Comparative study of feature selection methods for wind speed estimation at ungauged locations

Freddy Houndekindo[1*], Taha B.M.J. Ouarda[1]

[1]Canada Research Chair in Statistical Hydro-Climatology, Institut national de la recherche scientifique, Centre Eau Terre Environnement, 490 de la Couronne, Québec, QC, G1K 9A9, Canada

*Corresponding author: Freddy Houndekindo
490, Couronne street, Québec, QC, G1K 9A9, Canada
Tel: +1 418-654-3842
E-mail: freddy.houndekindo@inrs.ca

May 2023

Table of Contents

Highlights

- Six feature selection methods were evaluated for wind speed quantiles estimation at ungauged locations
- Feature selection enabled the identification of the most important predictors for various wind speed quantiles
- The most parsimonious feature selection methods led to the lowest generalization error
- The location distance from the coast, and the surface roughness were the most significant wind speed quantiles predictors

Abstract

Wind speed estimation at ungauged locations is one of the preliminary steps for wind resource assessment. With the availability of high-resolution Digital Elevation Models (DEM) and remote sensing data, the number of potential wind speed predictors has grown substantially. The adequate spatial scale of these predictors is unknown a priori, leading to the use of multiple spatial scales of predictors in wind speed estimation models. Implementing a feature selection method as a pre-processing step of the analysis is necessary to avoid overfitting and the resulting potential model underperformance. This paper evaluated six feature selection methods (forward stepwise regression, Least Absolute Shrinkage and Selection Operator (LASSO), Elastic Net, Maximum relevance Minimum redundancy (MRMR), Genetic algorithm, and recursive feature elimination using support vector regression) for the estimation of different wind speed quantiles across Canada. The selected features were used to fit a regression-kriging model, and the importance of the predictors was evaluated with their associated regression coefficients. The results of the study showed that LASSO and MRMR are the most efficient algorithms with the least number of parameters to tune and good generalization performance. The study found that some predictors were more important for specific exceedance probabilities. The most important predictors were the distance from the coast and surface roughness length, regardless of exceedance probability.

Keywords: Exceedance probability, Feature selection, Machine learning, Topographic feature, ungauged location, Wind speed.

1. Introduction

The global energy system significantly contributes to greenhouse gas emissions, with a share of approximately 34% (Lamb et al., 2021). Alternative energy sources, such as wind, can help mitigate the environmental footprint of our energy system (Jung et al., 2018; Shin et al., 2016). Wind energy production has experienced substantial growth during the last decades, accounting for 8% (594 GW) of the 7 400 GW of installed generating capacity worldwide as of 2019 (International Renewable Energy Agency, 2022). Unlike conventional energy sources such as coal and nuclear energy, wind energy is intermittent and heavily reliant on wind speed (WS). A sound understanding of the WS variability at a location of interest for wind energy production is necessary to integrate the energy source effectively into the energy mix (Aries et al., 2018). A significant step in wind energy planning is identifying a good location for resource exploitation. Potential sites of interest often do not coincide with a location where extensive WS measurements are available. Therefore, it is helpful to implement approaches that estimate wind resources at ungauged locations.

The challenge of WS estimation at ungauged locations has initially been tackled with spatial interpolation models. In recent studies, machine learning models have gained more popularity, and some researchers have suggested combining spatial interpolation models and machine learning (see Houndekindo and Ouarda (2023) for a detailed review of WS estimation at ungauged locations). These developments have led to the experimentation of new predictors, notably topographical features extracted from DEM. Many topographical features can be used for WS modelling (Maxwell and Shobe, 2022). One such feature is terrain curvature, which has been identified as one of the most effective WS predictors in regions with complex terrain, according to a study conducted in Switzerland by Robert et al. (2013). Several land surface parameters (ex.: plan curvature, gaussian curvature, minimum curvature) extracted from DEM can be used to describe the terrain curvature (Wilson, 2018), leading to several possible features to include in the model. Some of these features will undoubtedly be redundant (Maxwell and Shobe, 2022). The selection of the spatial scales of the topographical features represents another significant challenge. Two potential downsides of incorporating too many features into the model are overfitting the model's parameters to the training data and compromising the model's interpretability. To address this issue, FS can be used as a preprocessing step to build more accurate and concise models while minimizing computation time (Guyon and Elisseeff, 2003).

FS methods are often categorized as filter-based, wrapper, or embedded methods (Guyon and Elisseeff, 2003). Filter-based methods are more computationally efficient and less prone to overfitting compared to wrappers and embedded methods (Zhou et al., 2021). A drawback of most filter methods compared to wrappers and embedded methods is their inability to consider feature interactions (Urbanowicz et al., 2018). The filter approach selects predictors based on their relevance to the dependent variable. In the case of regression, the correlation coefficient can be used to assess the relevance of features.

On the other hand, wrappers and embedded methods rely on the model performance to select an optimal set of features. The wrapper methods search for the feature subset, which gives the best performance with a predefined learning algorithm. Wrapper methods can be used with any model, while embedded methods rely on models that inherently rank the features' importance (ex.: random forest) or eliminate irrelevant features (ex.: penalization methods).

Most studies have applied a data-driven approach to solving the feature selection challenge for WS estimation. For example, Robert et al. (2013) applied a modified version of the general regression neural networks to select the best spatial scale and topographical features for monthly WS interpolation. Jung (2016) employed feature importance ranking with random forest and a forward stepwise feature selection to identify suitable predictors for WS estimation. In the second step, the author used the variance inflation factor to evaluate feature redundancy in the study. For extreme WS mapping, Etienne et al. (2010) used the linear correlation between predictors to evaluate their redundancy and backward elimination to retain the most important predictors in the model. Foresti et al. (2011) applied a multiple kernel learning model for feature selection (FS) in WS mapping. Veronesi et al. (2016) employed the Least Absolute Shrinkage and Selection Operator (LASSO) technique to select relevant features to implement a statistical model for estimating WS distribution at ungauged sites.

To the best of our knowledge, no studies compared the performance of FS methods for WS estimation at ungauged locations. Nevertheless, such comparison is necessary as the number of available WS predictors increases, and so is the risk of redundancy and overfitting. Comparative studies are essential as they allow for a systematic comparison of various approaches with diverse complexity and performance levels. They serve as a basis to identify the strengths and weaknesses of each approach and better understand their performance in different conditions. Several comparative studies of features selections methods have been conducted in studies related to environmental variables. For instance, Carta et al. (2015) compared a wrapper method to a filter approach for FS for long-term WS prediction at locations with a short record. The authors found that the filter method produced sparser feature subsets, while the wrapper method had a better predictive ability. In that study, FS increased the interpretability of the final model while improving its performance. Seven FS methods were compared for river flow quantile estimation in ungauged basins (Fouad and Loáiciga, 2020). The authors found that the FS methods performed better than dimension reduction techniques (principal component analysis) to reduce multicollinearity in the feature subsets. The same study observed similar performance between FS using experts' knowledge and data-driven FS methods. Rodriguez-Galiano et al. (2018) evaluated the performance of various FS methods to predict the probability of the occurrence of nitrates above a threshold value in groundwater. The study revealed that FS helped isolate and identify the main drivers of nitrate pollution in groundwater. Chen et al. (2019) conducted a comparative study of statistical models with various FS methods to predict fine particles and nitrogen dioxide concentration across Europe. The study found that regularization algorithms such as LASSO and Elastic Net (ENET) efficiently selected relevant predictors despite high multicollinearity in the feature set. Also, the regularization algorithms had the additional benefit of model interpretability.

This study compared six different FS methods for WS quantile estimation. These methods included forward stepwise regression (FSWR), LASSO, ENET, Maximum relevance Minimum redundancy (MRMR), Genetic algorithm (GALG), and recursive feature elimination using support vector regression (RFES). The selected algorithms are composed of filter-based (ex.: MRMR), wrappers (ex.: FSWR, GALG), and embedded methods (ex.: LASSO, ENET, RFES). The selected predictors or features were used with a regression kriging (RK) model (Hengl et al., 2007) to estimate various WS quantiles. The RK model has previously shown promising results for WS estimation (Alsamamra et al., 2010; Lee, 2022). Reinhardt and Samimi (2018) also found that RK performed better than Artificial Neural Networks (ANN) and Support Vector Machines (SVM) for WS interpolation. RK is an attractive approach for interpolating environmental variables (Hengl et al., 2007). It allows the use of relevant predictors, and unlike universal kriging and kriging with external drift, RK can be adapted with various types of regression models (ex.: Random Forest, Generalized Additive Models).

The study also evaluated the importance of various predictors for estimating WS quantiles with different exceedance probabilities. Most features used in previous studies were derived and compared within the same framework. In addition, alternative features related to conventional WS predictors used in the literature were also evaluated. These alternative features may provide additional information and insights into WS behaviour at different exceedance probabilities and could improve the accuracy of WS predictions.

The paper is organized as follows. The dataset used is described in Section 2. In section 3, the six FS methods evaluated are presented. Section 4 presents the results of the analysis. The discussion and the conclusion are given in sections 5 and 6, respectively.

2. Data

2.1. Wind speed data

The data analyzed in the study are hourly WS data at 10m above ground from measurement stations across Canada. The data were obtained from Environment and Climate Change Canada (ECCC) historical climate database. Stations with at least 20 years of record available until 2010 were selected, and only those with at least ten years of record with less than two months of missing data were used. Figure 1 shows the spatial distribution of the selected stations, which amounted to 207.

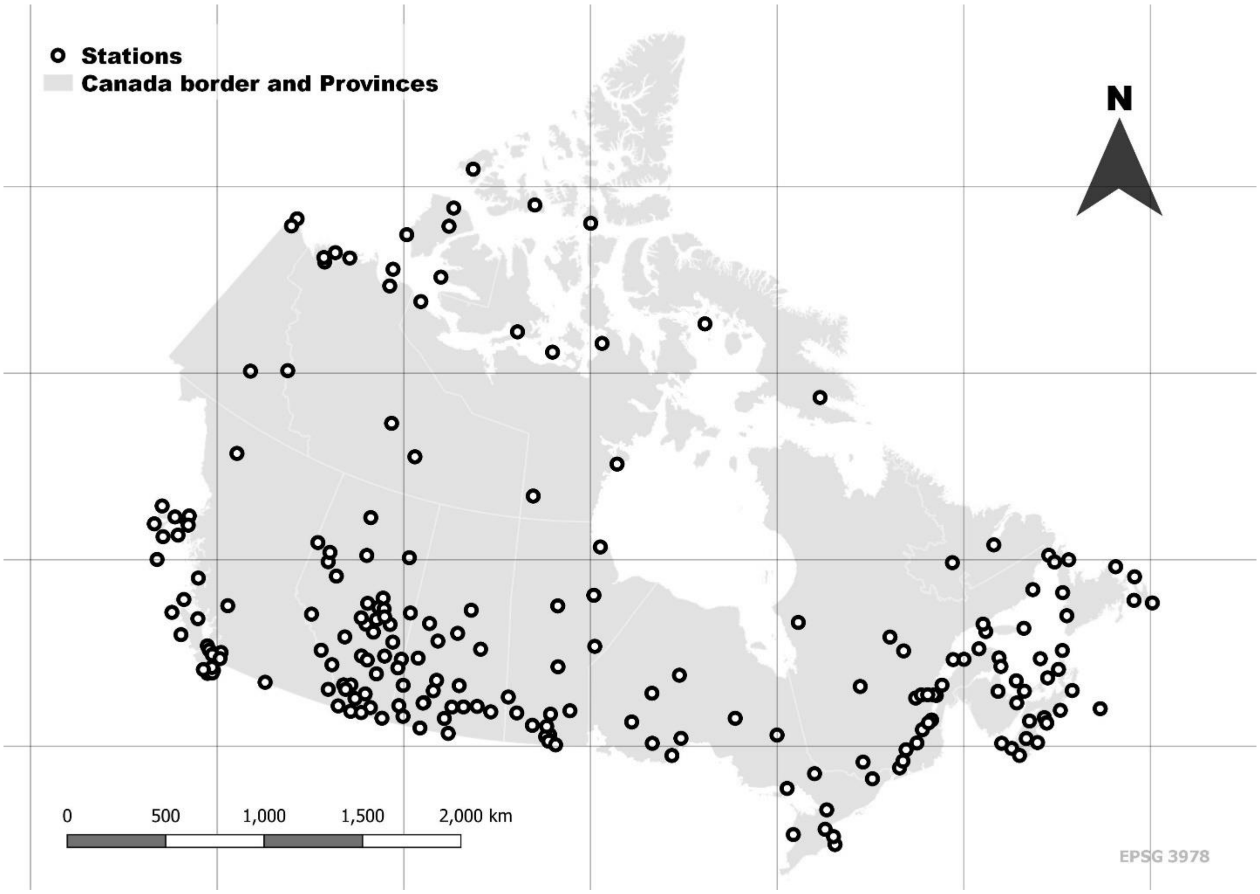

Figure 1: Study region and locations of the 207 selected stations

From the hourly records, empirical WS quantiles were estimated using the Weibull plotting position formula:

$$P_i = P(Ws > Ws_i) = \frac{i}{n+1} \quad (1)$$

Where:

P_i is the probability of exceedance associated with the observed hourly wind speed (Ws_i). i is the rank of the observed wind speed Ws_i sorted in descending order. $i = 1$ corresponds to the highest observed WS and $i = n$ corresponds to the lowest observed WS, with n the number of observations.

Monotonic decreasing penalized splines (P-Splines: Paul and Marx, 1996; Pya and Wood, 2015) were fitted between the exceedance probabilities and their associated observed WS quantiles to construct the empirical complementary cumulative distribution function (survival function). The fitted curve was used to estimate WS quantiles at 14 fixed percentile points at each location in the study area. The following 14 fixed percentile points were selected: p = 0.01%, 0.1%, 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% to cover an extensive range of WS quantiles. The P-Splines is a non-parametric model that allows fitting a smooth and flexible curve to data. Monotonic decreasing constraints were imposed on the P-Splines to respect the monotonic nature of complementary cumulative distribution functions.

2.2. Predictors

The predictors used in the study are topographical, surface roughness length, geographical coordinates, and the location distance from the coast. Table 1 provides more details on these predictors. The topographical variables were extracted from a resampled (100m spatial resolution) ALOS DEM (Tadono et al., 2014) and computed with the WhiteboxTools (Lindsay, 2014) developed at the University of Guelph, Canada. Information on the land cover type obtained from a 2015 land use map of Canada (Latifovic et al., 2017) was used to estimate the surface roughness length according to Wiernga (1993). The land use map was resampled to produce multiple spatial resolutions, with majority resampling (mode) providing information on the most common land use type for the given spatial scale.

Some of the features selected for the study were previously studied because they describe physical processes that influence wind movement. This study also introduced alternative features describing similar physical processes. For instance, Jung (2016) used slope (SLPE), curvature, aspect (ASPC), roughness length (RGLH) and relative elevation for WS mapping in Germany. In the present study, relative elevation measures used were deviation and difference from mean elevation (DVME and DFME), relative topographical position (RTGP) and elevation percentile (ELVP). Also, seven surface curvature measures (gaussian, maximal, mean, minimal, plan, tangential, and total curvature) were extracted from the DEM and used as WS predictors. In Switzerland, Foresti et al. (2011) used altitude (ELVT), geographic coordinates (XGEO and YGEO), and Differences of Gaussians (DOGS) to map WS. DOGS serves as a measure of terrain convexity and approximates the Laplacian of Gaussian (LPGS: Lowe, 2004). In the current study, DOGS and LPGS were both evaluated. Veronesi et al. (2015) employed topographical surface roughness from a DEM to interpolate the parameters of the Weibull distribution for wind resource mapping. Alternative topographical surface roughness measures employed in the present study were the ruggedness index (RUGI), the surface area ratio (SART) and the standard deviation of the slope (STDS). Etienne et al. (2010) generated landform classes (ex: canyons, ridges, valleys) from a DEM to model WS. Geomorphologic phonotypes (GMPG) and the Pennock landform class (PNCL) were two alternative landform classifications used in the present study. The distance from the coast (DSEA) was also used as a WS predictor in the current study, as done by Aniskevich et al. (2017).

Table 1: Description of the predictors and their spatial scale

Predictor	Abbreviation	Description	Spatial scale
Altitude	ELVT	Altitude of the location in m.	
Aspect	ASPC	Slope orientation in degree.	100m, 500m, 1000m, 1500m, 2000m
Deviation from mean elevation	DVME	Difference between the grid cell elevation and the mean of its neighbouring cells normalized by the standard deviation.	100m, 500m, 1000m, 1500m, 2000m
Difference from cell mean elevation	DFME	Difference between the grid cell elevation and the mean of its neighbouring cells.	100m, 500m, 1000m, 1500m, 2000m
Difference of Gaussian	DOGS	Difference between two copies of the DEM smoothed with two different gaussian kernel. Measure land surface curvature.	(100m, 500m), (100m, 1000m), (500m, 1000m), (300m, 500m), (1000m, 2000m), (1000m, 1500m), (100m, 2000m), (500m, 2000m)
Distance to coast	DSEA	The location distance to the coast	
Elevation percentile	ELVP	Percentile of the grid cell elevation relative to the neighbouring cells.	100m, 500m, 1000m, 1500m, 2000m
Gaussian curvature	GSCV	Product between the maximal and the minimal curvature. Measure of surface curvature (Florinsky, 2017).	100m, 500m, 1000m, 1500m, 2000m
Geographical coordinates	XGEO, YGEO	Geographical coordinates of the location.	
geomorphologic phonotypes (geomorphons)	GMPG	Landform element classification with the geomorphons-based method (Jasiewicz and Stepinski, 2013).	
Laplacian of Gaussian	LPGS	Derivative filter used to highlight location of rapid elevation change.	100m, 500m, 1000m, 1500m, 2000m

Maximal curvature	MXCV	Measure of surface curvature (Wilson, 2018).	100m, 500m, 1000m, 1500m, 2000m
Mean curvature	MNCV	Measure of surface curvature (Wilson, 2018).	100m, 500m, 1000m, 1500m, 2000m
minimal curvature	MICV	Measure of surface curvature (Florinsky, 2017).	100m, 500m, 1000m, 1500m, 2000m
Pennock landform class	PNCL	Landform classification based on the slope and curvature of the grid cell (Pennock et al., 1987).	
plan curvature	PLCV	Measure of surface curvature (Florinsky, 2017).	100m, 500m, 1000m, 1500m, 2000m
Relative topographical position	RTGP	Normalized measure of the grid cell elevation relative to its neighbouring cells.	100m, 500m, 1000m, 1500m, 2000m
Ruggedness index	RUGI	A measure of the local terrain heterogeneity (Jasiewicz and Stepinski, 2013; Riley et al., 1999)	100m, 500m, 1000m, 1500m, 2000m
Slope	SLPE	Slope at the grid cell.	100m, 500m, 1000m, 1500m, 2000m
Standard deviation of slope	STDS	Measure of surface roughness (Grohmann et al., 2011).	100m, 500m, 1000m, 1500m, 2000m
Surface area ratio	SART	Measure of the surface roughness (Jenness, 2004).	100m, 500m, 1000m, 1500m, 2000m
Surface roughness length	RGLH	Surface roughness length estimated from land use map.	100m, 500m, 1000m, 1500m, 2000m
tangential curvature	TGCV	Measure of surface curvature (Florinsky, 2017).	100m, 500m, 1000m, 1500m, 2000m
Total curvature	TLCV	Measure of surface curvature.	100m, 500m, 1000m, 1500m, 2000m

3. Materials and method

3.1. Feature selection methods

3.1.1 Forward stepwise regression

The stepwise regression is a greedy FS algorithm extensively covered in the literature. Three variants of the method exist backward, forward, and bi-directional stepwise regression. Backward stepwise regression builds a model with all potential predictors and eliminates the least relevant predictors at each iteration. Forward selection begins with a “null” model containing only a constant term and adds the most relevant predictors to the regression model at each iteration. Bi-directional stepwise regression combines backward and forward stepwise regression. Various criteria have been used in the literature to measure the predictors’ relevancy (ex.: AIC, P-value, R^2-adjusted).

There is a thorough discussion in the literature about the shortcomings of stepwise regression (Whittingham et al., 2006), with Smith (2018) advising against its use. The author found that stepwise regression underperformed as potential predictors increased. However, the method remains widely used in the scientific community. In this paper, a forward stepwise regression (FSWR) was applied as a benchmark. The algorithm was initiated with the null model, and potential predictors that led to the most significant increase in R^2-adjusted were added at each iteration. This procedure is repeated until no candidate variables left could improve the R^2-adjusted. A similar forward stepwise regression approach was implemented by Chen et al. (2019) and performed better than backward stepwise regression for annual average fine particle ($PM_{2.5}$) and nitrogen dioxide (NO_2) concentrations prediction.

3.1.2 Least Absolute Shrinkage and Selection Operator

LASSO algorithm is a penalty-based linear model developed by Tibshirani (1996), which imposes an L1-norm penalization on the regression coefficient forcing some coefficients to zero and thus producing a sparse solution. The LASSO regression coefficient estimates are given by:

$$\hat{\beta} = argmin_{\beta}(Y - X\beta)^T(Y - X\beta) + \alpha \sum_{j=1}^{p} \left|\beta_j\right|^1 \quad (2)$$

Where:

Y: is the response vector

X: is the matrix of predictors

β: are the regression coefficient

p: is the number of predictors

α : is a tuning parameter that controls the degree of penalization

$\alpha \sum_{j=1}^{p} \left|\beta_j\right|^1$: is the penalization term

$|.|^1$: represents the L1-norm of a vector

Zou and Hastie (2005) discussed some limitations of LASSO regression, which renders the algorithm inappropriate for FS in some situations. A particularly relevant limitation in this study is the inferior prediction performance of LASSO regression compared to Ridge regression when there is a high correlation between the predictors.

3.1.3 Elastic Net

LASSO regression can be seen as a particular case of the Bridge regression introduced by Frank and Friedman (1993). In Bridge regression, the penalization term in equation 2 becomes $\alpha \sum_{j=1}^{p} \left|\beta_j\right|^{\gamma}$ $with\, \gamma \geq 0$. LASSO regression is equivalent to Bridge regression when $\gamma = 1$. Another well-known case of Bridge regression is Ridge regression with $\gamma = 2$. With Ridge regression, the regression coefficients are shrunk depending on the predictors' importance, but they are not set to zero if the variables are irrelevant to the regression.

The ENET model combines the Ridge and the LASSO penalty. The Elastic net algorithm minimizes the following equation:

$$min_{\beta}(Y - X\beta)^T(Y - X\beta) + \alpha\lambda \sum_{j=1}^{p} \left|\beta_j\right|^1 + \alpha(1-\lambda) \sum_{j=1}^{p} \left|\beta_j\right|^2 \qquad (3)$$

With $0 \leq \lambda \leq 1$

$\alpha\ and\ \lambda$ are two hyperparameters of the model that can be selected using cross-validation.

3.1.4 Genetic Algorithms

GAGL is an optimization algorithm that emulates natural evolution and selection to find an optimal solution. It has been implemented in several studies for FS (Amini and Hu, 2021; Eseye et al., 2019; Gokulnath and Shantharajah, 2019; Leardi et al., 1992). The algorithm starts with a population of solutions (individuals) initialized randomly. A fitness measure is defined to evaluate every solution in the population. A new population is formed by producing offspring from the best solutions of the old population (by reproduction and genetic mutation). This procedure is repeated until a stopping criterion is reached. Several variations of the algorithm control, among others, how the offspring of the population are bred. The different steps of the genetic algorithm implemented in this study are described as follows:

Step 1: A population was initialized randomly with 50 potential solutions. The solutions were encoded as a sequence of binary strings (the genes), with each gene associated with a particular feature among the candidate features. A selected gene (a feature) was represented by “1” and a none selected gene by “0”. The population is represented by a binary matrix where the rows represent the potential solutions, and an entry represents a feature or a gene.

Step 2: The 50 solutions in the population were evaluated (fitness score), and the best solution was copied without modification to the next generation.

Step 3: The next generation's parents were selected with the roulette wheel selection method: the solutions with the highest fitness score have more chances to be selected as parents for reproduction to produce offspring. The reproduction process was performed through two genetic operators, uniform crossover and mutation.

Step 4: Step 3 was repeated until the new population size equalled the initial population size.

Step 5: Steps 2 to 4 were repeated until the maximum number of iterations was reached.

Table 2 presents different parameters of the algorithm used in this study. The performance of the solutions was evaluated with a 10-fold cross-validation root mean squared error (RMSE) estimated with a simple linear regression model:

$$RMSE(CV) = \frac{1}{10}\sum_{k=1}^{10}\sqrt{\frac{1}{n_k}\sum_{i=1}^{nk}(y_i - \hat{y}_i)^2} \quad (4)$$

Where n_k is the size of the kth fold, and y_i and $\hat{y}_i$ are the observed and predicted WS values.

The fitness score was estimated as a weighted sum of the solution performance (RMSE) and its cardinality (Card) as follows:

$$F_i = w_1/RMSE_i + (1 - w_1)/Card_i \quad (5)$$

Where:

$0 < w_1 < 1$

The probability of selection of a solution for the reproduction process was assigned based on equation 6:

$$Psel_i = F_i/\sum_{i=1}^{50} F_i \quad (6)$$

Table 2: Selected parameters of the genetic algorithm

GA parameter	Value/method
Initial population size	50
Crossover type	Uniform
Crossover probability	0.9
Mutation probability	0.05
Selection process	roulette wheel selection
Maximum number of iterations	100
w_1	0.1, 0.5, 0.7

3.1.5 Minimum redundance — Maximum relevance

Filter-based FS approaches such as maximal relevancy (ex: correlation) do not require the regression model to be evaluated multiple times (ex: in cross-validation); they are relatively computationally efficient and less prone to overfitting. One of their drawbacks is their failure to ignore redundant predictors correlated to the response variable.

The MRMR algorithm is an iterative approach developed by Ding and Hanchuan (2005) to improve conventional filter-based FS approaches. MRMR benefits from the advantages of the filter-based FS approach while ignoring redundant features in the process. At each iteration of the MRMR algorithm, a function measuring the redundancy and relevancy is computed, and the feature that maximizes this function is selected. Several measures of relevancy and redundancy have been proposed in the literature depending on the type of variables (discrete vs. continuous), the desired level of trade-off between relevancy and redundancy (Zhao et al., 2019), and the type of relationship (linear or nonlinear). In this study, the relevancy is measured with the F-statistic ($F(y, x_i)$). The redundancy of a non-selected feature is measured as the inverse of the sum of the correlation between the feature and the selected features (Ding and Hanchuan, 2005), and the MRMR optimization criterion function is:

$$f(x_i) = \frac{F(y,x_i)}{\frac{1}{s}\sum_{j=1}^{s} \rho(x_s,x_i)} \tag{7}$$

Where:

$$F(y,x_i) = \frac{\rho(y,x_i)^2}{[1-\rho(y,x_i)^2]} \times (n-2) \tag{8}$$

$\rho(x_1, x_2)$ is the Pearson correlation coefficient between features x_1 and x_2

$n-2$ is the degree of freedom of a simple linear regression model fitted with n samples, one predictor and a constant term.

At each iteration, the algorithm seeks to find the feature (x_i) which maximizes $f(x_i)$. The stopping criterion of the algorithm (number of selected features to include in the model) is a hyperparameter that can be determined using cross-validation.

3.1.6 Recursive Feature Elimination Support Vector Regression

The RFES algorithm (Guyon et al., 2002) is a backward elimination algorithm. The model is fitted to the data at each iteration, and the least important predictor is removed from the feature set. This process is repeated until a stopping criterion (ex.: minimum size of feature set) is reached. The stopping criterion can be determined through cross-validation. In the RFES algorithm, the importance of a predictor is measured by the square of its associated coefficient in the weight vector (w : equation 17) using the epsilon-insensitive SVR formulation (Vapnik, 2000), with epsilon the maximum tolerable deviation between the predictions and the observed values.

Let $f(x)$ be the linear function used to approximate the relationship between the predictors (x) and the response variables y:

$$f(x) = \langle w.x \rangle + \ b \tag{9}$$

In the epsilon-insensitive SVR formulation (Vapnik, 2000), the loss function is defined as follows:

$$Loss = \begin{cases} 0 & if\ |y - f(x)| \leq \varepsilon \\ |y - f(x)| - \ \varepsilon & otherwise \end{cases} \tag{10}$$

It is desirable to find a solution to equation 9 having w with minimum norm to reduce the model complexity. The optimization problem can be re-written as follows:

$$minimize\ J(w) = \frac{1}{2}\|w\|^2 \tag{11}$$

Subject to:

$$|y_i - \langle w, x_i \rangle + b| \leq \varepsilon \tag{12}$$

With noisy data, $f(x)$ may not satisfy the epsilon-insensitive constraint. Therefore, slack variables ($\xi_i \xi_i^*$) are introduced for each point to allow less restrictive constraints leading to the following formulation (Vapnik, 2000) :

$$minimize\ J(w) = \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{13}$$

$$subject\ to: \begin{cases} y_i - \langle w. x_i \rangle - b & \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i & \leq \varepsilon + \xi_i^* \\ \xi_i \xi_i^* & \geq 0 \end{cases} \tag{14}$$

C: is a regularization parameter

From the objective function and the constraints (Equations 13 and 14), a Lagrange function L is defined by introducing non-negative Lagrange multipliers $\alpha_i \alpha_i^*, \eta_i \eta_i^*$:

$$L = \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{n}(\xi_i + \xi_i^*) - \sum_{i=1}^{n}(\eta_i \xi_i + \eta_i^* \xi_i^*) - \sum_{i=1}^{n} \alpha_i(\varepsilon + \xi_i - y_i + \langle w, x_i \rangle + b) -$$
$$\sum_{i=1}^{n} \alpha_i^*(\varepsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b) \tag{15}$$

At the saddle point, the partial derivatives of L in all directions are null, giving the following equation in w direction:

$$\partial L / \partial w = w - \sum_{i=1}^{n}(\alpha_i + \alpha_i^*) x_i = 0 \tag{16}$$

And

$$w = \sum_{i=1}^{n} (\alpha_i + \alpha_i^*) x_i \tag{17}$$

3.2. Performance evaluation

The RK model was implemented to estimate the WS quantiles using the selected predictors. The RK model can be expressed as follows (Hengl et al., 2007):

$$\hat{y}(s_0) = \sum_{k=0}^{p} \beta_k \times x_k(s_0) + \sum_{i=1}^{n} \lambda_i \times \varepsilon(s_i) \tag{18}$$

Where $\hat{y}(s_0)$ is the estimated WS quantile at the target location (s_0), $x_k(s_0)$ are the values of the predictors at the target location, and β_k are the regression coefficients. λ_i are the ordinary kriging weights, and $\varepsilon(s_i)$ are the regression residuals at the sampled locations.

From the available data (207 samples), 155 samples (training set) were randomly selected for FS and fitting the RK model. The remaining 57 samples (test set) were used for the model evaluation. This procedure is a common practice in statistical modelling for the validation of the results (for instance, Qiu et al. (2022); Sun et al. (2023)). It helps ensure unbiased assessment and generalization of the model's predictive capability. In addition, 10-fold cross-validation was performed on the training set, and the results were presented. Veronesi et al. (2016) used a similar validation procedure to validate their models for predicting WS distribution parameters at unsampled locations in the UK.

The coefficient of determination (R^2), the RMSE, the Relative Root Mean Squared Error (RRMSE), and the Mean Absolute error (MAE) were computed separately for each percent point considered in the study to evaluate the performance of the RK models during the cross-validation and with the test set:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{19}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{20}$$

$$RRMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{\bar{y}}\right)^2} \tag{21}$$

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \tag{22}$$

4. Results

4.1. Wind speed quantiles

WS quantiles corresponding to 14 fixed percentile points for each location were estimated using shape-constrained P-Splines and the Weibull plotting position formula. Table 3 illustrates some statistics of the estimated WS quantiles in the training set.

Table 3: WS quantile statistics (P-Splines)

Percentile	Abbreviation	mean	std	min	25%	50%	75%	max
%		m/s	m/s	m/s	m/s	m/s	m/s	m/s
0.01	P1	19.68	5.72	7.92	15.49	19.54	23.29	45.58
0.1	P2	18.42	5.16	7.67	14.81	18.24	21.55	40.75
1	P3	12.72	3.92	5.75	9.90	12.17	15.07	29.75
5	P4	10.05	3.16	3.79	7.83	9.80	12.14	20.65
10	P5	8.63	2.65	3.15	6.84	8.43	10.48	17.51
20	P6	6.96	2.15	2.39	5.43	6.85	8.44	14.11
30	P7	5.85	1.84	2.00	4.61	5.73	7.08	11.88
40	P8	4.98	1.59	1.75	3.92	4.91	5.93	10.11
50	P9	4.24	1.39	1.56	3.33	4.17	5.12	8.61
60	P10	3.57	1.21	1.34	2.79	3.46	4.35	7.30

70	P11	2.92	1.02	1.07	2.28	2.77	3.53	6.08
80	P12	2.28	0.82	0.80	1.74	2.15	2.73	4.94
90	P13	1.55	0.58	0.51	1.19	1.45	1.87	3.41
95	P14	1.08	0.42	0.35	0.79	1.03	1.30	2.34

4.2. Model performances

The average R^2, RMSE, RRMSE, and MAE of the cross-validation with the training and the test set are listed in table 4 and table 5, respectively. When evaluated by cross-validation, the average R^2 ranges between 0.18 and 0.50, and the average RRMSE ranges between 22.4% and 33.1%. On the test set, the average R^2 ranges between 0.14 and 0.60, and the average RRMSE ranges between 20.7% and 35.5%. Model performance measured by cross-validation showed that GAGL was the best-performing FS algorithm, followed by MRMR, ENET and LASSO. On the test set, ENET, LASSO, and MRMR were the best-performing FS methods, and GALG and RFES had relatively medium performances. FSWR was the worst-performing FS method during cross-validation and with the test set.

A two-sample t-test ($H_0: \mu_{\Delta RRMSE} \geq \mu_0$, $H_1: \mu_{\Delta RRMSE} < \mu_0$) was conducted to assess the difference between the expected RRMSE ($\mu_{\Delta RRMSE} = \mu_{1RRMSE} - \mu_{2RRMSE}$) of pairs of FS methods on the test set. The results are presented in table 6. The expected RRMSE of FSWR is significantly superior to the expected RRMSE of all the other FS methods. Also, ENET, LASSO, and MRMR performances were not significantly different when considering the RRMSE. However, ENET, LASSO, and MRMR performances were significantly superior (lower RRMSE) to GALG and RFES at the significance level of α = 0.05. There was no statistically significant difference between the expected RRMSE of GAGL and RFES.

Table 4: Performance of FS methods with cross-validation on the training set

FS method	R^2	RMSE	RRMSE	MAE
	-	m/s	-	m/s
ENET	0.410	1.668	0.246	1.238
FSWR	0.125	2.978	0.347	1.699
GALG	0.510	1.432	0.222	1.120
LASSO	0.408	1.659	0.246	1.239
MRMR	0.417	1.645	0.244	1.230
RFES	0.317	1.702	0.272	1.238

Table 5: Performance of FS methods on the test set

FS method	R^2	RMSE	RRMSE	MAE
	-	m/s	-	m/s
ENET	0.559	1.312	0.21	0.911
FSWR	0.137	2.307	0.353	1.555
GALG	0.491	1.438	0.233	1.002
LASSO	0.596	1.231	0.207	0.869
MRMR	0.602	1.226	0.211	0.847
RFES	0.459	1.506	0.24	1.053

Table 6: Results of the t-test between the expected RRMSE of pairs of FS methods

		μ_{2RRMSE}					
	FS method	**ENET**	**FSWR**	**GALG**	**LASSO**	**MRMR**	**RFES**
μ_{1RRMSE}	**ENET**		-4.69	-2.03	0.59	-0.18	-1.88
	FSWR	4.69*		3.46*	4.77*	4.51*	3.14*
	GALG	2.03*	-3.46		2.26*	2.21*	-0.78
	LASSO	-0.59	-4.77	-2.26		-0.91	-2.09
	MRMR	0.18	-4.51	-2.21	0.91		-1.99
	RFES	1.88*	-3.14	0.78	2.09*	1.99	

***: $\mu_{1RRMSE} - \mu_{2RRMSE}$ is significantly less than 0 at α = 0.05**

The RRMSE of the FS methods is presented in figure 2 for the standalone multilinear regression model (REG) and the regression-kriging model (ROK). The kriging of the regression model residuals led to a slight improvement in the performance metric. On average, the residual kriging decreased the RRMSE by 4%.

Figure 3 presents the RRMSE of the different WS quantiles. The model performance deteriorated as the probability of exceedance increased. For example, the mean RRMSE for the estimation of P1 is 17.0% (excluding FSWR), 21.4% for P9 (excluding FSWR), and 29.3% for P14 (excluding FSWR). FSWR performed relatively poorly for the estimation of P1 to P8 and improved for the estimation of P9 to P14.

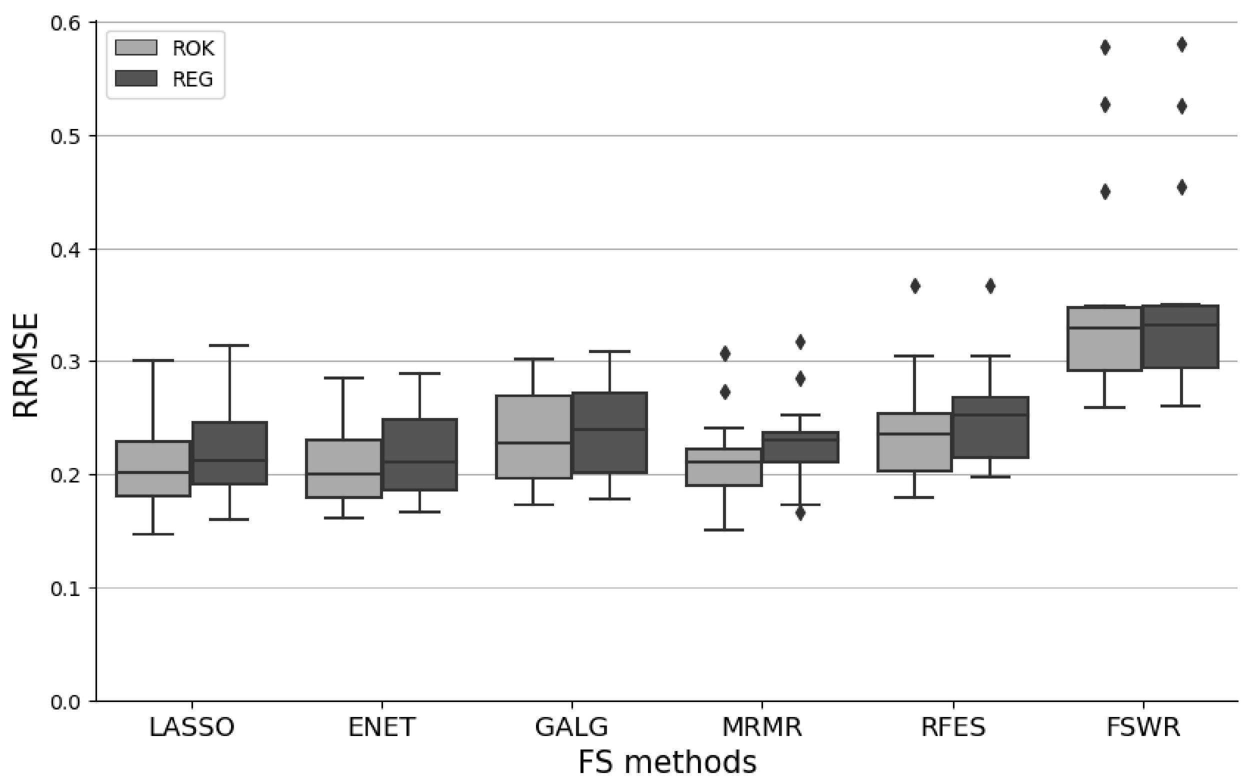

Figure 2: RRMSE of the standalone multilinear regression model (REG) and the regression-kriging model (ROK)

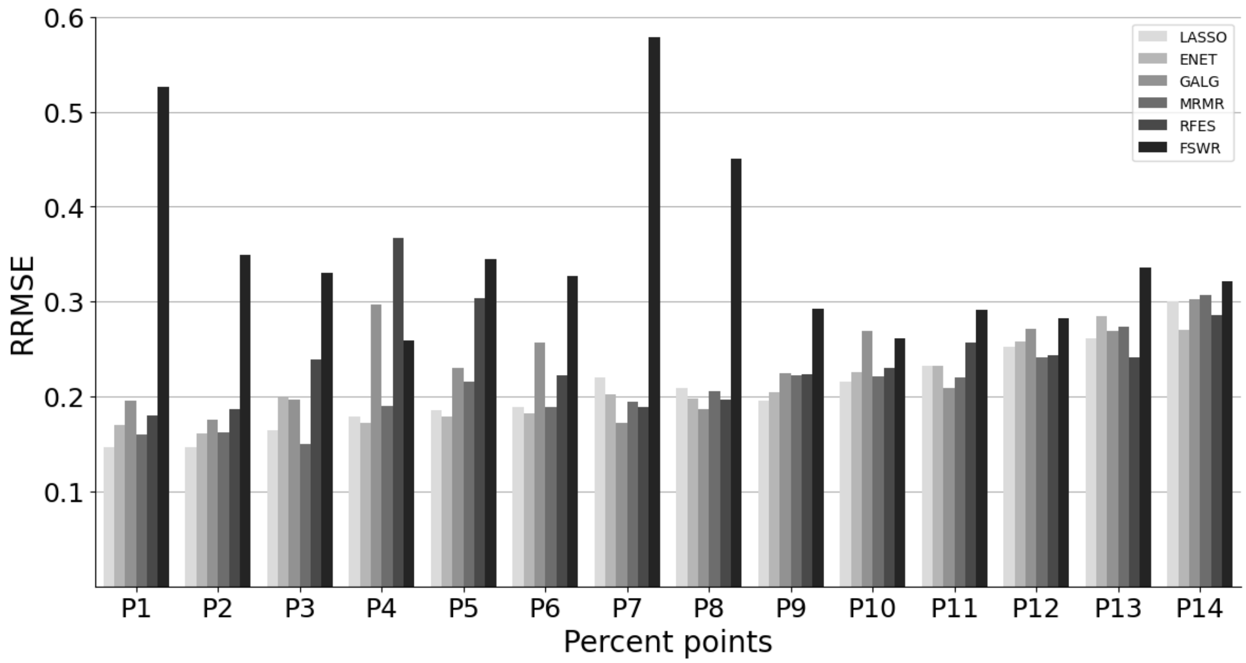

Figure 3: RRMSE of the FS methods for the estimation of different WS quantiles

4.3. Parsimony and multicollinearity

Figure 4 presents the mean number of selected features for each FS method. On average, the FSWR (44) method was the least sparse of the algorithm, followed by RFES (18) and GAGL (17). LASSO selected, on average, five features and was the sparsest FS method, followed by ENET (9) and MRMR (11). Figure 5 illustrates the mean number of selected features against the mean RRMSE. In general, the performance of the FS methods decreased (the RRMSE increased) as the number of selected features increased. Although FSWR selected many features, the model's performance remained relatively poor. As seen previously, the performance of ENET, LASSO and

MRMR were not statistically different (two-sample t-test of the expected RRMSE), but LASSO was, on average, slightly more parsimonious.

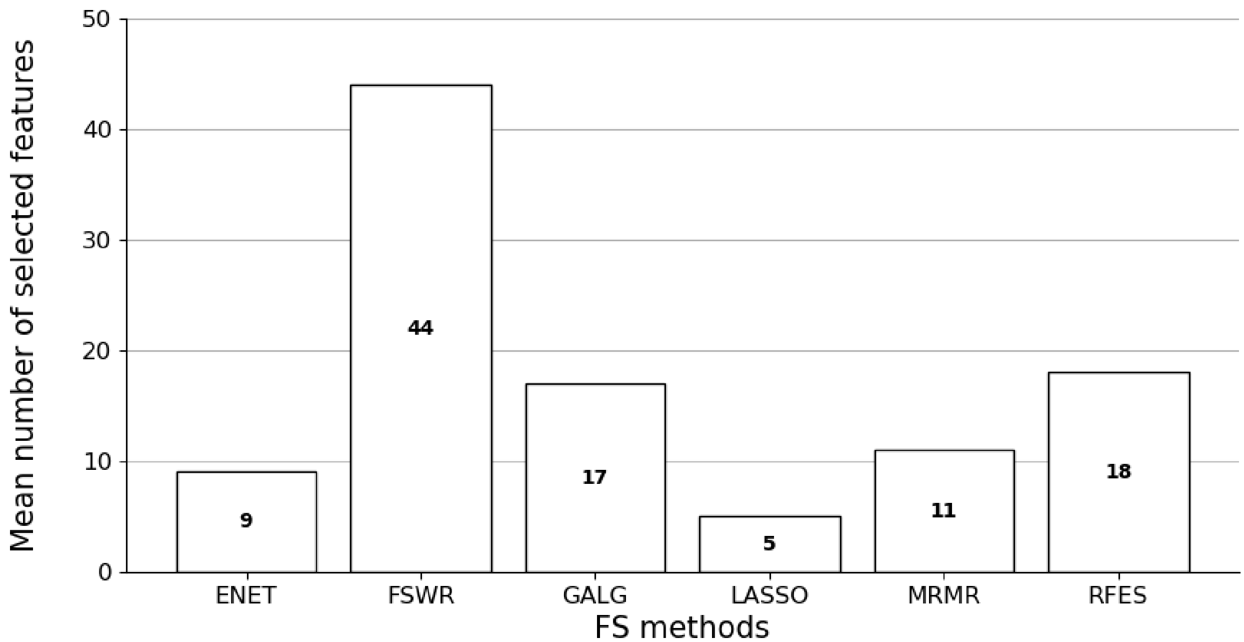

Figure 4: Mean number of selected features of each FS method

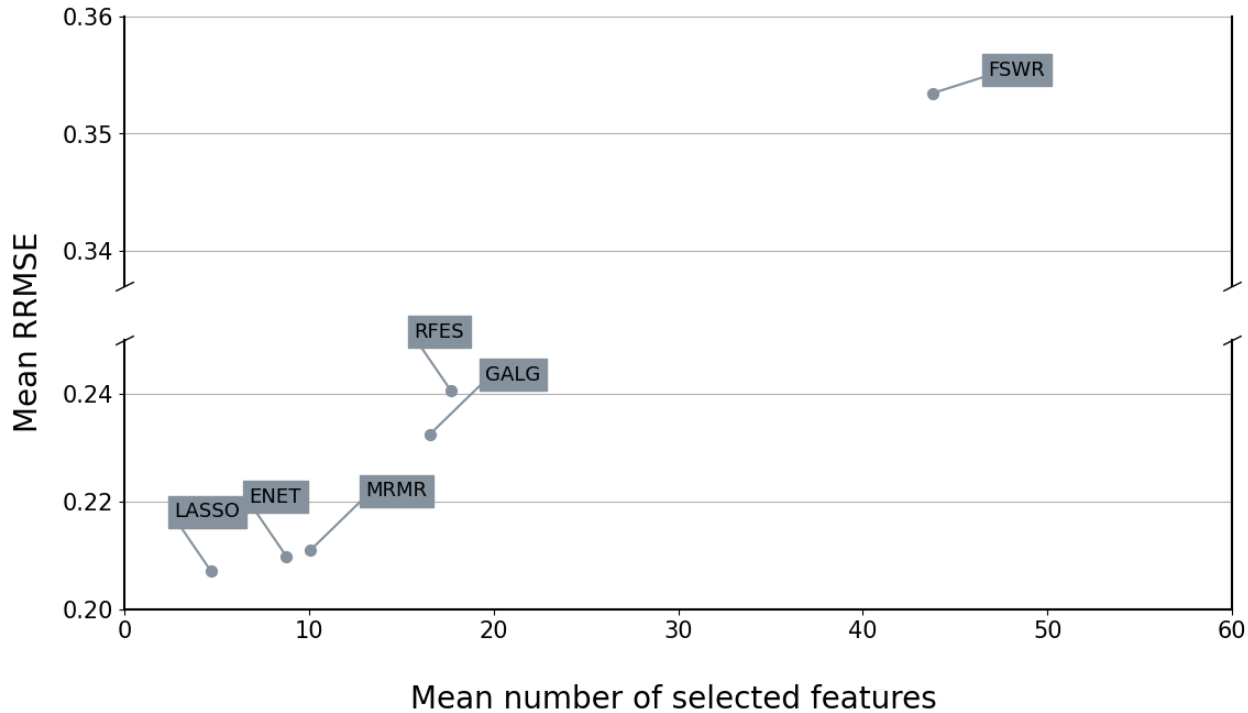

Figure 5: Mean number of selected features vs. mean RRMSE

The condition number (C) is a measure used to evaluate the presence of multicollinearity in a set of predictors. It is defined as the square root of the ratio between the maximum and the minimum eigenvalue of the predictor's correlation matrix. It is a single value summarising the likelihood of multicollinearity. Figure 6 shows the condition number estimated from the correlation matrix of the selected feature sets. From empirical observations, Chatterjee and Hadi (2013) suggested a cut-off of 15 to detect multicollinearity and recommended corrective action if C exceeds 30. All the feature sets estimated with LASSO had a condition number below 15. In the case of MRMR, the condition numbers were less than 15 in 13 cases out of 14 (92.8%) and were consistently below 30. For ENET and GAGL, the condition number was less than 15 in 11 cases out of 14 (78.6%). RFES condition numbers were inferior to 15 in 8 cases out of 14 (57.1%), and FSWR condition numbers consistently exceeded 15.

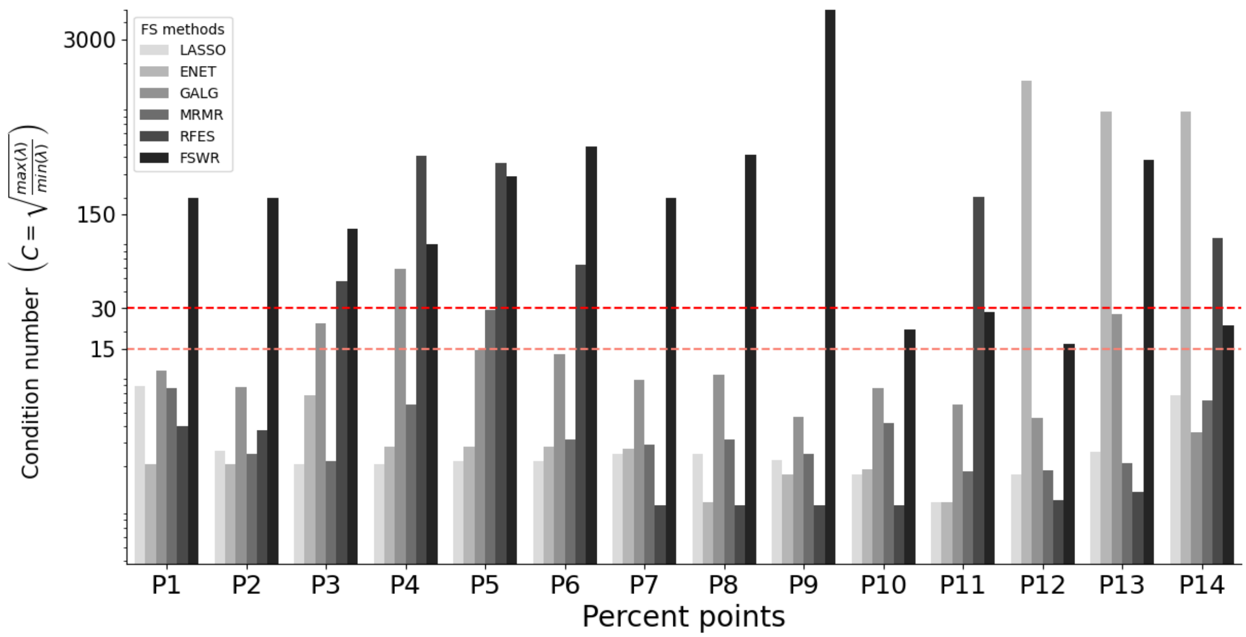

Figure 6: Condition number (C) of the selected feature sets

4.4. Residual analysis and visual inspection

Model residual analyses compare observed data with predicted values to evaluate a model's precision and reliability. Examining residuals can reveal patterns, outliers, and areas for improvement in the model's assumptions. Figure 7 compares the observed and predicted WS quantiles for the top-performing FS methods (MRMR and LASSO), indicating a strong agreement between the observed and estimated quantile for both methods with an R^2 of approximately 0.92. LASSO performed slightly better than MRMR, as indicated by the RRMSE. Two outliers were identified in the bottom-right section of the plots, with an underestimation of the WS quantiles for both outliers. The residual plot in Figure 8 confirmed that the models did not perform as well for high exceedance probabilities as they did for the lower ones.

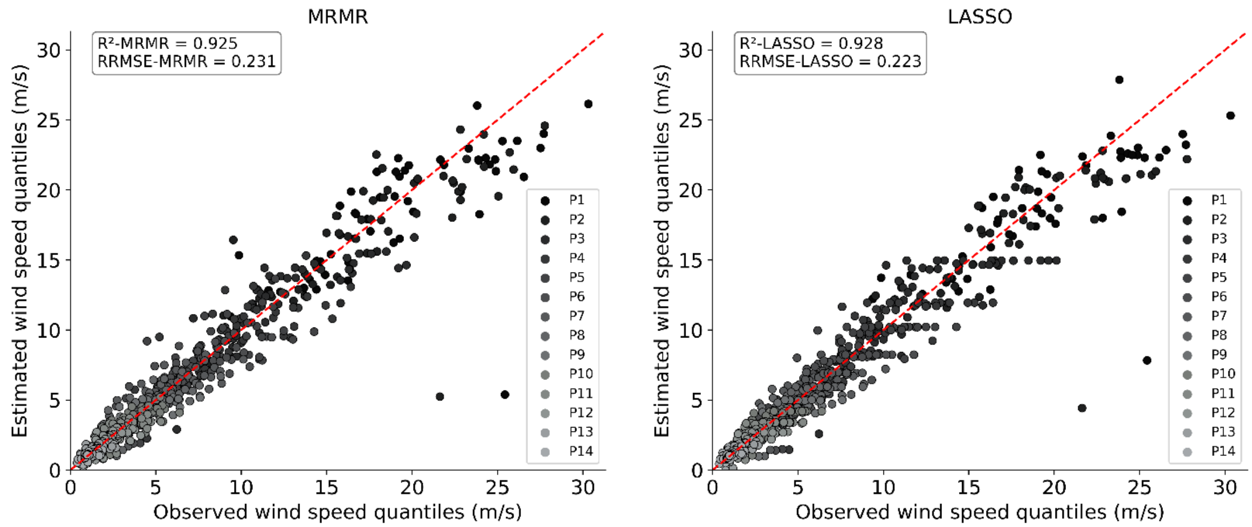

Figure 7: Plot of observed vs. estimated WS quantiles for MRMR and LASSO. The R^2 was calculated without averaging across the percent points.

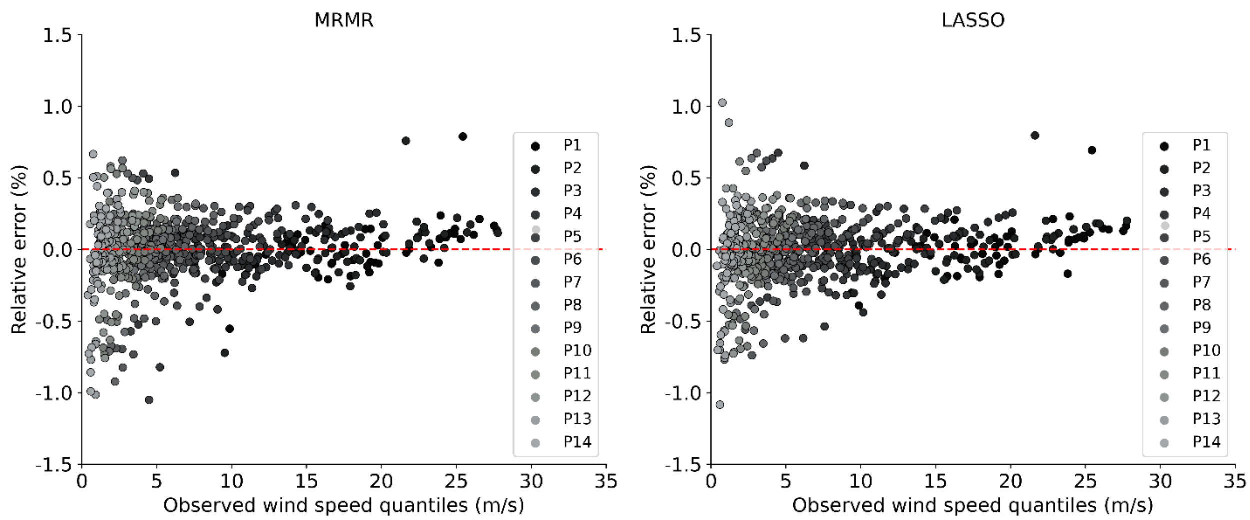

Figure 8: Plot of observed WS quantiles vs. the residuals for MRMR and LASSO.

4.5. Predictor importance

Figure 9 shows the ten most selected features for each WS quantile and the number of times they were selected. Overall, the most selected features were RGLH and DSEA. DSEA was consistently selected by every FS method. For the surface roughness length (RGLH), 2000m and 1000m (RGLH_2000m and RGLH_1000m) were the most selected spatial scales. RGLH at 100m spatial scale (RGLH_100m) was mostly selected for medium to high exceedance probabilities. DVME at a spatial scale of 2000m (DVME_2000m) was often selected for high exceedance probabilities (P10 – P14) and less often selected for lower exceedance probabilities (P1 to P9). Predictors describing the land surface curvature (MXCV, MNCV, TLCV, TGCV, GSCV) seemed important for predicting WS quantiles corresponding to very low exceedance probabilities (P1 to P5) and less important for medium and high exceedance probabilities. PNCL was also among the most selected predictors, especially class 7 of PNCL (PNLC_7), which indicates a level terrain at the grid cell with a low slope gradient. The location coordinates (XGEO and YGEO) were also often selected for different WS quantiles in the region.

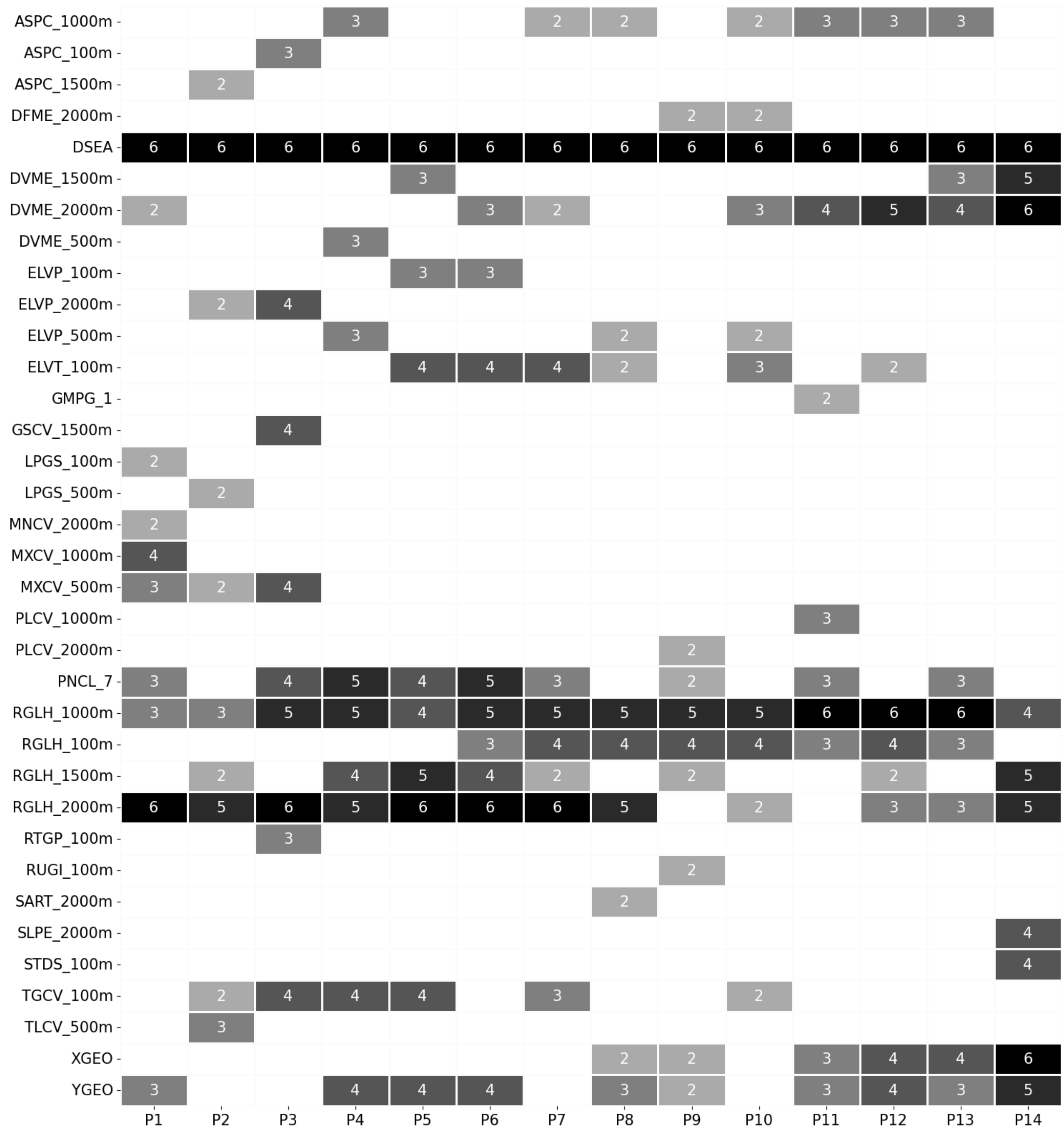

Figure 9: Selected predictors for each WS quantile

The predictors selected by the FS methods were used to fit a simple linear regression model. An advantage of the simple linear regression model is the interpretability of the model. Without multicollinearity, the regression coefficient magnitude and direction provide useful information to assess the relationship between the predictors and the dependent variable. Figure 10 shows the regression coefficient of the predictors selected with LASSO. The predictors were standardized to a zero mean and a unit variance prior to fitting the regression model. LASSO was the most parsimonious FS method with good predictive ability. In addition, the estimated condition numbers of all the feature sets selected by LASSO were below 15, indicating the absence of multicollinearity. It is observed that DSEA regression coefficients were often the strongest and were always negative. DSEA represents the location distance from the coast; the direction of the regression coefficient showed that WS quantiles, irrespective of their exceedance probabilities, were higher near the coast than inland. The surface roughness length (RGLH) showed relatively high regression coefficients with every WS quantile. The negative direction of the regression coefficient of RGLH is intuitive. An increase in surface roughness results in more friction between the land surface and the wind, decreasing WS near the ground. For P1 and P2, the maximum curvature (MXCV) had the second-highest regression coefficient with a positive direction. Note that higher values of MXCV correspond to elongated convex landforms such as ridges, and negative values are associated with concave landforms (Florinsky, 2017). The positive magnitude of the MXCV regression coefficient showed that the WS quantiles P1 and P2 were higher at locations where the landforms are convex and decreased as the landform concavity increased.

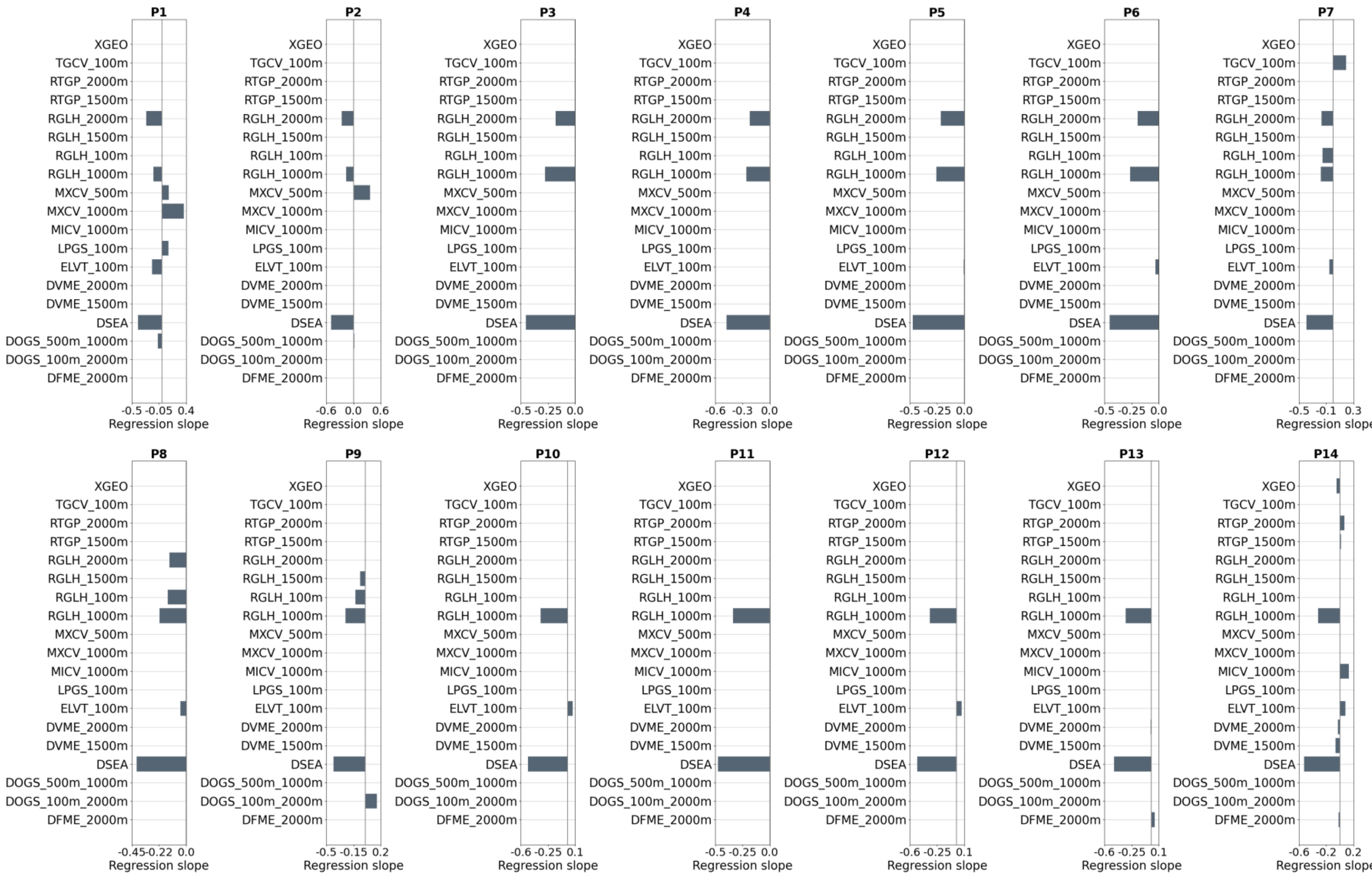

Figure 10: Regression coefficients of the WS quantile predictors

5. Discussion

This study compared six FS methods for WS quantile estimation in Canada. The results showed that LASSO, MRMR, and ENET had comparable performances on the test set and were the most effective FS methods. GAGL and RFES performed slightly worse than LASSO, MRMR, and ENET but outperformed FSWR. The FSWR method does not seem to ignore redundant features, leading to an unstable estimation of regression coefficients and poor performance during testing. This situation seems more pronounced for low than high exceedance probabilities (P10 to P14). There was less collinearity among the relevant predictors associated with high exceedance probabilities than for lower ones. Kriging of the regression residual slightly improved the model performances (4%), indicating that the selected predictors and the linear regression model could account for a significant portion of the spatial variability of WS quantiles in the region.

The models' performances were higher for low to medium exceedance probabilities and declined for high exceedance probabilities. This decline in performance could be attributed to several factors. One possible explanation is that there is a significant non-linear relationship between high exceedance probabilities WS and the predictors, requiring the implementation of non-linear models for improved performance. Another possible explanation is the exclusion of significant predictors of high exceedance probabilities from the models. For example, the models did not include climate-related predictors such as mean temperature or pressure. Climatic variables are often collected at meteorological stations where WS is also measured; thus, they should also be missing at locations with unavailable WS data. The results highlight the need for further research to enhance the performance of models in predicting high exceedance probability WS.

LASSO was found to produced, on average, the sparsest feature sets, followed by ENET and MRMR. In addition, LASSO could select relevant predictors without multicollinearity as evaluated by the feature set correlation matrix condition number. MRMR also eliminated multicollinearity in most cases (13 out of 14 cases), while ENET, RFES, GAGL, and FSWR were less effective at solving the issue of multicollinearity in their selected feature sets. These findings are consistent with existing literature on RFES: Xie et al. (2006) showed that this implementation does not consider feature redundancy. Overall, LASSO and MRMR were the most effective FS methods due of the following reasons:

- LASSO and MRMR exhibited high predictive ability, with no significant difference in performance between the two methods based on t-test results and residual analysis.
- Both FS methods could select relevant predictors while also reducing multi-collinearity within the feature subset.
- LASSO and MRMR are attractive because they are efficient to implement with a single parameter to tune, unlike ENET, which produced comparable performance. In the case of LASSO, the degree of penalization (α) is the only parameter that needs to be tuned. With MRMR, the number of features to select is the single tuning parameter of the algorithm. ENET requires the tuning of two parameters.

LASSO and MRMR have different approaches to feature selection. However, their good performance in the study could be explained by their inbuilt capability to select relevant features while ignoring redundant ones. LASSO is a penalization algorithm based on linear regression that promotes sparsity by imposing a penalty on the sum of the absolute values of the feature coefficients. In a group of redundant predictors, LASSO chooses one predictor among the group and shrinks towards zero the coefficients of the other predictors (Hammami et al., 2012; Zou and Hastie, 2005) , making it effective in dealing with collinear features.

On the other hand, MRMR ranks features from the most relevant and least redundant to the least relevant and most redundant, allowing for efficient selection of the smallest subset of the most relevant and least redundant features that provides the best cross-validation score. In addition, MRMR is a filter-based approach that is agnostic to any specific regression model, as it is based on the correlation coefficient. This coefficient is well suited for the linear regression model used in this study. However, other correlation metrics, such as mutual information, can be used for nonlinear models.

It is worth noting that GAGL showed superior performance during cross-validation on the training sets, and there was no significant decline in its performance on the test set. However, in some feature subsets selected by GAGL, the issue of multicollinearity remained unresolved. In addition, compared to LASSO and MRMR, GAGL has more parameters that require tuning, making it less efficient to implement.

In the present study, the location distance from the coast (DSEA) and the surface roughness length (RGHL) were the two most significant predictors of WS quantiles. The regression model coefficients for both DSEA and RGHL were physically consistent. In the case of DSEA, the regression coefficients were negative, indicating a decrease in the WS quantiles from coastal to inland areas. Few studies have used the distance from the coast to estimate WS, but it could be a valuable addition to models, particularly in larger study areas. For low exceedance probabilities (ex: 1%), surface convexity (concavity) was a significant predictor of WS, but it was less relevant for higher exceedance probabilities.

There are some limitations to this study. The dataset contained only 207 samples (155 training and 52 testing samples), and some regions of Canada were naturally less densely represented (see figure 1). Consequently, some results could be particular to the studied region or the analyzed dataset and may only be generalized after extensive analysis.

Among the various feature selection (FS) methods examined, the FSWR approach was the least effective. It is possible to improve the FSWR method performance by adding the variance inflation factor as a post-processing step. It should be noted that, the FSWR method in this study was mainly used as a benchmark for assessing the performance of other proposed FS methods as it remains one of the most common FS methods.

In the present study, the time series were considered stationary when estimating WS quantiles. Nevertheless, increased evidence points to non-stationarities in WS series and the importance of incorporating them in the analysis (see, for instance, Ouarda and Charron (2021)). For instance, several authors observed significant correlations between low-frequency climate oscillation indices and annual mean WS in different regions of the world (see, for instance, Naizghi and Ouarda (2017); Woldesellasse et al. (2020)); Including these climate oscillation indices in quantile estimation or regional transfer models could significantly improve their performances. Indeed, in a given region, WS stations are impacted by the same climate oscillation indices, and their incorporation in the models used to estimate WS at ungauged locations should lead to performance improvements. The issue of incorporation of teleconnections in WS estimation models is an important one but remains mainly unexplored in the literature. Future efforts should focus on incorporating non-stationarities in regional WS estimation models.

6. Conclusions

This paper evaluated six FS methods for WS quantile estimation. LASSO and MRMR were the most efficient algorithms in the study. It was found that the importance of some WS quantile predictors depends on their exceedance probability. The location distance from the coast and the surface roughness length were significant WS quantile predictors irrespective of the exceedance probability.

Future research should focus on the extrapolation of this study to other geographic regions, databases with different characteristics, and other FS methods. The diversity in the characteristics needs to be ensured to obtain guidelines for the relative performance and the applicability of different techniques based on such considerations as the number of sites, the length of the series, the number of features, the types of wind, the data variability and quality, etc.

Acknowledgements

The authors would like to thank Environment and Climate Change Canada for providing free access to the wind speed time series and the Japan Aerospace Exploration Agency for providing free access to ALOS Global Digital Surface Model. The authors would like to extend their gratitude to Prof. Marc Rosen, the Editor, and two anonymous reviewers for their comments and suggestions, which significantly enhanced the quality of the paper.

Funding

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chairs Program.

References

Alsamamra, H. et al., 2010. Mapping surface wind speed in Andalusia (Southern Spain) based on residual kriging, 10th EMS Annual Meeting, pp. EMS2010-124.

Amini, F., Hu, G., 2021. A two-layer feature selection method using Genetic Algorithm and Elastic Net. Expert Systems with Applications, 166: 114072. DOI:https://doi.org/10.1016/j.eswa.2020.114072

Aniskevich, S., Bezrukovs, V., Zandovskis, U., Bezrukovs, D., 2017. Modelling the Spatial Distribution of Wind Energy Resources in Latvia. Latvian Journal of Physics and Technical Sciences, 54(6): 10-20. DOI:https://doi.org/10.1515/lpts-2017-0037

Aries, N., Boudia, S.M., Ounis, H., 2018. Deep assessment of wind speed distribution models: A case study of four sites in Algeria. Energy Conversion and Management, 155: 78-90. DOI:https://doi.org/10.1016/j.enconman.2017.10.082

Carta, J.A., Cabrera, P., Matías, J.M., Castellano, F., 2015. Comparison of feature selection methods using ANNs in MCP-wind speed methods. A case study. Applied Energy, 158: 490-507. DOI:https://doi.org/10.1016/j.apenergy.2015.08.102

Chatterjee, S., Hadi, A.S., 2013. Regression analysis by example. WILEY SERIES IN PROBABILITY AND STATISTICS. John Wiley & Sons, Hoboken, New Jersey, USA, 424 pp.

Chen, J. et al., 2019. A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide. Environment International, 130: 104934. DOI:https://doi.org/10.1016/j.envint.2019.104934

Ding, C., Hanchuan, P., 2005. Minimum redundancy feature selection from microarray gene expression data. J Bioinform Comput Biol, 3(2): 185-205. DOI:10.1142/s0219720005001004

Eseye, A.T., Lehtonen, M., Tukia, T., Uimonen, S., Millar, R.J., 2019. Machine Learning Based Integrated Feature Selection Approach for Improved Electricity Demand Forecasting in Decentralized Energy Systems. IEEE Access, 7: 91463-91475. DOI:10.1109/ACCESS.2019.2924685

Etienne, C., Lehmann, A., Goyette, S., Lopez-Moreno, J.-I., Beniston, M., 2010. Spatial Predictions of Extreme Wind Speeds over Switzerland Using Generalized Additive Models. Journal of Applied Meteorology and Climatology, 49(9): 1956-1970. DOI:10.1175/2010jamc2206.1

Florinsky, I.V., 2017. An illustrated introduction to general geomorphometry. Progress in Physical Geography: Earth and Environment, 41(6): 723-752. DOI:10.1177/0309133317733667

Foresti, L., Tuia, D., Kanevski, M., Pozdnoukhov, A., 2011. Learning wind fields with multiple kernels. Stochastic Environmental Research and Risk Assessment, 25(1): 51-66. DOI:10.1007/s00477-010-0405-0

Fouad, G., Loáiciga, H.A., 2020. Independent variable selection for regression modeling of the flow duration curve for ungauged basins in the United States. Journal of Hydrology, 587: 124975. DOI:https://doi.org/10.1016/j.jhydrol.2020.124975

Frank, L.E., Friedman, J.H., 1993. A statistical view of some chemometrics regression tools. Technometrics, 35(2): 109-135.

Gokulnath, C.B., Shantharajah, S.P., 2019. An optimized feature selection based on genetic approach and support vector machine for heart disease. Cluster Computing, 22(6): 14777-14787. DOI:10.1007/s10586-018-2416-4

Grohmann, C.H., Smith, M.J., Riccomini, C., 2011. Multiscale Analysis of Topographic Surface Roughness in the Midland Valley, Scotland. IEEE Transactions on Geoscience and Remote Sensing, 49(4): 1200-1213. DOI:10.1109/TGRS.2010.2053546

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. Journal of machine learning research, 3(Mar): 1157-1182.

Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene Selection for Cancer Classification using Support Vector Machines. Machine Learning, 46(1): 389-422. DOI:10.1023/A:1012487302797

Hammami, D., Lee, T.S., Ouarda, T.B.M.J., Lee, J., 2012. Predictor selection for downscaling GCM data with LASSO. Journal of Geophysical Research: Atmospheres, 117(D17). DOI:https://doi.org/10.1029/2012JD017864

Hengl, T., Heuvelink, G.B.M., Rossiter, D.G., 2007. About regression-kriging: From equations to case studies. Computers & Geosciences, 33(10): 1301-1315. DOI:https://doi.org/10.1016/j.cageo.2007.05.001

Houndekindo, F., Ouarda, T.B.M.J., 2023. Statistical approaches for wind speed estimation at ungauged or partially gauged locations, review, and open questions (Under review). Institut national de la recherche scientifique, Centre Eau Terre Environnement.

International Renewable Energy Agency, 2022. World Energy Transitions Outlook 2022: 1.5°C Pathway, Abu Dhabi.

Jasiewicz, J., Stepinski, T.F., 2013. Geomorphons — a pattern recognition approach to classification and mapping of landforms. Geomorphology, 182: 147-156. DOI:https://doi.org/10.1016/j.geomorph.2012.11.005

Jenness, J., 2004. Calculating Landscape Surface Area from Digital Elevation Models. Wildlife Society Bulletin, 32: 829-839. DOI:10.2193/0091-7648(2004)032[0829:CLSAFD]2.0.CO;2

Jung, C., 2016. High Spatial Resolution Simulation of Annual Wind Energy Yield Using Near-Surface Wind Speed Time Series. Energies, 9(5): 344. DOI:doi:10.3390/en9050344

Jung, C., Schindler, D., Laible, J., 2018. National and global wind resource assessment under six wind turbine installation scenarios. Energy Conversion and Management, 156: 403-415. DOI:https://doi.org/10.1016/j.enconman.2017.11.059

Lamb, W.F. et al., 2021. A review of trends and drivers of greenhouse gas emissions by sector from 1990 to 2018. Environmental Research Letters, 16(7): 073005. DOI:10.1088/1748-9326/abee4e

Latifovic, R., Pouliot, D., Olthof, I., 2017. Circa 2010 Land Cover of Canada: Local Optimization Methodology and Product Development. Remote Sensing, 9(11): 1098.

Leardi, R., Boggia, R., Terrile, M., 1992. Genetic algorithms as a strategy for feature selection. Journal of Chemometrics, 6(5): 267-281. DOI:https://doi.org/10.1002/cem.1180060506

Lee, C., 2022. Long-term wind speed interpolation using anisotropic regression kriging with regional heterogeneous terrain and solar insolation in the United States. Energy Reports, 8: 12-23. DOI:https://doi.org/10.1016/j.egyr.2021.11.285

Lindsay, J.B., 2014. The Whitebox Geospatial Analysis Tools project and open-access GIS, GIS Research UK 22nd Annual Conference. The University of Glasgow, University of Glasgow. DOI:10.13140/RG.2.1.1010.8962

Lowe, D.G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, 60(2): 91-110. DOI:10.1023/B:VISI.0000029664.99615.94

Maxwell, A.E., Shobe, C.M., 2022. Land-surface parameters for spatial predictive mapping and modeling. Earth-Science Reviews, 226: 103944. DOI:https://doi.org/10.1016/j.earscirev.2022.103944

Naizghi, M.S., Ouarda, T.B.M.J., 2017. Teleconnections and analysis of long-term wind speed variability in the UAE. International Journal of Climatology, 37(1): 230-248. DOI:https://doi.org/10.1002/joc.4700

Ouarda, T.B.M.J., Charron, C., 2021. Non-stationary statistical modelling of wind speed: A case study in eastern Canada. Energy Conversion and Management, 236: 114028. DOI:https://doi.org/10.1016/j.enconman.2021.114028

Paul, H.C.E., Marx, B.D., 1996. Flexible Smoothing with B-splines and Penalties. Statistical Science, 11(2): 89-102.

Pennock, D.J., Zebarth, B.J., De Jong, E., 1987. Landform classification and soil distribution in hummocky terrain, Saskatchewan, Canada. Geoderma, 40(3): 297-315. DOI:https://doi.org/10.1016/0016-7061(87)90040-1

Pya, N., Wood, S.N., 2015. Shape constrained additive models. Statistics and Computing, 25(3): 543-559. DOI:10.1007/s11222-013-9448-7

Qiu, R. et al., 2022. Generalized Extreme Gradient Boosting model for predicting daily global solar radiation for locations without historical data. Energy Conversion and Management, 258: 115488. DOI:https://doi.org/10.1016/j.enconman.2022.115488

Reinhardt, K., Samimi, C., 2018. Comparison of different wind data interpolation methods for a region with complex terrain in Central Asia. Climate Dynamics, 51(9): 3635-3652. DOI:10.1007/s00382-018-4101-y

Riley, S.J., DeGloria, S.D., Elliot, R., 1999. Index that quantifies topographic heterogeneity. intermountain Journal of sciences, 5(1-4): 23-27.

Robert, S., Foresti, L., Kanevski, M., 2013. Spatial prediction of monthly wind speeds in complex terrain with adaptive general regression neural networks. International Journal of Climatology, 33(7): 1793-1804. DOI:https://doi.org/10.1002/joc.3550

Rodriguez-Galiano, V.F., Luque-Espinar, J.A., Chica-Olmo, M., Mendes, M.P., 2018. Feature selection approaches for predictive modelling of groundwater nitrate pollution: An evaluation of filters, embedded and wrapper methods. Science of The Total Environment, 624: 661-672. DOI:https://doi.org/10.1016/j.scitotenv.2017.12.152

Shin, J.-Y., Ouarda, T.B.M.J., Lee, T., 2016. Heterogeneous mixture distributions for modeling wind speed, application to the UAE. Renewable Energy, 91: 40-52. DOI:https://doi.org/10.1016/j.renene.2016.01.041

Smith, G., 2018. Step away from stepwise. Journal of Big Data, 5(1): 32. DOI:10.1186/s40537-018-0143-6

Sun, Y., Zhu, D., Li, Y., Wang, R., Ma, R., 2023. Spatial modelling the location choice of large-scale solar photovoltaic power plants: Application of interpretable machine learning techniques and the national inventory. Energy Conversion and Management, 289: 117198. DOI:https://doi.org/10.1016/j.enconman.2023.117198

Tadono, T. et al., 2014. Precise Global DEM Generation by ALOS PRISM. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, II4: 71-76. DOI:10.5194/isprsannals-II-4-71-2014

Tibshirani, R., 1996. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1): 267-288. DOI:https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Urbanowicz, R.J., Meeker, M., La Cava, W., Olson, R.S., Moore, J.H., 2018. Relief-based feature selection: Introduction and review. Journal of Biomedical Informatics, 85: 189-203. DOI:https://doi.org/10.1016/j.jbi.2018.07.014

Vapnik, V.N., 2000. Methods of Function Estimation. In: Vapnik, V.N. (Ed.), The Nature of Statistical Learning Theory. Springer New York, New York, NY, pp. 181-224. DOI:10.1007/978-1-4757-3264-1_7

Veronesi, F., Grassi, S., Raubal, M., 2016. Statistical learning approach for wind resource assessment. Renewable and Sustainable Energy Reviews, 56: 836-850. DOI:https://doi.org/10.1016/j.rser.2015.11.099

Veronesi, F., Grassi, S., Raubal, M., Hurni, L., 2015. Statistical Learning Approach for Wind Speed Distribution Mapping: The UK as a Case Study. In: Bacao, F., Santos, M.Y., Painho, M. (Eds.), AGILE 2015: Geographic Information Science as an Enabler of Smarter Cities and Communities. Springer International Publishing, Cham, pp. 165-180. DOI:10.1007/978-3-319-16787-9_10

Whittingham, M.J., Stephens, P.A., Bradbury, R.B., Freckleton, R.P., 2006. Why do we still use stepwise modelling in ecology and behaviour? Journal of Animal Ecology, 75(5): 1182-1189. DOI:https://doi.org/10.1111/j.1365-2656.2006.01141.x

Wierenga, J., 1993. Representative roughness parameters for homogeneous terrain. Boundary-Layer Meteorology, 63(4): 323-363. DOI:10.1007/BF00705357

Wilson, J.P., 2018. Environmental applications of digital terrain modeling. John Wiley & Sons.

Woldesellasse, H., Marpu, P.R., Ouarda, T.B.M.J., 2020. Long-term forecasting of wind speed in the UAE using nonlinear canonical correlation analysis (NLCCA). Arabian Journal of Geosciences, 13(18): 962. DOI:10.1007/s12517-020-05981-9

Xie, Z.-X., Hu, Q.-H., Yu, D.-R., 2006. Improved Feature Selection Algorithm Based on SVM and Correlation. In: Wang, J., Yi, Z., Zurada, J.M., Lu, B.-L., Yin, H. (Eds.), Advances in Neural Networks - ISNN 2006. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1373-1380. DOI:10.1007/11759966_204

Zhao, Z., Anand, R., Wang, M., 2019. Maximum Relevance and Minimum Redundancy Feature Selection Methods for a Marketing Machine Learning Platform, 2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 442-452. DOI:10.1109/DSAA.2019.00059

Zhou, Y., Liu, Y., Wang, D., Liu, X., Wang, Y., 2021. A review on global solar radiation prediction with machine learning models in a comprehensive perspective. Energy Conversion and Management, 235: 113960. DOI:https://doi.org/10.1016/j.enconman.2021.113960

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2): 301-320. DOI:https://doi.org/10.1111/j.1467-9868.2005.00503.x