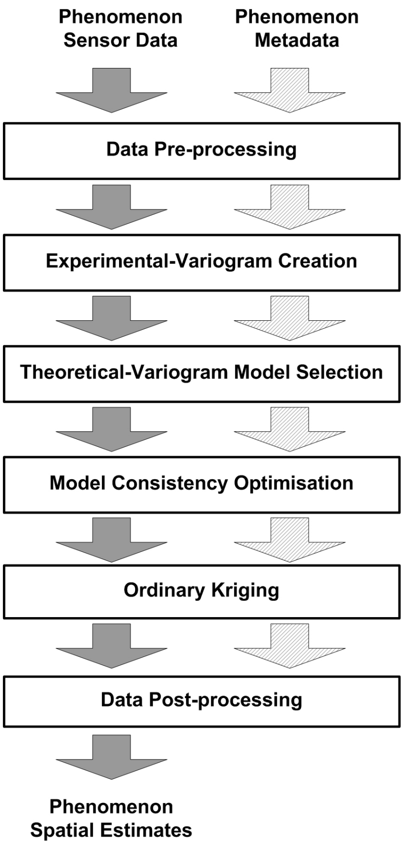

Figure 5. Ordinary kriging with automated variogram model selection procedure.

The first stage is the data pre-processing stage, where data cleaning, normalisation and necessary data transformations are performed (e.g. forward transformations necessary for estimating a periodic variable discussed in Section 2). Accordingly, in the last stage, data post-processing, data de-normalisation and reverse transformations are performed (e.g. backward transformations necessary for estimating a periodic variable discussed in Section 2). The core ordinary kriging with AVMS stages are the experimental-variogram creation, theoretical-variogram model selection, model-consistency optimisation and ordinary kriging.

The most critical part of the experimental-variogram creation stage is the selection of lags. Lags need to be selected so they contain an optimal number of points in a way that physical phenomenon characteristics are not smoothed out but that noise is not modelled. Generally the initial slope of the variogram needs to be well estimated so the first few lags shall contain smaller number of points. If no hole-effect is expected the following lags may contain a large number of points, but if hole-effect is expected the lags shall contain lower number of points so the effect is not smoothed out. The relative number of points in a lag is specified in the metadata supplied to the interpolation procedure. This relative number can be set by a phenomenon expert or pulled from an expert system listing known phenomena (an expert system is currently under design).

The next stage is the theoretical-variogram model selection. Currently, we have eight models implemented: spherical, exponential, Gaussian, linear, power, generalised Bessel, sine hole-effect and cosine hole-effect. The model shape is governed by a subset of the following parameters: nugget, range, power, hole and sill. We use least-squares fitting method to select a model that best fits the experimental variogram. We can introduce background information about the phenomenon by constraining the fitted model types and the parameter values and then, in effect, a variogram model reflecting the characteristics of the phenomenon of interest will be selected.

For wind velocity we expect the point measurements of wind velocity to have low spatial correlation because of the rapidly fluctuating component previously discussed, so we set the upper bound for the nugget parameter to be relatively high, e.g. 1/3 of the sill. Also, we expect spatial correlation of the average wind velocity to decrease very slowly with increasing the distance, so we set the lower bound of the range parameter to be relatively high (so relatively more observations influence the estimate), e.g. 1/3 of the maximum distance. Next, we don't expect a hole effect, so the models with hole effect are not selected. As the observed area is relatively large we expect that the phenomenon causes are to change within it and so the variogram should approach a horizontal asymptote, a sill, at some distance, so the power model is not selected.

In general, nugget reflects the magnitude of rapid fluctuations (assuming that the measuring error is negligible in comparison to these fluctuations), range reflects the scale of the phenomenon (for wind possibly reflects the surface topology too) in comparison to the size of the observed area, the presence and the magnitude of sill depends on the phenomenon scale in relation to the size of the observed area, and finally hole effect exists or not depending on the essence of the phenomenon and the characteristics of the environment (for wind hole effect may be caused by the surface topology).

We are positive that phenomenon experts will be able to select models and parameters' boundaries that reflect the characteristics of the phenomenon of interest.

After selecting the theoretical-variogram model, model parameters optimisation is performed in order to improve the internal consistency of the model. Kitanidis [4] suggests two statistics, Q_1 and Q_2, that need to be as close to their expected values as possible in order for the model to be consistent with the ordinary kriging inductive bias. We use quadratic-sequential-programming to tune the model parameters, subject to the parameter constraints discussed above, with a loss function proportional to the squared differences between Q_1 and Q_2 and their respective expectations. Additionally, we try to keep the nugget of the model minimal but as a secondary objective, i.e. it is included in the loss function with a smaller weight. Unrealistically high nugget will over-smooth the estimates. Kriging over-smoothing will cause a very high standard deviation in an estimate.

After the variogram model optimisation stage standard ordinary kriging is performed using the optimised variogram model and estimate's mean and standard deviation are computed. We take the mean as the area average wind. In general, the produced estimate standard deviation by kriging shall be utilised depending on the purpose of the interpolation. When kriging is used for spatial smoothing the use of the estimates' standard deviation can be subtle. We take the area average wind to be the mean of the estimated wind at a particular time point, given the correct variogram model for the phenomenon. The produced standard deviation by kriging is pertinent to the distribution of the estimated wind at a particular time point but not the estimated mean, taken to be the estimated area average wind. A very high kriging standard deviation indicates possible over-smoothing and the estimation results should be questioned. A very low standard deviation indicates that the estimate was possibly influenced by a spatially close observation, i.e. it is not an average value but a time point value, which according to the wind phenomenology is affected by the rapid fluctuations component. In such cases the estimate may be rejected and re-estimation performed with the spatially closest observation ignored. Currently we have not experimented with this idea but it seems viable.

When using ordinary kriging for spatial smoothing it is not possible to perform experimental procedure validations such as cross-validation. We can only make sure that the internal structure of the method is as consistent as possible - i.e. theoretical model optimisation is performed. In addition we have performed numerous tests with real data and visually inspected the results for presence of spurious results.

4. Conclusions.

We have implemented ordinary kriging with AVMS as an OGC compliant web processing service (WPS). It is a generic fusion service for grid interpolation on-demand from in-situ sensor data, self-configuring itself based on the OGC SWE self-described input data provided.

The ordinary kriging with AVMS takes phenomenon metadata as an input and selects a variogram model that reflects the characteristics of the phenomenon of interest, making the service generic. This also enables using the service as a spatial smoothing procedure as it is in the case of estimating average wind from time point wind measurements.

We require on-demand results to provide input into a bathing water quality model that allows decision support regarding the water quality at beaches. We thus have a simple yet effective solution based on vector rotation and stochastic simulation for estimating periodic random variables, which produces practical on-demand execution speeds for spatial grid estimates. Running on standard hardware, simulation with sample size of 100,000 to obtain estimates on a grid 50 by 50, takes about 5 minutes to execute.

References

1. Piironen AK, *Analysis of Volume Imaging Lidar Signals, Ph.D. Thesis*, University of Joensuu, Department of Physics, Vaisala Laboratory, Finland, 1994.

2. Kitanidis RK, *Introduction to Geostatistics: Applications to Hydrogeology*, Cambridge University Press, New York, 1997.

3. Marsaglia G, Ratios of normal variables and ratios of sums of uniform variables, *J. Amer. Statis. Assoc. 60*, p. 193-204, 1965.

4. Bishop CM, and Legleye, C., *Estimating Conditional Probability Densities for Periodic Variables*, Neural Computing Research Group, Department of Computer Science and Applied Mathematics, Aston University, Birmingham, 1994.

5. SANY project, website: *sany-ip.eu*

6. OGC SWE, Open Geospatial Consortium Sensor Web Enablement working group, website: *http://www.opengeospatial.org/projects/groups/sensorweb*