UTRECHT UNIVERSITY

Department of Information and Computing Science

Applied Data Science master thesis

Stability and performance evaluation of different windspeed spatial interpolation models to assist pesticide dispersion estimates.

First examiner:

Erik-Jan van Kesteren

Candidate:

Arsalan Anwari (1261711)

Second examiner:

Daniel M. Figueiredo

In cooperation with:

IRAS (institute of risk assessment sciences)

June 25, 2023

Abstract

Health risks to respiratory, reproductive, neurological, endocrine, and circulatory systems are known to be caused by prolonged pesticide exposure. An existing mixed model was developed (OBOmod) by Utrecht University's Institute of Risk Assessment Sciences (IRAS) which considers many variables to estimate pesticide exposure near households in the Netherlands. The effect meteorological estimates (specifically windspeed) using a less biased spatial interpolation model alongside the Gaussian plume-based pesticide dispersion model part of OBOmod has not been studied yet. This paper compared seven spatial interpolation models using a total of ten metrics and recommended the use of a hyperbolic trend surface model to minimize bias caused by random error and trends in measurements which Gaussian plume models are known to be most sensitive to. Ideal annual model hyperparameter where trained using Bayesian optimization with a logarithmic loss function.

Contents

1. Introduction

Pesticide exposure, whether direct or indirect, is known to raise the risk of various health consequences owing to interference with the respiratory, reproductive, neurological, endocrine, and circulatory systems (Rani L. et al, 2021). At Utrecht University's Institute of Risk Assessment Sciences (IRAS) a model was developed to estimate pesticide dispersion in several parts of the Netherlands. This model named OBOmod can predict pesticide exposure for households in the Netherlands.

While this model considers many different variables which influence pesticide dispersion, one neglected aspect of the model is the use of high spatial resolution meteorological data. It is unclear, for example, how different spatial interpolation models affect final pesticide exposure estimates (Figueiredo D. M. et al, 2022).

This paper investigates the stability and accuracy of different spatial interpolation models for windspeed estimation using a variety of metrics. The results of these metrics are evaluated to recommend a model to be used for future research which analyzes the effect of interpolated windspeed estimates on pesticide exposure estimates using the dispersion model part of OBOmod.

Reduction of seasonal and random bias in the windspeed estimates are detrimental to improve the performance of gaussian plume models used by the dispersion model (see chapter 2.3).

Basic understanding of spatial interpolation and geographic information systems is expected by the reader, but a brief overview [9] covering these topics and other topics covered in this paper is available in the appendix.

1.1 Structure

First background information is provided, needed to understand the context of this paper (chapter 2). Here the problem and causes of pesticide drift are briefly mentioned and the role windspeed plays is described (chapter 2.1). Windspeed itself is also explained as a meteorological condition and what external processed can influence it (chapter 2.2). Then a brief overview is given about the dispersion model and what factors can influence final pesticide dispersion estimates (chapter 2.3).

Next (chapter 3) more detail is given about the procedures used to prepare the meteorological data (chapter 3.1), the considerations that have been taken into account to ensure windspeed estimates are compatible with the format expected by the dispersion model of OBOmod (chapter 3.2); the concessions that have been made to reduce computational complexity and time needed to train the models (chapter 3.3); the methods used to train and optimize spatial interpolation models with a description of the different metrics used to validate different models (chapter 3.4) and a description of the exact spatial interpolation models used with references to existing studies using these models (chapter 3.5)

Then the results (chapter 4) of the different metrics are presented where the recommended model is provided (chapter 5) based on the metric results.

Finally, the reasons to not include auxiliary variables like temperature and windspeed; the selection of hyperparameters and not use more sophisticated models are discussed (chapter 6).

1.2 Scope

A total of seven different models were compared covering a wide range of spatial interpolation categories, with a total of ten metrics being used which give a measure of stability, reproducibility, and similarity of model estimates. Recommendations are based on measurements of a single year (2017) and estimates of a surface are represented by a grid of 5 km^2 cells.

1.3 Considerations

Abbreviations are used for both the metrics and models in this paper. This was done to make referencing these metrics and model in figures and text more compact.

Stability metric		**Surface metric**	
Abbreviation	**Name**	**Abbreviation**	**Name**
SAV	Seasonal-annual variability	MaxSR	Maximum surface roughness
SSV	Seasonal-seasonal variability	MedSR	Median surface roughness
ARV	Annual-random variability	MadSR	MAD surface roughness
RRV	Random-random variability	MiMaSD	Minimum-maximum surface deviation
GLV	Global-local variability	MadSD	MAD surface deviation

Figure 1.3.1: Metrics used in the paper with their abbreviations.

Abbreviation	Model
TS1	Linear trend surface
TS3	Hyperbolic trend surface
IDW	Inverse distance weighted
MQ-RBF	Multi quadratic radial basis function
OK	Ordinary kriging with spherical variogram (vsph)
UK1	Universal kriging with linear regression + vsph
UK3	Universal kriging with hyperbolic regression + vsph

Figure 1.3.2: Spatial interpolation models used in the paper with their abbreviation.

To prevent confusion between an axis used to denote a 2d or 3d vectors (x, y, z) and a spatial feature (x_i, y_i), a custom notation scheme will used to denote features (v_i, z_i) throughout the paper.

- $v_i := location\ coordinate \rightarrow \{v_{ix}, v_{iy}, v_{iz}\}$
- $z_i := attribute\ at\ location\ coordinte\ v_i$

Throughout the paper a location coordinate will be referred to as a **location** and an attribute as a **variable**. An **observation** (v_j, z_j) refers to a feature with a known variable whilst a **receptor** $(v_i, \hat{z}_i)$ refers to a feature with an unknown variable that is to be predicted.

Finally, a shorter naming convention will be used to indicate the used CRS consisting of two parts: the first is the name of the data source, and the second is the date of the projection. In some cases, an additional name will be added which represents the projection name. Some examples include (WGS:84) and <Amersfoort / RD New> (EPSG:28992).

2. Conceptual framework

2.1 Pesticide exposure and drift

Within the agricultural sector, pesticide drift is a major contributor to off-target contamination. This can occur through the evaporation of pesticide droplets before they reach their intended target, or because of wind-blown soil particles.

Additionally, the inherent chemical properties of the pesticide itself can impact its volatility and therefore its tendency to drift (Schampheleire D. et al, 2009). Pesticides with higher volatility are more likely to evaporate into the atmosphere under warm and dry conditions, while increased humidity can reduce evaporation rates (Bish M. et al, 2021).

Precipitation can also impact pesticide behavior, with the potential to reduce vaporization while increasing leaching into soil and water sources (Zaller J. et al 2022).

Windspeed also influences the deposition of pesticide droplets on ground and air. As wind speed increases, the deposition of droplets decreases. This effect is more pronounced for smaller droplets than for larger droplets (Zhang H. et al 2017).

Therefore, incorporating meteorological data into exposure assessments is crucial to improve the accuracy and reliability of exposure estimates. One of the key factors affecting exposure estimates is the variability of meteorological conditions. Meteorological conditions, such as temperature, wind speed, and rainfall, can significantly influence the behavior of pesticides in the environment.

2.2 Windspeed as a meteorological condition

Wind speed is a fundamental atmospheric parameter measured in (m/s) caused by air moving from high to low pressure, most commonly due to temperature or topology variations.

Temperature changes between air masses cause pressure variances, which result in wind. Winter provides increased temperature gradients, especially when cold fronts move in from the polar regions, resulting in higher-than-normal wind speeds. As the air temperature drops, the chilling effect of any wind rises (Pryor S. C. et al, 2020).

The topology of a region can influence wind speed by altering the pressure gradient force. This is the force that moves air from high-pressure locations to low-pressure areas. Topology also affects wind speed by influencing local weather conditions. Mountains, for example, can force air to climb and chill, resulting in precipitation and lower wind speeds. Valleys, on the other hand, can cause air to sink and warm, resulting in increased wind speeds (Wever N., 2012).

2.3 Gaussian plume model

Gaussian plume models are commonly used to simulate the dispersion of air pollution. It is simple to use and can provide quick pollutant concentration estimates. It does, however, have certain restrictions. It presupposes, for example, that the plume is well-mixed and that the wind speed and direction are constant throughout the plume (Henderson S.B., 1987).

When adding external meteorological data to the model, it is critical to minimize bias in the external data. Extreme outliers, roughness, trends, local, and random error in measurements can generate biased results using these models. Which of these factors contribute the most to bias depends on the model and its exact implementation. Previous research has shown that for model implementations random error and trends was the biggest culprit to bias in estimates (Hosni Snoun et al., 2023).

3. Data & Methods

The following chapters will further elaborate on the source and nature of the wind speed data; what preprocessing steps to clean the data where utilized; the different methods used to reduce computational complexity; which interpolation models were used and how they pertain to wind speed interpolation and finally what metrics were used to give a measure of stability, reproducibility, and similarity of model estimates.

3.1 KNMI weather stations

The Royal Netherlands Meteorological Institute (KNMI) is the official weather agency of the Netherlands. It is also the national meteorological, climate, air quality, seismology research and information center. Their main responsibilities include weather forecasting and monitoring of weather, climate, air quality, and seismic activity. The KNMI provides both hourly and daily data [1] for a variety of meteorological conditions, windspeed being one of them. There a total of 47 weather stations available spaced relatively equally throughout the Netherlands, with some exception which are clustered around coastal regions. Some stations are also situated further away from the sea.

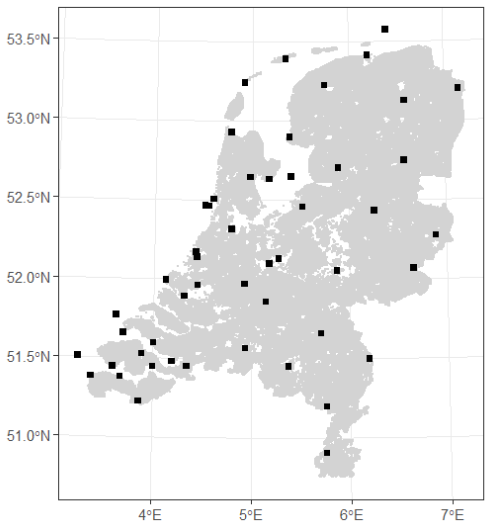

Fig 3.1.1: Map of the Netherlands with location of KNMI weather stations

The stations can be identified by their station code which is represented as a three-digit whole number in the range {209 ... 391}. There is no spatial location data provided in this dataset so an external data source [2] was merged with the data provided by the KNMI for each station code. This external data source uses (WGS:84).

Windspeed in this dataset is measured in the unit ($0.1/ms^{-1}$). Four variables were available which denote some function of windspeed but only the average windspeed (FG) was used for interpolation.

To align with the validation set available for the dispersion model, daily measurements in the year 2017 were chosen (see chapter 1.3).

Missing values

Three weather stations had missing values or incomplete data for the year 2017: Wilhelminadorp (323), Wijk aan Zee (257) and Huibertgat (285). The number of missing entries ranged between 232 and 365 days. Due to the high number of missing entries these stations were removed from the dataset instead of employing an imputing strategy. This would not cause negative side-effects for following two reasons:

1. The three stations were centered around a dense cluster of other stations where the difference in measurements (for the available days) between the closest neighbors were insignificant, with a minimum difference of 0.01% and a maximum of 0.06%.
2. Data reduction techniques would already have reduced the number of stations measuring a value at the same grid cell (see chapter 3.3.3).

Several stations had windspeed values which centered around a mean of 4.5 m/s without too many extremes. However, some stations alongside the coast had several high extremes up to 18 m/s, all showing an upwards seasonal trend. These extreme outliers might be caused by a diurnal pattern known to be present for windspeed along coastal areas. (Dennis Elliot et al, 2004; Semedo A., 2018; Miao, S., 2021).

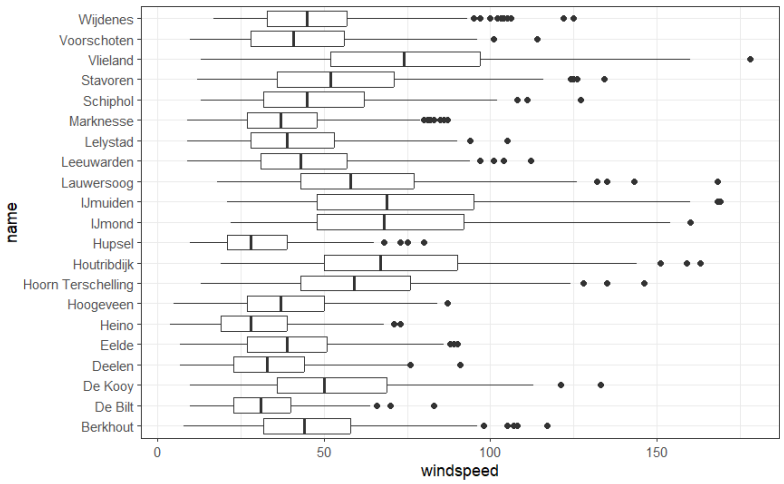

Fig 4.1.3: Random sample of stations with their data distribution in $(0.1/ms^{-1})$.

3.2 Dispersion model considerations

Windspeed estimates would be employed with a simplified version of the **OPS-st** model, which is one of five independent models employed by OBOmod in the future. OPS-st is an abbreviation for the **short term** (st) **Operational Priority Substances** model (OPS). This is a sophisticated Gaussian plume model that is used to calculate air transport, dispersion, and concentrations at receptor locations throughout the Netherlands. It computes concentrations that are used as input to the **gComis** ventilation model, which estimates concentrations in interior air based on outdoor concentrations by utilizing exchange rates between indoor and outdoor air

The innerworkings of the OPS-st model (even the simplified version) are beyond the scope of this paper; more information about this model can be found in the scientific paper about OBOmod [3]. This paper only concerns itself with the format the windspeed estimates need to have, to be compatible as an input parameter for the OPS-st model.

The OPS-st model uses the BRP Gewaspercelen (BRPG) [4] map as its source. This a geodatabase file format made by ESRI using <Amersfoort / RD New> (EPSG:28992) with an accuracy of 1 meter [5]. This map shows the crop fields in the Netherlands represented as a collection of non-uniform multi-polygon objects.

Different representations for the receptors can be used including polygon centroids, equal distance grid cells, hexagon grid cells and centroids of equal distance grid cells. Grids cells need to have minimum diameter of 1 km. In this paper centroids of equal distance grid cells with a diameter of 5 km were chosen to reduce computational complexity (see chapter 3.3.3).

Pesticide dispersion estimates can be generated on an hourly basis for simulations using the dispersion model.

3.3 Complexity reduction

Receptor reduction

By default, the BRPG map consists of 785′710 features represented as multi-polygons using between 3 and 8 coordinates per feature. Multi-polygons require each individual point to be predicted. This increases the receptor points far above 785′710 for a total of 3′928′550 receptors. To keep the computational complexity low a series of reduction techniques were utilized to lower the total number of receptors from 3′928′550 to 1′813.

1. A tessellation technique was used to generate a grid of 5 km^2 cells. This requires the use of a simplified reference map (see chapter 3.3.2) where the BRPG map acted as a source for the CRS. The diameter of 5 km was chosen as it produced a grid with a spatial resolution high enough to reduce the number of station collisions whilst minimizing the number of receptors generated (see chapter 3.3.3).
2. Next the same reference map was used again to intersect all grid cells which fall outside the boundaries of the Netherlands (cells representing the sea).
3. Then the centroids of the grid cells were calculated and used as receptors instead of the entire polygon.
4. Finally, the cells containing the stations were removed.

Reference map blending

Generation of a grid of cells using the BRPG map was ill suited as it contains many open holes and noisy surfaces alongside the coast. As the diameter used for the cell decreases the more cells that should be included in the grid will get omitted from the grid when removing unnecessary cells on the outlines of the map.

Although it is possible to reduce the features of the BRPG map to a single multi-polygon by using a variety of algorithms, each of these methods have their pros and cons further elaborated in a blog post using different maps in Germany [5]. The biggest problem with these methods is that they can skew the shape of the true map boundaries, essentially adding or removing new or existing plots of lands along the coast which are not actually present. Therefor the decision was made to use a simplified reference map which contained an accurate coastline of the Netherlands.

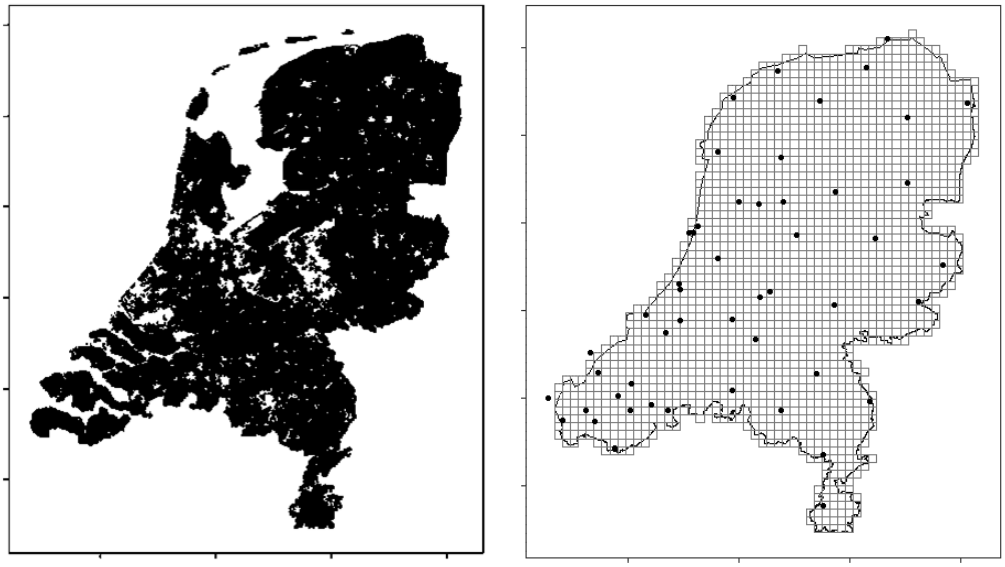

Fig 3.3.1: BRPG map (left) compared to reference map with KNMI stations (right).

Larger grid cell diameter

After some experimentation, it became apparent that using cells with a diameter of 5 *km* would be sufficient to cover most measurements in each cell individually, with only a single collision of two stations in the same cell and one instance of station intersecting with four cells. This was solved by taking the mean of the two stations and selecting the cell with the smallest distance.

The simplified dispersion model supports grid cell diameters up to 1 *km* which results in a total of 42'418 receptors being generated, but at a diameter of 5 *km* only 1'813 receptors were generated.

The total time needed to train multiple models, with different hyperparameters at different time (days) intervals and individually validating intermediate results could therefore be significantly reduced.

For reference, one iteration of training of an IDW model with 42'418 receptors took about 38 seconds whilst 1'813 receptors took about 7 seconds.

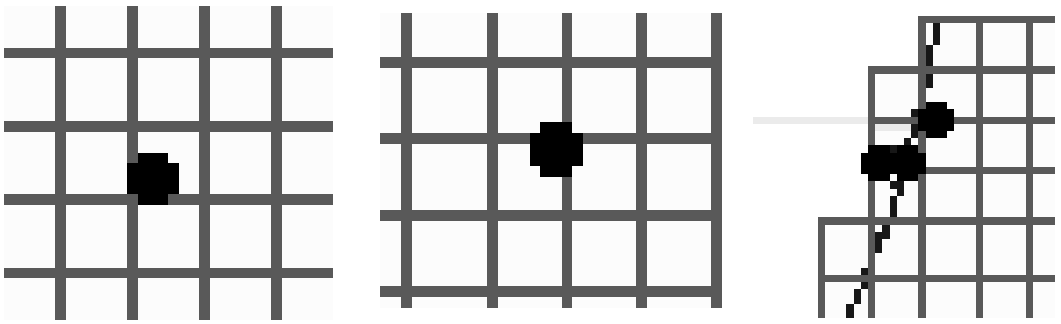

Fig 3.3.2: Example of a correct intersection (1), intersection with one station and multiple cells (2) and an intersection of multiple stations in a single cell (3).

3.4 Training and validation

In the previous chapter, different techniques were used to reduce the number of receptors. This was done to reduce the computational time needed for the **dynamic training routine**. This training routine required validation after each training iteration to determine the optimal next hyperparameter. This was done for each subsequent day to determine the **best annual parameters (BAP)** for each model. These parameters should lower the RMSE for any random day on average compared to not using any parameters at all. It should also reduce bias caused by not using optimally tweaked models in the metrics using the RMSE of several days.

Spatial-leave-one-out cross validation (SLOO-CV) was used to determine the RSME. The loss or gain in RMSE served as input for the dynamic model training routine to calculate loss values and parameters.

Finally, different metrics using models with BAP were utilized which give a measure of stability, reproducibility, and similarity of model estimates.

Dynamic training routine

Hyperparameters are not chosen randomly or sequentially, but rather a process of dynamic incrementation or decrementing of hyperparameters was used to reduce the amount of training iterations required.

After each training iteration the model was validated and the performance gain or loss in RSME from the previous iteration was measured. This performance gain or loss was fed to a **logarithmic loss function** which decided an appropriate parameter for the next iteration based on a unique **weight function** for each model.

This meant that models were trained bi-directionally to finally reach an equilibrium where the performance loss or gain would be below a minimum threshold which would subsequently halt the routine. This process was repeated for each model for each subsequent day and the mean parameters of all days were calculated as the BAP.

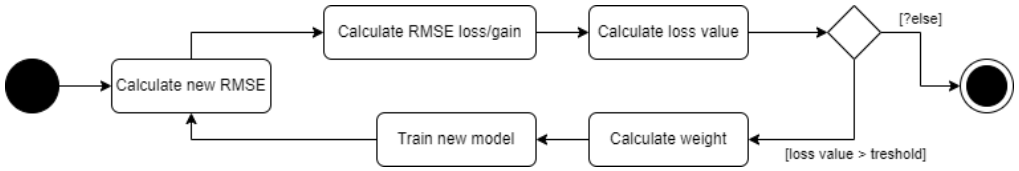

Fig 3.4.1: Abstract overview of the dynamic training routine process.

This is essentially a hyperparameter tuning mechanism different from existing methods like grid or random search which are **static training routines** where a fixed number of hyperparameters following some predefined distributions are used. A dynamic training routine generates new parameters on the fly based on some random or deterministic function.

The method used in the paper is comparable to **Bayesian optimization** but uses a logarithmic function instead of a random gaussian function.

This method allowed for much more efficient training and validation of models where optimal parameters between a wider range of values could be found. The total runtime was also much lower per model as only one single parameter needed to be calculated and compared for each new instance of the model per day instead of a distribution of parameters for each day.

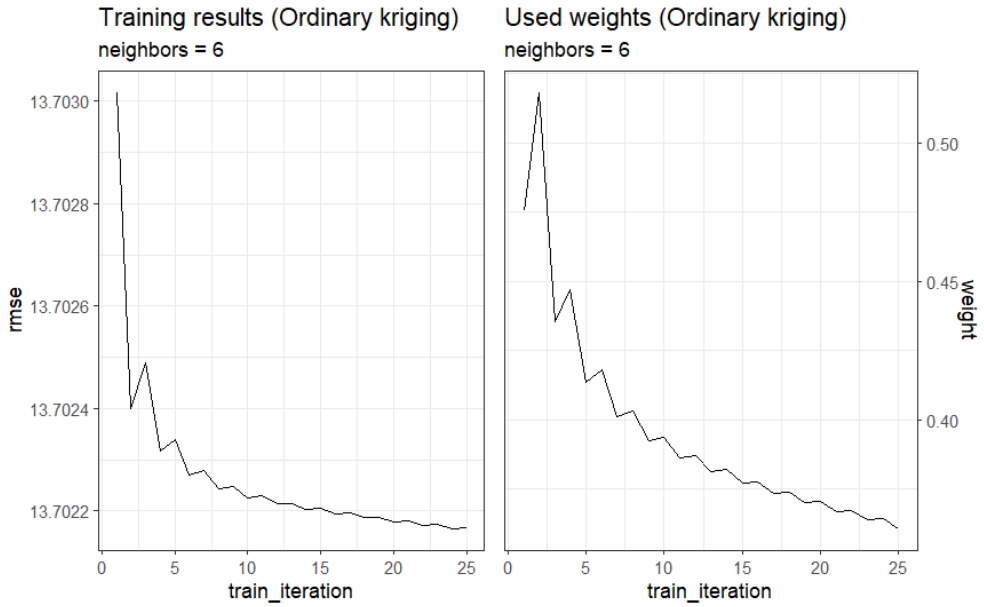

Fig 3.4.2: Training results (left) compared to sill parameter weights (right) using the dynamic training routine for one variation of an ordinary kriging model. Parameters are increased or decreased exponentially based on the loss or gain in RMSE until they reach an equilibrium where the loss or gain in RMSE is below a threshold at iteration 25.

The code used for the dynamic training routine is configurable to test different variations of a model (ex: linear or hyperbolic kriging models) each with their own tweaked settings. Parameters are generated by a set of different weight functions which can be used to train different parts of the model (ex: sill, range, and nugget values for kriging models). The logarithmic loss function can also be tweaked to calculate less or more drastically changing loss values. More information about how this routine works and what parts can be tweaked can be found on the repository [7] or on the wiki page [8].

Model metrics

To validate the interpolation performance, **stability metrics** and **surface metrics** were defined. The stability metrics cover annual, seasonal, and random accuracy at a global and local level. The surface metrics are related to the effect different interpolation methods have on the generated surface.

A total of 5 datasets samples were generated based on the windspeed dataset from KNMI. Sample sizes were divisible by 4, contained 12 days of data and compared sample sizes had the same **like-terms**. This was done to reduce bias in population **insensitivity** (Zhan S. et al, 2022). A total of 10 metrics were defined where 5 metrics were used covering local/global, annual/seasonal, and random variability, and the remaining 5 metrics covered the surface roughness and deviations at different time intervals.

Data set	Description
Seasonal set	The data is split into 4 seasons these being Spring (March to May), Summer (June to August), Autumn (September to November) and Winter (December to February). The first day of each week per month per season is picked for a total of 48 days with 12 days per season.
Annual set	The first 4 days of each month are picked for a total of 48 days per year.
Random set	48 random days are picked for a single year.
Clustered set	The data for 48 random days is taken for the 12 clustered station near the province of Zeeland.
Sparse set	The data for the 48 days of the clustered set is taken for 6 random stations in the center of the Netherlands and 6 stations in the north of the Netherlands where the distance between them was maximized around a center station.

Fig 3.4.3: List of different data samples sets with their descriptions used by the spatial interpolation performance metrics.

Stability metrics

To measure the stability of spatial interpolation methods the RMSE values were fed to a series of formulas representing different metrics. These metrics aim to provide a better overview of the strengths and weaknesses of each individual method in different scenarios and how stable the predictions are in these scenarios.

Some models might be more accurate at specific dates but struggle when the values of areas with densely clustered observations fluctuate more due to extremes (ex. this can happen in extreme weather conditions along the coastline). Greater importance is placed on precision and the degree of reproducibility.

Surface metrics

To measure how similar the surfaces generated by different interpolation methods are at different time intervals the random set was used explicitly for the observations where surfaces were generated for each day present in the random set. A total of five metrics were defined which covered the surface roughness and the difference between the minimum, maximum and mean absolute deviations between measurements at different time intervals.

Metric	Description
SAV	**Seasonal-annual variability** calculates the maximum difference in RSME between each seasonal set with the average RSME of the annual set. This metric describes how much prediction errors differ between each seasonal and annual trend. The lower this difference is, the less *sensitive* the interpolation method is for *temporal trends*. The mean RSME of each seasonal set is compared with the annual set to calculate 4 residuals. The maximum difference between these 4 residuals is used as the score.
SSV	**Seasonal-seasonal variability** calculates the maximum difference in RSME between seasons. This metric describes how much predictions can differ per season. The lower this difference is, the more *stable* predictions of this interpolation method are *between seasons*. The mean RSME of each seasonal set is calculated as 4 residuals. The maximum difference between these 4 residuals is used as the score.
ARV	**Annual-random variability** calculates the average difference in RSME of the random set and the annual set. This metric describes how much predictions can differ between a random set of days and the annual trend. The lower this difference is, the less *sensitive* the interpolation method is to *short term anomalies* in measurements. The mean RSME of the random and annual sets are calculated as 2 residuals. The difference between these 2 residuals is used as the score.
RRV	**Random-random variability** calculates the maximum difference in RSME between parts of the random set. This metric describes how the predictions can differ within the random set. The lower this difference is, the less sensitive the interpolation method is to *long term anomalies* in measurements. The mean RSME of 4 splits of the random set is calculated as 4 residuals. The maximum difference between these 4 residuals is used as the score.
GLV	**Global-local variability** calculates the difference in RSME between the clustered and sparse set. This metric describes how predictions can differ between samples of varying spatial densities. The lower this difference is the less spatially sparse predictions are *influenced* by spatially dense data. The RSME of the clustered and sparse sets are calculated as 2 residuals. The difference between these 2 residuals is used as the score.

Fig 3.4.4: List of stability metrics with their description and point distribution.

Metric	Description
MaxSR **MedSR** **MadSR**	The **surface roughness (SR)** at each point is calculated using the integrated squared second derivative standardized by its maximum. This formula expects data with some base line (ex. sea-level for topology or zero line for analogue signals). The standard deviation of the observations of that day is used as the base line. Here the maximum, median and MAD of the SR values are taken as three scores separately.
MiMaSD	**Minimum-maximum surface deviation** calculates the difference between the minimum and maximum value of a predicted surface for each day. Here the average deviation of all days is taken as the score.
MadSD	**MAD surface deviation** calculates the MAD of a predicted surface for each day. Here the average of all MAD values of all days is taken as the score.

Fig 3.4.5: List of surface metrics with their descriptions.

3.5 Windspeed interpolation

Linear and hyperbolic trend surface

Using trend surfaces for multiple types of climate data is commonly used as a baseline to capture a good mix of the local extremes and overall global trend when using meteorological data (Guo B. et al 2021). In an existing study a first order polynomial model (also referred as linear model) and a third order polynomial model (also referred as a cubic or hyperbolic model) were used for windspeed interpolation in Iraq (Ali S., 2012). Both linear and hyperbolic trend models were compared in this paper.

Multi quadratic radial basis function

An anisotropic radial basis function was used in an existing study for long term windspeed spatial interpolation for complex surfaces in the US. (Lee C. 2022). Another study explored both anisotropic and multi-quadratic radial basis functions where the latter generally performed better (Reinhardt K. et al 2018). As surface in the Netherlands can be considered flat, only multi quadratic radial basis functions were explored in this paper.

Inverse distance weighing.

There are several studies which cover the use of IDW for windspeed interpolation but one notable one compared a wide range of modified IDW models which consider both observation to observation and observation to receptor dependencies to improve the predictions and reduce sensitivity to severely clustered observations (Li Z., 2021). However, since observations of the KNMI weather stations are not this severely clustered, a regular IDW model was used.

Ordinary kriging with spherical variogram.

Multiple studies exist which tested the accuracy of ordinary kriging for spatial interpolation. Using a spherical variogram is preferable for data that has a short spatial correlation (Burrough et al., 2015; Cressie, 2015). Another study analyzing affected areas by dust storms in Iran using satellite images compared a wide variety of kriging methods and concluded that ordinary kriging with a spherical variogram had the lowest standard error compared to exponential or linear variograms (Ekhtesasi M. R. et al, 2012). Based on these two studies, the choice was made to only explore ordinary kriging and kriging methods in general with spherical variograms.

Universal kriging with linear and hyperbolic regression model.

Linear and gaussian based regression models have been used before for spatial interpolation of windspeed where one notable study concluded that hyperbolic and spherical models produce a lower RMSE than Gaussian based models (Wang, Y., 2020). Another notable study compared linear, spherical, and hyperbolic models against a neural kriging model for windspeed in Sicily and concluded that linear and hyperbolic models performed better against spherical models (Cellura, M. et al, 2008). Therefor the choice was made to explore both linear and hyperbolic regression models alongside a spherical variogram to explore the performance differences between the two models.

4. Results

The following chapter compares the performance of different spatial interpolation methods based on difference of models using daily, annual, and no optimized parameters. Next the generated surfaces are shown visually. Lastly the results of the stability and surface metrics of models using BAP are presented.

4.1 Daily, annual, and non-weighted errors

This dynamic training routine was executed twice, once to calculate ideal parameters for each single day (BDP) and once to calculate BAP. The idea behind this is to see how much BAP differs from BDP for any random day. Ideally this difference is minimized to reduce bias which can be present in the evaluation of the metrics using models with BAP. The performance of models without any parameters (NP) was also compared to be used as a baseline with BAP to evaluate how much BAP improves model performance on average. The decision to evaluate metrics using model with BAP over BDP was due to concessions in computation complexity and time complexity that had to be made for estimates to be viable to use with the dispersion model (see chapters 3.2 and 6.2)

As expected, models using BDP always had the lowest RMSE for a single day. Majority of the time models using BAP had a lower RMSE than those using NP, but some instances existed where NP produces a lower RMSE. For example, on January 1st, the RMSE of models using BAP compared to NP was increased by an average of 0.07308, whilst on October 18th, the RMSE was reduced by an average of 0.13749.

When calculating the average RMSE of all days in the year between models using BAP and NP, BAP on average reduced the RMSE by 0.11873 per day.

Model	Best annual parameters	Best daily parameters
TS1	case_weight = 1.151302	case_weight = 1.09743
TS3	case_weight = 1.00050	case_weight = 1.01732
MQ-RBF	alpha_seed = 1.166434 smoothing_factor = 90.770461	alpha_seed = 1.34922 smoothing_factor = 78.6283
IDW	nmax=6 idp=1.028791	nmax=6 idp=1.07923
OK	vgm="Sph" psill=348.3421 range=40943.12 nmax=6	vgm="Sph" psill=347.8823 range=40944.28 nmax=6
UK1	vgm="Sph" psill=188.709 range=13714.15 nmax=6	vgm="Sph" psill=188.653 range=13713.522 nmax=6
UK3	vgm="Sph" psill=188.7081 range=13714.12 nmax=6	vgm="Sph" psill=188.651 range=13713.517 nmax=6

Fig 4.1.1: Example of BAP and BDP used for January 1st

Model	RMSE (BAP)	RMSE (BDP)	RMSE (NP)
TS1	12.2824	10.8142	12.23348
TS3	12.92801	11.6253	12.92795
MQ-RBF	13.77851	13.37851	13.48196
IDW	13.98033	13.92033	13.96693
OK	13.70217	13.44217	13.62906
UK1	14.32642	14.26642	14.28966
UK3	13.66884	13.61531	13.62607

Fig 4.1.2: RMSE results on January 1st, RMSE(BAP) > RMSE(NP)

Model	RMSE (BAP)	RMSE (BDP)	RMSE (NP)
TS1	12.23751	10.6827	12.73446
TS3	12.38736	11.8154	12.62145
MQ-RBF	13.4674	13.39335	13.51874
IDW	13.9183	13.89374	13.97534
OK	13.6232	13.4182	13.71921
UK1	14.39231	14.37224	14.41833
UK3	13.67724	13.65232	13.67829

Fig 4.1.3: RMSE results on October 18th, RMSE(BAP) < RMSE(NP)

4.2 Generated surfaces

The surfaces generated by the different models where quite unique in their shapes, but one thing they had in common was that higher windspeed values were located more closely to coast whilst lower windspeed values were located closer to the border with Germany. Windspeed values ranged between a minimum of 3 (m/s) and a maximum of 10 (m/s). For readability the windspeed values were rounded upwards before creating the figures. Surfaces generated by models using BAP, BDP or NP looked very similar with some exceptions like the MQ-RBF model. In the figures the tag *'[best_case]'* refers to models using BAP and *'[avg_case]'* refers to models using NP. White grid cells represent the KNMI stations.

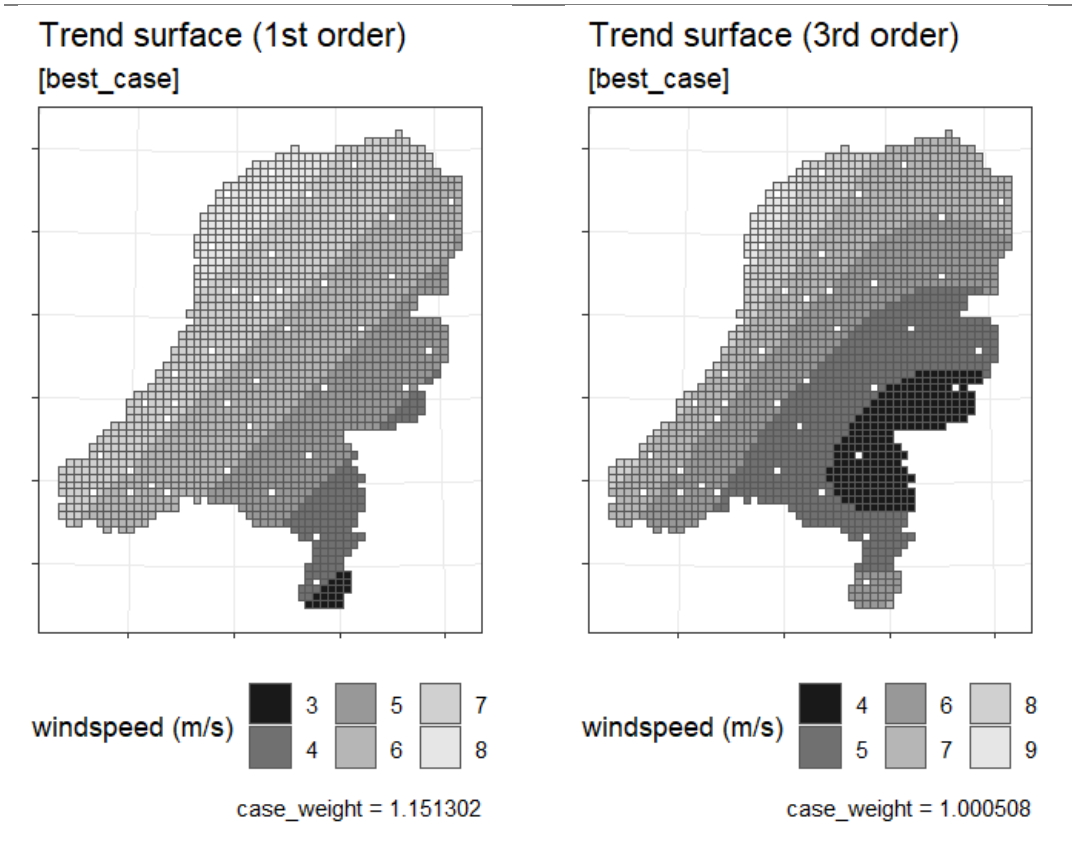

Fig 4.2.1: Surfaces generated by TS1 and TS3.

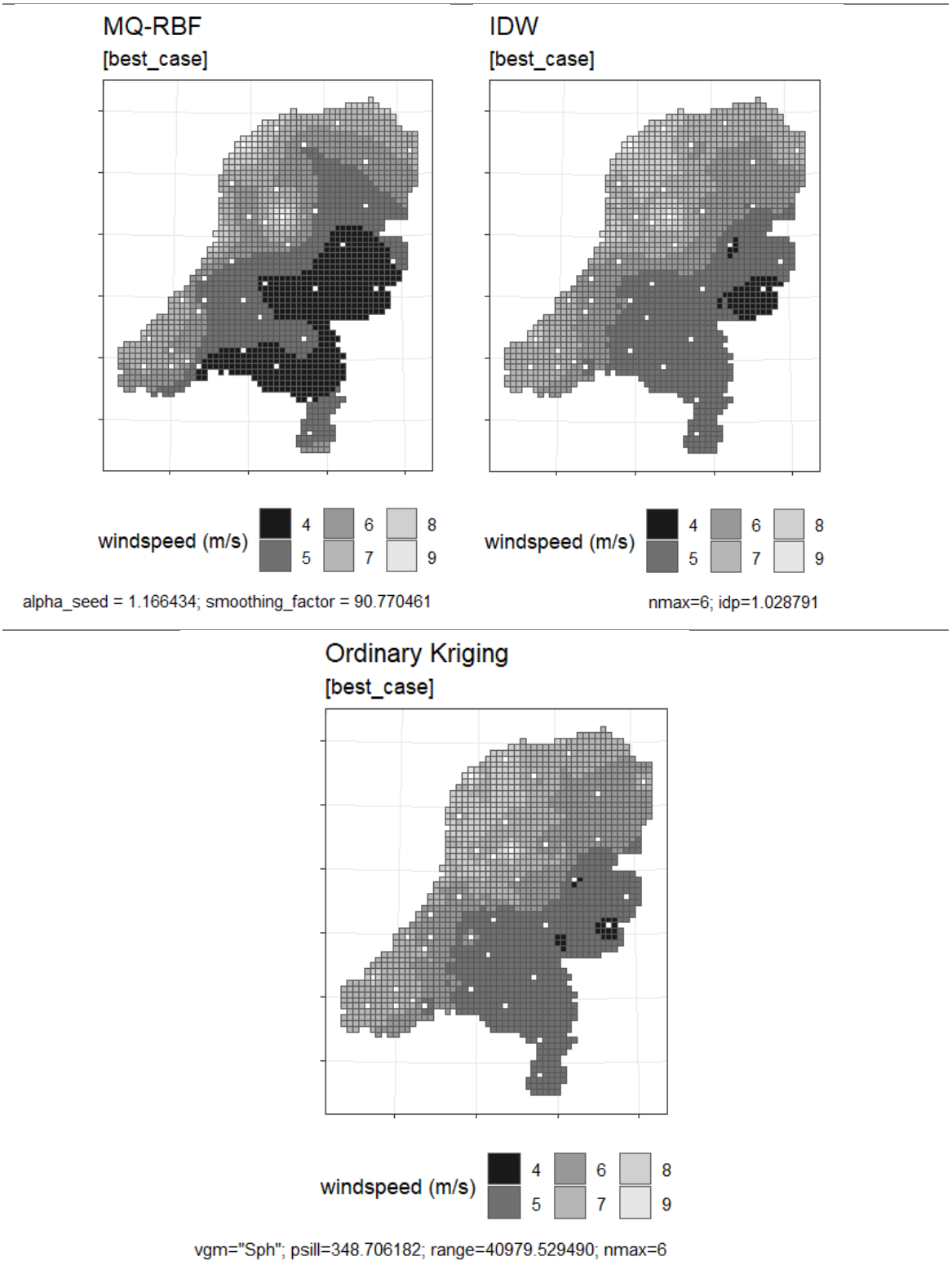

Fig 4.2.2: Surfaces generated by MQ-RBF, IDW and OK.

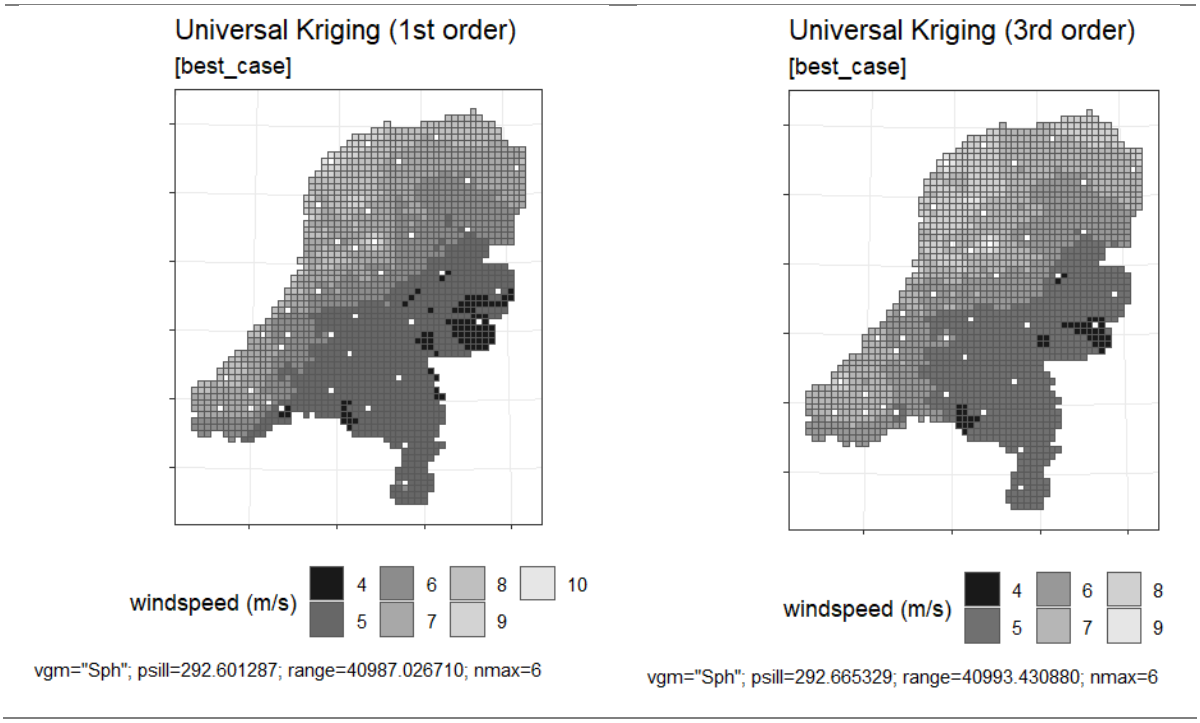

Fig 4.2.3: Surfaces generated by UK1 and UK3.

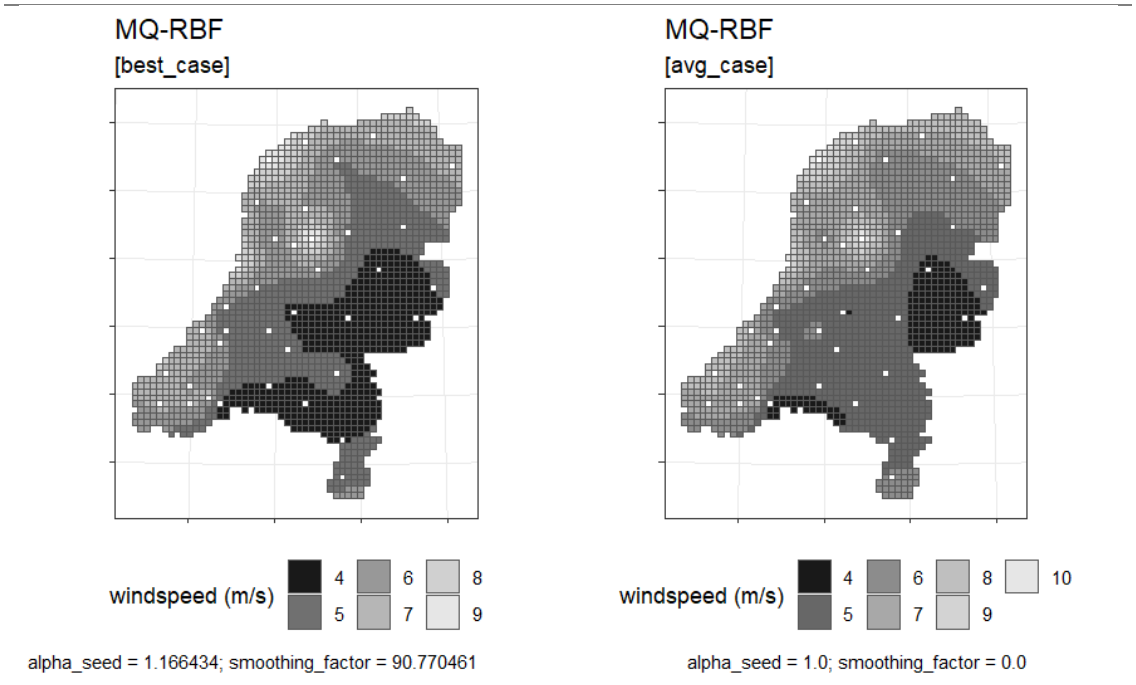

Fig 4.2.4: Surfaces compared between MQ-RBF models using BAP and NP.

4.3 Metric results

The results of the stability metrics showed that TS3 and TS1 were within the top 3 majority of the time whilst UK1 and UK3 were occasionally within the top 3. OK performed decently, averaging 4th place majority of the time whilst MQ-RBF performed poorly majority of the time. IDW performed well on the GLV metric but poorly in all other metrics. For most stability metrics the difference between the first two places and the last 5 places was highest whilst differences diminished after the 4th place (see figure 4.3.1).

For all surface roughness metrics UK3 scores the best followed by OK and IDW however this was not the case for the surface deviation metrics where IDW and OK scored significantly better than UK3. MQ-RBF performed decently averaging 4th place, majority of the time besides at metric MadSD where it took 1st place. TS1 performed the worst overall besides metric MiMaSD where it scored very close to UK3 for the 3rd place. Differences in scores were highest for the surface roughness metrics and lowest for the MadSD metric (see figure 4.3.1).

To quantify results a pointing system was introduced using a rating from 7 to 1 point. 7 points were given for the 1st place and 1 point for the last place.

UK3 performed best overall in all metrics. TS3 performed best overall for the stability metrics and UK3 performed best overall for the surface metrics. UK3 and TS3 performed better with seasonal or annual trends. UK1 performed best with global and local variability with UK3 performing slightly better than TS3. TS3 performed best overall for random variability in the data with UK3 performing significantly worse (see figure 4.3.2).

Metric	Leaderboard						
	1st	2nd	3rd	4th	5th	6th	7th
SAV	TS3 (1.3148)	UK3 (1.4789)	TS1 (1.5205)	OK (1.5882)	MQ-RBF (1.6909)	IDW (1.7082)	UK1 (1.7618)
SSV	UK3 (1.8719)	TS3 (2.1027)	MQ-RBF (2.1589)	OK (2.2040)	TS1 (2.3240)	UK1 (2.4070)	IDW (2.4161)
ARV	TS1 (0.4969)	TS3 (0.5005)	OK (0.6501)	UK3 (0.6724)	IDW (0.7787)	MQ-RBF (0.7825)	UK1 (0.909)
RRV	TS3 (1.8952)	UK1 (2.1186)	TS1 (2.2327)	MQ-RBF (2.4194)	OK (2.4428)	IDW (2.4963)	UK3 (2.7148)
GLV	UK1 (0.059)	IDW (0.3549)	TS1 (0.5003)	OK (0.8134)	UK3 (0.9943)	TS3 (1.3590)	MQ-RBF (3.1946)
MaxSR	UK3 (27.1947)	OK (30.4828)	IDW (31.0395)	MQ-RBF (33.4901)	UK1 (37.5314)	TS3 (38.6613)	TS1 (62.6854)
MedSR	UK3 (12.2785)	OK (13.8667)	IDW (13.7342)	MQ-RBF (14.3151)	TS3 (16.2958)	UK1 (17.8332)	TS1 (22.6922)
MadSR	UK3 (4.7468)	OK (5.1472)	IDW (5.2679)	MQ-RBF (5.9620)	TS3 (6.2661)	UK1 (6.4736)	TS1 (9.7963)
MiMaSD	IDW (48.5701)	UK3 (50.8512)	TS1 (50.8888)	OK (53.0801)	MQ-RBF (53.6334)	TS3 (54.5250)	UK1 (57.1220)
MadSD	MQ-RBF (10.5552)	OK (11.4156)	UK1 (11.7740)	IDW (11.9801)	UK3 (12.0932)	TS3 (12.4917)	TS1 (12.9181)

Fig 4.3.1: Stability and surface metric results for each model rounded to 4 decimals.

Model	Total metric points					
	All metrics	Stability metrics	Surface metrics	SAV + SSV	GLV	ARV + RRV
TS1	34	25	9	8	5	12
TS3	40	28	12	13	2	13
MQ-RBF	37	15	22	8	1	6
IDW	40	14	26	3	6	5
OK	48	20	28	8	4	8
UK1	39	17	13	3	7	7
UK3	51	21	30	13	3	5
Max	70	35	35	14	7	14

Figure 4.3.2: Point distribution of models with different section of metrics where the best models are highlighted for each section and the maximum obtainable points are defined in the last row.

5. Conclusion

This paper aimed to recommend a spatial interpolation model to be used for future research which analyzes the effect of interpolated windspeed estimates on pesticide exposure estimates using the dispersion model part of OBOmod. This dispersion model is based on a Gaussian plume model whose performance is known to be sensitive to bias in auxiliary variable used. Therefor it was recommended to use a model which minimized these forms of bias. Bias caused by random error or trends had the highest effect to model performance (see chapter 3.2).

To evaluate and measure this bias, several metrics which evaluate the stability and accuracy of different models were defined. These metrics required annual stable RMSE results based on optimized models. A dynamic training procedure was used to calculate BAP and the performance increase of BAP was measured by comparing it to BDP and NP. Purely looking at RMSE results one might assume that TS1 would be the best performing model but after comparing results using the different metrics it became apparent that TS3 and UK3 reduce bias in more different areas compared to TS1.

UK3 scored the highest on average on all metrics but TS3 scored significantly higher in metrics which measure bias in areas that have the highest effect on dispersion model performance. For example, TS3 scored significantly better than UK3 in metrics measuring random error, equivalently in metrics measuring trends and comparably in metrics measuring local error.

Therefor it can be concluded that TS3 is the most effective model to be used for windspeed estimated which has the lowest negative effect to pesticide dispersion estimates.

6. Discussion

6.1 Use of auxiliary variables

Windspeed as meteorological condition is affected greatly by temperature and topology of a surface (see chapter 2.2) therefor it would make sense to include such measurements as auxiliary variables in the models used. This commonly used in kriging models referred to as co-kriging. This, however, is only viable if the data of the auxiliary variables are at a higher spatial resolution than the to be estimated variable (Wan H. et al, 2021).

Therefor temperature measurements could not be included as they would be within the same spatial resolution as windspeed measurements, being that both sources are provided by the same KNMI weather stations. Other data sources for temperature measurements at a higher resolution do exist, but those only cover a small area of the Netherlands not the entire country at large.

Topology as an auxiliary variable does not have this problem, as many sources exists which provide height and depth measurements of the surface in the Netherlands, but these datasets often interpolate measurements themselves for larger areas, as measuring each individual square of land if practically impossible. Another problem is that many datasets covering the entire surface of the Netherlands are generated using lidar sensors underneath an UAV, lasers stations in the air or ground (Laser scanning confocal microscopy (LSCM)) or satellite images. All these techniques bring their own biases and measurements errors which would have to be accounted for. Finally, topology only effects windspeed drastically when difference is substantial (ex: in a mountainous area or hilly area), which is not the case for surface in the Netherlands being primarily flat lands around or slightly below the sea level. Elevation difference are only present for a small strip of land along the border with Germany and Belgium, so the effect topology plays can be considered insignificant.

6.2 Using BDP instead of BAP for metrics

Metrics defined to measure different form of bias in model performance were using annually optimized hyperparameters instead of daily optimized hyper-parameters (see chapter 3.4). In the results it became apparent that models using BDP always had a lower RMSE than models using BAP, so the choice to use BAP over BDP or the effect using BDP instead of BAP for the metrics might yield a different conclusion and outcome. The choice to use BAP over BDP was single done for computational complexity and time constraints. The simulation of the OBOmod model will in the future work on an hourly basis, so some concession had to be made to ensure that windspeed estimated could be generated within an hourly time frame.

The dynamic training routine (see chapter 3.4.1) took between 35 to 40 minutes to execute for all different variations and combination of the models. Even though this is below the 60-minute threshold, it can easily be assumed that the other parts of the OBOmod model require some time to execute.

Therefor the decision was made to calculate BAP which could be used for any random day throughout the year. BAP can easily be calculated for the next year and used for any random day or hour afterwards, essentially only requiring BAP to be calculated during the initialization phase of the simulation. BAP also showed significant RMSE reduction from NP, so including it is advisable.

6.3 Use of more sophisticated models

Existing research using more sophisticated models for windspeed interpolation is available, but such models where not considered due consideration of the simulation using the OBOmod model (see chapter 3.2 and 6.2).

For example, model which use both deterministic and stochastic methods, so-called combined models could be used which provide less bias in clustered measurements (Li J. et al 2014). Another type of model, so called neural kriging could be used for windspeed interpolation which yield in much lower RMSE than OK, UK1 or UK3 models, but such models tend to overfit estimates with spatial sparse measurements (Cellura, M. et al, 2008).

Another model which was considered is Taylor Kriging. Such models were used for windspeed interpolation in several studies. This model is a modification of kriging that uses a Taylor-based linearization approach to handle nonlinear trend functions, resulting in an iterative parameter estimation strategy. One downside of using such a model is that it requires significantly more computational resources and is more complex to implement (Liu H. et al, 2010).

Something to mention is that more complex models do not always yield better performing results or stable estimates as became apparent from the results in the conclusion and existing studies (Reinhardt K. et al, 2018).

A. Appendix

[1] KNMI weather station windspeed data:
https://www.daggegevens.knmi.nl/klimatologie/daggegevens

[2] KNMI weather station location data: **https://github.com/arsalan-anwari/thesis_project_f2/blob/main/Data/knmi_weather_stations.csv**

[3] OBOmod scientific paper:
https://doi.org/10.1016/j.scitotenv.2022.153798

[4] BRP Gewaspercelen 2017:
https://service.pdok.nl/rvo/brpgewaspercelen/atom/v1_0/downloads/brpgewaspercelen_definitief_2017.zip

[5] Amersfoort / RD New CRS information: **https://epsg.io/28992**

[6] Blog post about simplifying features of a map:
https://www.r-bloggers.com/2021/03/simplifying-geospatial-features-in-r-with-sf-and-rmapshaper/

[7] Source code dynamic training routine: **https://github.com/arsalan-anwari/thesis_project_f2/blob/main/dynamic-training-routine.R**

[8] Page to dynamic training routine wiki page:
https://github.com/arsalan-anwari/thesis_project_f2/blob/main/Docs/dynamic-training-routine.md

[9] Paper of spatial interpolation and GIS overview: **https://github.com/arsalan-anwari/thesis_project_f2/blob/main/Docs/Spatial%20statistics%20overview.pdf**

R. References

A

Amato F., Guignard F., Robert S., & Kanevski M. (2020). A novel framework for spatio-temporal prediction of environmental data using deep learning. Scientific Reports, 10(1).
https://doi.org/10.1038/s41598-020-79148-7

B

Bish M., Oseland E., & Bradley K. (2021, June 1). Off-target pesticide movement: A review of our current understanding of drift due to inversions and secondary movement. Weed Technology. Cambridge University Press.
https://doi.org/10.1017/wet.2020.138

Burrough P.A., McDonnell R.A., Lloyd C.D., 2015. Principles of Geographical Information Systems. Oxford University Press.
https://www.researchgate.net/publication/37419765_Principle_of_Geographic_Information_Systems

C

Cellura M., Cirrincione G., Marvuglia A., & Miraoui A. (2008). Wind speed spatial estimation for energy planning in Sicily: A neural kriging application. Renewable Energy, 33(6), 1251–1266.
https://doi.org/10.1016/j.renene.2007.08.013

Cressie N., 2015. Spatial Prediction and Kriging. (2015) (pp. 105–209).
https://doi.org/10.1002/9781119115151.ch3

D

Dennis Elliott, Marc Schwartz, George Scott (2004). Wind Resource Base.
https://doi.org/10.1016/B0-12-176480-X/00335-1

E

Ekhtesasi M.R., Gohari Z. (2012). Determining Area Affected by Dust Storms in Different Wind Speeds, Using Satellite Images.
https://jdesert.ut.ac.ir/article_32035_361cd104e809ec1db56bcb6b311a8321.pdf

F

Figueiredo D. M., Vermeulen R. C. H., Jacobs C., Holterman H.J. van de Zande, J. C. van den Berg, F., … Duyzer, J. (2022). OBOMod - Integrated modelling framework for residents' exposure to pesticides. Science of the Total Environment, 825.
https://doi.org/10.1016/j.scitotenv.2022.153798

G

Gräler B., Rehr M., Gerharz, L. & Pebesma E. (2013). Spatio-temporal analysis and interpolation of PM10 measurements in Europe for 2009 - revised version. ETC/ACM Technical Paper 2011/10 (p. 37).
https://www.researchgate.net/publication/236011362

Guo, B., Yang F., Wu H., Zhang R., Zang W., Wei C., … Zhang H. (2021). How the variations of terrain factors affect the optimal interpolation methods for multiple types of climatic elements? Earth Science Informatics, 14(2), 1021–1032.
https://doi.org/10.1007/s12145-021-00609-2

H

Henderson S.B. (1987). Gaussian Plume Model (pp. 93–97).
https://doi.org/10.1007/978-3-642-82976-5_4

Hosni Snoun, Moez Krichen & Hatem Chérif (2023). A comprehensive review of Gaussian atmospheric dispersion models: current usage and future perspectives.
https://doi.org/10.1007/s41207-023-00354-6

K

Kirkwood C., Economou T., Pugeault N. & Odbert H. (2022). Bayesian Deep Learning for Spatial Interpolation in the Presence of Auxiliary Information. Mathematical Geosciences, 54(3), 507–531.
https://doi.org/10.1007/s11004-021-09988-0

L

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.
https://doi.org/10.1016/j.egyr.2021.11.285

Li J. & Heap A. D. (2014, March). Spatial interpolation methods applied in the environmental sciences: A review. Environmental Modelling and Software.
https://doi.org/10.1016/j.envsoft.2013.12.008

Li Z. (2021). An enhanced dual IDW method for high-quality geospatial interpolation. Scientific Reports, 11(1).
https://doi.org/10.1038/s41598-021-89172-w

Liu D., Zhao Q., Fu D., Guo S., Liu P. & Zeng Y. (2020). Comparison of spatial interpolation methods for the estimation of precipitation patterns at different time scales to improve the accuracy of discharge simulations. Hydrology Research, 51(4), 583–601.
https://doi.org/10.2166/NH.2020.146

Liu H., Shi J. & Erdem E. (2010). Prediction of wind speed time series using modified Taylor Kriging method. Energy, 35(12), 4870–4879.
https://doi.org/10.1016/j.energy.2010.09.001

M

Miao S., Xiong H., Li D., & Gu Y. (2021). A two-phase wind speed simulation model considering diurnal and seasonal patterns and its application to adequacy assessment. Journal of Renewable and Sustainable Energy, 13(5), 1ENG.
https://doi.org/10.1063/5.0059911

P

Pryor S. C., Barthelmie R. J., Bukovsky M. S., Leung L. R. & Sakaguchi K. (2020, December 1). Climate change impacts wind power generation. Nature Reviews Earth and Environment. Springer Nature.
https://doi.org/10.1038/s43017-020-0101-7

R

Rani L., Thapa K., Kanojia N., Sharma N., Singh S., Grewal A. S., … Kaushal J. (2021, February 10). An extensive review on the consequences of chemical pesticides on human health and environment. Journal of Cleaner Production. Elsevier Ltd.
https://doi.org/10.1016/j.jclepro.2020.124657

Reinhardt K. & Samimi C. (2018). Comparison of different wind data interpolation methods for a region with complex terrain in Central Asia. Climate Dynamics, 51(9–10), 3635–3652.
https://doi.org/10.1007/s00382-018-4101-y

S

Schampheleire D.M., Nuyttens D., Dekeyser D., Verboven P., Spanoghe P., Cornelis W., … Steurbaut W. (2009). Deposition of spray drift behind border structures. Crop Protection, 28(12), 1061–1075.
https://doi.org/10.1016/j.cropro.2009.08.006

Semedo A. (2018). Seasonal variability of wind sea and swell waves climate along the Canary Current: The local wind effect.
https://dx.doi.org/10.3390/jmse6010028

Shrestha M.B. & Bhatta G. R. (2018). Selecting appropriate methodological framework for time series data analysis. Journal of Finance and Data Science, 4(2), 71–89.
https://doi.org/10.1016/j.jfds.2017.11.001

W

Wang Y., Wang D., Zhao J. & Zhu C. (2020). Wind speed spatial estimation using geostatistical kriging. In IOP Conference Series: Earth and Environmental Science (Vol. 619). IOP Publishing Ltd.
https://doi.org/10.1088/1755-1315/619/1/012049

Wan H., Li J., Shang S. & Rahman K. U. (2021). Exploratory factor analysis-based co-kriging method for spatial interpolation of multi-layered soil particle-size fractions and texture. Journal of Soils and Sediments, 21(12), 3868–3887.
https://doi.org/10.1007/s11368-021-03044-4

Wever N. (2012). Quantifying trends in surface roughness and the effect on surface wind speed observations. Journal of Geophysical Research Atmospheres, 117(11).
https://doi.org/10.1029/2011JD017118

Z

Zaller J.G., Kruse-Plaß M., Schlechtriemen U., Gruber E., Peer M., Nadeem I., … Landler L. (2022). Pesticides in ambient air, influenced by surrounding land use and weather, pose a potential threat to biodiversity and humans. Science of the Total Environment, 838.
https://doi.org/10.1016/j.scitotenv.2022.156012

Zhang H., Zheng J., Zhou H. & Dorr G.J. (2017). Droplet Deposition Distribution and Off-target Drift during Pesticide Spraying Operation. Nongye Jixie Xuebao/Transactions of the Chinese Society for Agricultural Machinery, 48(8), 114–122.
https://doi.org/10.6041/j.issn.1000-1298.2017.08.012

Zhan S. & Savani K. (2022). Relative Insensitivity to Sample Sizes in Judgments of Frequency Distributions: People are Similarly Confident in the Results From 30 Versus 3,000 Observations. Decision, 10(1), 61–80.
https://doi.org/10.1037/dec0000182