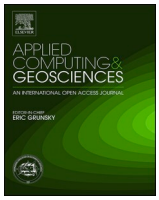
Applied Computing and Geosciences 27 (2025) 100280

Contents lists available at ScienceDirect

Applied Computing and Geosciences

journal homepage: www.sciencedirect.com/journal/applied-computing-and-geosciences



Assessing paleo channel probability for offshore wind farm ground modeling - comparison of multiple-point statistics and sequential indicator simulation

Lennart Siemann *, Ramiro Relanez

Fraunhofer Institute for Wind Energy Systems IWES, Am Fallturm 1, D-28359, Bremen, Germany

A B S T R A C T

The presented study investigates the prediction of buried paleo-channels for probabilistic ground modeling of offshore windfarm development areas using geostatistical methods. These channels, common in glaciogenic regions like the North Sea, can pose significant geohazards affecting turbine foundation stability. Conventional 2D seismic data interpretation provides the best estimate of the position but lacks probabilistic assessment, specifically at unexplored locations. Multiple-point statistics (MPS) and sequential indicator simulation (SIS) are applied to quantify the probability of channel features from seismic data, away from seismic lines. MPS utilizes training images to capture complex spatial structures, while SIS relies on variogram models for modeling spatial variability. Results demonstrate that denser seismic line spacing (150 m) yields higher accuracy compared to wider spacings (300 m and 600 m), underscoring the importance of data density in offshore subsurface site characterization. Additionally, the findings indicate that MPS provides lower errors, making it preferable for precise channel location prediction. The selected training image did not have a major impact on the outcome on the tested data. Conversely, SIS offers broader coverage of potential channel locations, which may be advantageous for further de-risking. This research contributes to more informed ground modeling by incorporating probabilistic approaches. Therefore, it supports in offshore wind farm site development by enhancing knowledge of the subsurface at an early stage of wind farm development to aid decisions in windfarm and further site investigation planning.

1. Introduction

In the planning of offshore wind farms and their turbine layout, ground modeling plays a crucial role as part of the subsurface site characterization of potential development areas. In addition to other important factors, such as wind resources and the restriction due to infrastructure or planning of adjacent areas, ground models contribute to determining the location of wind turbines, as well as their foundation type and depth. Despite its significance, probabilities in ground modeling are often not fully addressed. Probabilistic evaluation is essential for risk assessment and mitigation in the later installation process, ensuring long-term stability of offshore infrastructure.

Geospatial variability, both lateral and vertical, is a major geo-engineering constraint for offshore wind development, and accurately capturing this variability is essential. Among various geological features, buried channels are especially of interest because they can pose geohazards due to their often heterogeneous architecture (Coughlan et al., 2018; Velenturf et al., 2021). In the North Sea, such features are frequently encountered in areas influenced by glaciogenic processes, resulting in complex sedimentary structures due to repeated transgression and regression events, as well as dynamics of ice sheets over time (Schwarzer et al., 2008; Moreau et al., 2012). Identifying and accurately locating these geohazards is important, as they can adversely affect the stability and safety of wind turbine foundations since they are typically encountered within the installation depth of pile foundations (Coughlan et al., 2018; Petrie et al., 2022).

Offshore wind farm ground modeling commonly relies on 2D seismic data acquisition and interpretation to characterize the development area. Acquired 2D seismic lines typically follow a regular cartesian grid, but the inter-line spacing can differ based on requirements imposed by development stage, regulatory bodies or involved companies. Geological horizons are identified along the 2D seismic lines and interpolated onto a three-dimensional grid. However, conventional interpolation techniques only operate locally and remain agnostic of dataset-spanning correlations. In addition, these methods typically lack the possibility to estimate prediction uncertainties which are of particular importance for larger line separations. To effectively mitigate risks associated with channel features, it is necessary to properly quantify the likelihood to encounter those features away from measured locations.

The aim of this study is to quantify and evaluate the probabilities of channel features derived from the interpretation of seismic data in offshore wind farm developments by applying geostatistical techniques.

* Corresponding author.
E-mail address: lennart.siemann@iwes.fraunhofer.de (L. Siemann).

https://doi.org/10.1016/j.acags.2025.100280
Received 23 April 2025; Received in revised form 8 August 2025; Accepted 8 August 2025
Available online 9 August 2025

The focus is on 2D data, as it is currently the most common method of seismic data acquisition in offshore wind, although the authors acknowledge the increasing relevance of 3D data acquisition (Caselitz et al., 2025), where this type of challenge does not apply due to their area-wide coverage.

Conventional geostatistical interpolation techniques, such as Kriging, are primarily designed to provide the best linear unbiased estimates of spatial variables. However, they generally focus on deterministic predictions and do not explicitly quantify uncertainty or provide probabilistic assessments of spatial variability. This study, therefore, focuses on geostatistical simulation, which generates multiple realizations used to quantify the probability of specific geological features. Several methods exist for the simulation of categorical variables (Chiles and Delfiner, 2012), which are applied to predict facies or geological units. This research considers multiple-point statistics (MPS) and sequential indicator simulation (SIS). A comparative analysis of the application of both methods is available in the literature, illustrating different fields of application (Bastante et al., 2008; De Iaco and Maggio, 2011; Zhou et al., 2018). MPS employs training images (TIs) to capture complex spatial relationships and patterns found in geological structures. By generating multiple realizations based on these TIs, MPS allows us to derive probabilities of geological features, providing insights into where paleo-channels are likely to exist beyond the constraints of directly measured seismic lines. This approach enables to quantify uncertainty by comparing different realizations and assessing the variability in predictions. SIS, on the other hand, utilizes variogram models to simulate spatially correlated categorical variables. It generates multiple realizations that maintain the spatial dependencies observed in the conditional data (CD) obtained from seismic lines. Each realization represents a possible distribution of channel features, allowing us to assess uncertainty in predictions by analyzing the range of outcomes across simulations.

This study evaluates the applicability of MPS and SIS on offshore windfarm data to assess channel probability away from measured 2D seismic lines, considering variations in line spacing and different training images from diverse geographical areas.

2. Methodology

2.1. Multiple-point statistics by direct sampling

Modeling spatial heterogeneity is a fundamental challenge in geostatistics. Traditional geostatistical methods, such as variogram-based kriging techniques, often fail to capture the complexity of geological structures due to their reliance on two-point statistics (Strebelle, 2002). In contrast, MPS methods offer a more robust alternative by using training images (TIs) to reproduce complex spatial patterns (Mariethoz and Caers, 2014; Meerschman et al., 2013). In MPS statistical relationships from TIs are extracted, which serve as a reference for the expected spatial structures of the target domain. These approaches enable the generation of stochastic realizations that honor both the spatial connectivity observed in the training images and any available conditioning data (Straubhaar et al., 2016).

In this study the DeeSse algorithm as implemented in the Software Isatis.neo™ (Geovariances et al., 2024) was used. It is based on the direct sampling (DS) method, originally proposed by Marietholz et al. (Mariethoz et al., 2010). It differs from other MPS approaches by avoiding the pre-computation of pattern databases. Instead of storing all data events from the TI, DS directly scans the TI during the simulation process. DS is applicable to both categorical and continuous variables, as well as to multivariate simulations (Mariethoz et al., 2010).

In the DS framework, a simulation grid is populated sequentially by assigning values based on similarity measures between the local data and patterns within the TI (Fig. 1). The similarity is determined using a distance function, typically defined as the proportion of mismatching nodes for categorical data or the mean absolute error for continuous data (Straubhaar et al., 2020). Besides the TIs and the conditional data (CD) itself, key parameters controlling DS include the number of nearest neighbors (N), the distance threshold (t) for pattern matching, and the scan fraction (f) of the TI to be scanned before assigning a value (Meerschman et al., 2013). By calculating multiple stochastic realizations, probabilities can be derived from them and further used in risk assessment.

2.1.1. Training images and conditioning data

A TI serves as a spatial reference that captures expected patterns of subsurface features, providing a conceptual model that guides the MPS algorithms in generating plausible distributions of geological structures. In scenarios with limited input data, TIs help reconstruct subsurface structures beyond directly observed areas (Zhang et al., 2007). TIs can be derived from various data sources such as conceptual geological models, outcrop analogs, or geophysical datasets (Pyrcz et al., 2008). In this study, the TIs are based on conceptual images manually interpreted from 2D seismic data. Paleo-channel boundaries were first identified along seismic sections and manually digitized at the level of interpreted lithological horizons, considering the relative elevation, slope, and morphological characteristics typical of fluvial systems. These were then interpolated laterally between seismic lines, to generate a continuous 2D

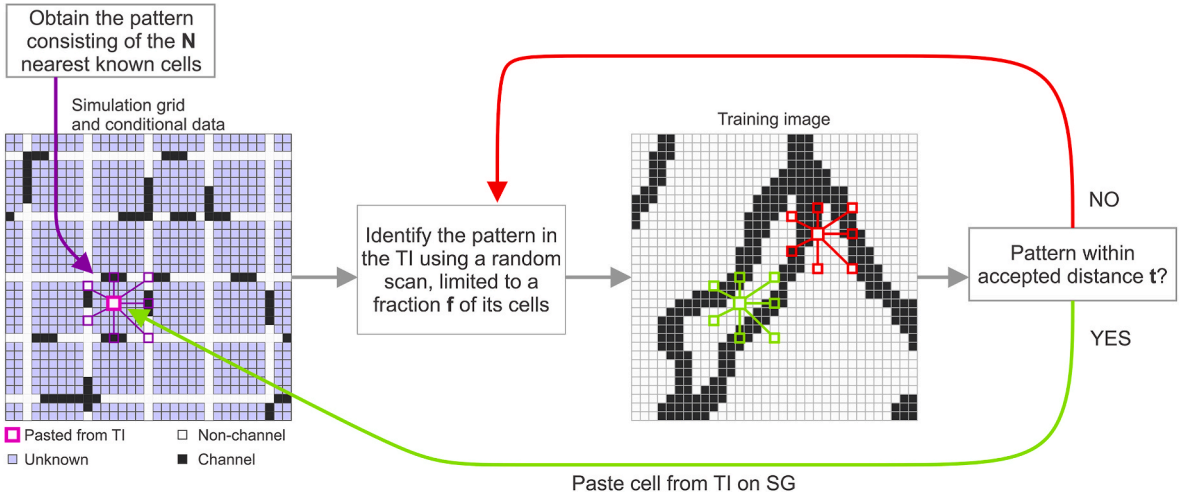


Fig. 1. Conceptual illustration of the general process of MPS by direct sampling using the DeeSse algorithm. (adapted after Straubhaar et al. (Straubhaar et al., 2020)).

surface. Each image was constructed over a predefined 5 km × 5 km area. The availability of such a geologically informed 2D surface serves as ground truth against which the performance of the simulation methods can be objectively evaluated. The image is then converted into a grid with a resolution of 10 m, containing categorical data (channel, non-channel), resulting in a 500 × 500 cell matrix. This grid defines both the training image and the simulation grid domain. From these gridded images, conditional data sets were generated by extracting categorical values along synthetic seismic survey lines that mimic actual acquisition geometries. These profiles were constructed with regular spacings of 150 m, 300 m, and 600 m, as well as in irregular configurations (see Supplementary Fig. S1). Fig. 2 shows an overview of the different images used in this study. The examples were taken from true windfarm development areas in the German North Sea, over Holocene and Upper Pleistocene deposits. However, due to confidentiality constraints and because the areas themselves as well as their geological background, are not the focus of the study, exact locations and detailed geological descriptions are not provided. Each image has a unique set of channel structures, with varying channel density, orientation and width. All images were selected from different locations but from the same regional geological context to investigate whether TI and CD sets that are distant from each other can still be used. These images serve as TIs and as the source of the conditional data for the different line spacings.

2.1.2. Direct sampling parameterization

The parameterization of MPS can be quite complex, as the outcome depends on several parameters. In Meerschman et al. (2013) and Juda et al. (2022), a description of the parameters and their effect on the prediction, along with suggestions for optimizing the outcomes can be found. After performing hyperparameter tuning through a sensitivity analysis of the main search pattern parameters, it was decided for this study to set the accepted distance threshold t to 0.02, the scanned fraction f to 50 percent and the number of cells for the nearest neighbor N to 50.

The number of realizations influences the uncertainty analysis. Montero et al. (2021) suggest performing at least 20 realizations, while Zhang et al. (2024) increased this number to 200. In this study, the number of realizations was determined based on a convergence criterion using the average absolute difference of probabilities (AADP). Stabilization was defined as a 95 % reduction in the slope of AADP over a sliding window, which results in a total amount of 200 realizations to be used. Full details are provided in supplementary material section S1.

2.2. Sequential indicator simulation

Sequential indicator simulation (SIS) is a geostatistical technique employed to model and simulate spatially correlated variables. This method is broadly applied in various domains, including mining, hydrology, and environmental science, particularly for resource estimation and risk assessment (Zhou et al., 2018; de Souza et al., 2013; Medina-Ortega et al., 2019; Madani, 2022). Simulations can be conditional, meaning that they are constrained to known data; in this case the information along the 2D seismic lines.

To model spatial categorical variables, continuous data are transformed into binary indicators, as in MPS. An essential aspect of SIS is the analysis of spatial correlation. Variograms are used to quantify the

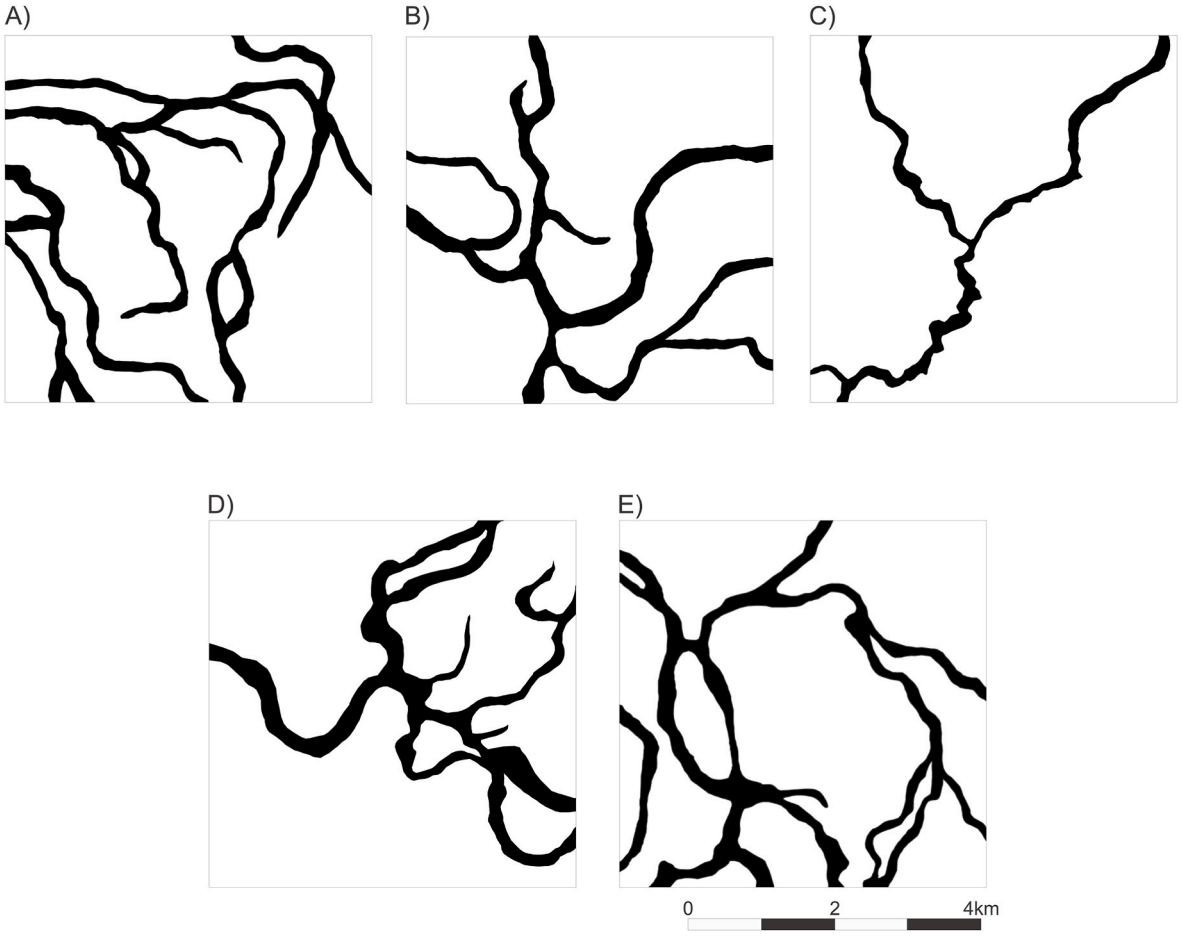


Fig. 2. Overview of the different images used for the creation of TIs and CD, with black highlighting channel locations. All images are based on interpretation from true offshore windfarm development areas and are taken from different regions, with the same overall geological context. All images cover an area of 25 km^2.

spatial structure of the indicators, providing insight into their relationships over distance. In SIS, multiple realizations of the indicators are generated sequentially, ensuring that the spatial correlation and dependency between indicators are preserved throughout the simulation process. Each simulation step depends on the outcome of the previously simulated data points in addition to the CD. By creating multiple realizations, SIS allows for assessing uncertainty in spatial data. By generating a range of possible scenarios for the distribution of variables, SIS enhances decision-making processes in resource management and environmental assessment ((Zhou et al., 2018) and references therein). More detailed information on the methodology is given by Chiles and Delfiner (2012).

In contrast to MPS, SIS simulates properties based on variogram models, making it independent of the selection of TIs. This can be advantageous when the overall structure of the outcome is unknown, or when no suitable TI is available. For each image presented in Fig. 2, SIS is run for different line spacings, as described previously. As with MPS, the AADP has been calculated to estimate the number of necessary realizations for a nearly steady-state condition. The results can be seen in supplementary material section S1 and indicate a necessary number of realizations of 200.

2.2.1. Variogram models for SIS

The variogram is a key component of geostatistical analysis, serving as a quantitative measure of spatial continuity between data points. It provides insights into how variables change over space, thereby allowing for the modeling of spatial relationships and making predictions at unsampled locations (Pyrcz and Deutsch, 2014).

The variogram is typically plotted as a function of the lag distance (h), to characterize spatial relationships. The variogram model can be described by three components: the nugget effect, sill and range. The nugget effect represents the variogram at h = 0 and reflects the variability of the data at a very small scale. It implies measurement errors or inherent micro-scale variability (Oliver and Webster, 2015). The sill is the value at which the variogram levels off, indicating that the spatial correlation has reached its maximum and therefore maximum variability, represented by the overall variance of the dataset. The range is the distance at which the variogram reaches the sill, suggesting the maximum distance of spatial correlation. Beyond this distance, the observations are considered spatially uncorrelated. A more rigorous description of the general concepts is provided by Chiles and Delfiner (2012) and Oliver and Webster (2015).

The variograms and the representative variogram model parameter for the different line spacings can be seen in Fig. 3 and Table 1, exemplary for image E. While the range and sill vary depending on the line spacing, the nugget is set to zero for all models, assuming no variability at zero distance. Two structure components are needed for a suitable fit of the data. The variograms and model parameters for the other images are found in the Supplementary Material Fig. S2 and Table S1.

Table 1
Variogram model parameters for the variograms shown in Fig. 3.

Spacing [m]	Structure type	Ranges [m]	Sill
150	Cubic	377	0.0335
	Exponential	425	0.1063
300	Cubic	388	0.0367
	Exponential	409	0.1041
600	Exponential	162	0.0408
	Cubic	415	0.0876

2.3. Assessment of output quality

To assess the quality of the output, the Mean Squared Error (MSE) is calculated for each set of TI and CD configurations. To facilitate a more direct comparison across different cases, the MSE values are normalized to a scale between zero and one, as described in Equation (1). This normalization is performed using the global minimum and maximum rather than those within each individual conditional dataset to allow for a standardized representation.

$$MSE_{normalized} = \frac{MSE - MSE_{min}}{MSE_{max} - MSE_{min}} \quad (1)$$

where MSE_{min} and MSE_{max} correspond to the minimum and maximum MSE values observed across all scenarios.

3. Results

The simulation results of both algorithms are derived from multiple stochastic realizations, in which successive possible scenarios are generated. Fig. 4 shows the results of different randomly picked realizations for MPS (above) and SIS (below). This comparison highlights the main differences between the outputs of the two approaches and their visual characteristics. MPS exhibits more continuous channel structures with a smooth appearance, while SIS appears patchy and discontinuous. Although the examination of individual realizations can aid interpretation, for the later results, only probabilities are considered. Channel probabilities are derived from counting all categorical variable 'channel' at each individual grid cell, divided by the total number of realizations.

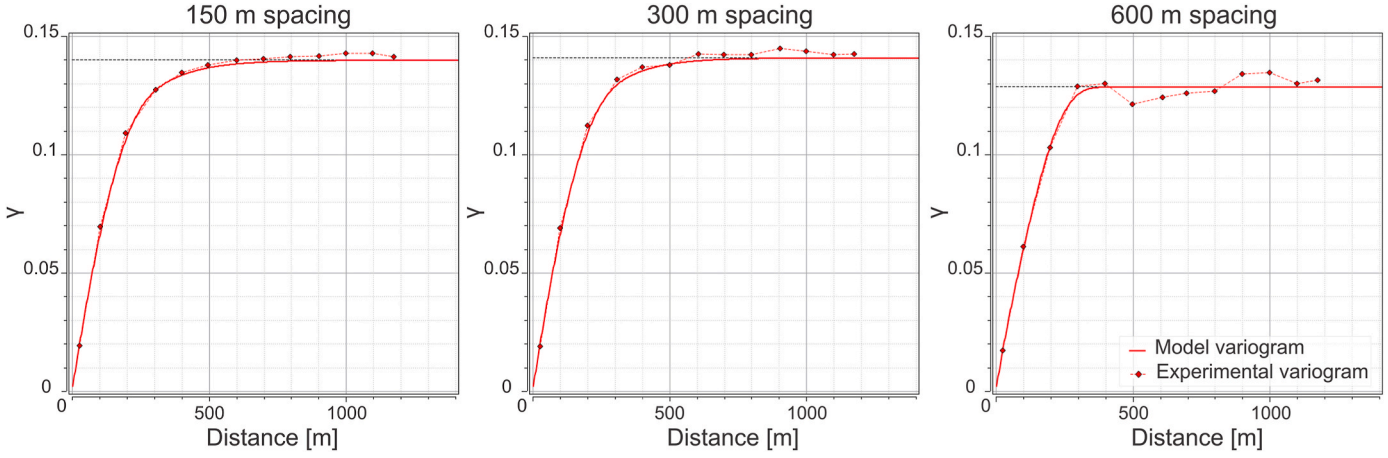


Fig. 3. Experimental variogram (dotted red line) and fitted variogram model (solid red line) used for the SIS calculation for the different line spacings on image E. The horizontal black dotted line indicates the sill. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

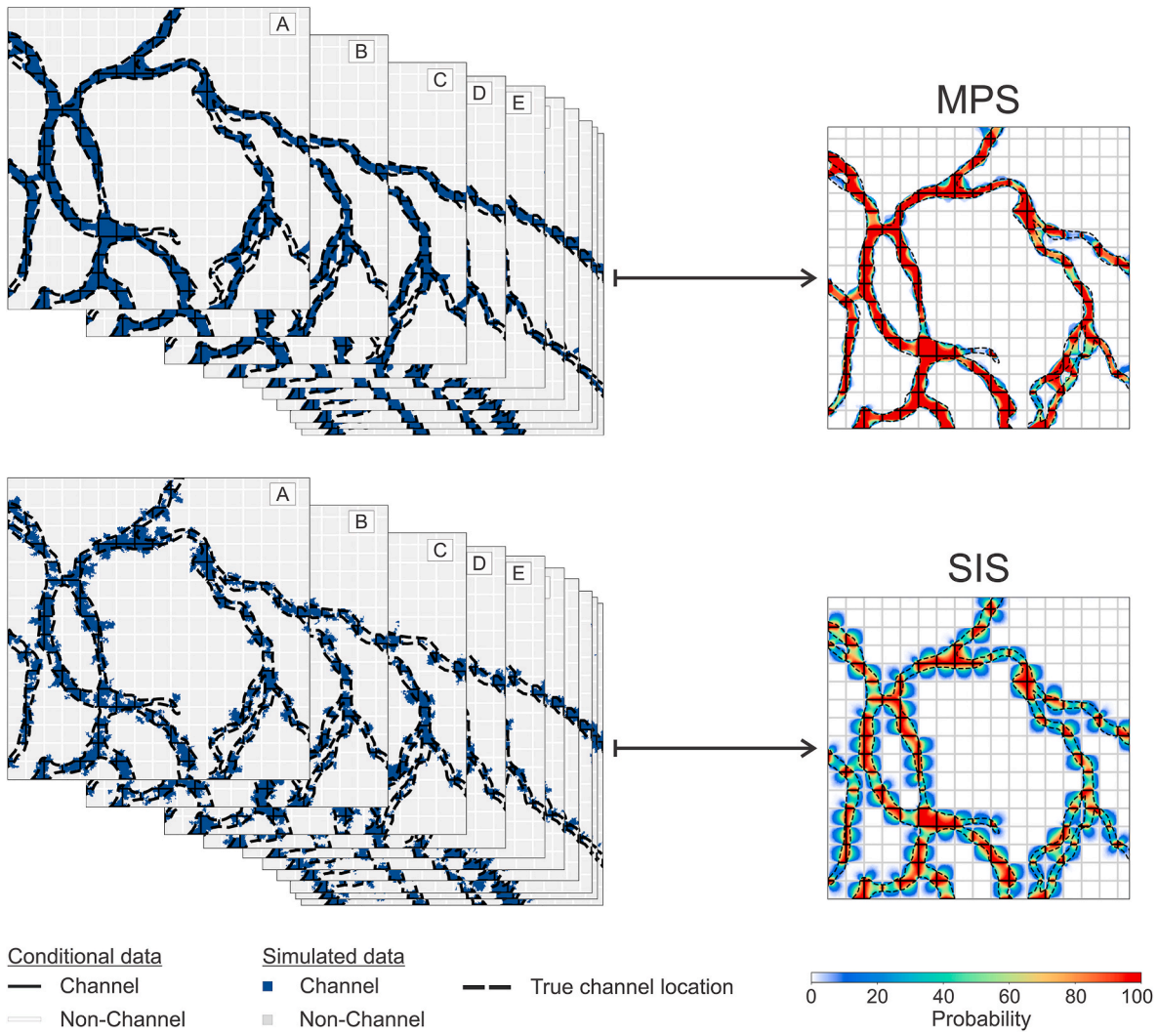


Fig. 4. Random realizations (A, B, C, D …) for MPS (top) and SIS (bottom), shown progressively overlapped to illustrate how multiple stochastic scenarios are stacked to derive the probability map. Blue indicates individual channel realizations; red and white in the probability map indicate high and low likelihood of channel presence, respectively. Black and white lines correspond to channel and non-channel conditioning data, while the true channel location is marked with a black dashed line. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.1. Multiple-point statistics

According to the previously described data, a matrix of different configurations of TI and CD can be calculated using MPS. Both TI and CD originate from the exact same area, as representatively shown in Fig. 5. Here, the result of different line spacings on the channel probability is shown using TI and CD both from image E. The results closely match the known channel location with a high probability. The overall outcome becomes less certain with increasing line spacing; however, it still indicates that the method is working properly. The MSE increases with line spacing but remains generally very low, for all configurations, where TI and CD are from the same area.

Table 2 presents the MSE results for all possible TI and CD sets, showing an overview of the possible outcomes color coded from low (green) to high (red). When TI and CD are from the same area, the MSE is lowest for the given line spacing. However, the MSE is not zero, indicating deviations in the prediction. It is evident as expected that the lowest MSE occurs for the narrowest line spacing, while the wider line spacing leads to increasing MSE due to less certain predictions caused by the sparser CD.

On average, considering all findings, there is no significant difference in the performance of the various TIs, despite their largely diverting shapes and patterns. This suggests that the calculation is not very sensitive to the selection of the TI for this particular data set. Therefore, it can be attested that for MPS knowledge transfers well to unseen datasets, which can support windfarm ground modeling in newly acquired areas based on prior knowledge derived from other areas.

Fig. 6 shows the channel probability estimations for image E using the different TIs (top to bottom) to test the various line spacings (left to right). Areas with high probabilities (red color) indicate that channels are likely to be present, whereas areas with low probabilities (purple color) suggest the presence of flood plains. The positions of the seismic lines are marked white, indicating that no channel has been interpreted on the 2D line, whereas black indicates a channel. For reference, the true channel position is marked by a dotted line in all sub-figures. The other TIs have also been tested for the remaining data sets, however, since there are no significant changes in the overall findings, the additional data is presented in the Supplementary Material Fig. S3–S6 and is not further described.

The effect of varying line spacing on the reconstruction of channel networks is evident in the probability maps presented in Fig. 6. An increase in line spacing leads to a decrease in probability matching with the true channel locations. At a spacing of 150 m, the conditional data provides sufficient constraints, allowing for a high degree of accuracy in

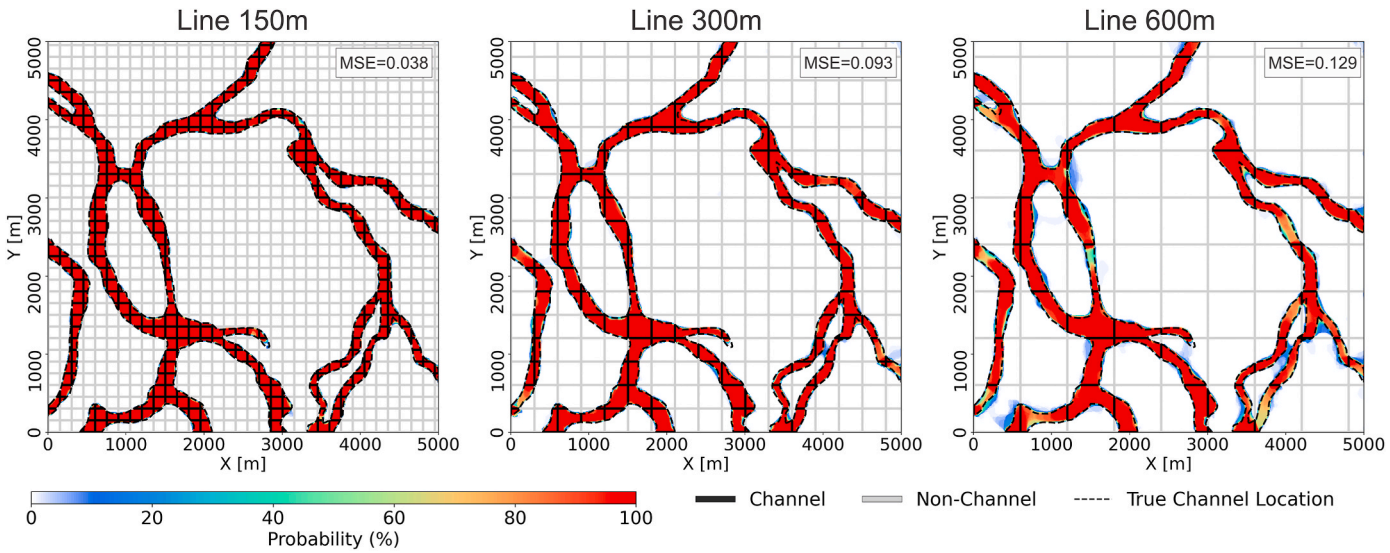


Fig. 5. Example of the channel probability result for the case when TI and CD are from the same image (compare E in Fig. 2), with increasing line spacing from left to right. The conditional data is defined as channel (black lines) or non-channel (white lines). The true channel location from the original data is indicated as black dotted lines. Red indicates high probabilities of channel locations, white indicates low probabilities of channel locations. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Overview of the normalized MSE for the different line spacings and TIs used for the CD that entered MPS.

Spacing [m]	Training image	Reference original data				
		CD-A	CD-B	CD-C	CD-D	CD-E
150	TI-A	0.051	0.031	0.016	0.048	0.058
	TI-B	0.090	0.020	0.020	0.051	0.073
	TI-C	0.100	0.053	0.000	0.060	0.086
	TI-D	0.105	0.044	0.018	0.018	0.081
	TI-E	0.083	0.032	0.021	0.046	0.038
300	TI-A	0.106	0.171	0.100	0.223	0.239
	TI-B	0.253	0.074	0.102	0.209	0.266
	TI-C	0.319	0.217	0.015	0.225	0.282
	TI-D	0.367	0.232	0.122	0.051	0.334
	TI-E	0.280	0.198	0.103	0.251	0.093
600	TI-A	0.164	0.478	0.297	0.622	0.732
	TI-B	0.824	0.161	0.272	0.528	0.687
	TI-C	0.846	0.483	0.048	0.510	0.707
	TI-D	1.000	0.579	0.331	0.073	0.712
	TI-E	0.854	0.501	0.326	0.608	0.129

channel reconstruction. Nevertheless, as the spacing increases to 300 m and 600 m, the probability of correctly identifying channels declines significantly, as indicated by the increasing MSE values across all considered TIs (Table 2).

At 150 m spacing, challenges arise when reconstructing thin channel structures. Even when the TI contains similarly thin channels, the reconstruction struggles due to a lack of sufficient matching points in the TI. Additionally, orientation mismatches become apparent, where the inferred channels deviate from their expected trajectories. This issue underscores the importance of incorporating orientation-sensitive constraints or more refined training datasets to improve accuracy in regions with narrow or intricate channel formations.

Channel continuity is strongly affected by the availability and density of conditional data. When data points are spaced further apart, the inferred spatial coherence of channels weakens. This is particularly noticeable at 600 m spacing, where previously well-connected channel networks begin to fragment, and some sections incorrectly connect, or new artificial channels emerge. In some cases, the model introduces artificial stretching effects, elongating channels in directions dictated by the available conditioning data rather than following the true channel

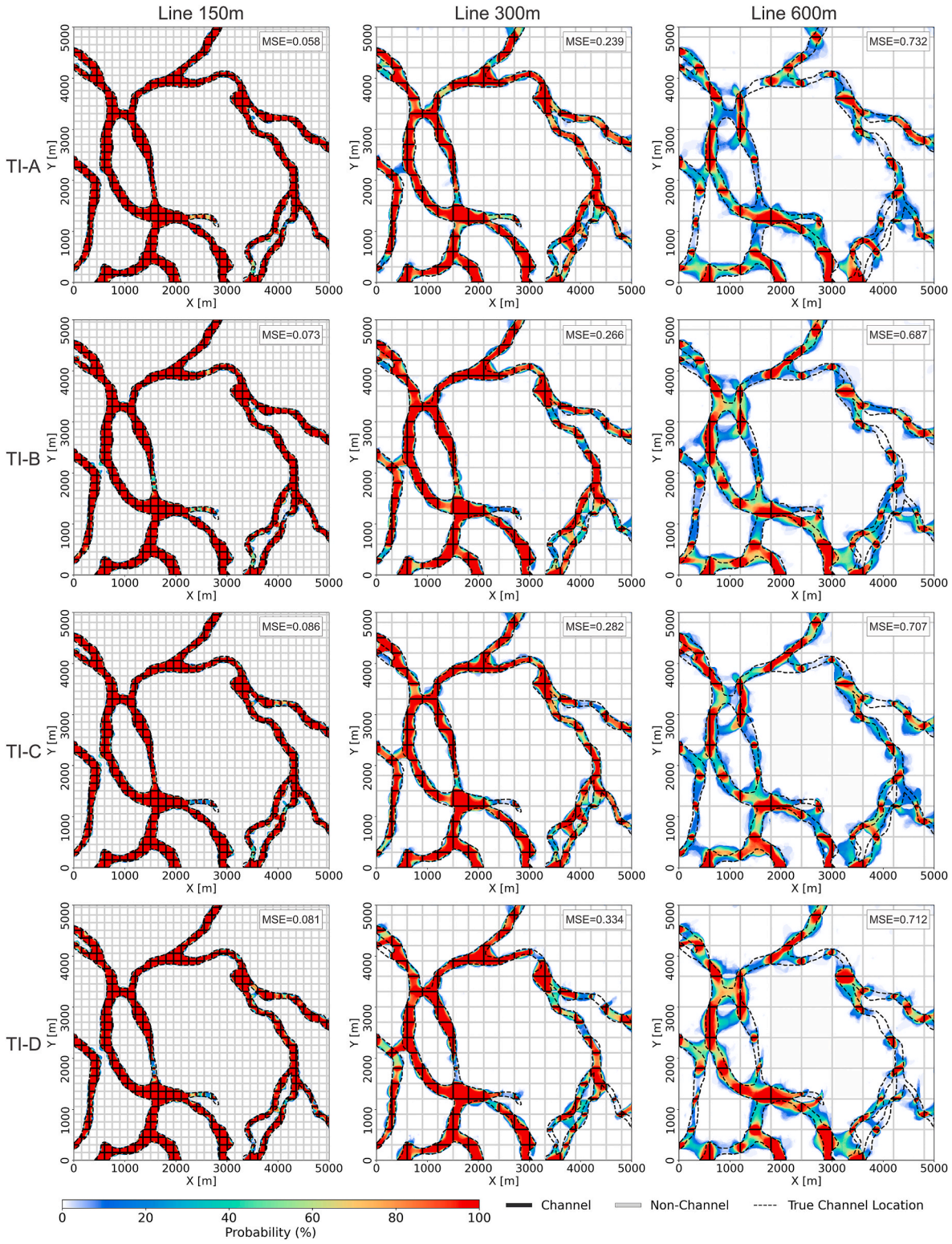


Fig. 6. Summarized results of the channel probability using MPS for the different TIs (A–D), where example image E served as CD. Fig. 2 serves as a reference for the different TIs. The TI name is given on the left. From left to right the results of the different line spacings are given (150 m, 300 m, 600 m). The MSE as per Table 2 is provided in the corresponding sub-figures as a cumulative measure for prediction accuracy. Channel (black) and non-channel (white) locations are indicated along the seismic line positions. For comparison, the true channel location is indicated by the black dotted lines. Red indicates high probabilities of channel locations, white indicates low probabilities of channel locations. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

structure. These misinterpretations highlight the limitations of sparse data in reliably capturing complex geological patterns.

Across different TIs, no major visual differences in performance are observed on a broad scale. However, finer-scale differences exist, indicating that the choice of TI influences specific reconstruction details. As seen in Fig. 6, performance varies locally, where one TI provides better results in one region, and another TI may perform better in a different region. This suggests that while the overall performance is comparable among TIs, specific characteristics of each TI may introduce subtle variations in the inferred structures.

3.2. Sequential indicator simulation

Since SIS is not based on TIs for sequential calculation, but on variogram models for each conditional data case, only one possible outcome is expected for each CD and line spacing, reducing the overall possible number of results. Only one variogram model is considered for each scenario, although the uncertainty of the variogram model and its effect on the results have not been investigated. Table 3 shows the results of the normalized MSE for the different line spacings and CD sets.

The overall MSE for SIS is higher compared to MPS indicating better numerical performance for the latter. Nevertheless, like MPS, an increasing MSE with increasing line spacing is observable for SIS, which is explained as well by the decreasing amount of CD, making the prediction less reliable. Particularly for 150 m line spacing, SIS has a significantly higher MSE than MPS. For 300 m, this discrepancy reduces but remains high. At 600 m, where the overall simulated results exhibit a less structured and more dispersed shape, the MSE shows good correspondence.

The difference in the MSE between SIS (Table 3) and MPS (Table 2) already suggests a variation in the performance of the two methods. Here, the TI based approach outperforms the variogram based method underlining the importance of selecting a suitable predictor.

Fig. 7 representatively shows the channel probabilities of the SIS results for the different line spacings and the example case of image E (Fig. 2). The general findings from the MPS results are supported by SIS. As the line spacing increases from 150 m to 600 m, the ability to accurately predict channel positions decreases, as reflected in the rising MSE values. The results for the other images can be found in Supplementary Material Fig. S7.

At a line spacing of 150 m, the probability maps indicate a high degree of agreement with the true channel locations. The channel structures are well captured, with minor deviations from the reference patterns. However, compared to the MPS results, the MSE at this spacing is higher. At 300 m spacing, a clear degradation in channel correlation is observed. The channels remain largely visible, but there is an increased presence of misclassified areas, where channel probabilities become more diffuse. At 600 m spacing, the connectivity of the channels breaks down significantly. While general channel structures remain somewhat discernible, there is considerable fragmentation and an increase in artificial channel connections. This further supports the conclusion that wider spacing leads to a significant loss of structural integrity in the inferred channels, likely due to the stretching effect and the lack of sufficient conditioning data to maintain accurate fit.

The direct comparison of the probability maps of MPS and SIS shows that the probability distribution of SIS is broader compared to MPS (Figs. 6 and 7, respectively). Additionally, probabilities are lower for SIS, while still covering the channel structures. The lower MSE for SIS can be explained by these observations (Table 3).

For a more detailed comparison of the different results of MPS and SIS and to support previous findings, Fig. 8 shows a close-up of the results for two example areas. One can observe that MPS exhibits a tendency to thin out certain channel segments, leading to incomplete coverage of the true channel extent. This is particularly critical as it may result in areas where channels exist but are not identified in the model. On the other hand, SIS shows a broader probability distribution around the true channels, which is also reflected in its higher MSE values (Table 3). SIS conservative appearance might be advantageous from an engineering perspective, as channels tend to have a "buffer zone" around them. This could help to avoid overestimating the extent of non-channel areas. In contrast, MPS offers a more constrained and definitive channel prediction, leading to greater certainty in channel placement but with a higher risk of underestimating channel extent.

3.3. Discussion

The presented results highlight the main advantages and disadvantages of the different techniques to be used in paleo-channel probability assessment for offshore windfarm ground modeling. MPS is pattern-based and heavily dependent on the chosen TI as input image. Nevertheless, it has been demonstrated that for this particular data set, for narrow line spacings MPS is not very sensitive to the selection of TI if plausible channel patterns are used. Known channel structures from adjacent windfarm areas can therefore be used, as they are based on a similar geological background. Nonetheless, it is necessary to understand the expected structures.

SIS is completely independent of TIs and only requires variograms for the definition of spatial distribution. This approach becomes particularly robust when a significant amount of conditional data is available which is narrowly distributed but becomes more challenging with increasing spacing. This is confirmed by Marietholz et al. (Mariethoz et al., 2010), who state that when a large amount of conditional data is available, it is possible to neglect a TI and rely purely on data driven approaches. Here, machine learning (ML) algorithms can also come into play, particularly convolutional neural networks (CNNs), which have shown potential in learning spatial patterns from labeled training data. In the context of 2D geostatistical simulation, CNNs could be trained to recognize and reproduce channel-like morphologies from sparse conditioning inputs, acting similarly to MPS in capturing complex spatial dependencies. CNNs have already been applied successfully in geoscientific problems such as fault detection in seismic images (Wu et al., 2018)). Studies by Bastante et al. (2008), De Iaco and Maggio (De Iaco and Maggio, 2011) and Zhou et al. (2018), have compared SIS and MPS, among others. While Bastante et al. (2008) and De Iaco and Maggio (De Iaco and Maggio, 2011) did not specially focus on channel prediction, they found that MPS is better in reproducing complex

Table 3
Overview of the normalized MSE for the different line spacings and TIs used for the CD for the SIS calculation. As before green and red colors indicate low and high MSE values, respectively.

Spacing [m]	CD-A	CD-B	CD-C	CD-D	CD-E
150	0.220	0.138	0.064	0.137	0.190
300	0.510	0.348	0.167	0.332	0.449
600	0.862	0.570	0.322	0.591	0.753

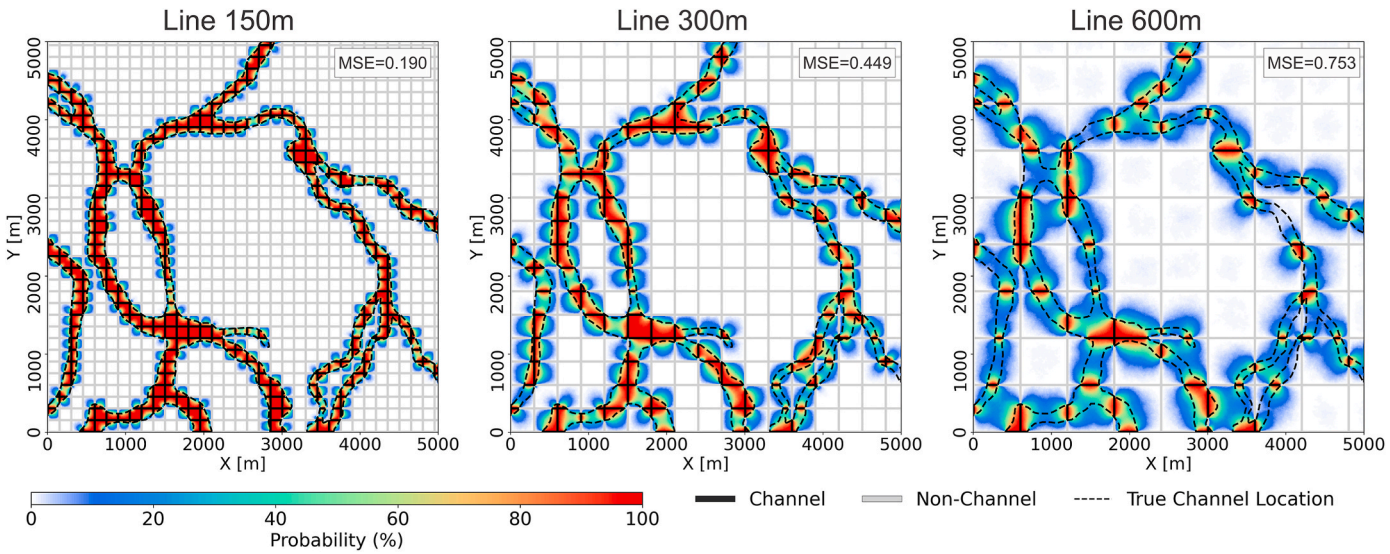


Fig. 7. Summarized results of the channel probability using SIS, tested for CD-E. From left to right the results of the different line spacings are given (150 m, 300 m, 600 m). The MSE as per Table 3 is provided in the corresponding sub-figures for a quantitative measure. Channel (black) and non-channel (white) locations are indicated along the seismic line positions. For comparison, the true channel location is indicated by the black dotted lines. Red indicates high probabilities of channel locations, white indicates low probabilities of channel locations. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

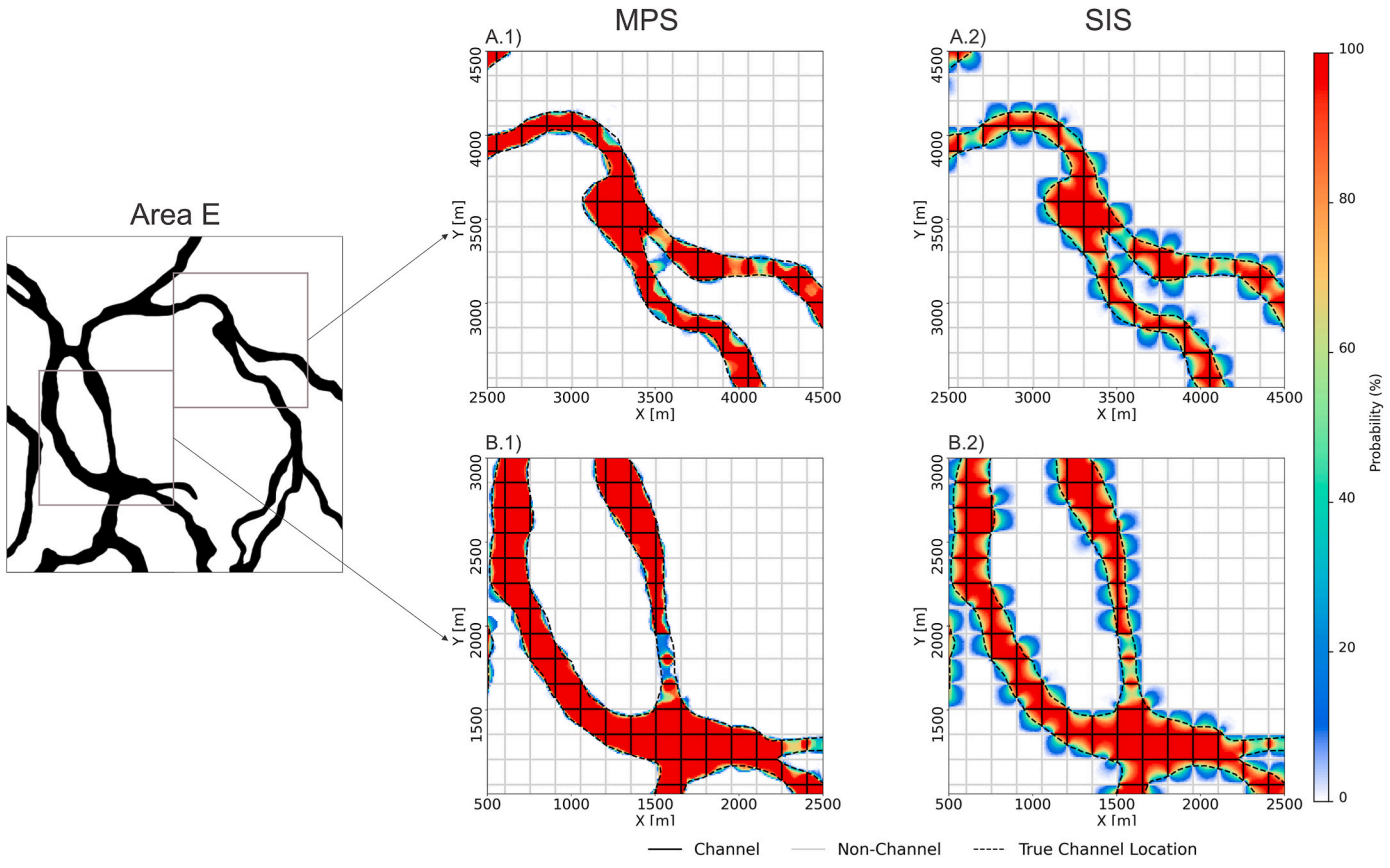


Fig. 8. Close-up of two examples showing details of the probability estimation comparing MPS (left) and SIS (right) results for a line spacing of 150 m. The figure highlights the major differences in the appearance of the probability estimates of the two tested methods. Channel (black) and non-channel (white) locations are indicated along the seismic line positions. For comparison, the true channel location is indicated by the black dotted lines. Red indicates high probabilities of channel locations, white indicates low probabilities of channel locations. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

patterns and their connectivity, which is supported by the findings of this study. Zhou et al. (2018) had similar findings while testing the methods particularly on channel features. They have also tested different conditional data densities based on aerial photos of a river, but the scale of several kilometers between the CD is not directly comparable to the study presented here.

Ultimately, the choice between the two methods depends on the acceptable level of risk for the end-user. If a conservative design approach is required, MPS may be preferred due to its lower overall error. However, if ensuring maximum channel coverage is more critical SIS can the better option. The close-ups in Fig. 8 further illustrate this trade-off, emphasizing the importance of defining an acceptable probability threshold based on risk tolerance and project-specific requirements. Nevertheless, both techniques and their findings support in the assessment and development of offshore wind farm ground models and can assist in decision making for further site investigation campaigns and wind farm planning. While they are not meant to replace true site investigation campaigns, they can be used for more informed decision making in the planning of campaigns or in windfarm layout design. Both methods can be potentially also applied to other types of morphologies like tunnel valleys or to estimate and quantify unit layer boundaries for probabilistic ground models. The assessment of connectivity of predicted channels was not the focus of the study. Nevertheless, it is recommended to address this in future studies. Guidelines to address it are found in e.g. Zhang et al. (Zhou et al., 2018).

It is not the intention of this study to provide a general recommendation of the line spacing necessary for achieving reliable results. The chosen grid depends on the specific need of the site survey and the subsequent ground modeling activities. Nevertheless, it is obvious that 150 m line spacing delivered the most accurate results independent of the chosen method. This corresponds to the current standard for the preinvestigation of offshore windfarm development areas commissioned by the German government agency BSH. While the overall channel structures remain interpretable even at 600 m spacing, the reproducibility diminishes visibly. The provided comparison emphasizes the importance of maintaining denser conditioning data to achieve robust reconstructions. According to investigations by Lohrberg et al. (2022), a line spacing of less than 800 m for 2D seismic data is sufficient to assess buried tunnel valleys in the shallow subsurface. The findings of this study indicate that such spacing would be too coarse to adequately capture channel distribution between the lines. However, it must be considered that the scope of the study was different and that TVs are commonly wider compared to the paleo-channel structures being investigated here, making them likely to be less sensitive to broader line spacings.

In the studies of Sauvin et al. (2023) and Forsberg et al. (2022) a 3D UHR seismic survey was used from which artificially 2D line spacings with increasing line separation was introduced. The final aim of the studies was to investigate the uncertainties of predicted cone penetration tests with increasing line spacing. Within that, the horizon depth for the interpreted units were interpolated between the lines of a decimated 3D seismic survey. In contrast to this study, the focus was not specifically on channel structure and instead of categorical variables, continuous variables (horizon depth) were investigated. Still, findings indicate that the error of horizon interpolation increases with increasing decimation, which is in alignment with the findings of this study.

After all, 3D seismic surveys become increasingly relevant in site characterization for offshore windfarms. Here, the questions of the necessary density of line spacing is rather obsolete through minimizing knowledge gaps in site investigation data. Nevertheless, 2D data acquisition still plays a major role in ongoing and future windfarm development projects, making the question of sufficient line spacing a relevant and ongoing topic.

4. Conclusions

In this study, geostatistical methods, specifically multiple-point statistics (MPS) and sequential indicator simulation (SIS), have been successfully applied to assess paleo-channel probabilities for geological ground models in offshore windfarm developments. The findings indicate that MPS generally outperforms SIS in terms of accuracy, particularly at denser line spacings. Specifically, a line spacing of 150 m yielded the lowest MSE across all tests, indicating a high degree of accuracy in channel reconstruction. In contrast, as line spacing was increased from 300 to 600 m, the MSE values rose significantly, reflecting a decline in the reliability of channel predictability.

While MPS demonstrated better numerical performance, SIS produced a more continuous spatial probability distribution across potential channel structures, but with a higher degree of uncertainty. This highlights the trade-off between precision and coverage, suggesting that the choice of method should depend on the specific risk tolerance and project requirements of the end-user. Furthermore, the ability to apply TIs and conditional data from geographically diverse areas enhances the robustness of both methods, contributing to improved risk assessment in offshore windfarm site development.

The study emphasizes the importance of the consideration of seismic line spacing in ground modeling efforts, supporting denser acquisition strategies to ensure accurate representation of subsurface features. With the increasing relevance of 3D seismic data acquisition the question of line spacing might become less relevant in future offshore windfarm site characterization.

Although the findings do not replace true site investigation coverage, they contribute to the improvement of ground modeling techniques and support improved decision-making in offshore wind farm planning, ultimately aiding in a safer and more efficient installation of turbine foundations.

CRediT authorship contribution statement

Lennart Siemann: Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Conceptualization. **Ramiro Relanez:** Writing – review & editing, Visualization, Validation, Methodology, Investigation.

Funding sources

This work was supported by the Federal Ministry of Economic Affairs and Climate Action Grant 03EE2050A (Research project ProbPerModel).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to acknowledge the support of Benjamin Schwarz and Tobias Mörz throughout the study. Many thanks to Alba de la Iglesia for giving valuable input. It is very much appreciated that Geovariances - a Datamine Company provided an academic license to conduct this research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.acags.2025.100280.

Data availability

Data will be made available on request.

References

Bastante, F.G., Ordóñez, C., Taboada, J., Matías, J.M., 2008. Comparison of indicator kriging, conditional indicator simulation and multiple-point statistics used to model slate deposits. Eng. Geol. 98, 50–59. https://doi.org/10.1016/j.enggeo.2008.01.006.

Caselitz, B., McKay, A., Widmaier, M., Oukili, J., Davies, D., Pernin, N., 2025. Harnessing 3D ultra-high-resolution seismic technology for offshore wind farm development: advancements, challenges, and future prospects. Lead. Edge 44, 170–177. https://doi.org/10.1190/tle44030170.1.

Chiles, J.-P., Delfiner, P., 2012. Geostatistics: Modeling Spatial Uncertainty. John Wiley & Sons.

Coughlan, M., Fleischer, M., Wheeler, A.J., Hepp, D.A., Hebbeln, D., Mörz, T., 2018. A revised stratigraphical framework for the Quaternary deposits of the German North Sea sector: a geological-geotechnical approach. Boreas 47, 80–105. https://doi.org/10.1111/bor.12253.

De Iaco, S., Maggio, S., 2011. Validation techniques for geological patterns simulations based on variogram and multiple-point statistics. Math. Geosci. 43, 483–500. https://doi.org/10.1007/s11004-011-9326-9.

de Souza, L.E., Costa, J.F.C.L., 2013. Sample weighted variograms on the sequential indicator simulation of coal deposits. Int. J. Coal Geol. 112, 154–163. https://doi.org/10.1016/j.coal.2012.12.005.

Forsberg, C.F., Sauvin, G., Klingvort, R.T., Vardy, M.E., Vanneste, M., Kort, A., 2022. Integrated Ground Model - Ten noorden van de Waddeneilanden Wind Farm Zone (Ministerie van Economische Zaken en Klimaat, Rijksdienst voor Ondernemend Nederland) (No. 20190798-04-R). Norwegian Geotechnical Institute, Oslo, Norway.

Geovariances, a Datamine Company, 2024. Isatis.neoTM.

Juda, P., Renard, P., Straubhaar, J., 2022. A parsimonious parametrization of the Direct Sampling algorithm for multiple-point statistical simulations. Applied Computing and Geosciences 16, 100091. https://doi.org/10.1016/j.acags.2022.100091.

Lohrberg, A., Schneider von Deimling, J., Grob, H., Lenz, K.-F., Krastel, S., 2022. Tunnel valleys in the southeastern North Sea: more data, more complexity. E&G Quaternary Science Journal 71, 267–274.

Madani, N., 2022. Revisited Bayesian sequential indicator simulation: using a log-linear pooling approach. Mathematics 10. https://doi.org/10.3390/math10244669.

Mariethoz, G., Caers, J., 2014. Multiple-point geostatistics: stochastic modeling with training images. https://doi.org/10.1002/9781118662953.

Mariethoz, G., Renard, P., Straubhaar, J., 2010. The Direct Sampling method to perform multiple-point geostatistical simulations. Water Resour. Res. 46. https://doi.org/10.1029/2008WR007621.

Medina-Ortega, P., Morales-Casique, E., Hernández-Espriú, A., 2019. Sequential indicator simulation for a three-dimensional distribution of hydrofacies in a volcano-sedimentary aquifer in Mexico City. Hydrogeol. J. 27, 2581–2593. https://doi.org/10.1007/s10040-019-02011-1.

Meerschman, E., Pirot, G., Mariethoz, G., Straubhaar, J., Van Meirvenne, M., Renard, P., 2013. A practical guide to performing multiple-point statistical simulations with the Direct Sampling algorithm. Comput. Geosci. 52, 307–324. https://doi.org/10.1016/j.cageo.2012.09.019.

Montero, J.M., Colombera, L., Yan, N., Mountney, N.P., 2021. A workflow for modelling fluvial meander-belt successions: combining forward stratigraphic modelling and multi-point geostatistics. J. Petrol. Sci. Eng. 201, 108411.

Moreau, J., Huuse, M., Janszen, A., van der Vegt, P., Gibbard, P.L., Moscariello, A., 2012. The Glaciogenic Unconformity of the Southern North Sea, 368. Geological Society, London, Special Publications, pp. 99–110. https://doi.org/10.1144/SP368.5.

Oliver, M.A., Webster, R., 2015. Basic Steps in Geostatistics: the Variogram and Kriging. Springer.

Petrie, H.E., Eide, C.H., Haflidason, H., Watton, T., 2022. A conceptual geological model for offshore wind sites in former ice stream settings: the Utsira Nord site, North Sea. J. Geol. Soc. https://doi.org/10.1144/jgs2021-163.

Pyrcz, M.J., Deutsch, C.V., 2014. Geostatistical Reservoir Modeling. Oxford University Press, USA.

Pyrcz, M.J., Boisvert, J.B., Deutsch, C.V., 2008. A library of training images for fluvial and deepwater reservoirs and associated code. Comput. Geosci. 34, 542–560. https://doi.org/10.1016/j.cageo.2007.05.015.

Sauvin, G., Vardy, M., Dujardin, J., Vanneste, M., Klinkvort, R., 2023. Geospatial Interpolation and Geological Complexity in Data-Driven Ground Models, 2023, pp. 1–5. https://doi.org/10.3997/2214-4609.202320162.

Schwarzer, K., Ricklefs, K., Bartholomä, A., Zeiler, M., 2008. Geological development of the North Sea and the Baltic Sea. Die Küste 74, 1–17.

Straubhaar, J., Renard, P., Mariethoz, G., 2016. Conditioning multiple-point statistics simulations to block data. Spatial Statistics 16, 53–71. https://doi.org/10.1016/j.spasta.2016.02.005.

Straubhaar, J., Renard, P., Chugunova, T., 2020. Multiple-point statistics using multi-resolution images. Stoch. Environ. Res. Risk Assess. 34. https://doi.org/10.1007/s00477-020-01770-8.

Strebelle, S., 2002. Conditional simulation of complex geological structures using multiple-point statistics. Math. Geol. 34, 1–21. https://doi.org/10.1023/A:1014009426274.

Velenturf, A., Emery, A., Hodgson, D., Barlow, N., Mohtaj Khorasani, A.M., Van Alstine, J., Peterson, E., Piazolo, S., Thorp, M., 2021. Geoscience Solutions for Sustainable Offshore Wind Development, 1, 10042. https://doi.org/10.3389/esss.2021.10042.

Wu, X., Shi, Y., Fomel, S., Liang, L., 2018. Convolutional neural networks for fault interpretation in seismic images. https://doi.org/10.1190/segam2018-2995341.1.

Zhang, T., McCormick, D., Hurley, N., Signer, C., 2007. Applying multiple-point geostatistics to reservoir modeling – a practical perspective. Petroleum Geostatistics 2007. https://doi.org/10.3997/2214-4609.201403044.

Zhang, C., Gravey, M., Mariéthoz, G., Irving, J., 2024. Reconstruction of high-resolution 3D GPR data from 2D profiles: a multiple-point statistical approach. Remote Sens. 16. https://doi.org/10.3390/rs16122084.

Zhou, F., Shields, D., Tyson, S., Esterle, J., 2018. Comparison of sequential indicator simulation, object modelling and multiple-point statistics in reproducing channel geometries and continuity in 2D with two different spaced conditional datasets. J. Petrol. Sci. Eng. 166, 718–730. https://doi.org/10.1016/j.petrol.2018.03.043.