2021

A machine learning approach for modeling irregular regions with multiple owners in wind farm layout design

Reddy, Sohail R.

Elsevier

Reddy, Sohail R. "A machine learning approach for modeling irregular regions with multiple owners in wind farm layout design." Energy 220 (2021): 119691.
http://hdl.handle.net/10945/68901

Energy 220 (2021) 119691

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

A machine learning approach for modeling irregular regions with multiple owners in wind farm layout design☆

Sohail R. Reddy

Department of Applied Mathematics, Naval Postgraduate School, Monterey, CA, United States

ARTICLE INFO

Article history:
Received 17 September 2020
Received in revised form
22 November 2020
Accepted 21 December 2020
Available online 31 December 2020

Keywords:
Wind farm layout optimization
Wind energy
Complex terrain
Support vector machine
Machine learning

ABSTRACT

Wind farm development projects require a detailed survey of the eligible land. The land selected is often segmented into different region, each owned by different landowners with different land pricing. These regions are often complex shaped with irregular boundaries. Therefore, an efficient method for numerically modeling such irregular domains is needed. This work uses support vector data description (SVDD) to generate an analytical, continuous description of the irregular regions. Whereas other methods typically work well for modeling convex domains, the SVDD approach can be used to model irregular regions as a spherical boundary using various kernel mapping. It was demonstrated that the SVDD approach can be used to model any number of complex regions. An error analysis showed that the SVDD approach can construct accurate descriptions using a relatively small data set. The applicability of SVDD method in wind farm layout optimization is also demonstrated. The wind farm optimization study considered that the terrain is divided into several regions each owned by a different owner offering the land at a different price. Two different methods for considering the cost of the land are presented. The differences in optimal farm layouts using the two land cost models were also presented. In each case, the optimized wind farm layouts resulted in lower cost-of-energy relative to the reference wind farm. It was shown that the SVDD approach can also be used to restrict the placement of wind turbines in infeasible/restricted regions. The library for support vector data description was also made available to the public.

1. Introduction

With depleting fossil fuel reserves, there has been a tremendous push towards energy independence. In recent years, a strong push towards clean, renewable energy has resulted in significant advances in solar, wind and nuclear energy systems. There are over 57,000 wind turbines and over 200 wind farms in the United States alone. The United States has shifted its focus on wind energy with a goal of generating 20% of the country's energy from wind power alone by 2030 [1] and 35% by 2050 [2].

Much work has been done on wind farm layout design and optimization. It is a challenging area due to the computationally expensive model, large number of design variables and constraints, and extreme multimodality of the function space. Often in literature, the wind farm layout optimization problem is focused more on maximizing the power generation and less on accounting for the cost of the wind farm. The overall cost of the wind farm can be attributed to cost of the wind turbine and its components, cost of labour, transportation and land, to name a few. The cost of wind power is generally higher than other conventional sources of renewable energy. Efforts have been made to design farms for minimum levelized cost of energy [3,4] but did not take into account the cost of land used.

Several works have attempted to include land costs into the optimization framework. One approach simply assumes that the cost of the land associated with the wind farm is directly proportional to the area of the farm [5]. Under this assumption, the wind farm layout design is solved as a multi-objective optimization problem where the objectives are to maximize the power generated and minimize the area occupied (cost). Another approach assumes the land lease cost is proportional to the annual power generation of the farm [6,7]. This approach does not directly take into account the arrangement of the wind turbines in the farm. When accounting for the land lease costs, previous works have assumed a regular square farm boundary and not a more realistic irregular boundary.

In addition to determining the number and layout of wind

☆ libSVDD can be downloaded from https://github.com/sohailrreddy/libSVDD.

E-mail address: sredd001@fiu.edu.

https://doi.org/10.1016/j.energy.2020.119691
0360-5442/

turbines, wind farm developers must also take into account the land geography and zoning [8,9]. Most literature on wind farm layout optimization assumes that the all regions of the land are feasible for wind turbine placement [10,11]. Often the land selected for the wind farm project is divided into multiple separate regions each owned by a different entity/individual [12,13]. It is possible that not all landowners are willing to participate and lease their land for the wind farm project. It is also possible that certain conditions, such as soil quality, proximity to features (cities, schools, etc.), can also make certain regions of the land infeasible for wind turbine placements.

Ryberg et al. [14] evaluated the eligibility of land over Europe for wind farm development. They investigated the different factors that influences the eligibility of land for wind farm development. Although the land eligible and available for wind farm development can be identified, an efficient and accurate approach is still needed to incorporate such constraints within an optimization framework. Restrictions placed on land available for wind farm development often lead to a complex-shaped eligible/feasible regions. Shapes of such complex regions are difficult to define analytically and, therefore, need special treatment to be incorporated into the optimization framework. These irregular boundaries are often modeled as a convex hull or a collection of convex hull of points within the boundaries [15–18]. This approach works well for very simple convex shaped boundaries but cannot model irregular, non-convex boundaries. Grid based methods have also been used to determine whether turbines lie within a boundary [6,19]. These methods divide the entire region into smaller quadrilaterals and use the row and column indices to determine where the turbine lies. The boundary of the complex region has also been modeled as a series of straight-line [20] or interpolated curves [21]. These methods however cannot accurately model complex, disconnected shapes and require additional ray-boundary intersection algorithms to determine whether a turbine lies within the region. Such methods also require identification of the points that lie only on the boundary of the region, and not on the interior. Table 1 presents recent efforts to model irregular domain boundaries. These approaches also represent the domain using smaller polygonal or regular rectangular grids, both of which are unable to model complex domains. Therefore, a method that does not require any additional sorting or analytical expressions of the boundary and with the ability to capture complex irregular boundary is needed.

One such method for modeling non-linear data is through machine learning. The recent advances in the field of machine learning and deep learning has allowed for developments in facial recognition, autonomous vehicles, and artificial intelligence, to name a few. Recently, such methods have also been applied in wind turbine and wind farm analysis. Steco et al. [26] performed a thorough review of various machine learning approaches that have been used for wind turbine condition monitoring. They stated that almost two-thirds of machine learning methods use a classification model based on artificial neural networks (ANN) or support vector machines (SVM). Ti et al. [27] demonstrated the use of ANN based machine learning for wake modeling of wind turbines. Their work trained the multi-layer neural network using simulated data from RANS/LES modeling. Clifton et al. [28] used a regression tree model to predict the power output of individual turbines. Over the years, machine learning has also be applied to wind farm analysis. Howland and Dabiri [29] used a physics-informed statistical model and a two-layer neural network model to accurately predict power production. Torres et al. [30] made use of a recurrent neural network to perform power production estimation.

Table 1
Methods used by other authors to model irregular domains.

Authors	Year	Method
Wang et al. [22]	2017	Discrete grid parameterization
Gonzalez et al. [23]	2017	Discrete grid parameterization
Perez-Moreno et al. [24]	2018	Discrete grid parameterization
Feng et al. [17]	2018	Polygon representation
Sorkhabi et al. [16]	2018	Polygon representation
Mittal and Mitra [18]	2019	Polygon representation
Stanley and Ning [25]	2019	Boundary-grid parameterization

Although machine learning approaches has been heavily used in analysis-based models such as RANS/LES, they have not been heavily applied in the wind farm planning. For example, they have not yet been applied to model the complex boundaries arising in wind farm projects. Machine learning approaches seem ideal for these problems since they can accurately classify points within sets and can determine which region a turbine lies in. This work presents an approach to efficiently model complex irregular boundaries using points sampled from a region. It uses the support vector domain description (SVDD) to generate an analytical description of the region where the complex boundaries can be represented simply as a hyperspherical boundary. Since the resulting description is analytical and continuous, both gradient based and non-gradient based optimization algorithms can be used to solve the optimization problem. The accuracy of the SVDD approach was demonstrated on domains with multiple complex regions with irregular, non-convex and complex boundaries. The SVDD model was then used to perform wind farm layout optimization on terrain with regions owned by different landowners with different land prices. This demonstrated the ability of the SVDD approach to not only model complex boundaries but also model several distinct regions simultaneously. Two different approaches for incorporating land costs into the optimization framework were also presented. It was shown that the SVDD model can efficiently model complex regions and drastically simplify the terrain modeling problem in wind farm layout design. The library developed to train the SVDD model [31] is released for public use.

2. Support vector domain description

The support vector data description [32], also sometimes called support vector domain description, is a technique inspired by the support vector machines (SVM) of Vapnik [33] for defining an optimal description of the objects. Whereas SVM identifies a separating hyperplane that classifies objects, the SVDD approach identifies a hypersphere to perform the classification. This method has been successfully used to define the boundary of the data set and can be used to model complex, irregular, non-convex boundaries.

2.1. Mathematical background

The SVDD method computes a sphere of radius R and center $\vec{a}$ that contains the data set such that the radius is minimized

$$F(R, \vec{a}, \xi_i) = R^2 + C\sum_i \xi_i \tag{1}$$

where C represents the trade-off between the simplicity and the number of objects rejected (outliers), and ξ_i is the slack variable, without which the description becomes sensitive to the outliers. In Eq. (1), $C\sum_i \xi_i$ is the penalty term associated with outlier detection.

The radius is minimized under the following constraint

$$\left|\left|\vec{x}_i - \vec{a}\right|\right|^2 \leq R^2 + \xi_i \quad \forall_i, \xi_i \geq 0 \tag{2}$$

where $\vec{x}_i$ are the objects (points) whose description is being constructed. The Lagrangian $\mathscr{L}$ can be constructed using Eq. (1) and Eq. (2) as

$$\mathscr{L}(R, \vec{a}, \alpha_i, \xi_i) = R^2 + C\sum_i \xi_i - \sum_i \alpha_i \left(R^2 + \xi_i - \left|\left|\vec{x}_i - \vec{a}\right|\right|^2\right) - \sum_i \gamma_i \xi_i \tag{3}$$

where the Lagrange multipliers α and γ satisfy $\alpha_i \geq 0$ and $\gamma_i \geq 0$. Differentiating Eq. (3) yields a new set of constraints

$$\sum_i \alpha_i = 1 \quad \forall_i \tag{4}$$

$$\vec{a} = \frac{\sum_i \alpha_i \vec{x}_i}{\sum_i \alpha_i} = \sum_i \alpha_i \vec{x}_i \quad \forall_i \tag{5}$$

$$C - \alpha_i - \gamma_i = 0 \quad \forall_i \tag{6}$$

It can be seen from the second constraint (Eq. (5)) that the center of the hypersphere is simply a linear combination of weights α_i and the objects $\vec{x}_i$. Since both $\alpha_i \geq 0$ and $\gamma_i \geq 0$, the third constraint (Eq. (6)) can be rewritten as $0 \leq \alpha_i \leq C$.

Substituting Eq. (5) and Eq. (6) into Eq. (3) yields a function to maximize given by

$$\mathscr{L} = \sum_i \alpha_i \left(\vec{x}_i \cdot \vec{x}_i\right) - \sum_{i,j} \alpha_i \alpha_j \left(\vec{x}_i \cdot \vec{x}_j\right) \tag{7}$$

with constraints $0 \leq \alpha_i \leq C$ and $\sum_i \alpha_i = 1$. It can be seen from Eq. (5) that only those objects $\vec{x}_i$ with $\alpha_i > 0$ are needed to define the hypersphere. These objects are referred to as support vectors. Objects with $\alpha_i = 0$ are within the hypersphere and objects with $\alpha_i = C$ are considered outliers. The radius of the sphere (R) can be obtained by taking the distance between the center ($\vec{a}$) and any support vector (since all support vectors lie on the boundary). This procedure of identifying the support vectors, the Lagrange multipliers α_i and the radius R by minimizing Eq. (7) is referred to as the training phrase and is similar to that used in SVM.

An object $\vec{z}$ is on or within a hypersphere with the center at $\vec{a}$ and radius R if $||\vec{z} - \vec{a}||^2 \leq R^2$. If the center is expressed using the support vectors then the object $\vec{z}$ is on or within the sphere if

$$(\vec{z} \cdot \vec{z}) - 2\sum_i \alpha_i \left(\vec{z} \cdot \vec{x}_i\right) + \sum_{i,j} \alpha_i \alpha_j \left(\vec{x}_i \cdot \vec{x}_j\right) \leq R^2 \tag{8}$$

An object is "accepted" and has the same classification as the training set if it satisfies this acceptance criteria (Eq. (8)). Fig. 1 shows the standard SVDD approach. It can be seen that any point inside the hypersphere is considered to be inside the region formed by the training points.

In real-world applications, the data sets are not spherically distributed and, therefore, classification using hyperspheres are not an accurate description of the data set. This issue can be addressed by mapping the data set (objects) into a feature space where a hypersphere is a better approximation. This is done using a "kernel-trick", where a kernel $K(\vec{x}_i \cdot \vec{x}_j)$ that satisfies Mercer's theorem [34] is used to map the data set. This SVDD procedure using kernel mapping is shown in Fig. 2. It can be seen that the approach constructs an irregular boundary that conforms to the training points in the input space and represents this irregular boundary as a hyperspherical boundary in an $n-$ dimensional feature space. If a point is within the hyperspherical boundary in the feature space, then it is also within the irregular boundary in the input space.

Replacing all inner products $(\vec{x}_i \cdot \vec{x}_j)$ by the kernel representation $K(\vec{x}_i \cdot \vec{x}_j)$, the domain description problem is now given by

$$\mathscr{L} = \sum_i \alpha_i K\left(\vec{x}_i \cdot \vec{x}_i\right) - \sum_{i,j} \alpha_i \alpha_j K\left(\vec{x}_i \cdot \vec{x}_j\right) \tag{9}$$

where constraints $0 \leq \alpha_i \leq C$ and $\sum_i \alpha_i = 1$ are enforced. Then, the acceptance criteria is given as

$$K(\vec{z} \cdot \vec{z}) - 2\sum_i \alpha_i K\left(\vec{z} \cdot \vec{x}_i\right) + \sum_{i,j} \alpha_i \alpha_j K\left(\vec{x}_i \cdot \vec{x}_j\right) \leq R^2 \tag{10}$$

2.2. Numerical implementation

The libSVDD library [31] is used to train the support vector data description. The library uses the sequential least squares quadratic programming (SLSQP) algorithm to find the optimum values for the Lagrange multipliers. There are several kernels that are available in libSVDD, some of which are given in Table 2. All three kernel functions were investigated, with the Gaussian kernel outperforming the Exponential and Laplace kernels.

Using the Gaussian kernel $K_G(\vec{x}_i \cdot \vec{x}_j) = \exp(-(||\vec{x} - \vec{y}||/\sigma)^2)$, the domain description problem can be written as

$$\mathscr{L} = 1 - \sum_i \alpha_i^2 - \sum_{i \neq j} \alpha_i \alpha_j K_G\left(\vec{x}_i \cdot \vec{x}_j\right) \tag{11}$$

with constraints $0 \leq \alpha_i \leq C$ and $\sum_i \alpha_i = 1$. The acceptance rule then becomes

$$-2\sum_i \alpha_i K_G\left(\vec{z} \cdot \vec{x}_i\right) \leq R^2 - C_X - 1 \tag{12}$$

where $C_X = \sum_{i,j} \alpha_i \alpha_j K(\vec{x}_i \cdot \vec{x}_j)$. The standard deviation σ determines the number of support vectors and the tightness of the fit, where a smaller σ results in a tighter fit and more support vectors. It should be mentioned that once an SVDD model is trained, the classification of new points into their respective regions is exceptionally fast. This is because the acceptance criteria simply compares the distance of the point from the center with the radius of the mapped hypersphere.

2.3. Validation of the support vector data description model

There are several area/terrain factors that can constrain the wind farm design problem. Such conditions include price of land, proximity to features (cities, schools, etc.) and soil quality to name a few. These cases not only require that the terrain be separated into feasible and infeasible areas but also require an efficient method to

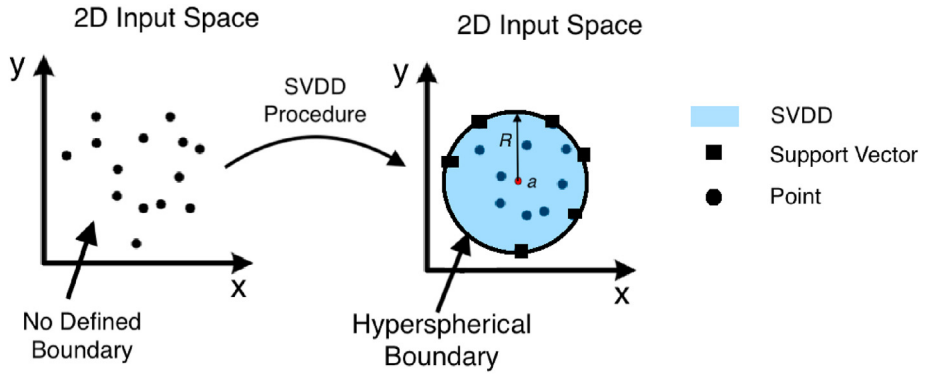


Fig. 1. Visualization of the SVDD procedure without using kernel mapping for a two-dimensional case.

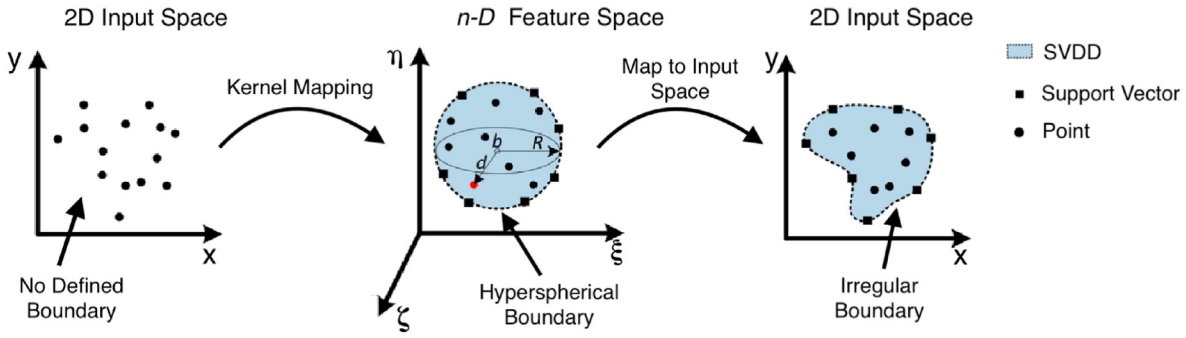


Fig. 2. Visualization of the SVDD procedure using kernel mapping for a two-dimensional case.

Table 2
Kernels available in libSVDD.

Name	$K(\vec{x} \cdot \vec{y})$
Gaussian	$\exp(-(\|\vec{x}-\vec{y}\|/\sigma)^2)$
Exponential	$\exp(-\|\vec{x}-\vec{y}\|/\sigma^2)$
Laplace	$\exp(-\|\vec{x}-\vec{y}\|/\sigma)$

Table 3
Number of training points, number of support points and the definition for terrain division.

	N	N_{sv}	Definition
Region A	744	276	$z > 240$
Region B	1407	520	$180 < z \leq 240$
Region C	1582	639	$130 < z \leq 180$
Region D	705	359	$100 < z \leq 130$
Region E	1454	628	$z \leq 100$

numerically represent such regions in an optimization framework. The traditional approach represent the region as a convex hull. This approach can be used for convex, regular domains but cannot model irregular non-convex shapes. Such regions can, however, be represented using the SVDD approach. An SVDD model can be built for each separate region of a terrain. Once trained, the SVDD model can then be used to identify which region a new point (turbine) belongs to.

The ability of SVDD to classify points into their respective regions is investigated. The terrain used for this study was assumed to be that of the AWEC wind farm (Fig. 7a) in Mojave, CA, USA and was obtained using the TouchTerrain program [35]. This program extracts terrain profile from satellite images and represents the terrain profile as elevation at scattered locations. The AWEC terrain was then separated into five distinct regions according to their elevation, each with a highly irregular boundary. An SVDD model was then trained for each region using the points that belonged only to that individual region. Table 3 shows the criteria for segmentation for each region of the AWEC farm. An SVDD model is built for a terrain with four distinct regions and five distinct regions. It should be mentioned that in the case with only four separate regions, Regions D and E are combined to form a single region. Table 3 also shows the number of points N used to train the SVDD and the number of support vectors N_{sv} in the trained SVDD model. It should be mentioned that the SVDD models for these examples were built using only the x and y coordinates of the points in the terrain. The z coordinate can also be included if the complex boundary (description of the data set) is three-dimensional. The SVDD approach can be used in any number of dimensions since the kernels are only a function of the Euclidean distance.

An additional testing set of 40,000 points uniformly distributed in $x, y \in [0, 2000]$ was generated to test the accuracy of the trained SVDD models. The constructed SVDD models were then used to place all points in the testing set into their respective regions.

Fig. 3 shows the exact description (Fig. 3a and Fig. 3b) of the terrain and the SVDD description of the terrain for the four region and five region case. The SVDD descriptions (Fig. 3c and Fig. 3d) show that the testing set of 40,000 points were accurately classified into their appropriate regions. This is seen in the similarity between the exact segmentation of the regions and the SVDD segmentation of the regions. It can be seen that the SVDD approach can accurately model and create accurate descriptions of segmented regions in the domain. Therefore, the SVDD approach can be used to identify the

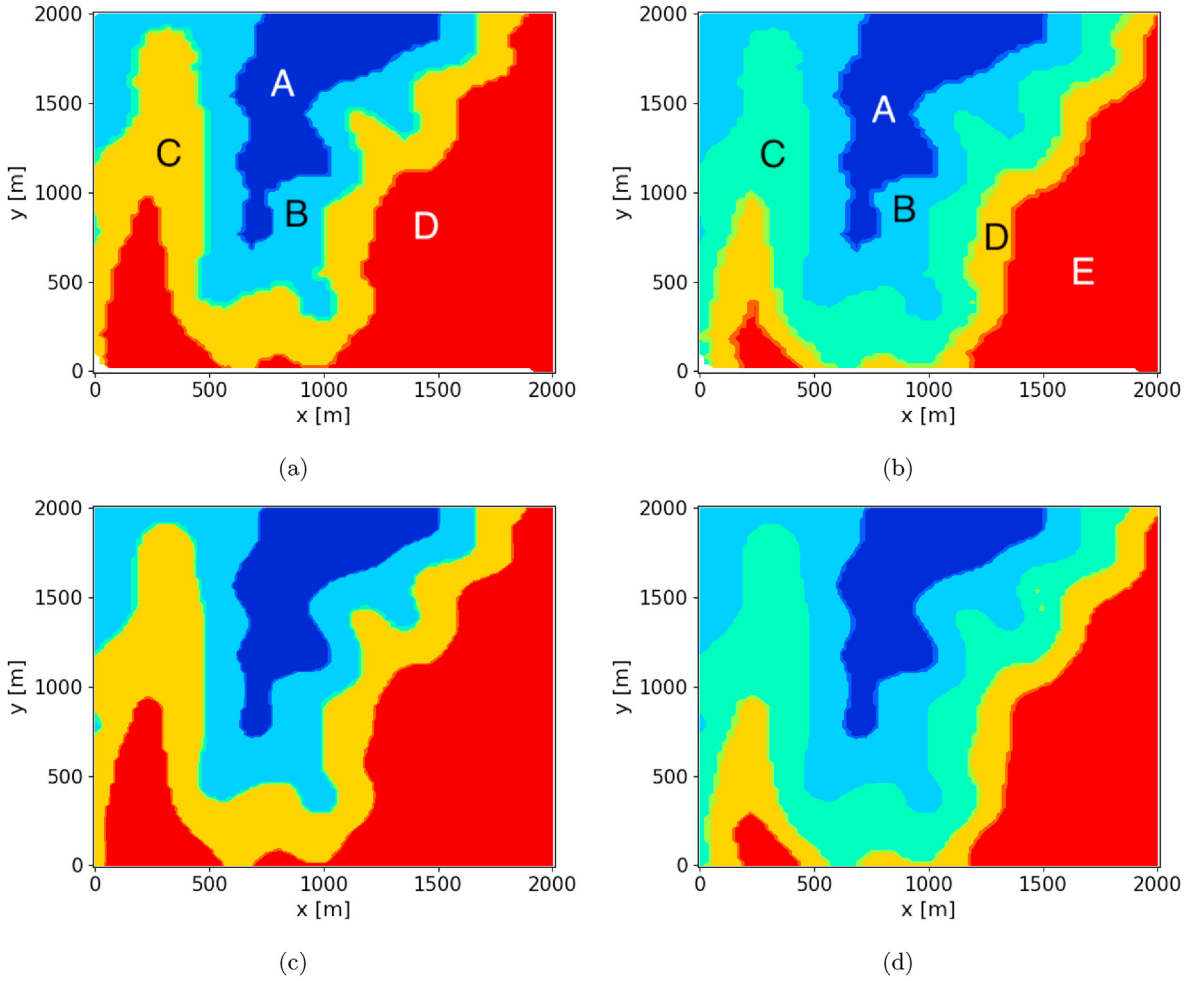


Fig. 3. Set used to train the SVDD for modeling a: a) four regions and b) five regions; testing set classified by the trained SVDD model (i.e. constructed description) for a: c) four region model and d) five region model.

region that a point belongs to so that certain constraints can be enforced. Standard methods such as polygonal representation and grid-parameterization are unable to capture such complex boundaries since their representations are linear. The nonlinear kernel mapping in the SVDD approach addresses this issue and allows for accurate modeling of the complex regions. The SVDD approach without kernel mapping would also fail to model such irregular regions. It should again be mentioned that the SVDD model was trained using only sampled data points and did not require any analytical description of the terrain. The SVDD approach does not depend on the number of separate regions of the terrain since a single SVDD model is trained for each separate region.

The description in Fig. 3 was constucted using a training set with a total of 5892 points. It is also important to quantify the accuracy of the description constructed using training sets of different size. For this reason, the SVDD model was trained for the five-region problem (Fig. 3b) using a varying sized training set. The points in each training set were uniformly distributed throughout the domain. Fig. 4 shows the relative error in the description as a function of the training set size. It can be seen that an accurate SVDD description can be constructed using a relatively small training set. It also shows that the error asymptotically converges. Therefore, the SVDD model can be accurately trained even when a large training data set is not available.

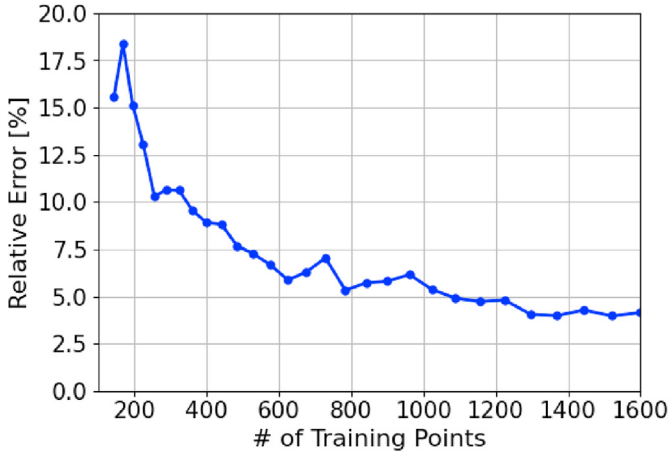


Fig. 4. Relative error in the constructed SVDD description as a function of training set size.

Fig. 5 shows the description constructed using various sizes of the training set. It can be seen as the size of the training set increases, the description becomes more accurate. However, a sufficiently accurate description was obtained using only 400 evenly distributed points.

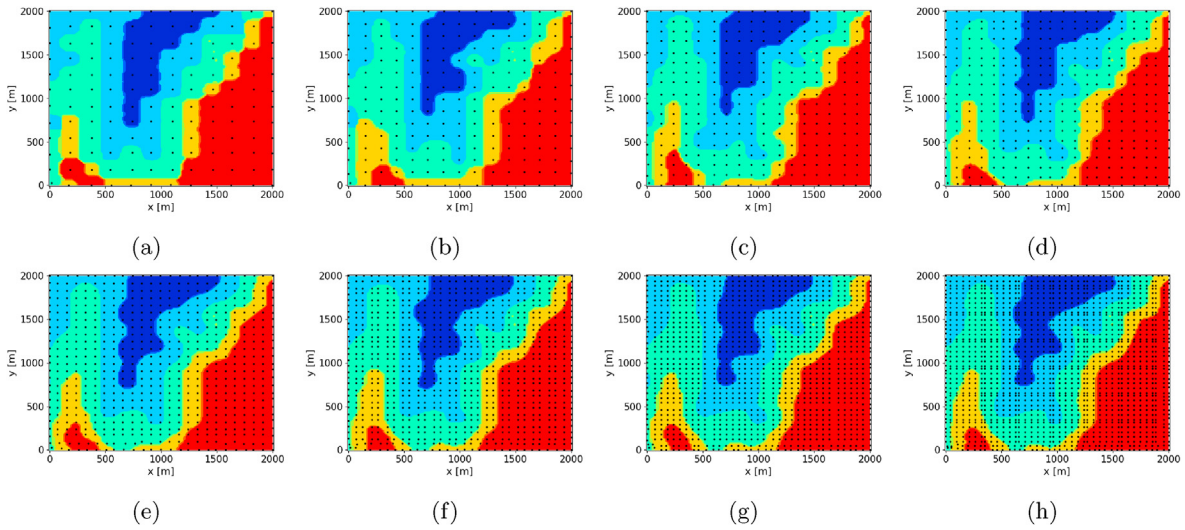


Fig. 5. Description constructed by an SVDD model trained using a grid of: a) 12 × 12 points, b) 15 × 15 points, c) 18 × 18 points d) 20 × 20 points, e) 25 × 25 points, f) 30× 30 points, g) 35 × 35 points, and h) 40 × 40 points (Black dots represent the training points.).

3. Wind farm layout optimization

The application of the SVDD approach in the context of wind farm layout optimization is presented.

3.1. Wind farm analysis

Each wind farm configuration in this work was analyzed using the WindFLO framework [36,37]. WindFLO is a publicly available software that can be used to analyze the power generation, land-usage and cost of wind turbines in farms. It uses analytical wake models, wake superposition schemes and ambient wind models to compute the wind velocity experienced by wind turbines in a farm. The framework can use six analytical wake models and four wake merge/superposition models. It can easily be expanded to include other models. WindFLO also accounts for the terrain profile/elevation and wake-terrain interactions. The schemes in WindFLO were validated using experimental data from a scaled wind farm to within 1% relative error.

The WindFLO framework also features uni-variate and bi-variate, linear and non-linear cost models for turbines and uses the convex hull approach to quantify the area of land occupied by the wind farm layout. For brevity, the specific numerical methods present in WindFLO are not discussed in this manuscript. For a detailed implementation and validation, readers are referred to Ref. [37].

3.2. Problem formulation

The wind farm in this study consists of 25, three-bladed Vestas V90-3 MW wind turbines featuring a rotor radius of $R = 45$m and a tower height of $H = 105$m. The power curve, coefficient of power (C_P) and coefficient of thrust (C_T) curves for the V90-3 MW turbines are shown in Fig. 6a. The terrain profile of the wind farm, shown in Fig. 7a was assumed to be that of the AWEC wind farm. Fig. 6b shows the wind conditions experienced at the AWEC wind farm from January 1, 2019 to December 31, 2019. The wind conditions were obtained using the cli-MATE software [38]. The wake in the wind farm is modeled using Frandsen's wake model [39]. Multiple-wake interaction is modeled using the quadratic superposition scheme. It was shown that the Frandsen's wake model with a quadratic superposition scheme yielded the most accurate results [37]. The objective is to minimize the cost-of-energy (CoE) of the wind farm. Since the type of wind turbine is kept constant, the change in cost of the wind farm is attributed solely to the change in cost of the land for the farm.

The WindFLO framework is able to account for terrain elevation and wake-terrain interaction which requires a continuous representation of the terrain. The AWEC wind farm terrain was modeled using a Radial Basis Function interpolation of the elevation at the scattered points obtained using the TouchTerrain software [35]. The terrain was then segmented into five different regions. Table 4 shows the segmentation criteria for each region. Each region was assumed to be owned by different individuals, each with a different land price or restriction.

Two different approaches for incorporating land costs into the optimization problem were considered. The first approach, called the per-area (PA) model, assumed that the cost is proportional to the area of the farm within a region and, therefore, their arrangement within a region. The area of the farm in a particular region was modeled as the area of the convex hull formed by the turbines within that region. In the event that a particular region only contains a single turbine, the area occupied by the farm in the region is taken to be the swept area of the rotor (πR^2). If the turbines in a region are distributed such that they form a line with length d, the area is taken to be the area of a rectangle with width of the rotor diameter and length d. For these two special cases, the convex hull approach cannot be used since the computed convex hull area will be zero.

The second approach, called the per-turbine (PT) model, assumes that the cost is proportional to the number of turbines (N_T) within a region and not the area occupied or their arrangement. The cost for each of the two approaches are given in Table 4. The range of land costs used were obtained from the work of Ratliff et al. [40], and Meyers and Meneveau [41]. In both land-based cost models, the trained SVDD model was used to identify the region that a turbine belonged to.

The AWEC farm terrain was assumed to be divided into five separate regions shown in Fig. 7b. The criteria for each region are shown in Table 4. It should be mentioned that the five regions were separated based on their elevation only for demonstration purposes since this segmentation is easier. The terrain can also be segmented using satellite imaging, land owner contracts, etc. Three different cases for

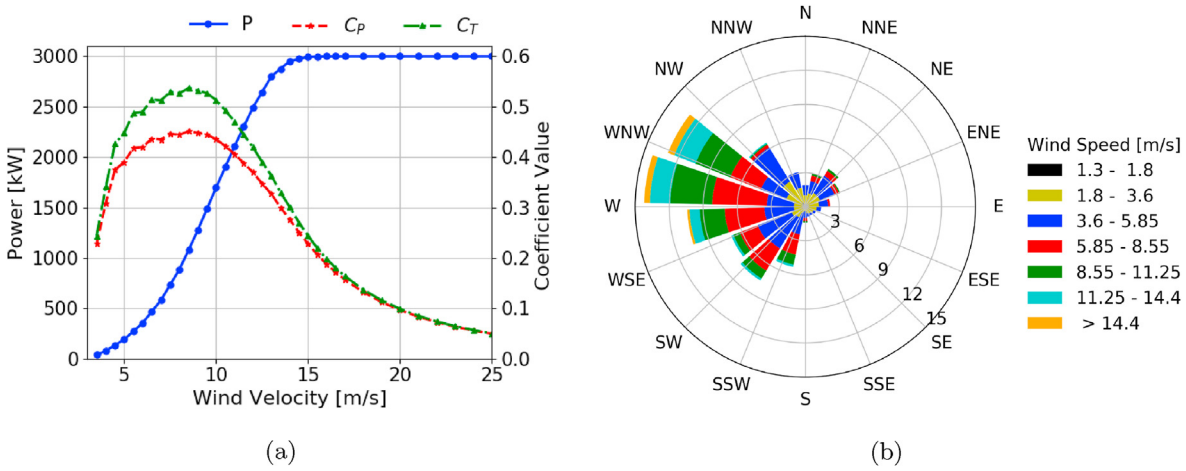


Fig. 6. a) power and turbine coefficients curve of the V90-3 MW turbines and b) windrose diagram for the AWEC wind farm for the year 2019.

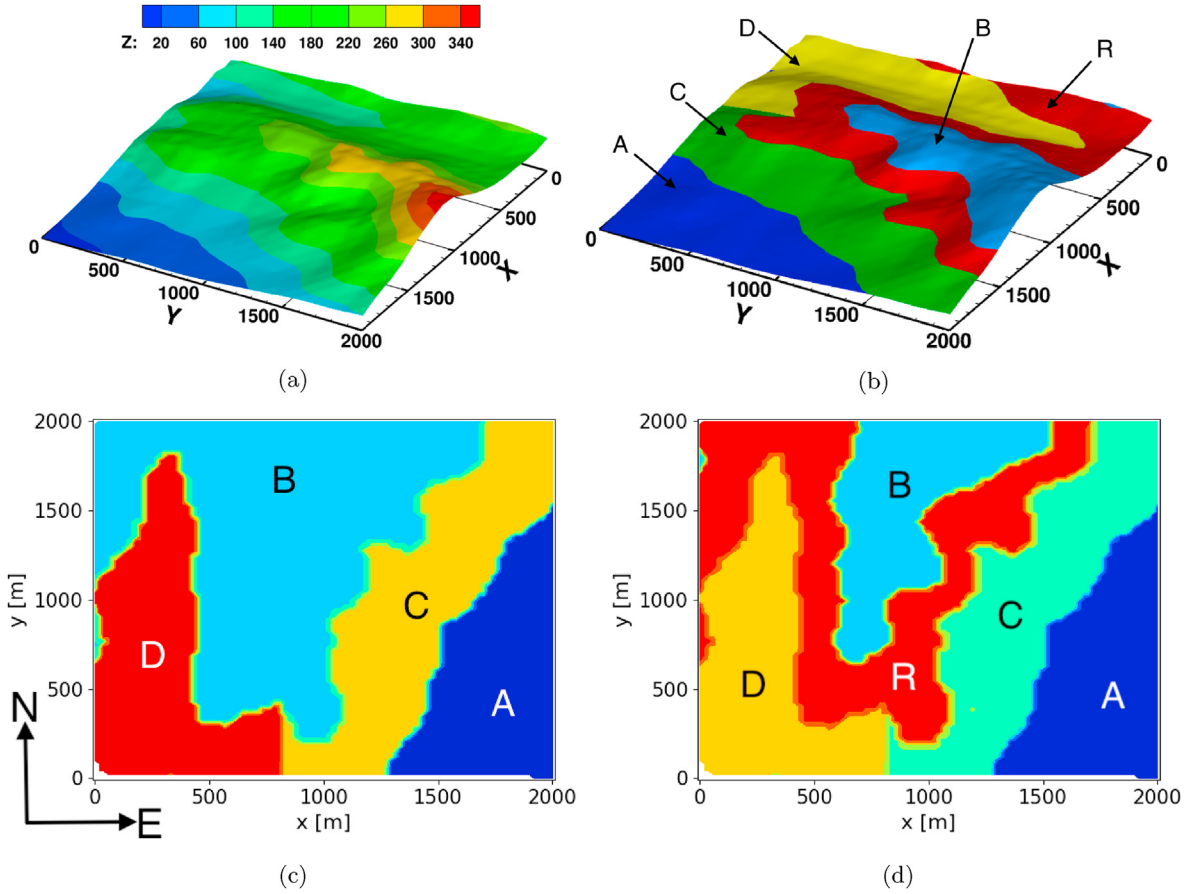


Fig. 7. a) Elevation of the AWEC wind farm terrain, b) division of the AWEC farm into five regions, c) terrain division for Case 2 optimization study and d) terrain division for Case 3 optimization study.

Table 4
Definition and cost of land in each region for the optimization study.

	N	N_{sv}	Definition	USD/m^2	USD/ N_T	Area (km^2)
Region A	1041	393	$z \leq 80$	0.25	2000	0.656
Region B	887	354	$z \geq 230$	0.35	3000	0.770
Region C	1260	512	$80 < z < 170$ *and* $x \geq 800$	0.45	4000	0.918
Region D	1100	428	$80 < z < 170$ *and* $x < 800$	0.55	5000	0.644
Region R	1604	618	$170 \leq z < 230$	∞	∞	1.133

optimization were considered. The first case (Case 1) neglected the cost of the land during the optimization. Under this assumption, minimizing the CoE is analogous to maximizing the annual energy production (AEP). In the second case (Case 2), regions R and B were combined to form a single region B (Fig. 7c). That is, all regions of the terrain were considered to be feasible for wind turbine placement where the cost of land in each of the four regions was different (Table 4). This case assumes that each region is owned by a different individual offering the land at a different price and that all individuals are participating in the project. The third case (Case 3) considers the same land prices as those in Case 2 but incorporates an infeasible/restricted region (Region R) where no turbines can be placed. This case considers five distinct regions (Fig. 7d) and assumes that the landowner for Region R is not participating in the wind farm project and, therefore, not selling/leasing the land. It can also be assumed that Region R is a body of water (i.e. river) that prohibits the placement of wind turbines.

3.3. Optimization algorithm

The typical multi-variate, single objective, constrained optimization can be written as

$$minf(\vec{x}) \tag{13}$$

$$\vec{x} = \{x_1, \ldots, x_m\}$$

$$\text{subject to} : x \in [a_i, b_i], \quad i = 1, \ldots, m$$

$$h_j(\vec{x}) = 0, \quad j = 1, \ldots, n$$

$$g_k(\vec{x}) \leq 0, \quad k = 1, \ldots, o$$

where $\vec{x}$ is a vector of design variables, h_j is the j^{th} equality constraint and g_k is the k^{th} inequality constraint. The optimization in this work is performing using the Single-Objective Hybrid Optimizer (SOHO). The SOHO algorithm is a hybrid evolutionary algorithm that actively switches between constitutive algorithms to accelerate convergence. A detailed description of SOHO is given in Ref. [42]. All optimization studies were ran for 500 generations with an initial population of 100 members.

This work uses the constraint violation [43] value approach to incorporate constraints within the optimization framework. The constraint violation value of a solution $\vec{x}$, denoted by $CV(\vec{x})$ is calculated as

$$CV(\vec{x}) = \sum_{j=1}^{n} |h_j(\vec{x})| + \sum_{k=1}^{o} \langle g_k(\vec{x}) \rangle \tag{14}$$

where the bracket operator $\langle \alpha \rangle$ returns the absolute value of α if $\alpha > 0$ and returns 0 otherwise. A lower value of $CV(\vec{x})$ is indicative of a smaller constraint violation. A $CV(\vec{x})$ value of 0 indicates that all constraints are inactive and the solution is feasible. This constraint violation value is used to modify the domination criteria.

Definition A solution $x^{(1)}$ is said to constraint-dominate another solution $x^{(2)}$, if any one of the following conditions is true:

1. if $x^{(1)}$ is feasible and $x^{(2)}$ is infeasible
2. if $x^{(1)}$ and $x^{(2)}$ are infeasible and $x^{(1)}$ has a smaller *constraint violation* value
3. if $x^{(1)}$ and $x^{(2)}$ are both feasible and $x^{(1)}$ dominates $x^{(2)}$ with the usual domination principle

The optimization study attempts to find a farm layout with minimal cost-of-energy (CoE = Cost/AEP) in USD/MWh. The annual energy production (AEP) is given by Eq. (15)

$$AEP_{\text{Farm}} = (365 \times 24) \int_{0^o}^{360^o} \int_{0}^{U_{\max}} P_{\text{Farm}}(U, \theta) p(U, \theta) \mathrm{d}U \mathrm{d}\theta \tag{15}$$

where $U_{\max}$ is the maximum wind speed experienced by the farm at that location, $P_{\text{Farm}}(U, \theta)$ is the power generated by the farm for a wind speed U and a wind direction θ, and $p(U, \theta)\mathrm{d}$ is the probability of the occurrence of the wind condition with speed U and direction θ. The AEP of the wind farm is also normalized by the AEP calculated if all turbines were operating year around at their rated wind speed (657,000 MWh).

The cost of the turbines in the wind farm was obtained using the non-linear, bivarate cost model in WindFLO [37]. The total cost of the farm was taken to be the sum of the cost of the turbines and the cost of the land (Table 4). The design variables were the x and y coordinates of each wind turbines. This resulted in a total of 50 design variables (two for each of the 25 turbines). The trained SVDD model was used to classify each turbine in the farm into one of the five regions based on its x and y coordinates.

Two constraints were included in the optimization study. The first constraint (Eq. (16)) required that the clearance around each turbine be greater than its diameter in order to avoid collision between turbines.

$$g_1(\vec{x}) = \max_{j=1,\ldots,N} \left(D - \left|\left| \vec{x} - \vec{x}_j \right|\right|_2 \right) \tag{16}$$

The second constraint was only implemented in Case 3 and required that each turbine be within the feasible region (Regions A-D). The SVDD model trained for Case 3 was used to determine whether a turbine was in the restricted region (Region R) by means of Eq. (12). Using the SVDD approach, the second constraint can be written as

$$g_2(\vec{x}) = C_X + 1 - 2\sum_i \alpha_i K_G \left(\vec{x} \cdot \vec{\chi}_i \right) - R^2 \tag{17}$$

where $\vec{\chi}_i$ are the support vectors, R is the radius of the trained hypersphere and $\vec{x}$ is the location of the turbines. Recently, Reddy [44] successfully performed constrained wind farm layout optimization using the SVDD approach to model regions restricted for wind turbine placement.

4. Results of optimization

4.1. Reference farm

For comparison purposes, the reference farm layout was considered to be a 5×5 array of wind turbines. Table 5 shows the wind farm performance and the cost of the reference wind farm using the two land cost models. The CoE was computed using the terrain division in Case 2 with the costs given in Table 4. It shows that the CoE under the two land price models are similar orders of magnitude. The 'Farm Area' is computed as the area of the convex hull of the entire wind farm, whereas the 'Area Convex Hull' is computed as the sum of the area of the convex hull formed by turbines in each region.

Fig. 8a and Fig. 8b show the reference farm layout with superimposed contours of terrain divisions in Case 2 and Case 3, respectively. Under the terrain divisions of Case 2, Regions A, B, C and D feature six, 11, four and four turbines, respectively. Under the terrain divisions of Case 3, there are nine turbines within the restricted region.

Table 5
Performance of reference wind farm layout.

	Per Area	Per Turbine
Normalized AEP	1.35	
Farm Efficiency	0.24	
Farm Area (in km^2)	4.0	
Area Convex Hull (in km^2)	2.63	
CoE (in USD/MWh)	70.85	69.88

Table 6
Performance of optimized wind farm in Case 1 and percent changes from reference farm (in brackets).

	Per Area	Per Turbine
Normalized AEP	1.37 (1.3)	
Farm Efficiency	0.24 (−0.6)	
Farm Area (in km^2)	3.72 (−6.9)	
Area Convex Hull (in km^2)	1.98 (−25)	
CoE (in USD/MWh)	69.78 (−1.5)	69.02 (−1.2)

4.2. Optimization study: case 1

In Case 1, the cost of the land was neglected. Therefore, minimizing the CoE is analogous to maximizing the AEP. Table 6 shows the performance of optimized wind farm under Case 1 assumptions. The CoE shown in Table 6 was computed a posteriori and also accounts for the cost of the terrain under Case 2 terrain division. It can be seen that maximizing the AEP also reduced the CoE. The AEP in the optimized wind farm layout increased by 1.3%. The CoE under PA and PT terrain cost models decreased by 1.5% and 1.2% respectively. The total farm area only decreased by 7%, whereas the combined area of the convex hulls decreased by 25%.

Fig. 9a and Fig. 9b show the optimized farm layout with superimposed contours of terrain divisions in Case 2 and Case 3, respectively. Under the terrain divisions of Case 2, Regions A, B, C and D feature five, 10, four and six turbines, respectively. Under the terrain divisions of Case 3, there are six turbines within the restricted region. It can be seen that the turbines in the layout are uniformly scattered throughout the domain with large inter-turbine distances. Such configurations are expected to change when the cost of the land is also incorporated into the optimization.

4.3. Optimization study: case 2

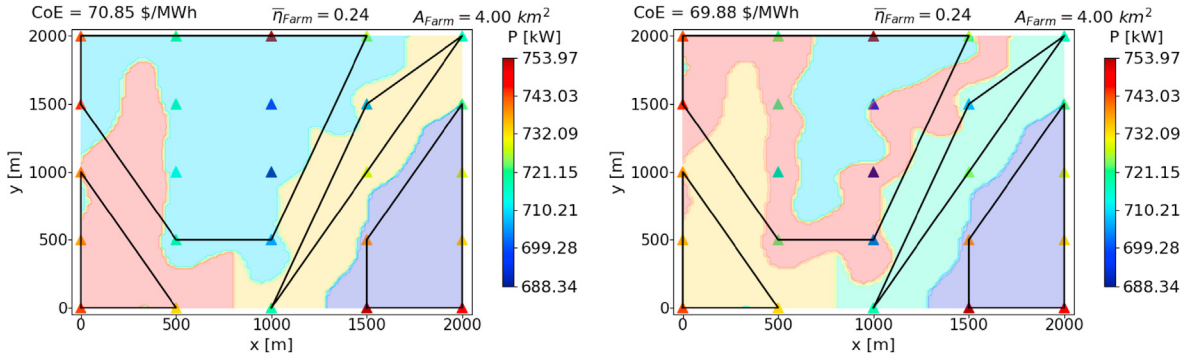
The Case 2 assumptions, which incorporate the cost of the land, should result in vastly different layouts than in Case 1 with the largest decrease in CoE. Table 7 shows the performance of the optimized wind farms. Optimization using the PA land cost model resulted in a 1.3% increase in AEP and an 2% decrease in CoE. It can be seen that the total land used in all the region decreased by 48%. Therefore, the significant reduction in CoE is attributed to the reduction in land costs and not necessarily an increase in AEP. The wind farm layout optimized using the PT terrain cost model resulted in a 1.3% increase in AEP and a 1.3% decrease in CoE. The total land used also decreased by 12%. This shows that the decrease in CoE is attributed to both the increase in AEP and decrease in land costs.

Fig. 10 shows the two optimized wind farm layouts under the PA and PT land cost model. It can be seen that the two layouts are vastly different. In the farm layout obtained under PA cost model (Fig. 10a), the layouts in each region occupy a smaller area than those obtained using the PT cost model (Fig. 10b). The wind turbines in Region D are distributed along a line, again occupying a minimum area in the most expensive region. Regions A, B, C and D contain five, 12, five, and three wind turbines, respectively. The resulting configurations feature arrangements where the wind turbines are clustered near one another to decrease the overall area occupied.

The layout obtained under the PT cost model (Fig. 10b) shows a more even distribution of the wind turbines throughout the four regions thereby occupying a larger area. This is because the farm cost is not dependent on the area occupied by the wind turbines. This allows for larger inter-turbine distances throughout the farm. The increased inter-turbine spacing in the wind farm reduces the velocity deficit due to wake effects and increases the effective velocity experienced by each turbine. This increased velocity leads to increased power generation by the turbine (Fig. 6a). Regions A, B, C and D contain six, 11, four and four turbines, respectively.

4.4. Optimization study: case 3

In Case 3, the terrain division is similar to that of Case 2 but also includes a restricted region (Region R) where no turbines can be placed. This restricted region can be due to geographical obstructions such as rivers, poor soil quality or it can be due to a landowner unwilling to sell/lease the land. Table 8 shows the performance of the optimized wind farms. Optimization using the PA terrain cost model results in an increase in AEP of 0.9% and a decrease in CoE by 1.8%. The wind farm layout optimized using the PT terrain cost model resulted in a 0.9% increase in AEP and a 0.9% decrease in CoE. Comparing Tables 7 and 8, it can be seen that when using the PT model, the percent decrease in CoE and the percent increase in AEP is the same.



(a) (b)

Fig. 8. The reference wind farm layout with superimposed regions from: a) Case 2 and b) Case 3.

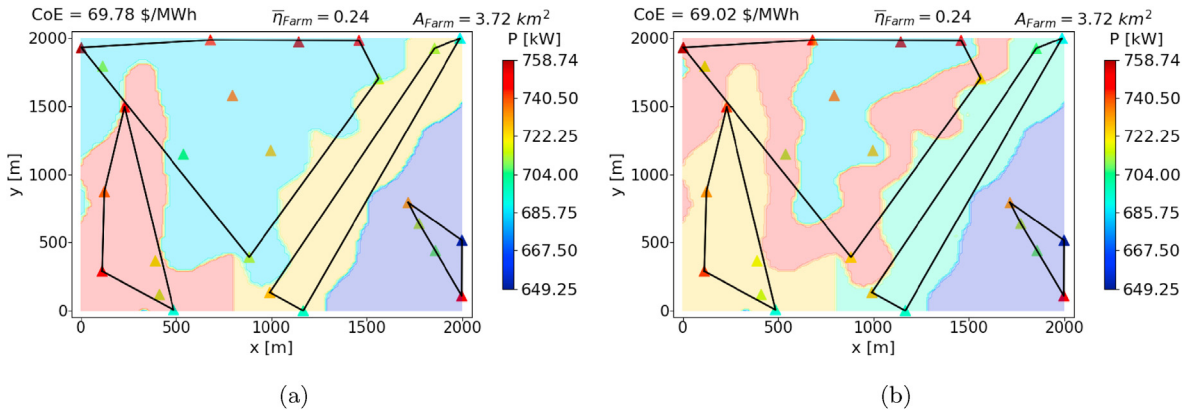


Fig. 9. The optimized wind farm layouts for Case 1 with superimposed regions from: a) Case 2 and b) Case 3.

Table 7
Performance of optimized wind farm in Case 2 and percent changes from reference farm (in brackets).

	Per Area	Per Turbine
Normalized AEP	1.37 (1.3)	1.37 (1.3)
Farm Efficiency	0.24 (−1.6)	0.24 (−0.8)
Farm Area (in km^2)	3.5 (−12.7)	3.7 (−8.2)
Area Convex Hull (in km^2)	1.37 (−48)	2.31 (−12)
CoE (in USD/MWh)	69.39 (−2.0)	68.98 (−1.3)

reduced the overall increase in AEP from 1.3% to 0.9%. Regions A, B, C and D contain seven, six, five and seven turbines, respectively.

In the case of the PT land cost model, the turbines are more uniformly distributed throughout the domain compared to the layout obtained using the PA land cost model. In Case 2 (Fig. 10b), it can be seen that seven turbines were within the restricted region. The SVDD constraint (Eq. (17)) in Case 3 displaced the seven turbines from their optimum location. This displacement reduces the overall increase in AEP from 1.4% in Case 2 down to 1.0% in Case 3. Regions A, B, C and D

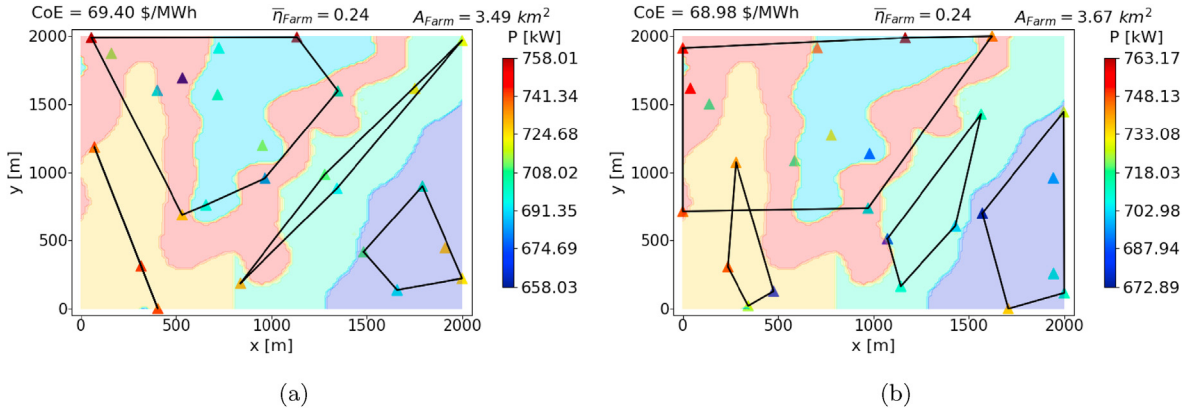


Fig. 10. The optimized wind farm layout for Case 2 under: a) per-area and b) per-turbine terrain cost model.

Fig. 11 shows the two optimized wind farm layouts under the PA and PT land cost model. In the case of PA cost model, the turbines in the most expensive region (Region D) are closely distributed along a line and occupy a smaller area. In Case 2 (Fig. 10a), it can be seen that seven turbines were within the restricted region. In Case 3, it can be seen that no turbines are placed in the restricted region. The displacement of these seven turbines from their optimum positions contain five, six, seven and seven turbines, respectively. This shows that the SVDD approach can also be used to enforce constraints and restrict placements of wind turbines within regions of complex shapes.

Table 8
Performance of optimized wind farm in Case 3 and percent changes from reference farm (in brackets).

	Per Area	Per Turbine
Normalized AEP	1.37 (0.9)	1.37 (0.9)
Farm Efficiency	0.24 (−1.6)	0.24 (−1.1)
Farm Area (in km^2)	3.23 (−19)	3.33 (−17)
Area Convex Hull (in km^2)	1.08 (−59)	1.77 (−33)
CoE (in USD/MWh)	69.59 (−1.8)	69.27 (−0.9)

5. Conclusion

This work presented an efficient method for modeling irregular terrain boundaries arising in wind farm projects. The proposed approach performs machine learning using support vector data description. The approach used support vector data description (SVDD) and kernel mapping to transform the irregular domain, encountered in wind farm development projects, into a spherical domain. This converts the irregular, complex boundary into a spherical boundary with an analytical, continuous description, thereby, allowing the use of either gradient-based or non-gradient based optimization algorithms. It was shown that an accurate SVDD model can be trained using a relatively small training set. The use of

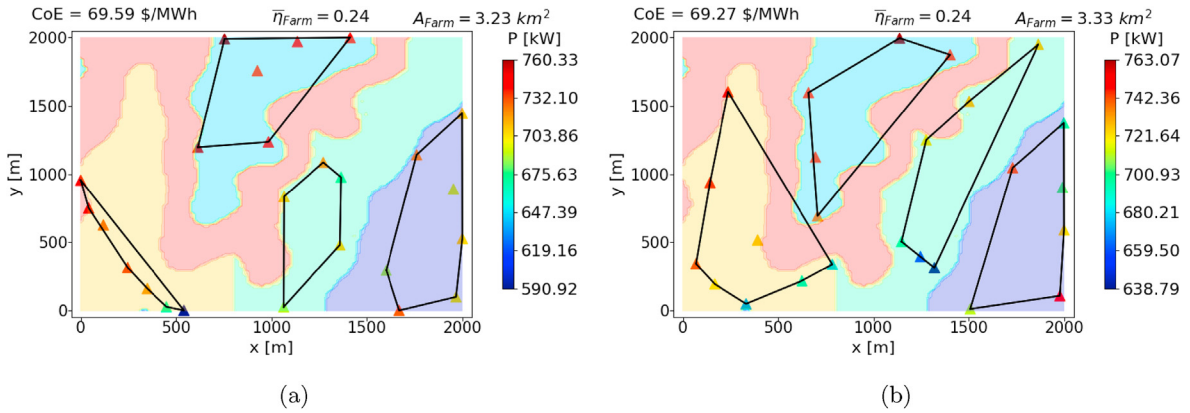


Fig. 11. The optimized wind farm layout for Case 3 under: a) per-area and b) per-turbine terrain cost model.

SVDD in wind farm development projects was demonstrated by optimizing the wind farm layout where the eligible land is owned by multiple landowners. The land for the wind farm project was assumed to be divided into multiple regions, each owned by a different individual at a different price. Two different methods for incorporating the cost of the land into the optimization framework were presented. The two approaches modeled the cost of the land on a per-area-used basis or a per-turbine-placed basis. It was shown that the final optimized layout is dependent on the land cost model used. The per-area cost model resulted in a layout that occupies the least land area, whereas the per-turbine cost model resulted in a layout with larger inter-turbine distances. The use of SVDD approach to restrict the placement of turbine in the infeasible region was also demonstrated. The results showed that the SVDD approach can not only model complex boundaries but also model several distinct regions simultaneously. The library to train the SVDD model was also released to the public.

CRediT

Sohail Reddy: Conceptualization, Methodology, Software, Data Curation, Visualization, Investigation, Validation, Writing-Original Draft Preparation, Writing-Reviewing and Editing.

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

This research was performed while the author held an NRC Research Associateship award at the Naval Postgraduate School.

References

[1] Lindenberg S. 20Contribution to US electricity supply. Diane Publishing; 2009.
[2] Wind and Water Power Technologies Office. Wind vision: a new era for wind power in the United States. US Department of Energy; 2015.
[3] Ashuri T, Zaaijer MB, Martins JR, van Bussel G, van Kuik G. Multidisciplinary design optimization of offshore wind turbines for minimum levelized cost of energy. Renew Energy 2014;68:893–905.
[4] Ashuri T, Zaaijer MB, Martins JR, Zhang J. Multidisciplinary design optimization of large wind turbines-technical, economic and design challenges. Energy Convers Manag 2016;123:56–70.
[5] Veeramachaneni K, Wagner M, O'Reilly U-M, Neumann F. Optimizing energy output and layout costs for large wind farms using particle swarm optimization. In: IEEE world congress on computational intelligence, (brisbane, Australia); 2012. p. 1–7.
[6] Chen L, MacDonald E. A system-level cost-of-energy wind farm layout optimization with landowner modeling. Energy Convers Manag 2014;77:484–94.
[7] Luo L, Zhang X, Song D, Tang W, Li L, Tian X. Minimizing the energy cost of offshore wind farms by simultaneously optimizing wind turbines and their layout. Appl Sci 2019;9:1–19.
[8] Rados K, Larsen G, Barthelmie R, Schlez W, Lange B, Schepers G, Hegberg T, Magnisson M. Comparison of wake models with data for offshore windfarms. Wind Eng 2001;25(5):271–80.
[9] Lydia M, Kumar SS, Selvakumar AI, Kumar GEP. A comprehensive review on wind turbines power curve modeling techniques. Renew Sustain Energy Rev 2014;30:452–60.
[10] McWilliam M, van Kooten G, Crawford C. A method for optimizing the location of wind farms. Renew Energy 2012;48:287–99.
[11] Song M, Chen K, He Z, Zhang X. Wake flow model of wind turbine using particle simulation. Renew Energy 2012;41:185–90.
[12] Schallenberg-Rodriguez J. A methodological review to estimate techno-economical wind energy production. Renew Sustain Energy Rev 2013;21:272–87.
[13] Carrillo C, Montano AO, Cidras J, Dias-Dorado E. Review of power curve modelling for wind turbines. Renew Sustain Energy Rev 2013;21:572–81.
[14] Ryberg DS, Robinius M, Stolten D. Evaluating land eligibility constraints of renewable energy sources in Europe. Energies May 2018;11:1–19.
[15] Sorkhabi SYD, Romero DA, Yan GK, Gu MD, Moran J, Morgenroth M, Amon CH. The impact of land use constraints in multi-objective energy-noise wind farm layout optimization. Renew Energy 2016;85:359–70.
[16] Sorkhabi SYD, Romero DA, Beck C, Amon CH. Constrained multi-objective wind farm layout optimization: novel constraint handling approach based on constraint programming. Renew Energy 2018;126:341–53.
[17] Feng J, Shen WZ, Li Y. An optimization framework for wind farm design in complex terrain. Appl Sci October 2018;8:1–18.
[18] Mittal P, Mitra K. Determination of optimal layout of wind turbines inside a wind farm in presence of practical constraints. In: 2019 fifth Indian control conference (ICC); 2019. p. 353–8 (New Delhi, India).
[19] Wang L, Tan AC, Gu Y, Yuan J. A new constraint handling method for wind farm layout optimization with lands owned by different owners. Renew Energy 2015;83:151–61.
[20] Chen Y, Li H, He B, Wang P, Jin K. Multi-objective genetic algorithm based innovative wind farm layout optimization method. Energy Convers Manag September 2015;105:1318–27.
[21] Wang L. Comparative study of wind turbine placement methods for flat wind farm layout optimization with irregular boundary. Appl Sci 2019;9:1–15.
[22] Wang L, Tan AC, Cholette ME, Gu Y. Optimization of wind farm layout with complex land divisions. Renew Energy 2017;105:30–40.
[23] Gonzalez JS, Garcia ALT, Payan MB, Santos JR, Rodriguez AGG. Optimal wind-turbine micro-siting of offshore wind farms: a grid-like layout approach. Appl Energy 2017;200:28–38.
[24] Perez-Moreno SS, Dykes K, Merz K, Zaaijer M. Multidisciplinary design analysis and optimization of a reference offshore wind plant. J Phys Conf 2018;1037:1–15.
[25] Stanley APJ, Ning A. Massive simplification of the wind farm layout optimization problem. Wind Energy Science 2019;4:663–76.
[26] Stetco A, Dinmohammadi F, Zhao X, Robu V, Flynn D, Barnes M, Keane J, Nenadic G. Machine learning methods for wind turubine condition monitoring: a review. Renew Energy 2019;133:620–35.
[27] Ti Z, Deng XW, Yang H. Wake modeling of wind turbines using machine learning. Appl Energy 2020;257:114025.
[28] Clifton A, Kilcher L, Lundquist J, Fleming P. Using machine learning to predict wind turbine power output. Enviornmental Research Letter 2013;8:024009.
[29] Howland MF, Dabiri JO. Wind farm modeling with interpretable physics-informed machine learning. Energies 2019;12:1–21.
[30] Torres J, Aguilar R, Zuniga-Meneses K. Deep learning to predict the generation

of a wind farm. J Renew Sustain Energy 2018;10(1):013305.
[31] Reddy SR. libSVDD: a library for support vector domain description. 2020. https://doi.org/10.5281/zenodo.3712443. https://github.com/sohailrreddy/libSVDD.
[32] Tax DM, Duin RP. Support vector domain description. Pattern Recogn Lett 1999;20:1191–9.
[33] Vapnik V. The nature of statistical learning theory. New York: Springer; 1995.
[34] Aronszajn N. Theory of reproducing kernels. Trans Am Math Soc May 1950;68:337–404.
[35] Hasiuk FJ, Harding C, Renner AR, Winer E. TouchTerrain: a simple web-tool for creating 3D-printable topographic models. Comput Geosci 2017;109:25–31.
[36] Reddy SR. Wind farm layout optimization (WindFLO): a framework for fast wind farm layout optimization. 2019. https://doi.org/10.5281/zenodo.3694675. https://github.com/sohailrreddy/WindFLO.
[37] Reddy SR. Wind farm layout optimization (WindFLO) : an advanced framework for fast wind farm analyis and optimization. Appl Energy May 2020;269. https://doi.org/10.1016/j.apenergy.2020.115090. 115090.
[38] Midwestern Regional Climate Center. cli-MATE. date retrieved December, 2019, https://mrcc.illinois.edu/CLIMATE/; 2000.
[39] Frandsen S, Barthelmie R, Pryor S, Rathmann O, Larsen S, Hājstrup J, ThÃgersen M. Analytical modelling of wind speed deficit in large offshore wind farms. Wind Energy 2006;9:39–53.
[40] Ratliff DJ, Hartman CL, Stafford ER. An analysis of state-level economic impacts from the development of wind power plats in summit county, Utah. In: Center for the market diffusion of renewable energy and clean technology: DOE/GO-102009-2918; 2009. p. 1–29.
[41] Meyers J, Meneveau C. Optimal turbine spacing in fully developed wind farm boundary layers. Wind Energy 2012;15(2):305–17.
[42] Reddy SR. PhD thesis. In: Many-objective hybrid optimization under uncertainty with applications. Miami, FL, USA: Florida International University; 2019.
[43] Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Comput Apr. 2002;6:182–97.
[44] Reddy SR. An Efficient Method for Modeling Terrain and Complex Terrain Boundaries in Constrained Wind Farm Layout Optimization. Renewable Energy 2021;165:162–73. https://doi.org/10.1016/j.renene.2020.10.076.