Spatio-temporal Kriging of Lower Caribbean Wind Data

Neil Ramsamooj

Faculty of Engineering, The University of the West Indies, Trinidad and Tobago, West Indies;
Email: Neil.Ramsamooj@sta.uwi.edu

Abstract: *Planning of a wind farm location requires significant data. However, wind speed data sets in the lower Caribbean are usually incomplete. This paper considers imputation by spatio-temporal kriging using data from neighbouring locations. Temporal basis functions with spatial covariates are used to model diurnal wind speed cyclicity. The residual set of our spatio-temporal model is modelled as a Gaussian spatial random field. Fitted models may be used for spatial prediction as well as imputation. Examples of predictions are illustrated using two months of hourly data from eight Caribbean locations with prediction accuracy being assessed by cross validation and residuals.*

Keywords: *Spatio-temporal modelling, wind speed, unbalanced data, imputation, temporal basis functions, kriging*

1. Introduction

Renewable energy (RE) sources are potential means of mitigating the reliance on petroleum and natural gas in the Caribbean. Wind energy has been identified regionally (Elliott *et al.*, 1987; Renewable Energy Committee, 2011) as a particularly viable RE source. The development of wind farms requires significant expenditure and therefore it is crucial to identify suitable wind farm locations in order to optimise energy production and reliability while minimising capital and working costs. Prediction methods for wind characteristics at a target wind farm location with minimal data from one or more reference locations are known in the wind engineering literature as measure-correlate-predict (MCP) algorithms (Carta, Velázquez and Cabrera, 2013). Initial MCP methods employed linear or quadratic regression techniques with data from a single reference location. Later MCP algorithms utilised multiple or vector regression methods with data from several reference locations.

Wind data is typically non-stationary and auto-correlated which implies that differencing or co-integration adjustments are required in multiple regression MCP algorithms. (Hunt and Nason, 2001) used a wavelet packet transform method to construct an MCP algorithm which accounts for non-stationarity; this method was only applied to a single reference location but may be extended to multiple reference locations. The above methods do not account for spatial correlation among multiple locations and do not provide a transparent method of trend removal. In this paper, we use the spatio-temporal model.

$$y(s,t) = \sum_{i=1}^{m} \beta_i(s) f_i(t) + v(s,t) \qquad (1)$$

where $y(s,t)$ is wind speed at location $s \in R^2$ specified by longitude and latitude coordinates and time $t \in R$. Temporal basis functions $\{f_i(t): i = 1, ..., m\}$ are used to capture the nonstationarity of wind data. The theory of spatial temporal basis functions is discussed in (Fuentes, Guttorp and Sampson, 2006) and (Wikle, Zammit-Mangion and Cressie, 2019). Hunt and Nason also use basis functions but our method is clearer as it does not rely on wavelet theory. The spatial fields $\beta_i(s)$ of our model (1) are readily tuned (by choosing latitude and longitude as geographical covariates) to allow for varying seasonality among the reference locations.

Similarly, there is flexibility in the choice of the covariance structure of the spatial random field in order to improve model fit. Furthermore, this model can be extended to include exogenous covariates (such as temperature and air pressure) in order to improve wind prediction accuracy (see Section 4). The model (1) is defined in Section 2.3. A case study illustrating spatio-temporal prediction of wind speeds using two months of hourly data from eight reference locations (shown in Figure 1) is given in Section 3. The prediction accuracy for this case study is assessed in Section 3.3 by the use of cross-validation groups. The R code for the case study is shown in Appendix A. A similar model was used by (Lindström *et al.*, 2013) in a pollution study of NO_x concentrations in metropolitan areas in the United States. Some code segments used in our case study are due to Lindström *et al.* (2013).

2. Problem Formulation
2.1 Data Sets

The wind speed data sets used in this paper were obtained as CSV files from the Iowa Environmental Mesonet (2021) and the National Oceanic and
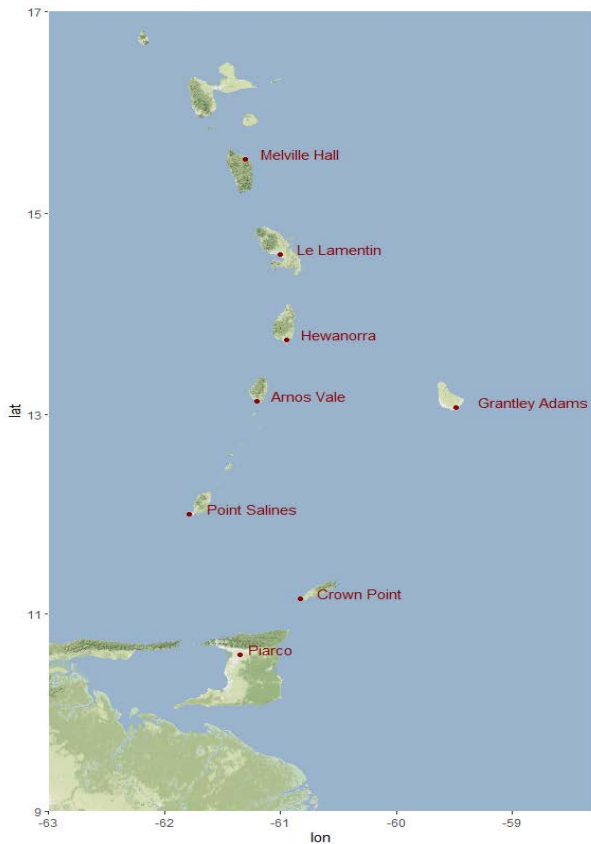
allows (as an example) for the imputation of missing Piarco data by the use of data from Point Salines and other nearby weather stations. Furthermore, kriging allows for wind speed prediction at any location within the lower Caribbean region.
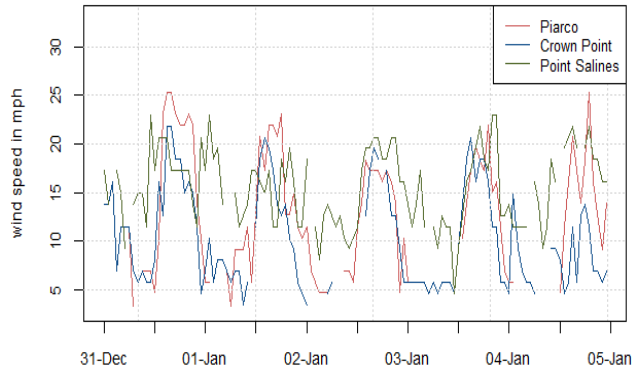

Figure 1. Eight reference locations for case study

Figure 2. Five day subset of hourly wind speed observations (from a two month data set)
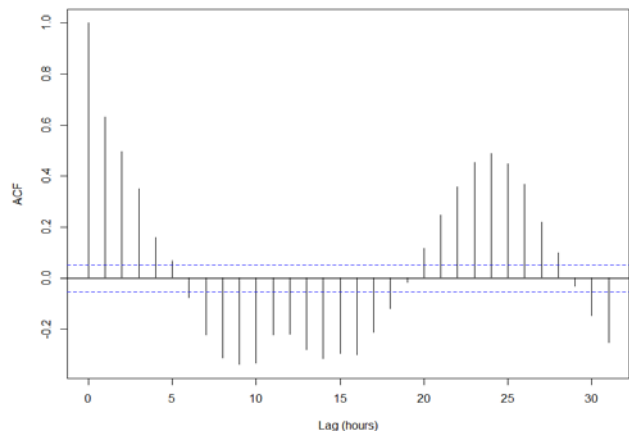

Figure 3. Percentages of missing values

Figure 4. Sample autocorrelation function of Piarco wind speed data

Atmospheric Administration (NOAA, 2021) and originated from Meteorological Terminal Aviation Routine Weather Report (METAR) (National Weather Service, 1998) hourly records from January to February 2015. The web addresses of the download interfaces of the Iowa Mesonet and NOAA are given in the references above. Iowa Mesonet data may also be accessed within R using the package riem (Salmon, 2016) and NOAA data may similarly be obtained from the R package rnoaa (Chamberlain, 2021). A code example using rnoaa is given in Appendix A.1. Note that the METAR wind speed measurements have a resolution of only knot.

Observations from eight weather stations (located at airports) in the lower Caribbean were used in our analysis. Boxplots were used to remove outliers from each weather station data set. Figure 2 shows time series plots of hourly wind speed observations for five days from the three of these locations: Piarco (Trinidad), Crown Point (Tobago) and Point Salines (Grenada). Note that these plots only show a small subset of our entire data set which consists of two months of hourly data from eight locations. A plot of this larger data set is given in Appendix A.1.

From Figure 2, we may observe that Point Salines and Piarco wind speeds are correlated. Note also that there is more available data at Point Salines than at Piarco (see Figure 3). The kriging method of this paper

The wind speeds illustrated in Figure 2 show a common diurnal cyclicity which is modelled by the use of temporal basis functions discussed in Section 2.3. A multivariate ARCH test (see (Tsay, 2013)) confirms that our wind speed data is heteroskedastic and so the data was log transformed in order to stabilise the variances; see (Paciorek *et al.*, 2009; Sampson *et al.*, 2011) for applications of spatio-temporal modelling with log transformed data.

2.2 Kriging

Let $D \subseteq \mathbb{R}^2$ be a domain in the plane and let $s \in D$ be a generic location. A *spatial random field* is a family of random variables which is indexed by the locations s in the domain D, see (Cressie, 1992). As an example, $Z(s)$ may model wind speeds at varying locations s at some fixed time.

$$\{Z(s) : s \in D\} \tag{2}$$

The *expected value* $\mu(s) = E(Z(s))$ of the random field $Z(s)$ may vary with location s. The *covariance function* $C(s_i, s_j)$ of the random field $Z(s)$ is defined as

$$C(s_i, s_j) = E\left((Z(s_i) - \mu(s_i))(Z(s_j) - \mu(s_j))\right) \tag{3}$$

and measures the dependence of Z between different spatial locations s_i and s_j.

Consider the case for which Z may only be sampled at known locations $s_1, s_2, \ldots, s_n$ (for example, wind data will usually only be available at a few weather stations in a given region). *Kriging* is the determination of a best unbiased linear estimator of Z at an unknown location s_0 in terms of random variables $Z(s_i)$ at known locations. In the case of *simple kriging* (when the random field $Z(s)$ is second order stationary with known constant mean $\mu(s) = \mu$) then the estimator at an unknown location s_0 is

$$Z(s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i) + \left(1 - \sum_{i=1}^{n} \lambda_i\right)\mu \tag{4}$$

where the *kriging weights* λ_i are obtained (see (Montero, Fernández-Avilés and Mateu, 2015)) by solving a linear system of n equations

$$\sum_{j=1}^{n} \lambda_j C(s_i, s_j) = C(s_i, s_0) \qquad i = 1, \ldots, n \tag{5}$$

with covariance coefficients. In the case of *universal kriging*, the mean $E(Z(s))$ is not assumed to be constant but is a linear combination of spatial functions $f_i(s)$. The kriging weights are again obtained from a linear system; see (Cressie, 1992). As an example, the functions $\beta_i(s)$ (used in our spatio-temporal model for wind speed given in Equation (1)) are spatial random fields with universal kriging structure in which the nonconstant means are assumed to be linear combinations of longitude and latitude.

A *spatio-temporal random field*

$$\{Z(s, t) : s \in D, t \in T\} \tag{6}$$

is a family of random variables indexed by locations s in a spatial domain D and by times t in a time interval T. The development of spatio-temporal kriging is similar to the cases of (spatial) kriging above where time may be regarded as a spatial dimension with the caveat that the spatio-temporal covariance functions (used to determine kriging weights) consist of distinct spatial and temporal components; see (Wikle, Zammit-Mangion and Cressie, 2019) for further details. The determination of a suitable covariance structure of a spatio-temporal random field is first approached by calculating the *sample spatio-temporal semivariogram*

$$\hat{\gamma}(h, \tau) = \frac{1}{2|N(h, \tau)|} \sum_{N(h, \tau)} \left(Z(s_i, t_i) - Z(s_j, t_j)\right)^2 \tag{7}$$

where $N(h, \tau)$ is the set of pairs of spatio-temporal locations (s_i, t_i), (s_j, t_j) that satisfy $s_i - s_j = h$ and $t_i - t_j = \tau$; see (Montero, Fernández-Avilés and Mateu, 2015; Wikle, Zammit-Mangion and Cressie, 2019). The calculated semivariogram is then fitted to a suitable covariance model. A wire plot of a sample semivariogram of hourly wind speeds from the Caribbean wind data set is shown in Figure 5. The R code for obtaining the semivariogram and directional spatial semivariogram (discussed below) is shown in Appendix A.1. Note that the cyclicity along the time axis corresponds to alternating positive and negative correlation with varying time lags (Pyrcz and Deutsch, 2014); this is supported by the cyclic sample autocorrelation function shown in Figure 4.

The variability of wind speeds in the north/south and east/west directions may be examined with the use of a *directional* spatial semivariogram

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{N(h)} \left(Z(s_i) - Z(s_j)\right)^2 \tag{8}$$

where $N(h)$ is the set of pairs of spatial locations that satisfy $s_i - s_j = h$ where the separation vectors h are partitioned into direction subsets; see (Pebesma, 2021). We first randomly sample 400 time instances from our spatio-temporal wind speed data frame and merge these into a single spatial data frame (with a time index) using the procedure from (Pebesma and Gräler, 2021). The directional spatial semivariogram shown in Figure 6 is formed by partitioning the direction vectors $h = s_i - s_j$ (for pairs of locations s_i, s_j) according to the directions of 0°, 45°, 90° and 135° that are measured clockwise from north. The direction h of a pair of locations is associated to the nearest of these four directions and so each point in the 45° panel in Figure 6 corresponds to a pair of locations with a relative direction that lies between 22.5° and 67.5°.

The variability shown in the 45° panel indicates that our wind data set contains information required to model variations in the northeast direction (each point in the

45° panel has a time resolution of 1391 hours). Variability in the east/west direction may also be examined by considering the variation of trend components with longitude as shown in Figure 9. This longitudinal trend is considered further in Section 2.3.
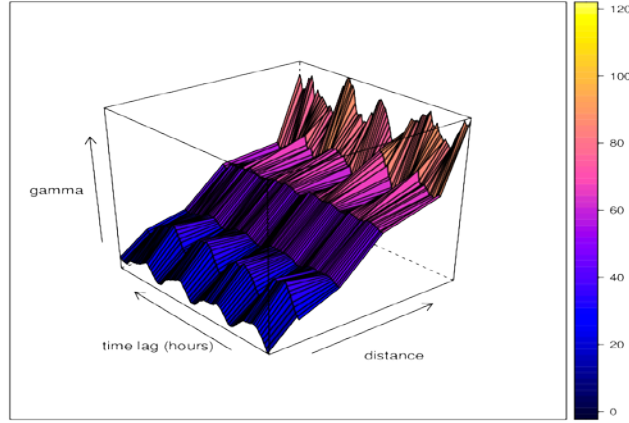

Figure 5. Sample semivariogram of hourly wind speeds from lower Caribbean data set during January and February 2015
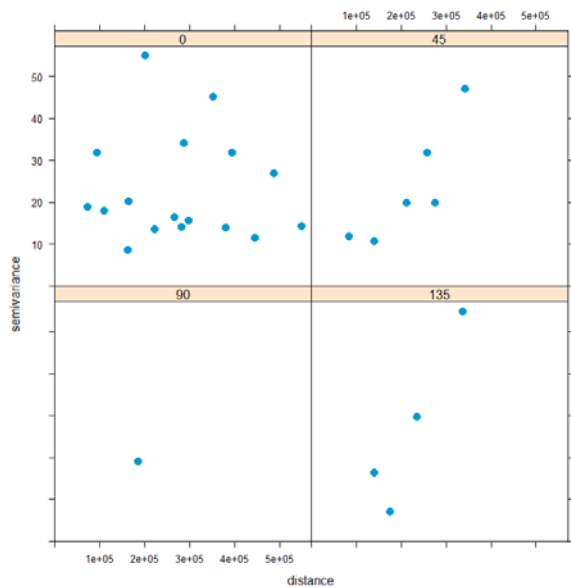

Figure 6. Directional spatial semivariogram of hourly wind speeds during January and February 2015

2.3 A Spatio-temporal Model

Wind speed $y(s,t)$ at location $s \in \mathbb{R}^2$ specified by longitude and latitude coordinates and time $t \in R$ is modelled as

$$y(s,t) = \sum_{i=1}^{m} \beta_i(s)f_i(t) + v(s,t) \qquad (9)$$

where $\{f_i(t): i = 1, \dots, m\}$ is a set of temporal basis functions with $f_1(t) \equiv 1$ and where $f_i(t)$ ($2 \leq i \leq m$) are

used to model non-stationary seasonal temporal behaviour. The basis functions $f_i(t)$ used in this paper are computed from incomplete data by using a singular value decomposition (Fuentes, Guttorp and Sampson, 2006; Lindström *et al.*, 2013). Other basis functions (Fourier, wavelet or Wendland) may be used (Wikle, Zammit-Mangion and Cressie, 2019). The number m of basis functions required to model temporal seasonality are determined by cross validation using the statistics AIC (Akaike information criterion), BIC (Bayesian information criterion), MSE (mean squared error) and R^2. A plot of these statistics (in the case of our Caribbean wind data) for an increasing number of basis functions is shown in Figure 7.
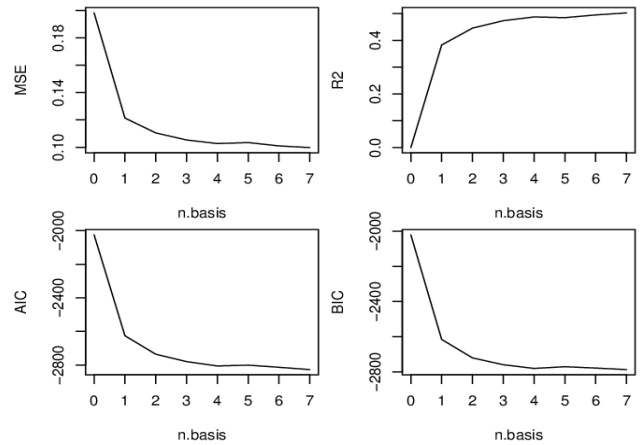

Figure 7. Facet plot of cross validation statistics for number of basis functions

In our case of strongly autocorrelated wind data together with persistent diurnal cyclicity (see the sample autocorrelation function in Figure 3), $m = 5$ temporal basis functions are used. A large number of basis functions is computationally impractical (particularly during cross validation) as parameter estimates are obtained by discrete optimisation routines. The basis functions are shown in Figure 8.
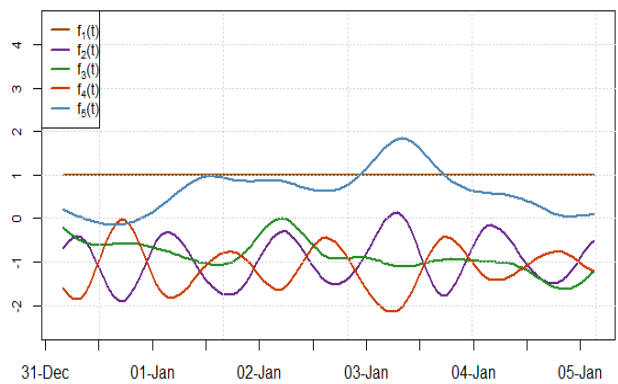

Figure 8. Temporal basis functions

The coefficients $\beta_i(s)$ in Equation (1) are spatial fields which allow for varying seasonal behaviour among sites. A plot of $\beta_1(s)$ versus latitude (at observed locations) together with a line of best fit is shown in Figure 8; these values are mean wind speeds as the corresponding temporal basis function $f_1(t) \equiv 1$. A similar plot of $\beta_1(s)$ versus longitude is shown in Figure 9. Figure 10 indicates a small latitudinal trend (which physically corresponds to a small decrease in wind speed with distance from the South American coast). It illustrates a significant longitudinal trend which corresponds to a nontrivial increase in mean wind speed in the eastern direction. This longitudinal trend is present in the wind speed predictions shown in Figure 12.
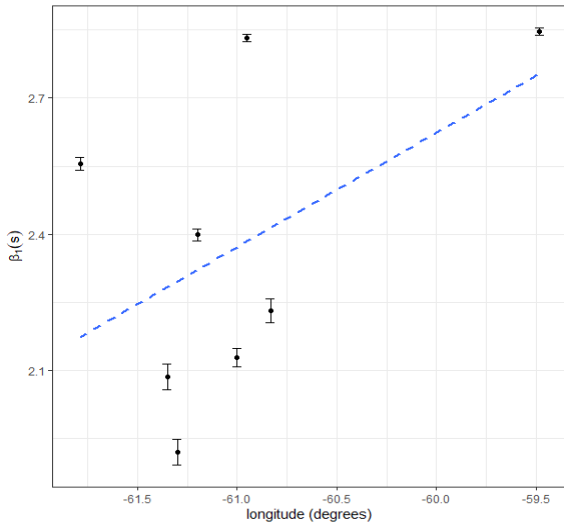

Figure 9. Longitudinal trend of spatial field estimates at station locations (with confidence intervals) together with a line of best fit
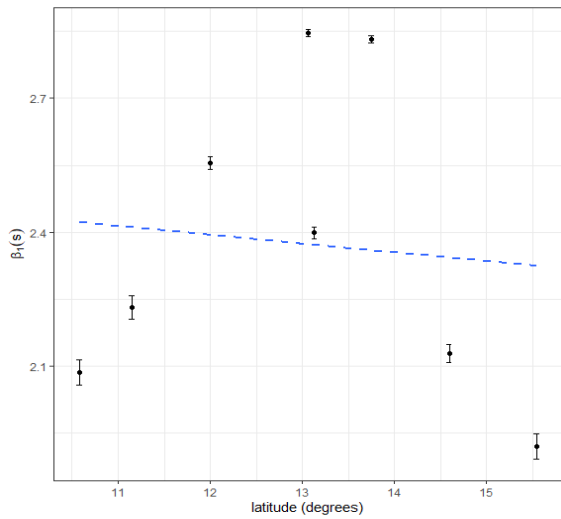

Figure 10. Latitudinal trend of spatial field estimates at station locations (with confidence intervals) together with a line of best fit

A covariate structure of the form

$$\beta_1(s) = a_1 + b_1 x + c_1 y + \eta(s) \qquad (10)$$

where $s = (x,y)$ and x,y represent longitude and latitude indicates that the trend component of the spatial field $\beta_1(s)$ is modelled as a linear function. Similar plots of $\beta_i(s)$ ($i = 2 \ldots 5$) versus both latitude and longitude are used to determine suitable covariate structures of the $\beta_i(s)$. The use of geographic covariates is called land use regression; see (Lindström et al., 2013). In this way, values of $\beta_i(s)$ at unobserved locations may be determined by universal kriging with longitude and latitude as spatial covariates and where the covariance functions $\text{cov}(\beta_i(s), \beta_i(s + h))$ are exponential. In our case, a nugget effect is assumed for the covariance models of $\beta_i(s)$.

Initial parameter estimates for these covariance functions are required for an optimisation procedure which determines the parameters of the spatio-temporal model (1). A fitted spatio-temporal model may be used to either predict wind speeds at unobserved locations or to interpolate missing data at observed locations. The former necessitates an initial grid which specifies the space and time coordinates for which we require wind speed predictions.

The spatio-temporal random field in the model (1) is assumed to be a zero mean Gaussian field which is spatially correlated but independent in time. It is used to model short-term random effects (not captured by the temporal basis functions) which affect medium to large spatial regions; see (Lindström et al., 2013).

$$v(s,t) \sim N\left(0, \begin{bmatrix} \sum_{v}^{1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sum_{v}^{\bar{\tau}} \end{bmatrix}\right) \qquad (11)$$

In this paper, the covariance function $c(h,\tau)$ of $v(s,t)$ is assumed to be exponential with a nugget effect, that is

$$c(h,\tau) = c_0(1 - \delta(h)) + b\exp\left\{-\frac{\|h\|}{a}\right\} \qquad (12)$$

where $h \equiv s' - s$, $\tau \equiv t' - t$, c_0 is the magnitude of the discontinuity at the origin, δ is the Dirac delta function, $b > 0$ is a variance parameter and $a > 0$ is a scale parameter (Wikle, Zammit-Mangion and Cressie, 2019; Hristopulos, 2020). Note that right of Equation (12) is independent of τ as $v(s,t)$ is assumed to be temporally uncorrelated.

2.4 Implementation

We briefly discuss the implementation of the model (1) to our data set in the programming language R. The CSV data for each weather station was converted to an xts

time series with POSIXct indices. xts objects are particularly useful and allow for the manipulation of irregularly spaced time series, interpolation of missing data, conversion of time zones, aggregation and merging of time series (Zhang, 2016). In our case, data from the various stations are collectively stored in a space-wide (Bivand *et al.*, 2008) xts object which has some missing data (see Figure 3). This data was used to form a STdata object with longitude and latitude covariates originating from the weather station locations. The temporal basis functions $f_i(t)$ and corresponding spatial fields $\beta_i(s)$ were then determined from this STdata object. Plots of these spatial fields (see Figure 8 for example) were used to specify their spatial covariates. A STmodel was then initialised from the wind STdata object, the covariance functions for the residual field $v(s,t)$ and spatial fields $\beta_i(s)$ together with their spatial covariates.

Initial values were then provided for the estimation of the parameters of the STmodel via optimisation. The fitted spatio-temporal model was used to predict wind speed on a three dimensional (two spatial and one temporal) grid which was constructed as a STdata object. In order to display these predictions, raster map tiles for the lower Caribbean were obtained from (Stamen Maps, 2021) via the R package ggmap. The predictions are then superimposed as a contour plot on the Caribbean map using the geom_point and geom_contour functions. A flowchart of our procedure is shown in Figure 11.
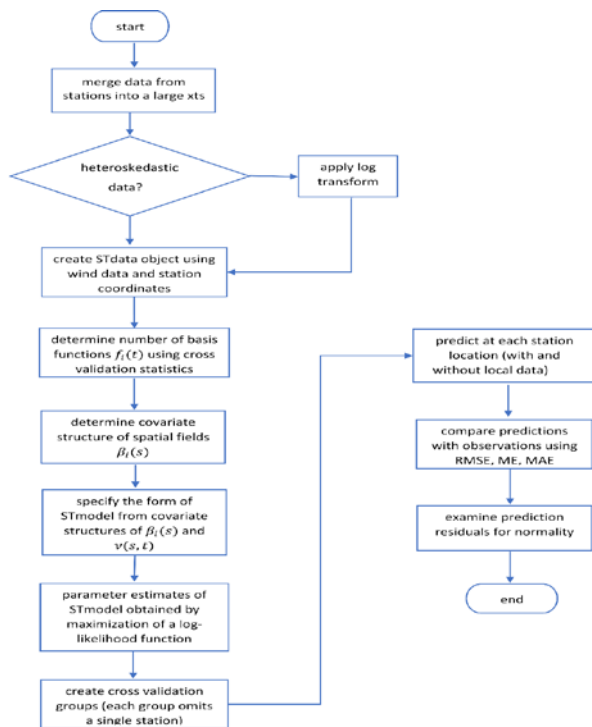

Figure 11. Procedure for predicting and validating wind speeds using the model (1)

3. Results

3.1 Spatial Prediction

Once the structure of the spatio-temporal model (1) is specified (see Section 2.3), its parameters are determined by likelihood estimation by calling the R function optim. The fitted model may then be used for spatial prediction, that is, the model may be used to predict at locations which have no measurement data. A 100 x 100 spatial grid of points was constructed within a bounding box between W57°, W67.2°, N5.8° and N15.7°. An hourly temporal sequence was also specified between midnight and noon of January 1 2015; note however that we may spatially predict at any time within the window of our measurement data. These spatial and temporal objects were combined into a STdata object (Lindström *et al.*, 2013) and the fitted model was used to predict wind speed at each point on this spatio-temporal grid. Contour plots of these predicted wind speeds at midnight and noon respectively are shown in Figure 12. The larger wind speeds at midnight are consistent with the diurnal cyclicity exhibited in Figure 2.

The contour topology also agrees with the direction of the north east trade winds which decrease as they move over the South American continent.

3.2 Imputation

Classical imputation methods only use temporal information (Yang *et al.*, 2018) to construct a time series model (such as ARIMA, support vector regression or hidden Markov models) which is then used to estimate missing values. Kriging utilises additional spatial information by taking a linear combination of nearby observations which is weighted according to covariance values (which depend on distance). Recall that our Caribbean wind observations have substantial missing data. As an example, we consider the Crown Point data set in late February 2015. These incomplete observations are plotted in red in Figure 13.

Recall from Figure 3 that more data is available at nearby Point Salines and Grantley Adams. The estimation of the spatio-temporal model (1) determines a collection of daily temporal trends $\{f_i(t): i = 1,…5\}$ using data from all locations. These temporal trends are then weighted by the spatial fields $\{\beta_i(s): i = 1,…5\}$. The evaluation of these spatial fields at the Crown Point location $s = s_0 = (x_0, y_0)$ determines the trend component of the predicted wind speed at Crown Point. The trend component for our Crown Point example is shown in green in Figure 13.

The component $v(s,t)$ of spatio-temporal model (1) interpolates the variation of wind speed observations from the trend component. The addition of the component $v(s,t)$ to the trend component results in the imputed (i.e. predicted) wind speeds illustrated in black in Figure 13. The variances of these predictions are used to form a confidence band for the predictions; this is
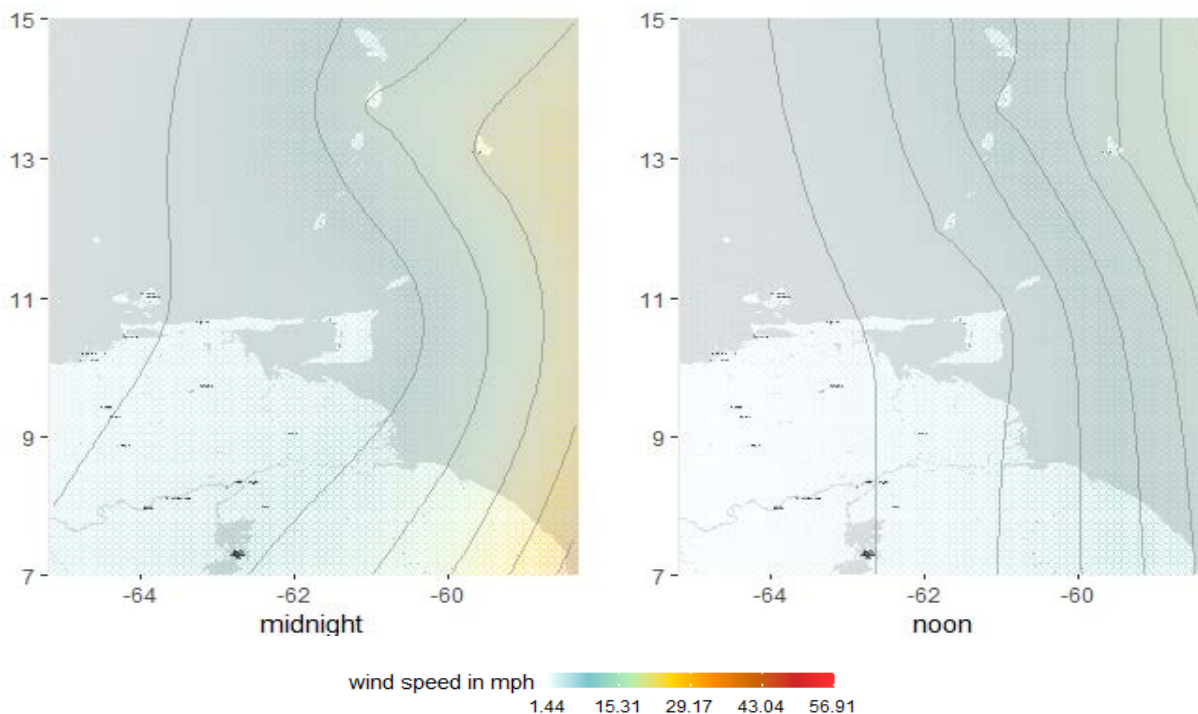
Figure 12. Facet plot of wind speeds in the lower Caribbean on January 1 2015.
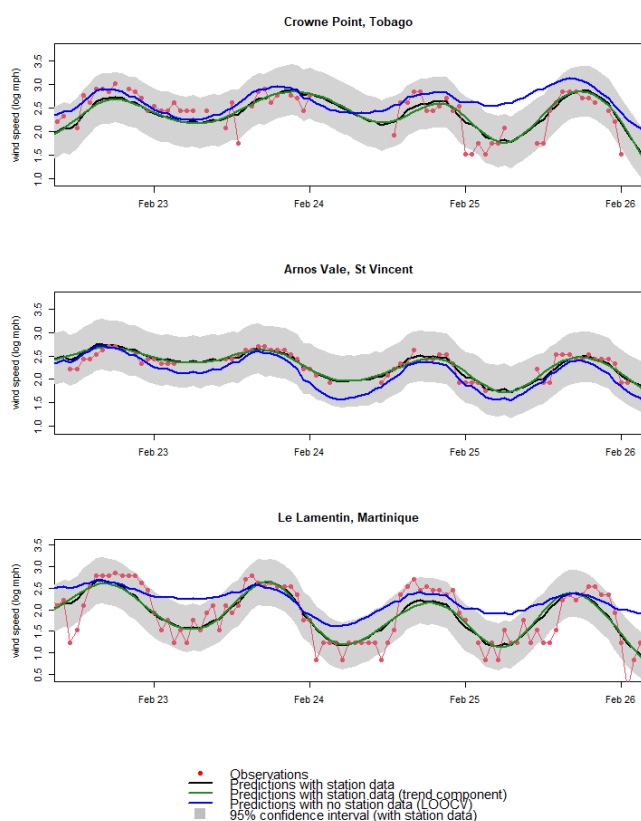The left and right panes show wind speeds at midnight and noon respectively.

Figure 13. Imputation with validation of Crown Point logarithmic
wind speed observations in late February 2015

illustrated as the grey areas in Figure 13.

3.3 Validation

Cross-validation is a method of assessing the predictive ability of a model by partitioning the available observations into *training sets* (used to estimate the parameters of the model) and *test sets*. The estimated model is used to form predictions at the locations of the test sets and so predictions may be compared with actual observations. Our case study is assessed by the use of eight cross-validation groups where the ith group consists of all reference locations except location i. The spatio-temporal model (1) is estimated using training data from the seven locations in the ith group and then used to predict wind speeds at location i. These predictions are then compared to the known test data at location i. Figure 13 shows results from three of the eight cross-validation groups. As an example, the blue plot in the Crown Point panel indicates predictions using data from all locations except Crown Point.

The red plot shows actual observations at Crown Point. As may be seen from this panel, the blue predictions are reasonable estimates which capture the cyclicity of the actual observations in red. However, there is a small bias caused by inaccurate estimation of the spatial coefficients $\beta_i(s)$ of the trend component.

Let $y_t, \hat{y}_t$ denote the observed and predicted values of wind speed at a fixed location and time t. The residuals

$$e_t = y_t - \hat{y}_t \tag{13}$$

are used to form accuracy measures of prediction models (Hyndman and Athanasopoulos, 2018). One such measure is the root mean squared error RMSE which is defined as

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N} e_t^2}{N}} \qquad (14)$$

where N is the number of observations. Other accuracy measures include the mean error (ME), mean absolute error (MAE), Pearson correlation coefficient and percent bias (Zambrano-Bigiarini, 2014). The accuracy measures in Table 1 show (as expected) that the predictions at Crown Point are improved when local data is available. Note that these measures indicate that predictions at Crown Point are meaningful even if no local data is used.

Table 1. Comparison of predictions with/without Crown Point (CP) data using goodness of fit measures

Measures	with CP data	without CP data
ME	0.0003	0.3155
MAE	0.2513	0.3654
RMSE	0.3302	0.4886
Pearson	0.7301	0.6505
Percent bias	0	13.4

The validity of our wind speed predictions is dependent on the assumption that the residual field $v(s,t)$ of our spatio-temporal model (1) is normally distributed. A check for normality of residuals may be done via a quantile-quantile (Q-Q) scatter plot which is obtained by plotting sample quantiles against theoretical quantiles of a normal distribution. For normal residuals, these points are expected to lie along a straight line. From the Q-Q plots in Figures 14 and 15, we see that the assumption of normal residuals is reasonable with or without the use of local Crown point data.
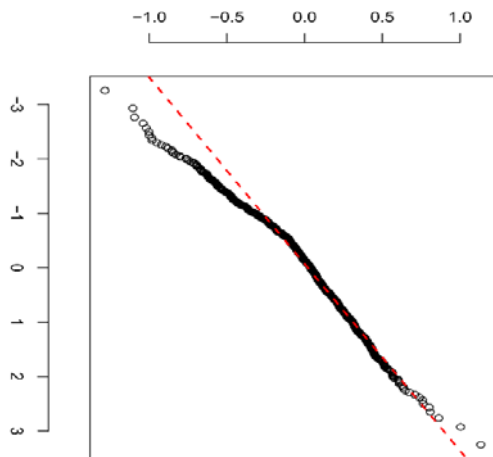

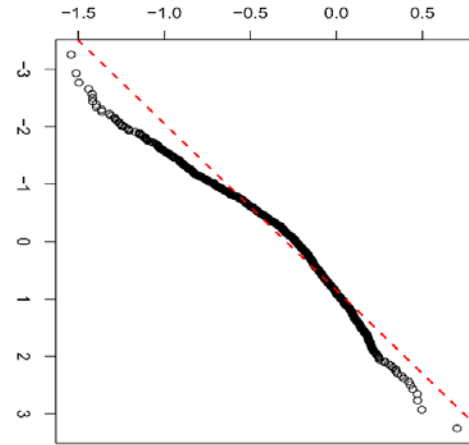**Figure 14.** Q-Q plot of residuals of prediction with Crown Point data

Figure 15. Q-Q plot of residuals of prediction without Crown Point data

4. Conclusion and Future Work

The particular kriging method used in this paper is suitable for imbalanced, trending data. Predictions are more accurate at locations which have some data points as these allow for better estimates of the local trend function. However, we have seen that estimates are still meaningful at locations with no data.

Predictions obtained from our spatio-temporal model given in equation (1) may be possibly improved by the use of a more general model.

$$y(s,t) = \sum_{l=1}^{L} \gamma_l \, \mathcal{M}_l(s,t) + \sum_{i=1}^{m} \beta_i(s) f_i(t) + v(s,t) \quad (15)$$

which now includes exogenous spatio-temporal covariates $\mathcal{M}_l(s,t)$ (see Lindström *et al.*, 2013). It is feasible that pressure and temperature covariates may improve our wind speed predictions (Şen, 1997); note also that pressure and temperature data are typically available in METAR records. Surface roughness and elevation are other (non-temporal) covariates which may be used although roughness data is less readily available (Wever and Groen, 2009). The use of roughness and elevation data may necessitate the use of bridges between R and geographic information software (GIS) (Lovelace, Nowosad and Muenchow, 2019).

Furthermore, the METAR data used in the paper may be supplemented with the ERA5-Land hourly dataset (CDS, 2021) which has a spatial horizontal resolution of 9 km. Finally, as direction may be numerically specified in degrees, our kriging method may also estimate wind directions which can then be combined with wind speed estimates to obtain velocity predictions.

.

References:

Bivand, R.S., Pebesma, E., and Gómez-Rubio, V. (2008), *Applied Spatial Data Analysis with R*, Springer, Switzerland.

Carta, J.A., Velázquez, S., and Cabrera, P. (2013) "A review of measure-correlate-predict (MCP) methods used to estimate long-

term wind characteristics at a target site", *Renewable and Sustainable Energy Reviews*, Vol.27, pp.362-400.

CDS (2021), 'ERA5-land hourly data from 1981 to present'. https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview.

Chamberlain, S. (2021), Rnoaa: 'NOAA' weather data from R. Available at: https://CRAN.R-project.org/package=rnoaa.

Renewable Energy Committee (2011), Framework for Development of a Renewable Energy Policy for Trinidad and Tobago. Ministry of Energy and Energy Affairs, Port of Spain, Trinidad; Tobago.

Cressie, N. (1992), "Statistics for spatial data", *Terra Nova. Wiley Online Library*, Vol.4, No.5, pp. 613-617.

Elliott, D.L., Holladay, C.G., Aspliden, C.I., Schwartz, M.N., and Gowerl, G.L. (1987), *Wind energy resource assessment of the Caribbean and Central America*. Pacific Northwest National Lab. (PNNL), Richland, WA (United States), April.

Fuentes, M., Guttorp, P., and Sampson, P.D. (2006), "Using transforms to analyse space-time processes", *Monographs on Statistics and Applied Probability*, Chapman and Hall, 107, p. 77.

Hristopulos, D.T. (2020), *Random Fields for Spatial Data Modeling: A Primer for Scientists and Engineers*, Springer.

Hunt, K. and Nason, G.P. (2001), "Wind speed modelling and short-term prediction using wavelets", *Wind Engineering*, SAGE Publications Sage UK: London, England, Vol.25, No.1, pp.55-61.

Hyndman, R. J. and Athanasopoulos, G. (2018) Forecasting: Principles and practice. OTexts.

Iowa Environmental Mesonet (2021), 'Iowa Environmental Mesonet, Iowa State University'. https://mesonet.agron.iastate.edu/request/download.phtml?network=AWOS.

Lindström, J., Szpiro, A., Sampson, P.D., Bergen, S., and Sheppard, L. (2013) "Spatiotemporal: An R package for spatio-temporal modelling of air-pollution", *Journal of Statistical Software*, pp.1-34 (https://mran.microsoft.com/snapshot/2016-09-08/web/packages/SpatioTemporal/vignettes/ST_intro.pdf).

Lovelace, R., Nowosad, J. and Muenchow, J. (2019) *Geocomputation with R*, CRC Press.

Montero, J.-M., Fernández-Avilés, G. and Mateu, J. (2015), *Spatial and Spatio-temporal Geostatistical Modeling and Kriging*, John Wiley & Sons.

NOAA (2021), 'National Oceanic and Atmospheric Administration, United States Department of Commerce' https://www.ncdc.noaa.gov/cdo-web/webservices/v2.

National Weather Service (1998), *Automated Surface Observing System (ASOS) User's Guide*, 2021), National Oceanic and Atmospheric Administration/US Government Printing Office, Washington, DC.

Paciorek, C.J., Yanosky, J.D., Puett, R.C., and Laden, F. (2009), "Practical large-scale spatio-temporal modeling of particulate matter concentrations", *The Annals of Applied Statistics, Institute of Mathematical Statistics*, Vol.3, No.1, pp.370-397.

Pebesma, E. (2021) 'The meuse data set: A brief tutorial for the gstat R package', ViennaR, March, https://cran.r-project.org/web/packages/gstat/vignettes/gstat.pdf

Pebesma, E. and Gräler, B. (2021), 'Introduction to spatio-temporal variography', March, https://cran.r-project.org/web/packages/gstat/vignettes/st.pdf

Pyrcz, M. J. and Deutsch, C.V. (2014), *Geostatistical Reservoir Modeling*, Oxford University Press.

Salmon, M. (2016), Riem: Accesses weather data from the Iowa environment, mesonet, Available at: https://CRAN.R-project.org/package=riem.

Sampson, P.D., Szpiro, A.A., Sheppard, L., Lindström, J., Kaufman, J.D. (2011), "Pragmatic estimation of a spatio-temporal air quality model with irregular monitoring data",

Atmospheric Environment, Elsevier, Vol.45, No.36, pp.6593-6606.

Şen, Z. (1997), "Statistical investigation of wind energy reliability and its application", *Renewable Energy*, Elsevier, Vol.10, No.1, pp.71-79.

Stamen Maps (2021), 'Stamen Maps'. http://maps.stamen.com/.

Tsay, R. S. (2013) Multivariate time series analysis: With R and financial applications. John Wiley & Sons.

Wever, N. and Groen, G. (2009), *Improving Potential Wind for Extreme Wind Statistics*, Koninklijk Nederlands Meteorologisch Institut.

Wikle, C.K., Zammit-Mangion, A. and Cressie, N. (2019), *Spatio-temporal statistics with R*, CRC Press.

Yang, H., Yang, J., Han, L.D., Liu, X., Pu, L., Chin, S., and Hwang, H. (2018), "A kriging based spatiotemporal approach for traffic volume data imputation", *PloS One*. Public Library of Science, Vol.13, No.4.

Zambrano-Bigiarini, M. (2014), 'hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series', R package version 0.3-8., https://CRAN.R-project.org/package=hydroGOF

Zhang, D. (2016), *R for Programmers: Mastering the Tools*, CRC Press.

Appendix A: R code for case study

A.1 Data acquisition, processing and preliminary analysis

```
Sys.setenv(TZ='UTC')
library(pacman)
pacman::p_load("rnoaa","readxl","dplyr","lubridate","stringr","xts","zoo","tidyr","knitr")
stations_near_Trinidad = isd_stations_search(lat = 10.6918, lon = -61.2225, radius = 600, bbox = NULL)
knitr::kable(head(select(stations_near_Trinidad,usaf,wban,station_name,icao,lat,lon)))
```

usaf	wban	station_name	icao	lat	lon
999999	11610	WALLER BWI		10.617	-61.217
789700	99999	PIARCO	TTPP	10.595	-61.337
789700	11634	PIARCO INTL AP	TTPP	10.583	-61.350
749040	99999	TOCO		10.833	-60.933
749041	99999	CHICKLAND		10.400	-61.400
999999	11621	TRINIDAD BWI	MCGU	10.683	-61.617

#first six rows of data frame of weather stations within 600 km of Trinidad
*#contains usaf and wban numbers for each station*

```
Piarco_noaa <- isd(usaf="789700", wban="11634", year=2015)
```
 # download Piarco weather data (for year 2015) from NOAA site
```
write.csv(Piarco_noaa, file="Piarco1.csv")
Piarco <- read.csv("Piarco1.csv", stringsAsFactors = FALSE,colClasses=c("NULL", "NULL", "NULL", "NULL",NA,NA,"NULL", "NULL","NULL", "NULL", "NULL", "NULL","NULL", NA,NA,NA,NA,NA,NA))
Piarco <- Piarco[c(1,2,3,4,6,7)]
Piarco$time <- formatC(Piarco$time, flag = 0, width = 4)
Piarco$date <- formatC(Piarco$date)
```

```
Piarco<-unite(Piarco, date_time, c(date,time), sep=":")
Piarco$date_time<-as.POSIXct(strptime(Piarco$date_time,for
mat = "%Y%m%d:%H%M",tz="UTC"))
Piarco$wind_direction[ Piarco$wind_direction == 999] <-NA
```
 # 999 represents NAs
```
Piarco<-drop_na(Piarco)
Piarco$wind_speed<-as.numeric(Piarco$wind_speed)*(2.237/1
0)
```
 # noaa wind speed is in m/s and scaled by 10
```
Piarco_speed<-Piarco[c(1,4)]
boxplot(Piarco_speed$wind_speed)$out
outliers <- boxplot(Piarco_speed$wind_speed, plot=FALSE)$o
ut outliers<-outliers[outliers >30]
```
 # remove outliers
```
Piarco_speed <- Piarco_speed[-which(Piarco_speed$wind_spe
ed  %in% outliers),]
```
 # convert Piarco_Speed from data frame to xts
```
Piarco_speed_xts<-xts(Piarco_speed[,-1], order.by = as.POSIX
ct(Piarco_speed$date_time))
plot(Piarco_speed_xts["20150101/20150227"],main=NULL)
```
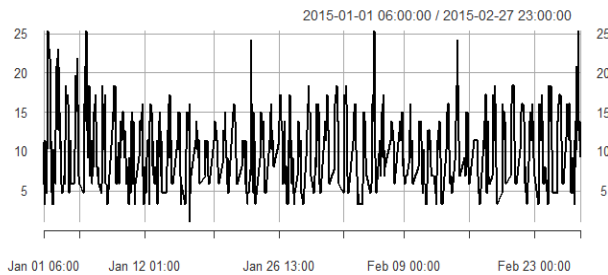

 # plot of Piarco wind speeds for Jan 1 2015 to Feb 27 2015
*#similar data processing done for other stations and data from
Iowa Mesonet*

 # merge data from all stations into a single xts
```
All_stations<-merge(Piarco_speed_xts,Crown_point_Tobago_
xts,Point_Salines_Grenada_xts,Grantley_Adams_Barbados_xts
,Arnos_Vale_St_Vincent_xts,Hewanorra_St_Lucia_xts,Maturi
n_VZ_xts,Cheddi_Jagan_xts,
Simon_Bolivar_xts,Melville_Hall_xts,Le_Lamentin_xts,join="
outer")
```
#remove the rows with duplicated indices
```
All_stations<-All_stations[!duplicated(index(All_stations))]
```
#remove rows with all NAs
```
All_stations<-All_stations[rowSums(is.na(All_stations)) != nco
l(All_stations),]
```
 # saveRDS(All_stations, file = 'All_stations.Rds')
```
Cbbn_wind <- readRDS('All_stations.Rds')
Cbstart="2015-01-01 00:00:00";Cbend="2015-02-25 23:00:00"
Cbbn_wind<-Cbbn_wind[paste(Cbstart,"/",Cbend,sep="")]
```
#truncate dataset to first two months
```
Cbbn_wind_2<-Cbbn_wind[seq(from=as.POSIXct(Cbstart, tz=
"UTC"),
to=as.POSIXct(Cbend, tz="UTC"),by="hour" )]
```
*# remove observations that occurred at half hours as such obse
rvations*
 # only occurred at Le Lamentin
```
plot(Cbbn_wind_2["20150201/20150228"],ylim=c(3,32))
```
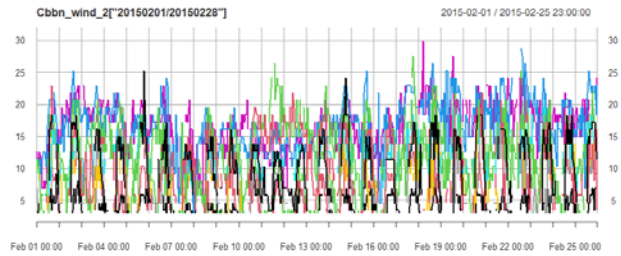

*# plot of wind speeds from 11 stations from Feb 1 2015 to Feb
28 2015*

 # MISSING VALUES DIAGRAM
```
Cbbn_wind<-Cbbn_wind[,c(1:6,10,11)]
names(Cbbn_wind)<-c("Piarco", "Crown Point","Point Salines
","Grantley Adams",
    "Arnos Vale","Hewanorra", "Melville Hall", "Le Lamentin")
Cbstart="2015-01-01 00:00:00";Cbend="2015-02-26 23:00:00
"
Cbbn_wind_2<-Cbbn_wind[seq(from=as.POSIXct(Cbstart, tz
="UTC"),
            to=as.POSIXct(Cbend, tz="UTC"),by="hour" )]
Cbbn_wind_3<-as.data.frame(Cbbn_wind_2)

mvals <- Cbbn_wind_3 %>%
  gather(key = "key", value = "val") %>%
  mutate(isna = is.na(val)) %>%
  group_by(key) %>%
  mutate(total = n()) %>%
  group_by(key, total, isna) %>%
  summarise(num.isna = n()) %>%
  mutate(pct = num.isna / total * 100)

levels <-
  (mvals  %>% filter(isna == T) %>% arrange(desc(pct)))$key

pplot<- mvals %>%
  ggplot() +
  geom_bar(aes(x = reorder(key, desc(pct)),
          y = pct, fill=isna),
        stat = 'identity', alpha=0.8) +
  scale_x_discrete(limits = levels) +    scale_fill_manual(name
= "",
        values = c('steelblue', 'tomato3'), labels = c("Present", "
Missing")) +
  coord_flip() +
  labs(title = "", x =
      'Location', y = "percentage of missing values")
pplot
```
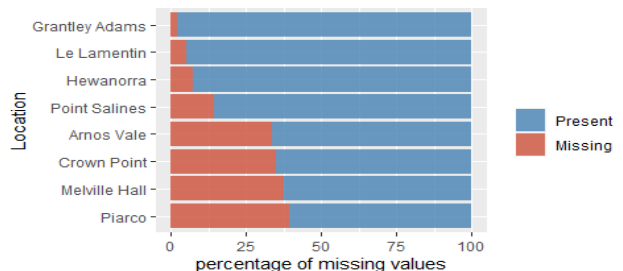
```
MarchTest(as.matrix(na.spline(Cbbn_wind_2)))
## Q(m) of squared series(LM test):
## Test statistic:  593.7194  p-value:  0
## Rank-based Test:
## Test statistic:  1403.636  p-value:  0
## Q_k(m) of squared series:
## Test statistic:  21008  p-value:  0
## Robust Test(5%) :  19562.9  p-value:  0
```

 # sample variogram
```
Cbbn_wind <- readRDS('All_stations.Rds')
Cbstart="2015-01-01 00:00:00";Cbend="2015-12-31 23:00:00"
Cbbn_wind<-Cbbn_wind[seq(from=as.POSIXct(Cbstart, tz="
UTC"),
     to=as.POSIXct(Cbend, tz="UTC"),by="hour" )]
Cbbn_wind <- Cbbn_wind[,1:7]
Cbbn_df<- data.frame(time=index(Cbbn_wind), coredata(Cbbn
_wind))
stations <- c("Piarco", "Crown Point","Point Salines","Grantley
 Adams",          "Arnos Vale","Hewanorra","Maturin")
lat<-c(10.583,11.150,12.004,13.067, 13.133,13.750,9.749)
lon<-c(-61.350,-60.833,-61.786,-59.483,-61.200, -60.950, -63.1
53)
Ccrds<-data.frame(stations,lon,lat)
coordinates(Ccrds) = ~lon+lat
proj4string(Ccrds) = "+proj=longlat +datum=WGS84"
Cpts = coordinates(Ccrds)
rownames(Cpts) = Ccrds$stations
Cpts = SpatialPoints(Cpts, CRS("+proj=longlat +datum=WGS
84 +ellps=WGS84"))
utm20 = CRS("+proj=utm +zone=20,21 +datum=WGS84 +ellp
s=WGS84")
Cpts = spTransform(Cpts, utm20)
Cwind.data = stConstruct(as.matrix(Cbbn_df[,2:8]), space = lis
t(values = 1:ncol(Cbbn_wind)),
        time = Cbbn_df[,1], SpatialObj = Cpts, interval = TRUE)
vargmCbbn<-variogramST(values~1,Cwind.data[, "20150101/
20150107"],tlags=0:100, cutoff=2000000,na.omit=TRUE)
plot(vargmCbbn,wireframe=TRUE)
```
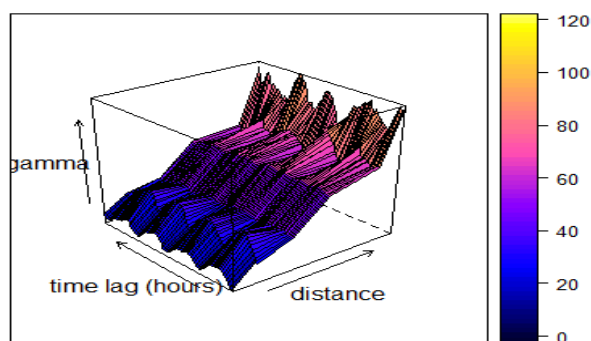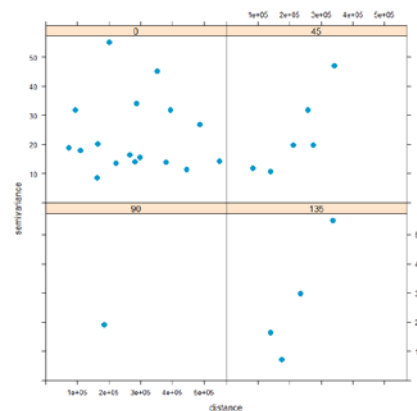

 # directional variogram
```
time_samples = sample(dim(Cwind.data)[2], 400)
lst2 = lapply(time_samples, function(i) { x = Cwind.data[,i]; x
$ti = i; rownames(x@coords) = NULL; x} )
pts2 = do.call(rbind, lst2)
v1.dir = variogram(values~ti, pts2[!is.na(pts2$values),], dX=0,
width=50,cutoff=2000000,alpha = c(0, 45, 90, 135))
plot(v1.dir,as.table = TRUE,pch=19, col="deepskyblue3", cex=
1.3)
```


A.2 Estimation, prediction and validation

 # use logs because of heteroskedasticity
```
Cbbn_wind<-log(Cbbn_wind)
```
 # use two months of hourly data
```
Cbstart="2015-01-01 00:00:00";Cbend="2015-02-25 23:00:00"

Cbbn_wind<-Cbbn_wind[paste(Cbstart,"/",Cbend,sep="")]
```
#truncate dataset

```
Cbbn_wind_2<-Cbbn_wind[seq(from=as.POSIXct(Cbstart, tz=
"UTC"), to=as.POSIXct(Cbend, tz="UTC"),by="hour" )]
```
*# remove observations that occurred at half hours as such obse
rvations*
*# only occured at Le Lamentin*

```
Cbbn_wind_2_core <-coredata(Cbbn_wind_2)
Cbbn_wind_daily_xts <- xts(Cbbn_wind_2_core,seq(as.Date("
2015-01-01"), length=nrow(Cbbn_wind_2), by="days"))
Cbbn_wind_matrix2<-as.matrix(Cbbn_wind_daily_xts)
Cbbn_STdata<- createSTdata(Cbbn_wind_matrix2, covars = C
crds)
```

 # facet plot of change in AIC, MSE, BIC, R^2 with n.basis
```
D <- createDataMatrix(Cbbn_STdata)
CV_facet<- SVDsmoothCV(D, 0:7)
plot(CV_facet)
```
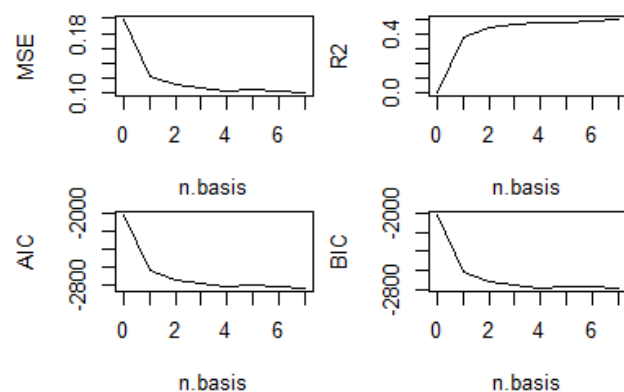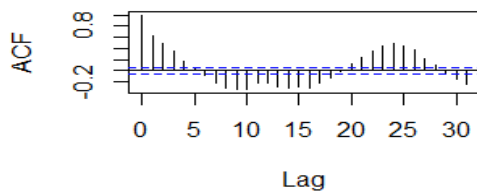

 # Piarco acf before trend removal
```
plot(Cbbn_STdata, "acf", ID = "1")
```
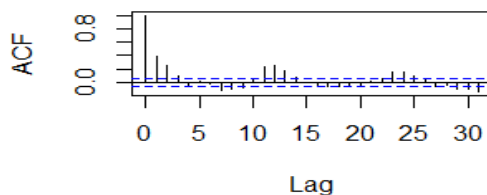
ACF for 1

Cbbn_STdata <- updateTrend(Cbbn_STdata, n.basis = 4)

 # Piarco acf after trend removal
plot(Cbbn_STdata, "acf", ID = "1")

ACF for 1

#plot of five basis functions
Cbbn_STdata_trend_xts <- xts(Cbbn_STdata$trend,seq(as.POS IXct("2014-12-31 20:00:00"), length=nrow(Cbbn_STdata$tren d), by="hour"))
col_ones_xts<-xts(rep(1, nrow(Cbbn_STdata$trend)),seq(as.P OSIXct("2014-12-31 20:00:00"), length=nrow(Cbbn_STdata$t rend), by="hour"))
Cbbn_STdata_trend_xts<-cbind(col_ones_xts,Cbbn_STdata_tr end_xts)
tt <- time(Cbbn_STdata_trend_xts)[seq(1, length(Cbbn_STdata _trend_xts), by = 12)]
cols <- c("darkorange4" ,"darkorchid4","forestgreen","oranger ed3","steelblue")
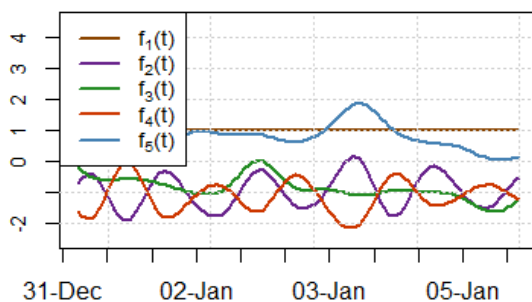plot(as.zoo(Cbbn_STdata_trend_xts["20150101/20150105",c(1 :5)]),col = cols,
    screens = 1,ylim=c(-2.5,4.5),ylab="",xlab="",cex.axis=0.8,c ex.lab=0.85,cex.main=0.95,
    main="",xaxt="n",lwd=2 ,font.main = 1)
grid()
legend(x = "topleft", legend = c(TeX('$f_1(t)$'),TeX('$f_2(t)$') , TeX('$f_3(t)$'),TeX('$f_4(t)$'), TeX('$f_5(t)$')),lty = 1,col = cols,pt.cex = 1,cex=0.9,bg="white",lwd=2)
axis(1, tt,format(tt, '%d-%b'),las=1)

beta_fields <- estimateBetaFields(Cbbn_STdata)
row.names(beta_fields$beta)
beta_fields$beta <- data.frame(beta_fields$beta)

beta_fields$beta.sd <- data.frame(beta_fields$beta.sd)
beta_fields$beta$ID <- row.names(beta_fields$beta)

 # using five basis funtions
merged_beta <- cbind(beta_fields$beta[,1:5], beta_fields$beta$ ID,beta_fields$beta.sd[,1:5])
#using first 4 columns of merged_beta etc
colnames(merged_beta) <- c("alpha1", "alpha2", "alpha3", "alp ha4",
        "alpha5", "ID",
        "alpha1_CI", "alpha2_CI", "alpha3_CI", "alpha4_CI",
        "alpha5_CI")
merged_beta <- left_join(merged_beta, Ccrds, by = "ID")

ggplot(merged_beta) + geom_point(aes(x = lat, y = alpha1)) +
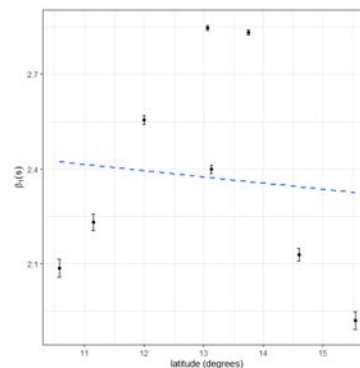   geom_smooth(aes(x = lat, y = alpha1),method = "lm", se = FALSE, linetype=2)+
  geom_errorbar(aes(x = lat,
        ymin = alpha1 - 1.96*alpha1_CI,
        ymax = alpha1 + 1.96*alpha1_CI)) +
  ylab(expression(beta[1](s))) +
  xlab("latitude (degrees)") + theme_bw()

 # beta1 essentially shows mean wind speed at each station
lcbeta <- list(covf = "exp", nugget = TRUE)
c_nu <- list(covf = "exp",nugget = ~1,random.effect = FALSE)
    _
locations <- list(coords = c("lon", "lat"))

 # land_use in the case of four basis function (well five when co nstant is included)
land_use <- list(~lat + lon, ~lat, ~lat , ~1 , ~1)
STmodel <- createSTmodel(Cbbn_STdata, LUR = land_use, co v.beta = lcbeta,
            cov.nu = c_nu, locations = locations)
 # init_ in the case of four basis functions
init_ <- matrix(1, 18, 1)
SpatioTemporalfit1 <- estimate(STmodel, init_)
pred_Cbbn <- predict(STmodel, SpatioTemporalfit1, pred.var= TRUE)
 # LOOCV cross validation
CV_groups <- createCV(STmodel, groups=8)   *# this will crea te 8 groups each omitting a single station*
init_ <- coef(SpatioTemporalfit1, pars="cov")[,c("init")]
est.cv.STmodel <- estimateCV(STmodel, init_, CV_groups)
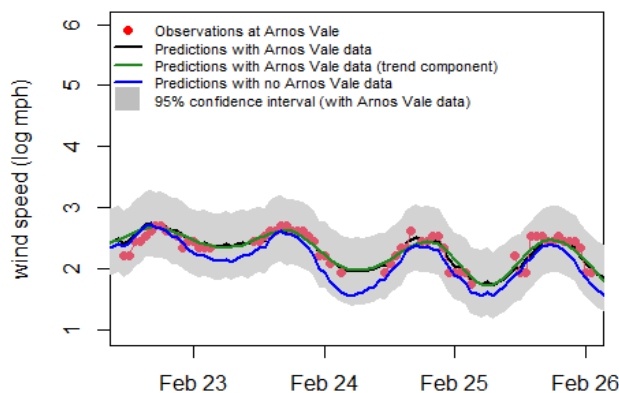pred.cv.STmodel <- predictCV(STmodel, est.cv.STmodel, LT A=TRUE)

```
par(mfrow=c(1,1))
plot(pred_Cbbn, ID=5, main="", font.main=1,
    xlab="", ylab="wind speed (log mph)", xaxt="n",
    xlim=as.Date(c("2016-10-20","2017-01-12")),
    ylim=c(0.95,6), pch=NA, pred.var=TRUE,STmodel=STmo
del,
    lty=NA, col=c(1,1,"lightgrey"))
plot(pred_Cbbn, ID=5, STmodel=STmodel, add=TRUE,
    lty=c(1,1),lwd=c(2,1), pch=c(NA,16),  cex=c(1.2,1),col=c(1
,2,NA))
axis.Date(1, at=rev(seq(as.Date("2017-01-12"), length=4, by="
-24 days")),labels=c("Feb 23","Feb 24","Feb 25","Feb 26"))
plot(pred_Cbbn, ID=5, STmodel=STmodel, add=TRUE,
    pred.type = "EX.mu.beta", lwd=2, col="forestgreen")
legend("topleft", c("Observations at Arnos Vale",
        "Predictions with Arnos Vale data",
        "Predictions with Arnos Vale data (trend component)",
        "Predictions with no Arnos Vale data",
        "95% confidence interval (with Arnos Vale data)"), bt
y="n",
    lty=c(NA,1,1,1, NA), lwd=c(NA,2,2,2, NA),
    pch=c(16,NA,NA,NA,15), pt.cex=c(1,1,NA,NA,2.5),
    col=c("red", 1,"forestgreen","blue", "grey"),cex=0.7)
plot(pred.cv.STmodel, ID="5",add=TRUE,col="blue",lwd=2)
```
 # blue curve obtained from LOOCV


```
rmse(as.vector(pred_Cbbn[["EX"]][,2]),as.vector(Cbbn_wind_
daily_xts[,2]),na.rm=TRUE)
```
 #   package: hydroGOF
```
## [1] 0.3335035
TEMP<- pred.cv.STmodel[["pred.obs"]][pred.cv.STmodel[["pr
ed.obs"]]$ID=="2",]
```
#restrict to ID=2
```
rmse(as.vector(TEMP$EX),as.vector(TEMP$obs),na.rm=TRU
E)
```
*# actually no need for na.rm here as STDF stripped missing val
ues*
```
## [1] 0.4986291
mae(as.vector(pred_Cbbn[["EX"]][,2]),as.vector(Cbbn_wind_d
aily_xts[,2]),na.rm=TRUE)
mae(as.vector(TEMP$EX),as.vector(TEMP$obs),na.rm=TRUE
)

me(as.vector(pred_Cbbn[["EX"]][,2]),as.vector(Cbbn_wind_da
ily_xts[,2]),na.rm=TRUE)
me(as.vector(TEMP$EX),as.vector(TEMP$obs),na.rm=TRUE)
 rPearson(as.vector(pred_Cbbn[["EX"]][,2]),as.vector(Cbbn_wi
nd_daily_xts[,2]),na.rm=TRUE)
rPearson(as.vector(TEMP$EX),as.vector(TEMP$obs),na.rm=T
RUE)
```

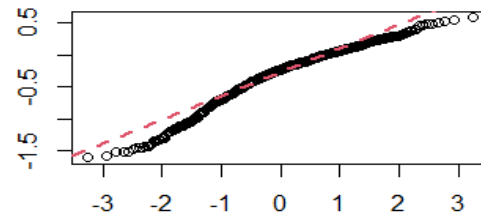*# QQ plots for residuals*

```
qqnorm(TEMP$res,main="",cex.main=1.5,cex.lab=1.5,xlab=""
, ylab="")
```
#without Crown Point data
```
qqline(TEMP$res, col = 2,lwd=2,lty=2)
```


Define space-time grid
```
spat_pred_grid <- expand.grid(lon = seq(-67.2, -57, length = 10
0),
            lat = seq(5.8, 15.7, length = 100))
spat_pred_grid$id <- 1:nrow(spat_pred_grid)
```
 # spat_pred_grid is simply a data frame with lat lon coords.

```
temp_pred_grid <- as.Date("2017-01-12") + seq(-11, 0, length
= 12)
```
#temp_pred_grid

Initialise data matrix
```
obs_pred_wide <- matrix(0, nrow = 12, ncol = 10000)
```

Set row names and column names
```
rownames(obs_pred_wide) <- as.character(temp_pred_grid)
colnames(obs_pred_wide) <- spat_pred_grid$id
covars_pred <- spat_pred_grid # covariates
STdata_pred <- createSTdata(obs = obs_pred_wide, # ST objec
t
            covars = covars_pred)
x.final <- coef(SpatioTemporalfit1, pars = "cov")$par
```
 # predict on a grid
```
E <- predict(STmodel, x.final, STdata = STdata_pred)
library(sp)
library(spacetime)
spat <- as.matrix(spat_pred_grid[,1:2])
colnames(spat) <- NULL
spat <- SpatialPoints(spat)
proj4string(spat) <- CRS("+proj=longlat +datum=WGS84")
times <- as.Date(temp_pred_grid)
ESX <-t(as.matrix(E$EX))
colnames(ESX) <- NULL
pred.st <- STFDF(spat, times, data.frame(vals = as.vector(as.m
atrix(ESX))))
day1<-as.data.frame(pred.st[,1,1] )
```
#extract day1 data from STFDF
```
day12<-as.data.frame(pred.st[,12,1] )  # 12 hours later
```

 # exponentiate the logs
```
day1$vals <- exp(day1$vals)
day12$vals <- exp(day12$vals)
vals <- attr(pred.st[,1,'vals'],'data')
```

```
library(rworldmap)
library(rworldxtra)
library(ggmap)
library(scales)
cbbn_stamen_toner <- get_stamenmap(bbox = cbbn_bb,mapty
pe = "toner-lite", zoom = 9)
breaks <- seq(min(day1$vals, na.rm = TRUE) * 0.95, max(day
1$vals, na.rm = TRUE) * 1.05, length = 21)
require('RColorBrewer')
ColorScale <-c("azure","cadetblue3","darkseagreen2","gold1",
"darkorange","darkorange3","firebrick3",
"firebrick1")

Labs <- round(breaks, 2)
iLabs <- floor(seq(1, length(Labs), length.out = 5))
iLabs[length(iLabs)] <- length(Labs)
Labs <- as.character(Labs[iLabs])
Labs[Labs == "0"] <- " 0.00 "

P <- ggmap(cbbn_stamen_toner)
P <- P + geom_point(aes(x = coords.x1, y = coords.x2, col = va
ls), alpha = 0.15,
     data = day1, na.rm = TRUE)
P <- P + geom_contour(data = day1, aes(x = coords.x1,
        y = coords.x2, z = vals), alpha = 0.5, colour = " gray4
8")
P <- P + scale_colour_gradientn(name = "wind speed in mph",
 colours =ColorScale ,
        breaks = breaks[iLabs], limits = range(breaks), labels
 = Labs)
P <- P + theme(legend.position = "none")
```
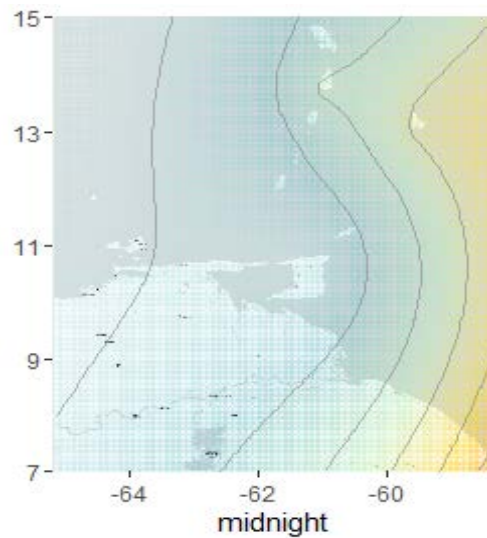
```
P <- P + labs(x = "midnight", y="")
P
```

Author's Biographical Notes:

Neil Ramsamooj *is a Lecturer at the Faculty of Engineering, The University of the West Indies, St Augustine. He has a Ph.D. in Mathematics from the University of Wisconsin-Madison.*

. ∎