Utrecht University

Using Wind Direction Interpolation Methods to Improve Pesticide Exposure Estimates

Student
Thomas Luke Nibbering

Thesis Supervisors
Erik-Jan van Kesteren
Daniel Figueiredo

MSc Applied Data Science
Faculty of Science
Department of Information and Computing Science
Utrecht University
The Netherlands

Abstract

Pesticides play an important role in modern-day agriculture by protecting produce from pests and diseases. Nonetheless, exposure to these biological agents poses a significant public health concern, highlighting the necessity for exposure assessments. In recent years, spatial simulation models have emerged as an effective approach to estimate the extent and distribution of pesticide drift due to their ability to consider a range of factors, such as wind directions. Here, the typical approach is to derive this type of information from meteorological stations closest to the pesticide application area, introducing considerable inaccuracies. In order to provide a more robust and versatile alternative, this study aimed to examine spatial interpolation methods that may improve pesticide exposure estimates using wind field records from the Netherlands in 2017. In doing so, five spatial interpolation models were adopted to estimate wind directions at unobserved sites, namely naïve interpolation, nearest neighbour, inverse distance weighting, universal kriging and random forest. Performance of these models was evaluated using an out-of-sample circular root-mean-squared error (CRMSE) that was obtained through spatial k-fold cross validation. A sensitivity analysis examined the influence of varying observations on the performance of each model. Results showed distinct visual patterns that aligned with previous studies. Nonetheless, limited variability in hourly wind field measures resulted in a relatively similar performance across the employed models. All in all, the inverse distance weighting demonstrated the lowest out-of-sample error for interpolating wind directions. This finding suggested that the adoption of this model in pesticide drift simulations provides a more valid representation of wind fields at the application areas compared to the approach often employed. In turn, this may improve the accuracy of pesticide exposure estimates obtained from these simulations.

Contents

1. Introduction

Pesticides play an important role in modern agriculture [1, 2]. As productivity enhancers, these biological agents are considered to help protect produce from pests and diseases [3–5]. Despite their notable advantages, exposure to pesticides poses a significant public health concern with negative health effects ranging from acute toxicity [6, 7] to neurodegenerative diseases [8–10]. As a result, assessments of exposure are critical to identify areas of high risk and informing decisions aimed at mitigating the adverse health effects of these chemical components.

In recent years, spatial simulation models have emerged as an effective approach to estimate the extent and distribution of pesticides [2, 11–15]. Among the various alternatives available, gaussian plume models are widely adopted due to their ability to incorporate a range of factors that might influence exposure to pesticides in ambient air, including wind directions [2, 13, 14]. Here, the common approach is to derive this type of information from meteorological stations closest to the pesticide application area [13]. However, this method may inaccurately represent the wind direction at the site of interest, resulting in an under- or overestimation of exposure.

To enhance these measures, several studies have incorporated wind direction parameters in their models, obtained from sensors mounted directly on the pesticide sprayer [2, 15]. Apart from this approach being considered practical in nature, it remains deficient in capturing the temporal and spatial variability inherent in wind directions. As a result, some studies proposed spatial interpolation of wind directions as a more versatile means of improving estimates of exposure [16, 17].

Spatial interpolation being applied to climatic observations has gained considerable attention in literature [18, 19]. These techniques aim to estimate variable values at unknown locations based on available samples obtained from climatic stations [20]. In this context, deterministic and geostatistical approaches are the predominant frameworks used [21–28]. Nonetheless, these methods also induce considerable interpolation error and may encounter challenges in adoption [29]. More recently, studies combined statistical learning with spatial interpolation [18, 30–32]. Despite the fact that these methods often outperform the more classical frameworks, literature is generally limited to wind speed or power rather than the direction of the field [32]. In contrast, approaches based on circular statistics focused more on interpolating wind directions by considering the periodic nature of the data [33–35]. In doing so, these studies provided more realistic interpolation results [36]. Nonetheless, empirical results in literature that adopted spatial interpolation of wind directions are limited, specifically in studies that employed gaussian plume models to simulate the drift of pesticides.

Therefore, this study aims to examine spatial interpolation methods that may improve pesticide exposure estimates using wind direction records from the Netherlands in 2017. In doing so, the study contributes to existing literature in two ways. First, it compares and evaluates methods suitable to interpolate wind directions. Second, the study examines the adoption of a robust and versatile alternative to incorporate wind directions in simulation models by accounting for the spatial variability of this climatic factor.

The study is structured in four parts. Section 2 defines the gaussian plume model and their application in simulating drift of pesticides. Moreover, the section examines relevant literature on spatial interpolation methods suitable for estimating wind directions at unsampled locations. Section 3 elaborates on the area under study, introduces the data and provides an outline of the methodology employed. The results are presented and discussed in Section 4. Lastly, Section 5 concludes the study by summarising the key findings and future work.

2. Literature Review

2.1 Pesticide Spray Drift

Pesticide spray drift refers to the airborne movement of spray droplets and particles towards sites other than the intended application areas [37]. Despite the fact that pesticide spraying systems are designed to provide optimal deposition of pesticides on the targeted crops to help protect produce from pests and diseases, a significant amount of these chemical components drifts to non-targeted areas during application [38, 39]. The resulting exposure poses a notable public health concern with negative health effects ranging from acute toxicity [6, 7] to neurodegenerative diseases [8–10]. Hence, several studies aimed to identify factors that play an important role in the drift of these biological agents, such as the spray system design [40–43], meteorological conditions at the site of interest [44–46] and crop characteristics [47, 48].

First, the size of droplets discharged from spray systems is recognised as an important element affecting drift [49]. Here, fine droplets tend to be airborne for prolonged periods of time, increasing their potential to be transported by wind [50, 51]. In contrast, more sizable droplets tend to retain their initial velocity, making these aerosols more likely to be deposited on the intended sites [51]. Despite this general agreement in literature, threshold values of droplet sizes highly susceptible to drift vary. Most studies found that aerosols with diameters less than 100 μm exhibit a tendency to drift [52, 53], while others proposed a broader range of 50 to 200 μm [42, 43]. All in all, the effect of droplet sizes in pesticide drift underscores the significance of spray system design, as the setup influences the granularity of aerosols released.

Second, meteorological conditions at the application site are considered to significantly impact pesticide spray drift. Here, wind speed and direction are important determinants along with the atmospheric stability [46]. Various studies found that greater wind velocities result in increased drift distances as they facilitate the movement of more chemical components towards non-targeted areas during application [45, 49, 54]. Furthermore, it was observed that wind fields influence drift by determining the trajectory of droplets, affecting the deposition of aerosols. In line with these findings, some studies demonstrated that spray droplets tend to exhibit a drift pattern along the mean wind direction under stable atmospheric conditions while dispersion becomes more extensive in unstable ones [44, 55]. The influence of atmospheric stability was found to be more substantial at greater downwind distances [44].

Lastly, spray retention is considered to be dependent on foliage density of crops, as several studies found that pesticide drift is inversely related to canopy density [47, 48]. However, another study also argued that dense foliage of grapefruit and orange trees actually increased spray drift due to the deflection over the top of this produce [56]. This underlines the complex and variable nature of this interaction, suggesting that the effect may not be uniformly across all types of crops.

2.2 Spatial Simulation Models to Estimate Pesticide Exposure

In recent years, spatial simulation models have emerged as an effective approach to estimate the extent and distribution of pesticide drift and the associated exposure [2, 11–15]. Apart from enhancing the understanding of drift phenomena, simulation models are also considered to complement more resource intensive field experiments prone to uncontrollable meteorological conditions [49]. A review of literature revealed that these simulation models broadly fall into either empirical or mechanistic categories.

Empirical simulation models are based on field measurements and statistical relationships without considering the inherent physical basis of spray drift [57]. In this context, regression models are widely adopted to estimate the association between input variables and the movement of pesticides [58–62]. For instance, based on a multiple regression procedure Smith et al. [62], concluded that downwind distance was the most influential variable to estimate drift. Despite the fact that this empirical approach provides valuable insights with substantial data available, it may encounter challenges in capturing the underlying trend [61].

Mechanistic models aim to address the fundamental physical processes associated with pesticide drift [2, 57]. Here, a distinction might be made between studies that employed drop trajectory [63–67] or gaussian models [2, 13, 14, 68–70]. The first category aims to estimate the position and movement of each individual droplet under external forces. In doing so, stochastic measures of pesticides at various locations may be obtained [2]. On the other hand, a gaussian approach describes pesticide levels as a function of time and space originating from a singular point [71]. In this context, plume concentrations are modelled using a gaussian distribution in both the horizontal and vertical direction, as illustrated in Figure 1.

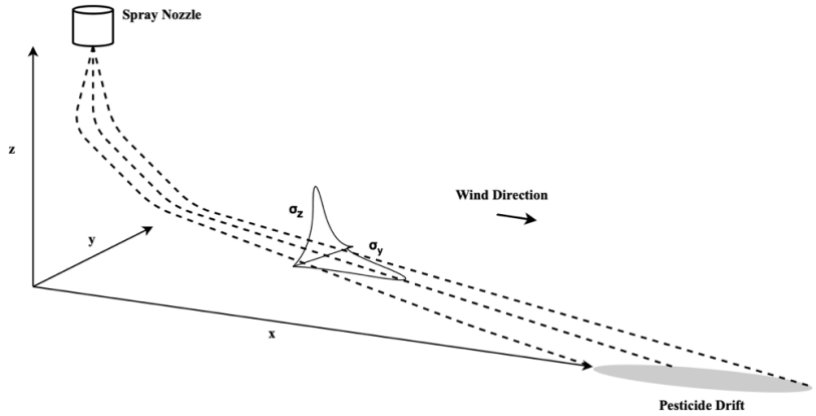


Figure 1 Representation of pesticide dispersion from an isolated spray nozzle based on a gaussian plume model.

Among the literature that explored pesticide drift and the associated exposure more extensive, most adopted a gaussian plume model to simulate the movement of pesticides released from aerial [68–70] or surface sprayers [2, 13, 14]. Although several studies have proven its limitations in exposure assessment due to its relative simplicity, the method is widely accepted and can be considered almost a standard approach. This broad adoption might be attributed to the ability of the model to consider a range of factors that have been found to influence drift of pesticides in ambient air, including meteorological parameters [2, 13, 14]. Here, the common approach is to derive these inputs from weather stations closest to the pesticide application area [13]. However, this may inaccurately represent conditions at the site of interest, resulting in an under- or overestimation of pesticide drift and the associated exposure.

In response, some studies incorporated climatic parameters into their models, obtained from sensors mounted directly on the pesticide sprayer in an effort to enhance estimates [2, 15]. Nevertheless, this approach remains deficient in capturing the spatial and temporal variability inherent in particular meteorological parameters, such as wind direction. The reason for this is that these sensors primarily focus on local wind directions measured at time and location of pesticide application. Hence, some studies proposed spatial interpolation of this climatic condition as a more versatile means of improving measures of exposure [16, 17]. Nonetheless, empirical findings based on this approach are limited, specifically in studies that employed gaussian plume models to simulate drift of pesticides.

2.3 Methods to Interpolate Wind Directions

Spatial interpolation techniques being applied to meteorological observations have received considerable attention in literature [18, 19]. In this context, spatial interpolation refers to the process of estimating variable values at unknown locations based on a collection of available samples [20]. Among the various approaches available, the prevailing ones mainly originate from either a deterministic or geostatistical framework. Deterministic techniques are solely based on the geometric properties of samples while geostatistical methods consider both geographical attributes as well as the spatial autocorrelation among target variables [72].

Common deterministic models to interpolate wind measurements include inverse distance weighting [21–23] and nearest neighbour approaches [21, 24]. For instance, Swain et al. [23], aimed to interpolate ocean surface wind fields using inverse distance weighting. Although these types of models are widely adopted due to their relative simplicity, they also induce considerable interpolation error. Hence, some studies proposed geostatistical methods based on kriging in an effort to enhance estimates [25–27]. However, adopting these models for wind field interpolation may be challenging due to the requirement of defining variable distributions a priori and the inherent non-stationarity of this climatic condition [29].

In recent years, the advancement in computational power and data availability has facilitated the emergence of studies combining statistical learning with spatial interpolation of wind observations. Here, artificial neural networks [30, 31] and random forests are widely adopted [18, 32]. The first category dates back to the 1950s and encompasses a collection of processing elements that receive inputs and deliver outputs based on predefined activation functions [73]. On the other hand, random forests were first proposed in 2001 and are based on the aggregation of individual decision trees that aim to produce a robust ensembled output [74]. In doing so, it was found that this approach provides spatial interpolation results that often outperform deterministic, geostatistical or other machine learning models [32]. Nevertheless, studies combining statistical learning with the interpolation of meteorological observations mostly focus on estimating wind speed or power rather than the direction of the field, highlighting an important area for further research.

Another strand of literature that placed more emphasis on interpolating wind directions is the field of directional statistics. In this context, studies employed either a wrapped or embedded approach to address the periodic nature of data represented in degrees or radians [33–35]. The first method transforms the density of linear variables into circular ones by reprojecting their values while the embedded approach aims to estimate the angle between two linear variables [75]. In doing so, studies employing circular statistics mitigate the limitation of interpolation techniques based on the arithmetic mean of neighbouring observations, as these often lead to unrealistic outcomes for circular data [36].

For instance, these classical methods may interpolate the wind direction as 170 degrees based on known values from stations measuring 320 and 20 degrees, respectively. Nonetheless, an appropriate interpolation would correspond to 350 degrees, considering the directional nature of the data. Notably, this limitation becomes particular evident for wind directions originating from the north. Highlighting the fact that depending on the distribution of wind fields, methods based on the arithmetic mean may still provide valid results.

3. Methods

3.1 Study Area

The geographical scope of this study was concentrated on the Netherlands. The country is situated in the North-Western part of Europe and is home to almost 17.9 million inhabitants [76], occupying a space of 41,541 km^2 [77]. This makes it one of the most densely populated areas in the European Union (429.5 inhabitants/km^2) [78]. Surrounding the large urban areas, the vast majority of land is used for agriculture [77]. Here, Dutch farmers have progressive pesticide application programs in place that aim to enhance crop yields [79]. In turn, exposure to these biological agents may give rise to significant public health concerns, especially in such a densely populated country. Hence, accurate exposure assessments are critical to identify areas of high risk and to inform decisions aimed at mitigating the adverse health effects associated with these chemical components.

3.2 Data Extraction

In this study, a total of 403,392 meteorological observations were obtained from the Royal Netherlands Meteorological Institute (KNMI), the Dutch national weather service [80]. In this context, data included hourly wind directions (in degrees) measured by 47 climatic stations spread across the Netherlands in 2017, as illustrated in Figure 2. Meteorologists from the institution validated and corrected each record [81], exempting this study from further data verification. Finally, administrative boundaries of the country were obtained from the Central Bureau of Statistics (CBS) [82]. This facilitated the creation of regular grids covering the area of interest, leading to a surface suitable for interpolating wind fields.

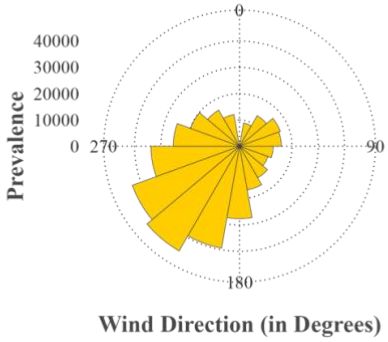


Figure 2 Distribution of hourly wind direction origins across the Netherlands in 2017.

3.3 Data Preparation and Enrichment

In an effort to supplement and ensure consistency of the extracted information, data preparation and enhancement procedures were performed, as illustrated in Figure 3. Here, the location of each meteorological station was obtained and transformed into a single coordinate reference system suitable for the Netherlands, i.e. EPSG 28992. Both stations and climatic records were combined into a single dataset to ease further analysis. Inspection of this information revealed some inconsistencies, including missing values and outliers. To address and resolve these types of records, exclusion criteria were in place.

First, observations with partial or complete absence of wind field measurement during 2017 were excluded from the analysis. This was done to ensure consistent data coverage across the study area for the interpolation models. Moreover, sensor failures were the primary source of absence, indicating that records were missing completely at random [83]. As a result, inferences obtained from the remaining data were still considered valid [84].

Second, exclusion criteria were in place for measurements representing calm (value of 0) or highly changeable (value of 880) wind conditions, as specified by KNMI [80]. The main rationale behind this exclusion was that these conditions were not considered significant or accurate contributors to simulate pesticide drift given the setup of the gaussian plume model, i.e. exposures are estimated based on the assumption that sources of drift are upwind [13], making it challenging to simulate dispersion in stagnant or unstable wind conditions.

In turn, this allowed the analysis to solely focus on interpolating wind directions that were considered to be influential and appropriate for simulating pesticide drift. In doing so, the exclusion procedures resulted in 382,521 records that were available for interpolation. Finally, administrative boundaries were used to obtain regular grids covering the area of interest, leading to a surface suitable to interpolate wind fields at a resolution of 1 km. The adoption of this spatial scale provided a balance between capturing relative fine-grain variations in wind fields while maintaining computational efficiency [85].

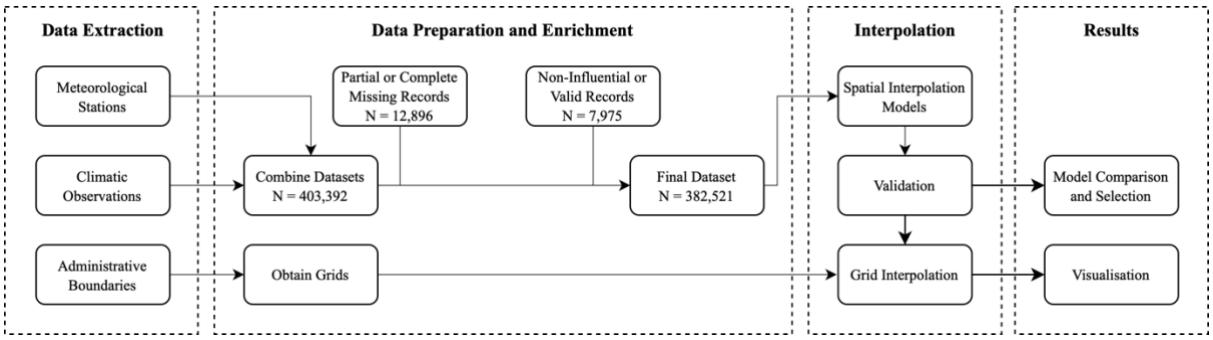


Figure 3 Schematic overview of the methodological procedures.

3.4 Spatial Interpolation

3.4.1 Spatial Interpolation Methods

Based on the examined literature, five spatial interpolation models to estimate wind fields at unknown locations were employed, namely naïve interpolation, nearest neighbours, inverse distance weighting, universal kriging and random forest. In this context, the primary objective was to obtain models capable to accurately predict wind directions at unobserved sites based on available samples, as presented in Figure 4. In doing so, the spatial interpolation problem considered by each of the models could be represented as a linear combination or average of neighbouring observations, as depicted in Equation 1 [36]. Here, $\hat{z}(x)$ represents the estimation of the target variable at location x, being influenced by coefficients $\lambda_i(x)$ that determine the weighting of the observed values $z(x_i)$.

$$\hat{z}(x) = \sum_{i=1}^{N} \lambda_i(x) z(x_i) \quad (1)$$

First, a naïve interpolation approach was employed to serve as a baseline for interpolating wind fields. This model provided an estimation of the target variable $\hat{z}(x)$ by assigning all the weight $\lambda_i(x)$ to the mode of the observed values $z(x_i)$. In doing so, the naïve method provided a point of reference in absolute sense, enabling the comparison and evaluation of alternative spatial interpolation approaches. In essence, any model demonstrating inferior performance compared to the naïve method could be deemed as an underperforming spatial interpolator for wind fields.

Figure 4 Overview of meteorological stations considered in the study.

Second, a nearest neighbour method was employed to represent the common approach of incorporating wind direction records in pesticide drift simulations. This type of model provided an estimation of the target variable $\hat{z}(x)$ by assigning all the weight $\lambda_i(x)$ to the observed value $z(x_i)$ that was nearest to the site of the unobserved point based on some distance function d [86]. Given the relative limited study area, the curvature of the earth was not considered to significantly affect this metric. Therefore, the Euclidean flat-surface distance was adopted to measure the proximity between locations with and without wind direction records.

Third, inverse distance weighting was adopted to further refine the spatial interpolation of wind directions. This type of model is based on the principle that the influence $\lambda_i(x)$ of available samples $z(x_i)$ on the unmeasured locations $\hat{z}(x)$ is inversely related to the distance between these points, as reflected in the distance decay function p [86]. To determine the optimal value of this power function, grid search hyperparameter tuning was performed. Moreover, the selection of records in this model are often limited to k-nearest neighbours or a subset of records within a specified radius of point x [25–27]. Nonetheless, these approaches may introduce bias into the interpolation due to local variations being inaccurately accounted for, i.e. some areas are more densely clustered than others, which may result in an under -or overestimation of local characteristics [87]. Hence, all observations N were used in this study.

Fourth, universal kriging was employed to enhance interpolations of wind fields by considering the spatial dependence between observations. The main rationale behind the adoption of this model was that it allowed to account for the non-stationarity present in the data, i.e. wind mostly originated from the South-West [80, 88]. In this context, estimates of the target variable $\hat{z}(x)$ were based on weights $\lambda_i(x)$ obtained by combining coordinate pairs of each observed point $z(x_i)$ with a model representing the spatial autocorrelation; the semi-variogram [86]. Most often the selection of the theoretical semi-variogram model is based on a visual inspection of the empirical one [88]. In order to automate the fitting procedure, the Gauss-Newton algorithm was adopted [88, 89]. This approach provides an initial estimate of the theoretical semi-variogram after which it iteratively converges towards an optimal model that aims to capture the spatial dependence present in the data [89].

Fifth, random forests were adopted to capitalise on the spatial autocorrelation present in wind field measurements using a different approach. In this study, the influence $\lambda_i(x)$ of available samples $z(x_i)$ on the unobserved locations $\hat{z}(x)$ was determined by constructing multiple decision trees that were based on a random subset of the data, allowing for the aggregation of individual tree predictions [74]. In standard applications of the random forest algorithm, spatial dependence is not implicitly considered [90]. Given the fact that this autocorrelation may carry information about wind directions at unobserved locations, additional covariates were included in the model. These encompassed coordinate pairs of each site, records from the n nearest observed locations and the distances from these observed points to the unobserved sites k.

The primary parameters that influenced wind field interpolations included the number of trees B, the number of features considered at each node split m, the minimum node size N_{min} and the number of nearest observed locations n [18, 32, 90]. To determine the optimal values of these parameters, grid search hyperparameter tuning was performed for the number of trees B and the number of nearest observed locations n due to their considerable influence on model performance. As proposed by Hastie et al. [91], optimal parameters for m and N_{min} adopted corresponded to the square root of the number of variables p and 1, respectively.

Lastly, the main justifications for the adoption of the five distinct models stemmed from their wide usage across topical literature and their suitability for interpolating wind fields given the characteristics of the data, i.e. wind directions mostly originated from the South-West. Since the adopted models, that are mostly based on the arithmetic mean, exhibit a tendency to provide unrealistic estimates mainly for wind fields originating from the north, the adoption of these methods could still provide valid results for the majority of interpolations.

3.4.2 Validation of Spatial Interpolation Methods

To compare and evaluate the wind direction interpolation methods proposed, the root-mean-squared error (RMSE) was initially used. The main rationale behind this adoption originated from its prevalent usage across topical literature and relative ease of interpretation [29]. This evaluation metric is based on the difference between measured and predicted values, as depicted in Equation 2. Here, p_i denotes the prediction of some observed value o_i for N measurements.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2} \tag{2}$$

Although this metric provided an estimate of model performance, direct subtraction may lead to inaccurate evaluations due to the directional nature of the data [36]. To address this issue, an adjusted circular root-mean-squared error (CRMSE) was employed, as presented in Equation 3. In this context, C_i denotes the circular difference between the prediction p_i of some observed value o_i, as stated in Equation 4.

$$\mathrm{CRMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(C_i)^2} \tag{3}$$

$$C_i = |p_i - o_i|, \quad \text{If } |p_i - o_i| \leq 180°; \quad 360° - |p_i - o_i|, \quad \text{Otherwise} \tag{4}$$

Furthermore, spatial k-fold cross validation was used to obtain the out-of-sample circular root-mean-squared error of each spatial interpolation model. This was done to limit variabilities in the evaluation metric, providing a more stable and robust assessment of each interpolator. In doing so, climatic stations with their associated wind field measurements were partitioned into ten disjoint folds by applying K-means clustering on the spatial coordinates, as proposed by Pohjankukka et al. [92]. Each fold was sequentially left out to examine the model performance while records from the remaining stations $\mathrm{N} - k$ were used for model training. Hence, this allowed to account for spatial dependencies present in the data, reducing the possibility of data leakage between folds used for evaluation and training [92]. In turn, the approach provided a set of hyperparameters, obtained from a single randomly selected hour, that were used for the comparison and evaluation of models, as outlined in Table 1.

Table 1 Overview of adopted spatial interpolation models, associated hyperparameters and cross validation parameters.

Spatial Interpolation Model	Model Hyperparameters	Cross Validation Parameters
Naïve Interpolation	-	$K = 10$, K-means
Nearest Neighbour	$N_{max} = 1$	$\mathrm{K} = 10$, K-means
Inverse Distance Weighting	$\mathrm{p} = 3.441$, N = N	$\mathrm{K} = 10$, K-means
Universal Kriging	-	$\mathrm{K} = 10$, K-means
Random Forest	$\mathrm{B} = 150$, $\mathrm{m} = \sqrt{p}$, $N_{min} = 1$, $\mathrm{n} = 3$	$\mathrm{K} = 10$, K-means

3.4.3 Sensitivity Analysis

A sensitivity analysis was conducted to facilitate a reliable comparison and evaluation of the five spatial interpolation models adopted. This analysis encompassed wind direction records that were obtained from a collection of hours across 2017. Here, a simple random selection was performed, aiming to represent the characteristics of the entire dataset. In turn, this allowed to examine the influence of varying observations on the performance of each model. Estimation of the appropriate number of random selections was done using Slovin Formula due to its widespread adoption across literature [93]. In this context, n denotes the sample as a function of the finite population N and the margin of error e, as stated in Equation 5.

$$\mathrm{n} = \frac{N}{1 + Ne^2} \tag{5}$$

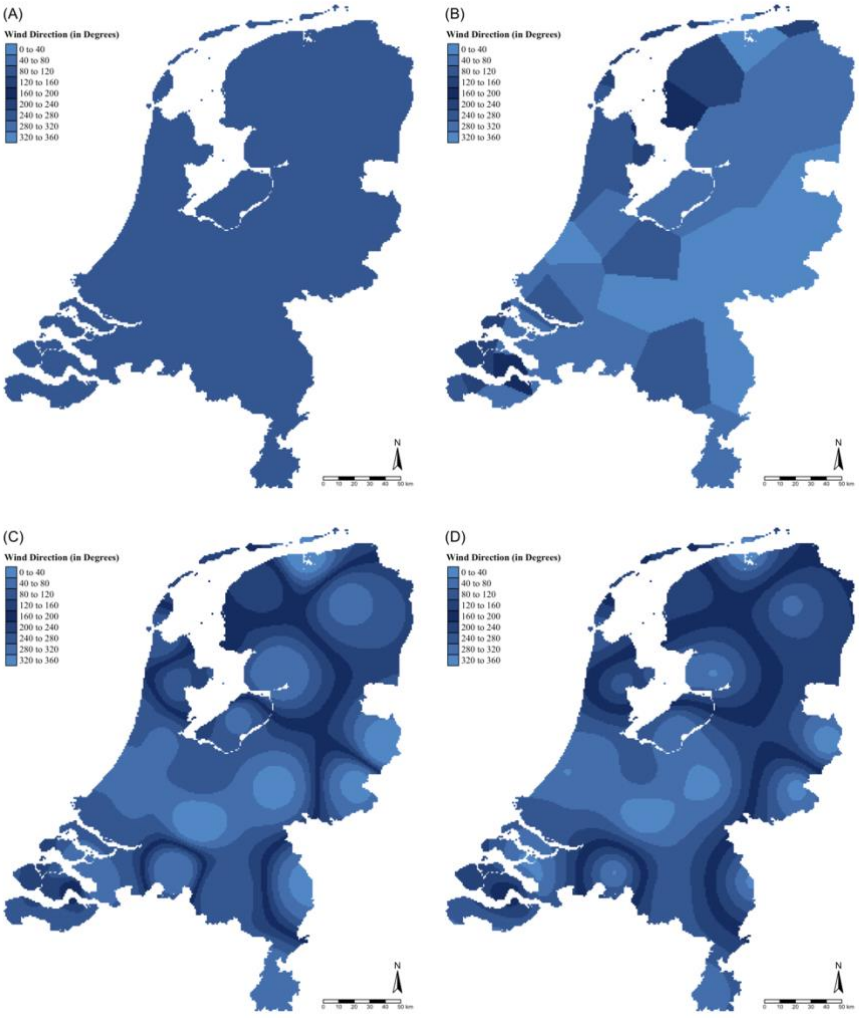
To enhance the reliability and validity of the findings while ensuring computational feasibility, the value of e corresponded to 0.05. This resulted in a total of 383 random selections n that were obtained from a finite population N of 8760, i.e. 365 days times 24 hours. Subsequently, spatial k-fold cross validation was employed to evaluate the performance of the interpolators on each of the 383 random selections.

Lastly, the statistical programming language R (version 4.3.1) was used throughout this study. Specifically, data extraction and enrichment procedures were performed by employing the httr [94], jsonlite [95], tidyverse [96] and sf packages [97, 98]. Interpolations of wind fields were estimated using the gstat [99, 100] and meteo [101, 102] packages while ggplot2 [103] and tmap [104] were used for visualisation purposes. Finally, an overview of scripts and data used in this study may be found on: github.com/thomasnib/Improving-Pesticide-Exposure.

4. Results and Discussion

The results showed that the employed spatial interpolation methods applied to wind direction records produced distinct patterns, as illustrated in Figure 5. In this context, a visual comparison of interpolations on a randomly selected hour revealed that the inverse distance weighting and universal kriging approaches resulted in a continuous surface while the nearest neighbour approach provided abrupt transitions across space. Interpolations obtained from the random forest model exhibited an intermediate behaviour, capturing both the more gradual and abrupt variations that were present in the wind field measurements. Conversely, the naïve approach provided a singular estimate across the study area, disregarding the spatial variations present.

Interpolation patterns produced by the employed models align with findings from previous studies. First, results of the naïve method underline the simplicity of this approach that involves assigning the mode of the observed locations to the unobserved ones. Second, interpolations of the nearest neighbour model highlight the polygon interpretation of this method, as described by Isaaks and Srivastava [28]. This explains the abrupt transitions of wind field interpolations across the study area. Second, estimates derived from the inverse distance weighting exhibited the isolated and concentric circular effects often associated with this type of model while universal kriging provided a slightly more gradual surface [105]. Finally, interpolations obtained by random forest revealed linear boundaries along the coordinate axes. Others found similar patterns, suggesting that the inclusion of covariates representing northing and easting coordinates may result in this phenomenon [18, 105].



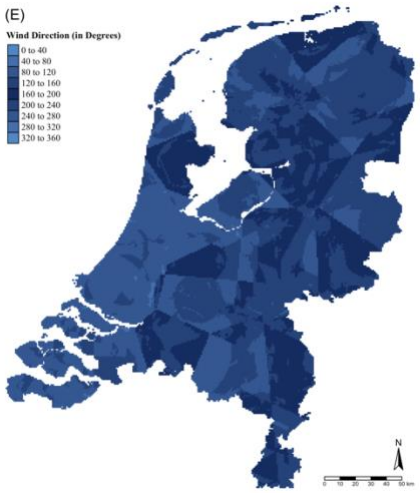


Figure 5 Spatial interpolations of wind directions based on the naïve interpolation (A), nearest neighbour (B), inverse distance weighting (C), universal kriging (D) and random forest (E) approaches at 12:00:00 on May 4th, 2017.

Furthermore, the sensitivity analysis revealed that the employed spatial interpolation methods performed relatively similar, as presented in Figure 6. In this regard, all models outperformed the baseline wind field interpolations of the naïve approach. Notably, universal kriging and inverse distance weighting exhibited superior performance compared to the nearest neighbour approach that is commonly used in pesticide dispersion simulations. In contrast, the random forest algorithm was found to perform slightly worse compared to the other interpolators, only outperforming the naïve method. Nonetheless, all models showed considerable variation in the out-of-sample error across the 383 randomly selected hours, underlining the challenges and complexities associated with interpolating wind direction records.

The inverse distance weighting and universal kriging methods demonstrated the lowest out-of-sample circular root-mean-squared errors across the proposed models with median error values corresponding to 15.214 (MAD = 9.396) and 16.225 (MAD = 11.097) degrees, respectively. Moreover, interpolations of the nearest neighbour model provided a median performance error of 18.432 (MAD = 10.925) degrees, making it a slightly more accurate method compared to the random forest algorithm that revealed a median out-of-sample error value of 18.458 (MAD = 11.427) degrees. Furthermore, all models outperformed the naïve approach that exhibited a median error value of 21.376 (MAD = 13.317) degrees. Overall, these results suggest a relative similar performance across the models. Nonetheless, substantial variation in the error metric was observed, as indicated by the wide spread of the boxplots in Figure 6.

A possible explanation for the superior performance of the inverse distance weighting and universal kriging methods may be related to the positive spatial autocorrelation present in wind fields across the study area. In this context, these distance-based algorithms have a competitive advantage, as was demonstrated by Ford and Quiring [106]. Interestingly, the inclusion of spatial covariates in the random forest algorithm, aiming to address the spatial dependence, revealed limited performance improvements compared to the nearest neighbour model that is typically used to include wind directions in pesticide drift simulations. This may be related to the restricted number of observations available at each randomly selected hour, as it was found that the algorithm generally achieves better performance with more substantial datasets [107]. However, the algorithm still outperformed the baseline interpolations of the naïve approach, underscoring its potential as a promising alternative for estimating wind fields at unobserved sites. Lastly, the convergence in performance across the models might be explained by the level of variability in wind directions, suggesting that the spatial structure may have played an important role [101, 108], i.e. less distinct wind patterns produce more similar model results.

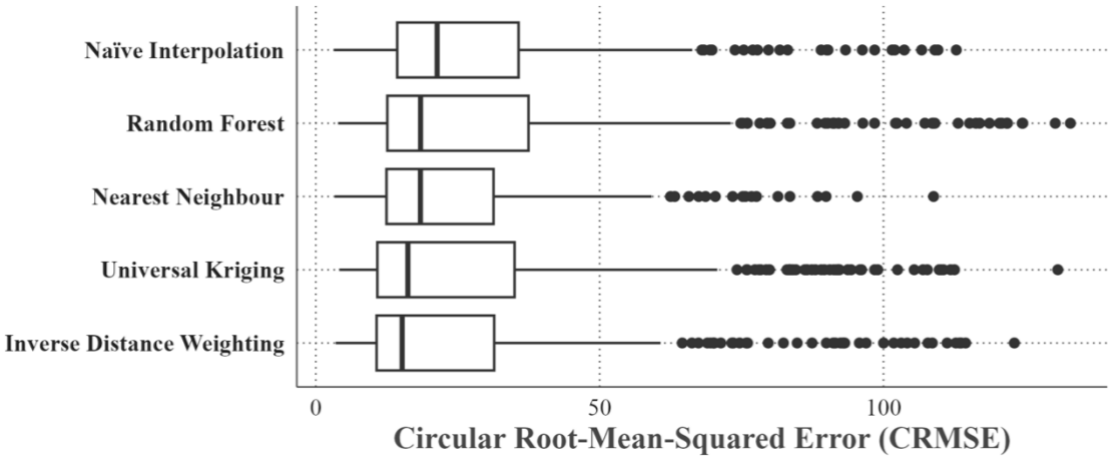


Figure 6 Mean out-of-sample circular root-mean-squared error of the proposed spatial interpolation models across 383 randomly selected hours.

Finally, limitations of this study were three-fold. First, external validation of each model was not performed, making it challenging to generalise the comparison and evaluation. Second, the temporal dimension of wind field observations was not considered, despite its demonstrated influence on interpolation accuracy [18, 100, 106]. Lastly, circular statistical models, aiming to address the directional nature of wind fields were not included in the evaluation due to their neglectable contribution to improved interpolation accuracy given the characteristics of the data. Nonetheless, it may be important to acknowledge that the inclusion of these types of models could have potentially revealed different results.

5. Conclusion

This study aimed to examine spatial interpolation models that may improve pesticide exposure estimates using wind direction records from the Netherlands in 2017. In doing so, five spatial interpolation approaches were adopted, namely naïve interpolation, nearest neighbour, inverse distance weighting, universal kriging and random forest. In this context, data included hourly wind field measures obtained from the Dutch national weather service. Furthermore, spatial k-fold cross validation was used to examine the circular root-mean-squared error of each model across a collection of randomly selected hours.

The results revealed that the adopted spatial interpolation approaches applied to wind direction records produced distinct visual patterns. Nonetheless, the performance among the five models was relatively similar. The main theoretical explanation for this finding was the limited variety observed in hourly wind field measures. Moreover, inverse distance weighting demonstrated the lowest median out-of-sample error for interpolating wind directions among the five distinct models. This may suggest that the adoption of this method in pesticide drift simulations might provide a more valid representation of wind directions at the application areas compared to the nearest neighbour approach that is often employed. In turn, this may improve the accuracy of pesticide exposure estimates obtained from these simulations.

Future research may focus on enhancing the generalisability of the comparison and evaluation by performing an external validation on each of the proposed models. Moreover, future studies might consider examining the effect of incorporating inverse distance weighting wind direction interpolations in pesticide dispersion simulations. In doing so, this might provide empirical evidence regarding the efficacy of combining this interpolation method with a drift simulation model in an effort to improve pesticide exposure estimates.

References

1. Tudi M, Daniel Ruan H, Wang L, et al (2021) Agriculture Development, Pesticide Application and Its Impact on the Environment. Int J Environ Res Public Health 18:1112. https://doi.org/10.3390/ijerph18031112

2. Lebeau F, Verstraete A, Stainier C, Destain M-F (2011) RTDrift: A real time model for estimating spray drift from ground applications. Comput Electron Agric 77:161–174. https://doi.org/10.1016/j.compag.2011.04.009

3. Mahmood I, Imadi SR, Shazadi K, et al (2016) Effects of Pesticides on Environment. In: Plant, Soil and Microbes. Springer International Publishing, Cham, pp 253–269

4. Cooper J, Dobson H (2007) The benefits of pesticides to mankind and the environment. Crop Protection 26:1337–1348. https://doi.org/10.1016/j.cropro.2007.03.022

5. Aktar W, Sengupta D, Chowdhury A (2009) Impact of pesticides use in agriculture: their benefits and hazards. Interdiscip Toxicol 2:1–12. https://doi.org/10.2478/v10102-009-0001-7

6. Zare S, Behzadi M, Tarzanan M, et al (2015) The impacts of pesticides on the health of farmers in Fasa, Iran. Electron Physician 7:1168–1173

7. Damalas C, Koutroubas S (2016) Farmers' Exposure to Pesticides: Toxicity Types and Ways of Prevention. Toxics 4:1. https://doi.org/10.3390/toxics4010001

8. Parrón T, Requena M, Hernández AF, Alarcón R (2011) Association between environmental exposure to pesticides and neurodegenerative diseases. Toxicol Appl Pharmacol 256:379–385. https://doi.org/10.1016/j.taap.2011.05.006

9. Baldi I (2003) Neurodegenerative Diseases and Exposure to Pesticides in the Elderly. Am J Epidemiol 157:409–414. https://doi.org/10.1093/aje/kwf216

10. Jokanović M (2018) Neurotoxic effects of organophosphorus pesticides and possible association with neurodegenerative diseases in man: A review. Toxicology 410:125–131. https://doi.org/10.1016/j.tox.2018.09.009

11. Costanzini S, Teggi S, Bigi A, et al (2018) Atmospheric Dispersion Modelling and Spatial Analysis to Evaluate Population Exposure to Pesticides from Farming Processes. Atmosphere (Basel) 9:38. https://doi.org/10.3390/atmos9020038

12. Butler Ellis MC, van den Berg F, van de Zande JC, et al (2017) The BROWSE model for predicting exposures of residents and bystanders to agricultural use of pesticides: Comparison with experimental data and other exposure models. Biosyst Eng 154:122–136. https://doi.org/10.1016/j.biosystemseng.2016.09.002

13. Figueiredo DM, Vermeulen RCH, Jacobs C, et al (2022) OBOMod - Integrated modelling framework for residents' exposure to pesticides. Science of The Total Environment 825:153798. https://doi.org/10.1016/j.scitotenv.2022.153798

14. Butler Ellis MC, Miller PCH (2010) The Silsoe Spray Drift Model: A model of spray drift for the assessment of non-target exposures to pesticides. Biosyst Eng 107:169–177. https://doi.org/10.1016/j.biosystemseng.2010.09.003

15. Fujimoto A, Satow T, Kishimoto T (2016) Simulation of spray distribution with boom sprayer considering effect of wind for agricultural cloud computing analysis. Engineering in Agriculture, Environment and Food 9:305–310. https://doi.org/10.1016/j.eaef.2016.04.001

16. Hong S-W, Zhao L, Zhu H (2018) SAAS, a computer program for estimating pesticide spray efficiency and drift of air-assisted pesticide applications. Comput Electron Agric 155:58–68. https://doi.org/10.1016/j.compag.2018.09.031

17. Kasner EJ, Prado JB, Yost MG, Fenske RA (2021) Examining the role of wind in human illness due to pesticide drift in Washington state, 2000–2015. Environmental Health 20:26. https://doi.org/10.1186/s12940-021-00693-3

18. Hengl T, Nussbaum M, Wright MN, et al (2018) Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6:5518. https://doi.org/10.7717/peerj.5518

19. Appelhans T, Mwangomo E, Hardy DR, et al (2015) Evaluating machine learning approaches for the interpolation of monthly air temperature at Mt. Kilimanjaro, Tanzania. Spat Stat 14:91–113. https://doi.org/10.1016/j.spasta.2015.05.008

20. Stein ML (1999) Interpolation of Spatial Data. Springer New York, New York, NY

21. Luo W, Taylor MC, Parker SR (2008) A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales. International Journal of Climatology 28:947–959. https://doi.org/10.1002/joc.1583

22. Loghmari I, Timoumi Y, Messadi A (2018) Performance comparison of two global solar radiation models for spatial interpolation purposes. Renewable and Sustainable Energy Reviews 82:837–844. https://doi.org/10.1016/j.rser.2017.09.092

23. Swain J, Umesh PA, Murty ASN (2017) Demonstration of an efficient interpolation technique of inverse time and distance for Oceansat-2 wind measurements at 6-hourly intervals. The International Journal of Ocean and Climate Systems 8:101–112. https://doi.org/10.1177/1759313117736596

24. Keskin M, Dogru AO, Balcik FB, et al (2015) Comparing Spatial Interpolation Methods for Mapping Meteorological Data in Turkey. pp 33–42

25. Guo B, Yang F, Wu H, et al (2021) How the variations of terrain factors affect the optimal interpolation methods for multiple types of climatic elements? Earth Sci Inform 14:1021–1032. https://doi.org/10.1007/s12145-021-00609-2

26. Friedland CJ, Joyner TA, Massarra C, et al (2017) Isotropic and anisotropic kriging approaches for interpolating surface-level wind speeds across large, geographically diverse regions. Geomatics, Natural Hazards and Risk 8:207–224. https://doi.org/10.1080/19475705.2016.1185749

27. Lin Q, Li C (2020) Kriging based sequence interpolation and probability distribution correction for gaussian wind field data reconstruction. Journal of Wind Engineering and Industrial Aerodynamics 205:104340. https://doi.org/10.1016/j.jweia.2020.104340

28. Isaaks EH, Srivastava RM (1989) An introduction to Applied Geostatistics. Oxford University Press, Oxford

29. Li J, Heap D (2008) A Review of Spatial Interpolation Methods for Environmental Scientist. Geoscience Australia, Canberra

30. Yun E, Hur J (2021) Probabilistic estimation model of power curve to enhance power output forecasting of wind generating resources. Energy 223:120000. https://doi.org/10.1016/j.energy.2021.120000

31. Philippopoulos K, Deligiorgi D (2012) Application of artificial neural networks for the spatial estimation of wind speed in a coastal region with complex topography. Renew Energy 38:75–82. https://doi.org/10.1016/j.renene.2011.07.007

32. da Silva Júnior JC, Medeiros V, Garrozi C, et al (2019) Random forest techniques for spatial interpolation of evapotranspiration data from Brazilian's Northeast. Comput Electron Agric 166:105017. https://doi.org/10.1016/j.compag.2019.105017

33. Marques I, Kneib T, Klein N (2022) A non-stationary model for spatially dependent circular response data based on wrapped Gaussian processes. Stat Comput 32:73. https://doi.org/10.1007/s11222-022-10136-9

34. Shuku T, Ropponen J, Juntunen J, Suito H (2022) Data-driven model of the local wind field over two small lakes in Jyväskylä, Finland. Meteorology and Atmospheric Physics 134:18. https://doi.org/10.1007/s00703-021-00857-3

35. Cremers J, Klugkist I (2018) One Direction? A Tutorial for Circular Data Analysis Using R With Examples in Cognitive Psychology. Front Psychol 9:. https://doi.org/10.3389/fpsyg.2018.02040

36. Grancher D, Bar-Hen A, Paris R, et al (2011) Spatial interpolation of circular data: application to tsunami of December 2004. In: Tsunarisque. Éditions de la Sorbonne, pp 49–58

37. Zhang B, Tang Q, Chen L, et al (2018) Numerical simulation of spray drift and deposition from a crop spraying aircraft using a CFD approach. Biosyst Eng 166:184–199. https://doi.org/10.1016/j.biosystemseng.2017.11.017

38. Delele MA, Jaeken P, Debaer C, et al (2007) CFD prototyping of an air-assisted orchard sprayer aimed at drift reduction. Comput Electron Agric 55:16–27. https://doi.org/10.1016/j.compag.2006.11.002

39. M. Farooq, M. Salyani (2004) Modeling of Spray Penetration and Deposition on Citrus Tree Canopies. Transactions of the ASAE 47:619–627. https://doi.org/10.13031/2013.16091

40. Balsari P, Gil E, Marucco P, et al (2017) Field-crop-sprayer potential drift measured using test bench: Effects of boom height and nozzle type. Biosyst Eng 154:3–13. https://doi.org/10.1016/j.biosystemseng.2016.10.015

41. Grella M, Gallart M, Marucco P, et al (2017) Ground Deposition and Airborne Spray Drift Assessment in Vineyard and Orchard: The Influence of Environmental Variables and Sprayer Settings. Sustainability 9:728. https://doi.org/10.3390/su9050728

42. Al Heidary M, Douzals JP, Sinfort C, Vallet A (2014) Influence of spray characteristics on potential spray drift of field crop sprayers: A literature review. Crop Protection 63:120–130. https://doi.org/10.1016/j.cropro.2014.05.006

43. Hilz E, Vermeer AWP (2013) Spray drift review: The extent to which a formulation can contribute to spray drift reduction. Crop Protection 44:75–83. https://doi.org/10.1016/j.cropro.2012.10.020

44. Bonds JAS, Leggett M (2015) A Literature Review of Downwind Drift from Airblast Sprayers: Development of Standard Methodologies and a Drift Database. Trans ASABE 58:1471–1477. https://doi.org/10.13031/trans.58.11057

45. Miller PCH, Ellis MB, Lane A, et al (2011) Methods for minimising drift and off-target exposure from boom sprayer applications. Asp Appl Biol 281–288

46. Baetens K, Nuyttens D, Verboven P, et al (2007) Predicting drift from field spraying by means of a 3D computational fluid dynamics model. Comput Electron Agric 56:161–173. https://doi.org/10.1016/j.compag.2007.01.009

47. Chen Y, Zhu H, Ozkan HE, et al (2013) Spray Drift and Off-Target Loss Reductions with a Precision Air-Assisted Sprayer. Trans ASABE 1273–1281. https://doi.org/10.13031/trans.56.10173

48. Chen Y, Ozkan HE, Zhu H, et al (2013) Spray Deposition inside Tree Canopies from a Newly Developed Variable-Rate Air-Assisted Sprayer. Trans ASABE 1263–1272. https://doi.org/10.13031/trans.56.9839

49. Hong S, Park J, Jeong H, et al (2021) Fluid Dynamic Approaches for Prediction of Spray Drift from Ground Pesticide Applications: A Review. Agronomy 11:1182. https://doi.org/10.3390/agronomy11061182

50. Dexter AG (1980) Herbicide Spray Drift. Agricultural Extension Service

51. Spillman JJ (1984) Spray impaction, retention and adhesion: An introduction to basic characteristics. Pestic Sci 15:97–106. https://doi.org/10.1002/ps.2780150202

52. Ferguson JC, O'Donnell CC, Chauhan BS, et al (2015) Determining the uniformity and consistency of droplet size across spray drift reducing nozzles in a wind tunnel. Crop Protection 76:1–6. https://doi.org/10.1016/j.cropro.2015.06.008

53. Xue S, Xi X, Lan Z, et al (2021) Longitudinal drift behaviors and spatial transport efficiency for spraying pesticide droplets. Int J Heat Mass Transf 177:121516. https://doi.org/10.1016/j.ijheatmasstransfer.2021.121516

54. Arvidsson T, Bergström L, Kreuger J (2011) Spray drift as influenced by meteorological and technical factors. Pest Manag Sci 67:586–598. https://doi.org/10.1002/ps.2114

55. Gil Y, Sinfort C, Brunet Y, et al (2007) Atmospheric loss of pesticides above an artificial vineyard during air-assisted spraying. Atmos Environ 41:2945–2957. https://doi.org/10.1016/j.atmosenv.2006.12.019

56. Stewart Agricultural Research Services (1997) A Summary of Airblast Application Studies. Macon

57. Baker RE, Peña J-M, Jayamohan J, Jérusalem A (2018) Mechanistic models versus machine learning, a fight worth fighting for the biological community? Biol Lett 14:20170660. https://doi.org/10.1098/rsbl.2017.0660

58. Bueno MR, Cunha JPAR da, de Santana DG (2017) Assessment of spray drift from pesticide applications in soybean crops. Biosyst Eng 154:35–45. https://doi.org/10.1016/j.biosystemseng.2016.10.017

59. Alves GS, Cunha JPAR da (2014) Field data and prediction models of pesticide spray drift on coffee crop. Pesqui Agropecu Bras 49:622–629. https://doi.org/10.1590/S0100-204X2014000800006

60. Holterman HJ, van de Zande JC, Huijsmans JFM, Wenneker M (2017) An empirical model based on phenological growth stage for predicting pesticide spray drift in pome fruit orchards. Biosyst Eng 154:46–61. https://doi.org/10.1016/j.biosystemseng.2016.08.016

61. Moges G, McDonnell K, Delele MA, et al (2022) Development and comparative analysis of ANN and SVR-based models with conventional regression models for predicting spray drift. Environmental Science and Pollution Research 30:21927–21944. https://doi.org/10.1007/s11356-022-23571-y

62. Smith DB, Bode LE, Gerard PD (2000) Predicting Ground Boom Spray Drift. Transactions of the ASAE 43:547–553. https://doi.org/10.13031/2013.2734

63. Mokeba ML, Salt DW, Lee BE, Ford MG (1997) Simulating the dynamics of spray droplets in the atmosphere using ballistic and random-walk models combined. Journal of Wind Engineering and Industrial Aerodynamics 67–68:923–933. https://doi.org/10.1016/S0167-6105(97)00129-3

64. Holterman HJ, van de Zande JC, Porskamp HAJ, Huijsmans JFM (1997) Modelling spray drift from boom sprayers. Comput Electron Agric 19:1–22. https://doi.org/10.1016/S0168-1699(97)00018-5

65. Zhu H, Reichard DL, Fox RD, et al (1994) Simulation of Drift of Discrete Sizes of Water Droplets from Field Sprayers. Transactions of the ASAE 37:1401–1407. https://doi.org/10.13031/2013.28220

66. Miller PCH, Hadfield DJ (1989) A simulation model of the spray drift from hydraulic nozzles. Journal of Agricultural Engineering Research 42:135–147. https://doi.org/10.1016/0021-8634(89)90046-2

67. Thompson N, Ley AJ (1983) Estimating spray drift using a random-walk model of evaporating drops. Journal of Agricultural Engineering Research 28:419–435. https://doi.org/10.1016/0021-8634(83)90134-8

68. Craig IP (2004) The GDS model—a rapid computational technique for the calculation of aircraft spray drift buffer distances. Comput Electron Agric 43:235–250. https://doi.org/10.1016/j.compag.2004.02.001

69. Woods N, Craig IP, Dorr G, Young B (2001) Spray Drift of Pesticides Arising from Aerial Application in Cotton. J Environ Qual 30:697–701. https://doi.org/10.2134/jeq2001.303697x

70. Tsai M, Elgethun K, Ramaprasad J, et al (2005) The Washington aerial spray drift study: Modeling pesticide spray drift deposition from an aerial application. Atmos Environ 39:6194–6203. https://doi.org/10.1016/j.atmosenv.2005.07.011

71. Snoun H, Krichen M, Chérif H (2023) A comprehensive review of Gaussian atmospheric dispersion models: current usage and future perspectives. EuroMediterr J Environ Integr 8:219–242. https://doi.org/10.1007/s41207-023-00354-6

72. Verma PA, Shankar H, Saran S (2019) Comparison of Geostatistical and Deterministic Interpolation to Derive Climatic Surfaces for Mountain Ecosystem. In: Remote Sensing of Northwest Himalayan Ecosystems. Springer Singapore, Singapore, pp 537–547

73. Eshragh F, Pooyandeh M, Marceau DJ (2015) Automated negotiation in environmental resource management: Review and assessment. J Environ Manage 162:148–157. https://doi.org/10.1016/j.jenvman.2015.07.051

74. Breiman L (2001) Random Forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324

75. Jona Lasinio G, Santoro M, Mastrantonio G (2020) CircSpaceTime: an R package for spatial and spatio-temporal modelling of circular data. J Stat Comput Simul 90:1315–1345. https://doi.org/10.1080/00949655.2020.1725008

76. Centraal Bureau voor de Statistiek (2023) Bevolking op eerste van de maand; geslacht, leeftijd, migratieachtergrond. In: CBS Statline. https://opendata.cbs.nl/#/CBS/nl/dataset/83482NED/table?ts=1685974361696. Accessed 5 Jun 2023

77. Centraal Bureau voor de Statistiek (2023) Bodemgebruik; uitgebreide gebruiksvorm, per gemeente. In: CBS Statline. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70262ned/table?fromstatweb. Accessed 5 Jun 2023

78. European Commission (2022) Population Density. In: EC Eurostat. https://ec.europa.eu/eurostat/databrowser/view/tps00003/default/table. Accessed 5 Jun 2023

79. Bakker L, Sok J, van der Werf W, Bianchi FJJA (2021) Kicking the Habit: What Makes and Breaks Farmers' Intentions to Reduce Pesticide Use? Ecological Economics 180:106868. https://doi.org/10.1016/j.ecolecon.2020.106868

80. Koninklijk Nederlands Meteorologisch Instituut (2017) Uurwaarnemingen. https://daggegevens.knmi.nl/klimatologie/uurgegevens. Accessed 6 Jun 2023

81. Koninklijk Nederlands Meteorologisch Instituut (2017) Uurgegevens van het weer in Nederland. https://www.knmi.nl/nederland-nu/klimatologie/uurgegevens. Accessed 6 Jun 2023

82. Centraal Bureau voor de Statistiek (2017) CBS Gebiedsindelingen. https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/cbs-gebiedsindelingen. Accessed 6 Jun 2023

83. van Buuren S (2018) Flexible Imputation of Missing Data, Second Edition. Chapman and Hall/CRC, Second edition. | Boca Raton, Florida : CRC Press, [2019] |

84. Rubin DB (1976) Inference and Missing Data. Biometrika 63:581. https://doi.org/10.2307/2335739

85. Hengl T (2006) Finding the right pixel size. Comput Geosci 32:1283–1298. https://doi.org/10.1016/j.cageo.2005.11.008

86. Hsieh WW (2023) Introduction to Environmental Data Science. Cambridge University Press

87. Xu C, Wang J, Li Q (2018) A New Method for Temperature Spatial Interpolation Based on Sparse Historical Stations. J Clim 31:1757–1770. https://doi.org/10.1175/JCLI-D-17-0150.1

88. Hiemstra PH, Pebesma EJ, Twenhöfel CJW, Heuvelink GBM (2009) Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network. Comput Geosci 35:1711–1721. https://doi.org/10.1016/j.cageo.2008.10.011

89. Cressie N (1993) Statistics for spatial data. John Wiley & Sons

90. Sekulić A, Kilibarda M, Heuvelink GBM, et al (2020) Random Forest Spatial Interpolation. Remote Sens (Basel) 12:1687. https://doi.org/10.3390/rs12101687

91. Hastie T, Tibshirani R, Friedman JH, (2009) The Elements of Statistical learning: Data Mining, Inference and Prediction. Springer, New York

92. Pohjankukka J, Pahikkala T, Nevalainen P, Heikkonen J (2017) Estimating the prediction performance of spatial models via spatial k-fold cross validation. International Journal of Geographical Information Science 31:2001–2019. https://doi.org/10.1080/13658816.2017.1346255

93. Slovin E (1960) Slovin's Formula for Sampling Technique

94. Wickham H (2023) httr: Tools for Working with URLs and HTTP

95. Ooms J 1 (2014) The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects

96. Wickham H, Averick M, Bryan J, et al (2019) Welcome to the Tidyverse. J Open Source Softw 4:1686. https://doi.org/10.21105/joss.01686

97. Pebesma E, Bivand R (2023) Spatial Data Science. Chapman and Hall/CRC, Boca Raton

98. Pebesma E (2018) Simple Features for R: Standardized Support for Spatial Vector Data. R J 10:439. https://doi.org/10.32614/RJ-2018-009

99. Pebesma EJ (2004) Multivariable geostatistics in S: the gstat package. Comput Geosci 30:683–691. https://doi.org/10.1016/j.cageo.2004.03.012

100. Gräler B, Pebesma E, Heuvelink G (2016) Spatio-Temporal Interpolation using gstat. R J 8:204. https://doi.org/10.32614/RJ-2016-014

101. Sekulić A, Kilibarda M, Heuvelink GBM, et al (2020) Random Forest Spatial Interpolation. Remote Sens (Basel) 12:1687. https://doi.org/10.3390/rs12101687

102. Kilibarda M, Hengl T, Heuvelink GBM, et al (2014) Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. Journal of Geophysical Research: Atmospheres 119:2294–2313. https://doi.org/10.1002/2013JD020803

103. Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York

104. Tennekes M (2018) tmap: Thematic Maps in R. J Stat Softw 84:1–39. https://doi.org/10.18637/jss.v084.i06

105. Li J, Heap AD, Potter A, Daniell JJ (2011) Application of machine learning methods to spatial interpolation of environmental variables. Environmental Modelling & Software 26:1647–1659. https://doi.org/10.1016/j.envsoft.2011.07.004

106. Ford TW, Quiring SM (2014) Comparison and application of multiple methods for temporal interpolation of daily soil moisture. International Journal of Climatology 34:2604–2621. https://doi.org/10.1002/joc.3862

107. Mariano C, Mónica B (2021) A random forest-based algorithm for data-intensive spatial interpolation in crop yield mapping. Comput Electron Agric 184:106094. https://doi.org/10.1016/j.compag.2021.106094

108. Betzek NM, Souza EG de, Bazzi CL, et al (2019) Computational routines for the automatic selection of the best parameters used by interpolation methods to create thematic maps. Comput Electron Agric 157:49–62. https://doi.org/10.1016/j.compag.2018.12.004