A Novel Framework for Spatiotemporal Analysis of Temperature Profiles Applied to Europe

Jamaer, S.; Allaerts, D.; Meyers, J.; Van Lipzig, N. P.M.

DOI
10.1175/JAMC-D-22-0205.1

Publication date
2023

Document Version
Final published version

Published in
Journal of Applied Meteorology and Climatology

Citation (APA)
Jamaer, S., Allaerts, D., Meyers, J., & Van Lipzig, N. P. M. (2023). A Novel Framework for Spatiotemporal Analysis of Temperature Profiles Applied to Europe. *Journal of Applied Meteorology and Climatology*, *62*(12), 1855-1873. https://doi.org/10.1175/JAMC-D-22-0205.1

A Novel Framework for Spatiotemporal Analysis of Temperature Profiles Applied to Europe

S. JAMAER,[a] D. ALLAERTS,[b] J. MEYERS,[c] AND N. P. M. VAN LIPZIG[a]

[a] *Department of Earth and Environmental Sciences, KU Leuven, Leuven, Belgium*
[b] *Faculty of Aerospace Engineering, TU Delft, Delft, Netherlands*
[c] *Department of Mechanical Engineering, KU Leuven, Leuven, Belgium*

(Manuscript received 11 January 2023, in final form 3 October 2023, accepted 4 October 2023)

ABSTRACT: Vertical temperature profiles influence the wind power generation of large offshore wind farms through stability-dependent effects such as blockage and gravity waves. However, numerical tools that are used to model these effects are often computationally too expensive to cover the large variety of atmospheric states occurring over time. Generally, an informed decision about which representative nonidealized situations to simulate is missing because of the lack of easily available information on representative vertical profiles, taking into account their spatiotemporal variability. Therefore, we present a novel framework that allows a smart selection of vertical temperature profiles. The framework consists of an improved analytical temperature model for the atmospheric boundary layer and lower troposphere, a subsequent clustering of these profiles to identify representatives, and last, a determination of areas with similar spatiotemporal characteristics of vertical profiles. When applying this framework on European ERA5 data, physically realistic representatives were identified for Europe, excluding the Mediterranean. Two or three profiles were found to be dominant for the open ocean, whereas more profiles prevail for land. Over the open ocean, weak temperature gradients in the boundary layer and a clear capping inversions are widespread, and stable profiles are absent except in the region of the East Icelandic Current. Interestingly, according to the ERA5 data, at its resolution, coastal areas and seas surrounded by land behave more similar to the land areas than to the open ocean, implying that a larger set of model integrations are needed for these areas to obtain representative results for offshore wind power assessments in comparison with the open ocean.

SIGNIFICANCE STATEMENT: Numerical tools used to simulate the effect of large, offshore wind farms on neighboring farms and the atmosphere are very expensive. Therefore, they can only be computed for a limited number of cases. As temperature is one of the most important parameters in these kinds of simulations, this work provides a new vertical temperature model and an analysis framework that allows for a smart selection of these cases such that they ideally represent the full variation of the atmosphere's temperature profiles.

KEYWORDS: Europe; Boundary layer; Temperature; Renewable energy; Clustering

1. Introduction

Wind energy is projected to play an important role in the transition toward a sustainable energy mix [International Energy Agency IEA (2021)]. Therefore, accurate wind resource assessments and power forecasting are crucial to develop new wind farms and to balance the electricity grid. The emergence of large, offshore wind farms introduces new challenges. More specifically, to accurately predict the energy yields and the influence that different wind farms have on the atmosphere and neighboring farms, effects such as wind farm wakes (Emeis et al. 2016; Platis et al. 2018), blockage (Bleeg et al. 2018; Branlard et al. 2020), gravity waves (Allaerts and Meyers 2017; Allaerts et al. 2018; Ollier et al. 2018), and entrainment from above (Andersen et al. 2017; Kelly et al. 2019) should be taken into account.

Currently, different models are used to study this interaction of wind turbines and farms on the one hand, and the atmosphere on the other hand. Large-eddy simulations (LES) are frequently used to simulate the effect that a wind turbine or a single wind farm has on the atmosphere and energy production in detail (Porté-Agel et al. 2020; Allaerts and Meyers 2015; Lanzilao and Meyers 2022). Mesoscale models are used for studies over larger areas with multiple wind farms and possibly over longer periods (Fitch et al. 2013; Chatterjee et al. 2016; Porchetta et al. 2021; Fischereit et al. 2022; Docquier et al. 2016). Both LES and mesoscale models are computationally expensive, making it difficult to simulate all occurring atmospheric conditions. Therefore, they are typically only applied to a limited number of atmospheric states or time periods. Mesoscale simulations usually are integrated over long time periods without making an a priori subselection of states, although there are studies that are based on weather type approaches using the surface level pressure to find suitable states (Reyers et al. 2015; Lamb 1972; Souverijns et al. 2016; Demuzere et al. 2009; Brisson et al. 2011). LES typically needs vertical profiles of physical quantities that are selected based on theoretical assumptions and model limitations (Porté-Agel et al. 2020; Allaerts and Meyers 2015; Lanzilao and Meyers 2022). For wind speed profiles, advancements in smart selection have already been made by Schelbergen et al. (2020), however, this kind of selection is still missing for temperature profiles.

Corresponding author: S. Jamaer, sebastiaan.jamaer@kuleuven.be

DOI: 10.1175/JAMC-D-22-0205.1

Vertical temperature profiles are a key parameter for these simulations since large-scale effects influencing energy production depend on the stability of the surface layer (SL) and the atmospheric boundary layer (ABL) (Platis et al. 2022; Argyle and Watson 2014; Platis et al. 2018). For example, a stably stratified ABL is less turbulent, and therefore the wind farm wakes dissipate slower, while the increased turbulence levels in an unstable ABL dissipates wakes more efficiently, reducing their length (Platis et al. 2021, 2022). Furthermore, wind farm blockage and gravity waves (GWs) are strongly linked to the capping inversion (CI), as it traps the air in the ABL and prevents entrainment from the troposphere. The strength of GWs also depends on the stratification of the lower troposphere (Schneemann et al. 2021; Devesse et al. 2022).

A smart selection of the vertical temperature profiles for wind farm simulations can therefore be beneficial for future studies regarding farm–farm and farm–atmosphere interactions. However, because of a lack of easily accessible data and data mining techniques, the choice of these background profiles is rarely informed by a temporal analysis of the ABL temperature profiles in the area of interest. This implies that the background temperature profiles for expensive LES simulations are possibly not ideal for offshore wind energy applications and extrapolating location-specific results to other locations is difficult. Hence, a comprehensive overview of these temperature profiles, and more in general the stability of the ABL and lower troposphere, support an improved selection procedure of cases, more representative for the conditions occurring at the specific site.

Previous temperature profile studies are generally focused on one aspect of the profile or on one type of profile. Studies focus either on the surface layer stability (Baas et al. 2016; Argyle and Watson 2014), CI characteristics (Rampanelli and Zardi 2004), the boundary layer height (BLH; Dang et al. 2019; Seibert et al. 2000), the stratification of the free atmosphere (Schneider 2007), or studies focus on just stable (Mahrt and Acevedo 2023) or neutral/unstable profiles (Liu and Stevens 2022). Studies rarely treat the entire profile in an unknown stability state (with a notable exception in Lang et al. 2018). One reason for this is the absence of a generally applicable temperature model for the lower atmosphere. Therefore, an existing analytical temperature model (Rampanelli and Zardi 2004) was further developed to allow for a spatiotemporal data analysis of temperature profiles up to the lower troposphere regardless of the atmospheric stability of the ABL.

Currently, analytical temperature models can broadly be divided into two kinds: piecewise linear (Deardorff 1979; Fedorovich and Mironov 1995; Batchvarova and Gryning 1994) and continuously differentiable fits (Rampanelli and Zardi 2004; Frei 2014). For the piecewise linear fits, the temperature profile is assumed to consist of approximately piecewise linear parts, possibly with a discontinuous jump (zeroth-order jump) at the CI. These piecewise linear models are often combined with a constant temperature for the mixed layer to model the neutral and unstable ABLs. Generally, unconstrained piecewise linear fits are suitable for all kinds of stability criteria. However, they often can be difficult to interpret and, because they are not continuously differentiable, they more easily end up in a local minimum during optimization.

The continuously differentiable Rampanelli and Zardi model (RZ) (Rampanelli and Zardi 2004) is, to the authors' knowledge, the only developed continuously differentiable model for potential temperature profiles. RZ uses a linear combination of basis functions that each model a part of the temperature profile. By defining each of these functions over the entire height of the profile, they ensure a continuously differentiable function, allowing for higher-order optimization algorithms. Furthermore, their basis functions are based on physical insight and, therefore, they are suited for physical interpretation of the fits. However, the RZ model is only usable for neutral boundary layers. Therefore, it is not suitable to use in a general setting where the stability of the surface layer is unknown. In this work, we extend the original RZ model with an extra basis function to create the surface extended Rampanelli and Zardi model (SERZ), which is applicable to both stable and unstable boundary layers.

With this improved model, a spatiotemporal analysis of the vertical temperature profiles of Europe (excluding the Mediterranean area) is carried out to create a reference atlas with representative temperature profiles, with a focus on offshore applications. The reanalysis data ERA5 (Hersbach et al. 2020) are used as input data for this work, as it has a relatively high resolution, both in time and space, for a global multiyear dataset. Our approach follows three steps: First, the ERA5 data are fitted with the analytical SERZ model up to a height of 5000 m. Second, the resultant fitting parameters are clustered to create several representative temperature profiles that ideally span the atmospheric variation. Third, with these representatives, each point in space, which represents the weighted volume averages of the data in the spatial discretization, is assigned a fingerprint and these fingerprints are analyzed with a second clustering algorithm. These points in space will further be referred to as pixels and their fingerprints as pixel fingerprints.

The resulting product of this multiclustering approach is a temperature profile atlas for Europe with clusters of pixels that behave similarly in their atmospheric stratification up to 5000 m. Each of those clusters has a different prevalence of each representative profile, showing the wide variation of temperature profiles over Europe. This atlas is useful for subselecting atmospheric conditions for computationally expensive models, such as LES or mesoscale atmospheric model studies for wind energy and can be used to compare simulations from different sites.

This paper is structured as follows. In the data and methods section (section 2), the domain and variables of the data, the new SERZ model, and the general fingerprinting framework are presented. Afterward, in the results and discussion section (section 3), the SERZ model and fingerprinting algorithm are applied to Europe to obtain representative profiles and clusters of similar points. A section with brief conclusions and future work is added to complete this paper (section 4).

2. Data and methods

a. Data

The virtual potential temperature profiles were extracted from ERA5 data on their native hybrid vertical discretization (Hersbach et al. 2020), which are assumed to be good

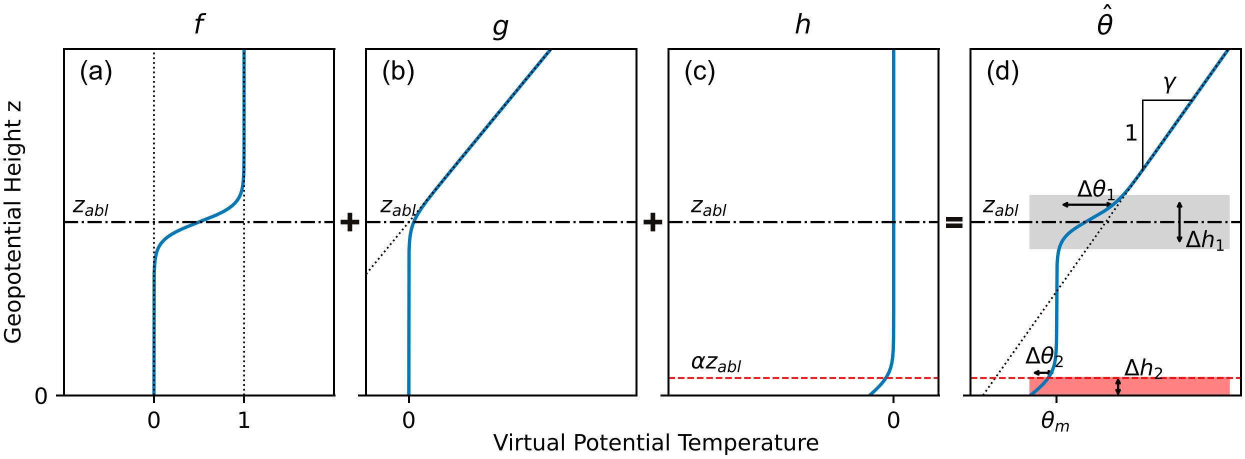

FIG. 1. Construction of (d) an example temperature profile $\hat{\theta}$ using the vertical temperature model consisting of the linear combination of the three basis functions (a) f, (b) g, and (c) h. Physical parameters that can be extracted from the SERZ model are indicated in (d): the stratification of the free atmosphere γ, ABL height z_{abl}, CI width Δh_1, CI strength $\Delta\theta_1$, SL height Δh_2, and SL strength Δh_2.

approximations for the real temperature profiles of the atmosphere. Because of computational constraints, a subset of only one year of a window centered over the North Sea was used. The window spans 25°W–25°E and 40°–70°N over the year 2020 at a resolution of 0.28° (see appendix B for a map). This domain contains both parts of the Atlantic Ocean as well as the majority of Europe, which allows for a comparison between the types of surface. However, the focus will be on offshore and open ocean areas, since this study is mainly tailored toward offshore wind energy.

Since we are only interested in the lower atmosphere, the height of the profiles was limited to 5000 m, which allows for a characterization of the ABL stratification as well as the stratification of the lower troposphere. The ERA5 data up to 5000 m contain, on average, 40 height levels, where the levels are more densely spaced close to the surface. These height levels (shown in Fig. B2) are vertical hybrid coordinates from ERA5 (Hersbach et al. 2020) that follow the orography close to the surface and converge to pressure levels higher in the atmosphere. The temporal resolution was the native hourly resolution of ERA5, resulting in approximately 166×10^6 profiles.

b. SERZ

The stability of the ABL and the lower troposphere is mainly determined by the vertical potential temperature gradient. Traditionally, the stratification of the ABL is split into three different kinds: stable, neutral, and unstable (Stull 1988). The stable ABL typically has a positive potential temperature gradient, which is often assumed to be constant. The conventionally neutral and unstable ABL are relatively similar and are typically modeled with a constant temperature in the ABL, topped by a capping inversion. Additionally, the unstable ABL has a negative potential temperature gradient close to the surface (typically the lowest 5%–10% of the ABL).

As most clustering algorithms that generate representative cases have difficulties with high-dimensional data, the raw ERA5 data with their 40 height levels need to be preprocessed to reduce the number of variables. Therefore, an extension to the Rampanelli and Zardi model was implemented that results in a set of seven parameters for each profile. The RZ model was chosen as it is still used in many wind energy applications (Lanzilao and Meyers 2022; Allaerts and Meyers 2017), and by using a physics-based analytical model, derivation of physical parameters such as the ABL height and CI strength is more natural than when using more abstract models such as autoencoders or piecewise linear fits.

The original vertical temperature model of Rampanelli and Zardi (2004) was developed for conventionally neutral and weakly convective boundary layers. The model consists of the linear combination of two basis functions, f (Fig. 1a) and g (Fig. 1b), that model the CI and the stratification of the troposphere, respectively. The main parameters of this model are the linear fit parameters, denoted with a, b, and θ_m, which mainly determine the CI strength, free lapse rate, and the temperature in the mixed layer, respectively, and two nonlinear parameters, the height and width of the CI (z_{abl} and Δh_1). The two nonlinear parameters z_{abl} and Δh_1 are used to create a nondimensional height η on which the basis functions f and g are applied. The original RZ model's formulation is summarized by the following equations (Rampanelli and Zardi 2004):

$$\hat{\theta}_{\text{RZ}}(z) := \theta_m + af(\eta) + bg(\eta), \tag{1}$$

where f, g, and η are defined as

$$f(\eta) = \frac{\tanh(\eta) + 1}{2},$$

$$g(\eta) = \frac{\ln[2\cosh(\eta)] + \eta}{2}, \quad \text{and}$$

$$\eta = \frac{z - z_{\text{abl}}}{C_1 \Delta h_1},$$

with $C_1 = 3/2$ being a constant; z is the geopotential height above the surface.

When $\eta \to -\infty$, or, equivalently, $z \ll z_{\rm abl}$, both f and g will approach zero. This means that the original RZ model is not able to model a boundary layer with surface layer that has a nonzero temperature gradient. Therefore, to accommodate for atmospheres with a nonconstant surface layer temperature or a stable boundary layer, an extra basis function h was included in the model:

$$\hat{\theta}(z) := \theta_m + af(\eta) + bg(\eta) + ch(\eta'), \quad (2)$$

where

$$h(\eta') = -\frac{\ln[2\cosh(\eta')] - \eta'}{2},$$

$$\text{with } \eta' = \frac{z - \alpha z_{\rm abl}}{C_2};$$

c and α are trainable parameters, and $C_2 = 100$ m a constant that is related to the shape of the temperature profile of the surface layer and that was chosen such that the optimization remained well behaved [e.g., the function $h(\eta')$ remained smoothly differentiable at our resolution]. The extra parameter α determines the height relative to the boundary layer height where the boundary layer deviates from a constant temperature. The three basis functions of the SERZ model are shown in Figs. 1a–c.

To find the SERZ parameters from a vertical temperature profile, an optimization problem was solved. In this work, the optimal fit of a temperature profile is defined as the fit that minimizes the mean-squared error (MSE) between the data and the approximation model. However, a simple least squares optimization can sometimes result in an ill-conditioned problem that yields unrealistic results. As this ill conditioning is mainly caused by large values of $z_{\rm abl}$, two extra constraints are added to the optimization problem that limits this parameter. The full optimization problem is then given by

$$\underset{\Theta}{\text{minimize}} \quad S = \frac{1}{N}\sum_{i=1}^{N}[\theta_i - \hat{\theta}(z_i;\Theta)]^2, \quad (3)$$

subject to

$$0 \le \theta_m, a, b, z_{\rm abl}, \Delta h_1, \quad (4)$$

$$0 \le \alpha \le 1, \quad (5)$$

$$\alpha z_{\rm abl} < z_{\rm abl} - \frac{\Delta h_1}{2}, \quad \text{and} \quad (6)$$

$$z_{\rm abl} + \frac{\Delta h_1}{2} < 4000 \text{ m}, \quad (7)$$

where N is the number of height levels, Θ the set of fitting parameters ($\Theta = [a, b, c, \theta_m, z_{\rm abl}, \Delta h_1, \alpha]$), the first two constraints are simple bounds, and the last two constraints are there to ensure a limited Δh_1 and $z_{\rm abl}$. The constraint $\alpha z_{\rm abl} < z_{\rm abl} - (\Delta h_1/2)$ ensures that the surface layer and CI do not overlap, and the constraint $z_{\rm abl} + (\Delta h_1/2) < 4000$ m ensures that the maximal height of the CI is 4000 m. These constraints were effectively reached in 6% for our profiles, where the majority (>5%) of these cases reached the constraint $0 \le \alpha \le 1$ and the other constraints were only reached in less than 0.6% of the time. Further implementation details can be found in appendix A.

The parameters obtained by solving the optimization problem can in some cases be linked to physical quantities of the boundary layer and the lower troposphere, as shown in Fig. 1d. Since the SERZ model is based on the RZ model, many of the interpretations from Rampanelli and Zardi (2004) can be used for known convective and conventionally neutral boundary layers. Quantities that were derived by Rampanelli and Zardi (2004) are the stratification rate in the lower troposphere, the boundary layer height, the (constant) boundary layer temperature, the capping inversion width, and the temperature jump of the capping inversion (if present). The boundary layer height, boundary layer temperature, and capping inversion width follow straight from the optimization problem in the parameters $z_{\rm abl}$, θ_m, and Δh_1, respectively. The stratification coefficient γ and the temperature jump across the CI $\Delta\theta_1$ are given by

$$\gamma = \frac{b}{C_1 \Delta h_1} \quad \text{and} \quad (8)$$

$$\Delta\theta_1 = a + \frac{b}{2C_1}. \quad (9)$$

Note that the equation for $\Delta\theta_1$ differs from the one in RZ (Rampanelli and Zardi 2004); yet, it was confirmed that this is the correct expression for $\Delta\theta_1$ (D. Zardi 2021, personal communication).

The height and strength of the surface layer that is added in this work through the extra function h require more discussion. Since h is a continuous function that only reaches zero in the limit $\eta' \to +\infty$, a significant surface layer can be present even if $\alpha = 0$. In these cases, the (absolute) value of c will also be large, which, together with the fact that $h(0) = [\ln(2)]/2$, allows a substantial surface layer even if $\alpha = 0$. To account for these effects, the height of the surface layer Δh_2 is defined as the height on which the profile reaches a band of width $2k$ centered around θ_m. The height Δh_2 is found by solving

$$|h[\eta'(\Delta h_2)]| = k = \left|\frac{T_r}{c}\right|, \quad (10)$$

where T_r is a temperature threshold taken equal to 0.5 K. Solving this using the formulation of h gives

$$\Delta h_2 = \max\left[\alpha z_{\rm abl} - \frac{C_2}{2}\ln(e^{2k} - 1), 0\right]. \quad (11)$$

Note that there is a distinct difference between Δh_1 and Δh_2 in their meaning and the way they are computed: Δh_1 is the depth of the CI and is one of the parameters that is directly computed by the optimization. Δh_2 is the height of the surface layer and is not included in the parameters of the optimization but derived from its fitting parameters.

Last, the surface layer strength $\Delta\theta_2$ is given by

$$\Delta\theta_2 = -ch(\eta_0'), \tag{12}$$

where

$$\eta_0' = \frac{-\alpha z_{\mathrm{abl}}}{C_2}. \tag{13}$$

Since $h(\eta') < 0$ for all η', a positive c corresponds to an increase in temperature in the surface layer. This definition of $\Delta\theta_2$ can be used since the basis function h is the only significant contribution to the SL strength, given the constraint $\alpha z_{\mathrm{abl}} < z_{\mathrm{abl}} - (\Delta h_1/2)$.

ON THE USE OF THE PHYSICAL QUANTITIES

As explained above, physically meaningful quantities can be derived from the fits. However, it should be kept in mind that these quantities are purely derived from the mathematical fits, which can result in flawed physical values. This mainly follows from two limitations of our model.

First, because of the nonlinearity, the optimization problem is nonconvex, which means that there can be multiple local minima. Therefore, the optimization problem can in some cases depend on the choice of initial conditions.

Second, even if the optimization problem ends in the global minimum, it is not guaranteed to also be physically accurate, for example, the value of z_{abl} or Δh_1 may not correspond to their respective physical quantities. This is because vertical temperature profiles are seldom ideal. Hence, if the overall curvature of the profile is better fitted with a large Δh_1 when compared with a fit where Δh_1 models the actual capping inversion, the optimization problem will end with a large Δh_1.

Therefore, it is the authors' opinion that, whenever possible, the analysis should be carried out on the raw fitting parameters instead of analysis on the extracted physical values. However, if physical parameters are necessary (e.g., the ABL height for LES), the optimization problem can be made more robust by tailoring the constraints to the user's case. For example, a constraint on Δh_1 can be added when requiring a limited depth of the CI, or a constraint on α can be added to limit the height of the surface layer. Similarly, our code base allows users to include regularization on parameters in the optimization problem by adding a Tikhonov regularization term (see appendix A). This regularization allows the user to penalize unrealistically large values for the fitting parameters by adding a weighted, quadratic term to the cost function in Eq. (3). Tuning these weights then becomes an integral part of the application that we do not consider further.

The primary goal of this work is to create representatives that are as accurate as possible, thus, it omits any regularization and limits itself to the absolutely necessary constraints, as this will generate the best fits with respect to the mean squared error. Therefore, the resulting representative profiles are accurate, but also more difficult to interpret.

c. *Fingerprint analysis*

The SERZ fit of the ERA5 data results in a new dataset of fitting parameters (a, b, c, θ_m, z_{abl}, Δh_1, and α) that are defined over a spatial grid over a period of one year. The presence of temporal and spatial variation in the multidimensional data complicates its analysis and interpretation. To tackle this, a simple fingerprinting framework based on a two-stage clustering algorithm is proposed. This framework will result in representative profiles and the temperature profile atlas and offers several advantages over traditional analyses of the mean and standard variation of each parameter:

- By using a representative set, all aspects of the profile are combined and analyzed together. Therefore, the algorithm automatically incorporates correlations between different components of the profile, and it produces physically realistic representative profiles.
- The number of representatives can be chosen independently from the number of parameters in the original dataset. This is particularly useful when the number of parameters becomes too large to manually interpret and cross correlate.
- By automatically determining spatial clusters in the data, the fingerprinting framework eliminates possible bias relative to when the areas would be identified manually.

In our application, the inputs of the framework are the SERZ fit parameters without the parameter θ_m, as stability is mainly determined by the temperature gradient, and is therefore independent of the absolute temperature. With this choice of parameters, the temperature gradient of the SL (c and α), CI (a and Δh_1), and free atmosphere (b), as well as the boundary layer height z_{abl} are represented in the dataset.

The first part of this fingerprinting algorithm (shown in Fig. 2) computes the individual fingerprints of each pixel. To do this, the fingerprinting algorithm starts by flattening the data over time and space and normalizing it using a quantile transformation (QT) with 1000 bins, which maps the empirical quantiles on the quantiles of the normal distribution. Then, the Mini-batch kMeans++ algorithm (Sculley 2010), initialized using the kMeans++ procedure, finds cluster centers by iteratively updating the centers with batch size 2560 and a reassignment rate of 0.1. After convergence, each point in time and space is assigned to one of the representatives and these assignments are used to compute the occurrence rate of each representative for each point in space. This creates a frequency distribution for each pixel, which is referred to as the *fingerprint* of the pixel.

After removing the temporal component from the signal in the aforementioned way, the second step of our spatiotemporal data analysis algorithm groups the fingerprints to find areas with similar behavior, allowing for easier interpretation and visualization. To this end, a bottom-up agglomerative clustering with Ward linkage, implemented with the scikit-learn package (Pedregosa et al. 2011), was used that recursively merges points together based on the resulting variance reduction until the desired number of clusters is achieved. This algorithm was chosen because it allows for a direct comparison between instances with different numbers of clusters.

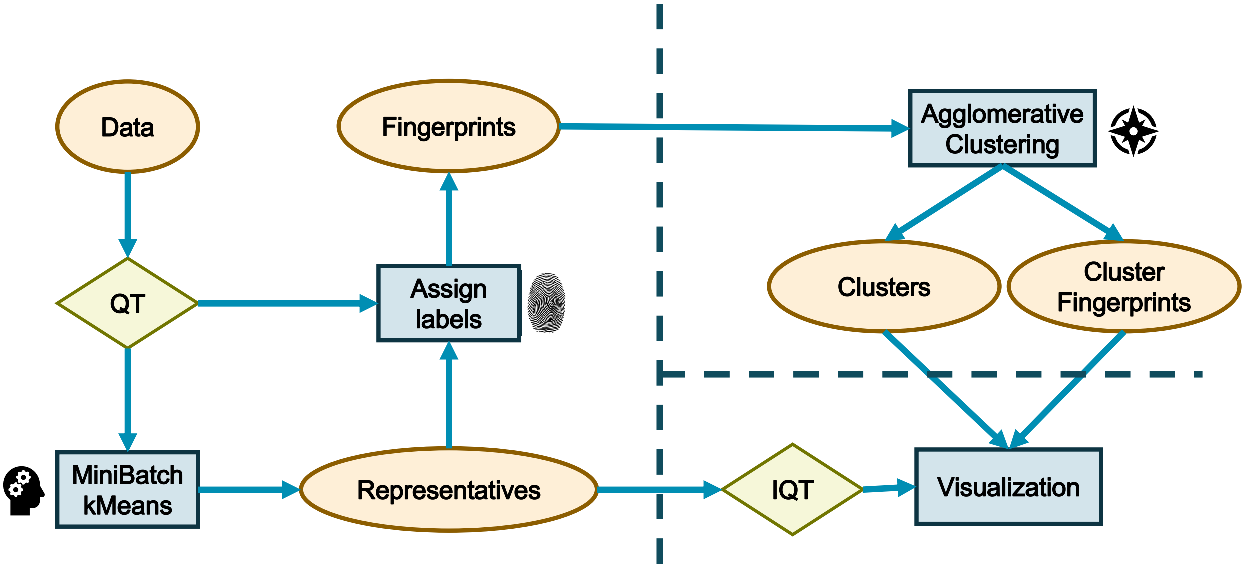

FIG. 2. Setup of the fingerprinting analysis. First, the data are preprocessed using a quantile transformer (QT). Second, the MiniBatch kMeans clustering algorithm is used to find the representatives and compute the fingerprints. Third, agglomerative clustering is applied to the fingerprints to generate clusters and cluster fingerprints. Fourth, the representative profiles are transformed back to the original space with the inverse quantile transformer (IQT) and the results are visualized.

Last, the inverse quantile transformation (IQT), that is, the reversed mapping of the bins of the distributions, is applied to the representatives and the results are visualized. The fingerprints of each pixel are visualized as bar plots (see Fig. 3a for an example) where each bar represents the occurrence rate of profiles similar to the relative representative. These representatives are ordered based on their value of *c*, which is similar to an ordering the temperature gradient of the surface layer. Profiles with very high and very low values for *c* correspond to boundary layers with surface layers that are stable and unstable, respectively, which are equivalent to large surface-induced cooling and heating, respectively.

When fingerprints are grouped in a cluster, an aggregated fingerprint is defined. This fingerprint is defined as a bar graph with error bars, where the error bars indicate the 90% confidence interval. As an example, Fig. 3 shows a hypothetical fingerprint and the corresponding aggregated fingerprints and guides the reader in interpreting the fingerprint plots. In this example, the fingerprint of a single pixel (Fig. 3a) shows that it is classified to P4 for more than 25% of the time. The aggregated fingerprint of the relevant cluster (Fig. 3b) then shows that 90% of pixels within that cluster have an occurrence rate for the fourth representative between 26% and 32%.

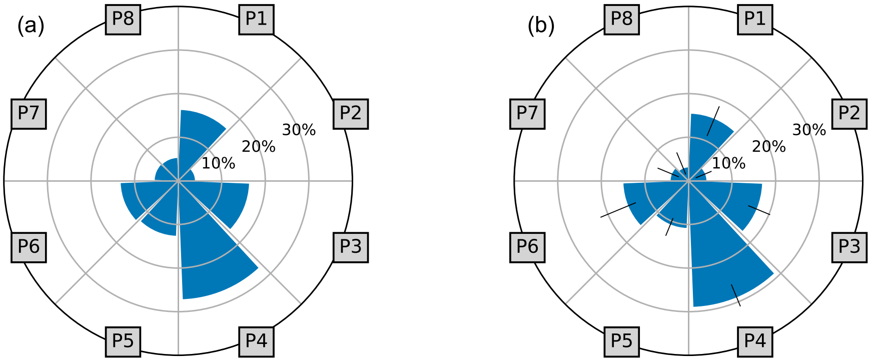

FIG. 3. (a) A fictional example of an individual fingerprint that was computed using eight representatives. Each bar represents the occurrence rate of the corresponding class of profiles that are associated with the representative, and each concentric circle is a 10% interval. (b) A fictional example of an aggregated fingerprint that was computed using eight representatives. Each bar represents the median occurrence rate of the corresponding class of profiles, and each error bar corresponds to the 90% confidence interval for the occurrence rate for points in the cluster. The PX labels in the squares refer to hypothetical representatives of the example.

3. Results and discussion

The results of this work are grouped into two parts. First, the performance of the SERZ model is compared with the original RZ model. Second, the fingerprinting analysis is applied for one year of SERZ data with which we identify areas with similar behavior in temperature profiles and link the physical parameters of these areas with physical systems. Additionally, an analysis of seasonal and diurnal differences is carried out, both to gain insight into the profile behavior and to see if the most well-known characteristics of the temperature profiles can be distinguished. To make the results from the complete period comparable with the seasonal and diurnal analysis, the representative profiles are always computed using the full dataset and only the clusters were computed separately for the seasons and time stamps and subsequently matched as well as possible to the clustering colors of the clustering with the full dataset. The seasonal analysis applies the fingerprint analysis to summer months (June–August) and winter months (December–February) separately. The diurnal analysis limits the fingerprint analysis to daily hours (1200–1700 UTC) and nightly hours (0000–0500 UTC). Please note that the diurnal division was made based on UTC and not on local time; however, since our domain is centered around the Greenwich meridian, the UTC is the mean time of our domain. Hence, since our window only spans up to longitudes with time zones UTC ± 2, the approximation with the UTC time for all pixels is justified.

a. *SERZ evaluation*

The original RZ model was not developed to be applicable for stable profiles. Therefore, the MSE was computed separately for stable and unstable profiles, where the stability parameter ζ (Stull 1988) was used to classify the profiles. The stability parameter ζ was computed using the expression:

$$\zeta = -\frac{\kappa z_{\mathrm{ref}} g_0 \left(\overline{w'\theta'}\right)_s}{T_{2\mathrm{m}} u_*^3}, \tag{14}$$

where κ is the von Kármán constant, $z_{\mathrm{ref}} = 2\ m$, u_* is the friction velocity, g_0 is the gravitational constant ($\approx$9.81 m s^{-2}), and $T_{2\mathrm{m}}$ is the 2-m temperature, which are parameters extracted from the ERA5 data. The $(\overline{w'\theta'})$ is the surface heat flux approximated by

$$(\overline{w'\theta'}) = -\frac{F_{\mathrm{sh}}}{\rho C_p} - \frac{R_v}{R_d - 1}\frac{T_{2\mathrm{m}} F_e}{\rho}, \tag{15}$$

where R_v and R_d are the gas constants for water vapor and dry air, respectively, C_p is the specific heat of dry air, and ρ, F_{sh}, and F_e are the air density, sensible heat flux, and moisture flux, respectively, which are all included in the ERA5 data. The sign of ζ determines the stability of a profile (Stull 1988): profiles with a positive value of ζ were classified as stable and profiles with a negative value were classified as unstable. This is equivalent to classifying stability based on the sign of the net heat flux from the 2-m level to the surface.

Figure 4 shows the MSE for both RZ and SERZ models over the entire period and domain. We see that the SERZ model performs much better than the original RZ model for stable profiles while causing a marginal decrease in performance for unstable profiles. While both these results are statistically significant ($p \ll 0.05$), the difference in performance for the stable profiles is much larger than for the unstable profiles. SERZ decreases the MSE for stable profiles by 43.2% relative to RZ, while the MSE for unstable profiles only increases by 1.6%. This implies that our extra basis function serves its purpose in modeling overall stable profiles, allowing it to be used without prior knowledge of the stability of the profile.

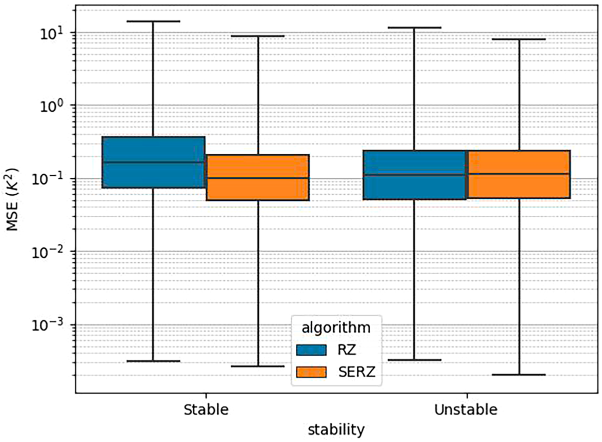

FIG. 4. Mean-square error comparison between RZ and SERZ, split for stable and unstable profiles, where the stability parameter was used for the classification. Note the logarithmic axis, which prevents the boxplots from being dominated by their upper whiskers.

b. *Fingerprints of the ABL*

To determine the number of representatives and the number of clusters, the Davies–Bouldin index (DBI; Davies and Bouldin 1979) was used. This index compares the inter- and intracluster variance to find the optimal number of clusters, where a low index implies a better clustering. Figure 5 shows that the optimal number of representatives is eight, and the optimal number of spatial clusters is two. The number of representatives is taken as such, however, the number of spatial clusters requires more attention. Because of their fundamental differences in ABL processes, the two clusters that optimally divide the data split the domain in land and open ocean. However, the difference in temperature profile behavior between land and open ocean is already well known and would therefore limit the usability of the atlas. Hence, a higher number of clusters is chosen in this work. This number determines the trade-off between the interpretability of the atlas and the possibility to recognize interesting systems. As a compromise, six clusters were chosen, as the resulting atlas already offered substantial physical insight, while not overfitting clusters by splitting them into multiple, similar, smaller clusters. Nevertheless, the fingerprinting analysis was repeated with 2, 4, and 8 clusters, which all gave similar results. These results are included in appendix C.

Figure 6 shows the representative profiles extracted from the data and Table 1 shows their corresponding physical

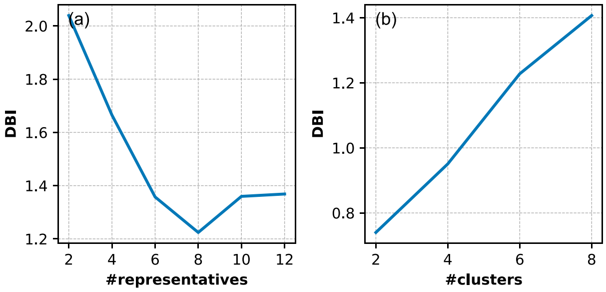

FIG. 5. The Davies–Bouldin index for the two clustering steps in the fingerprinting algorithm, where a low index corresponds to a better clustering: (a) the DBI for the clustering to find representative profiles in function of the number of representatives. (b) the DBI for the clustering of the fingerprints in function of the number of spatial clusters.

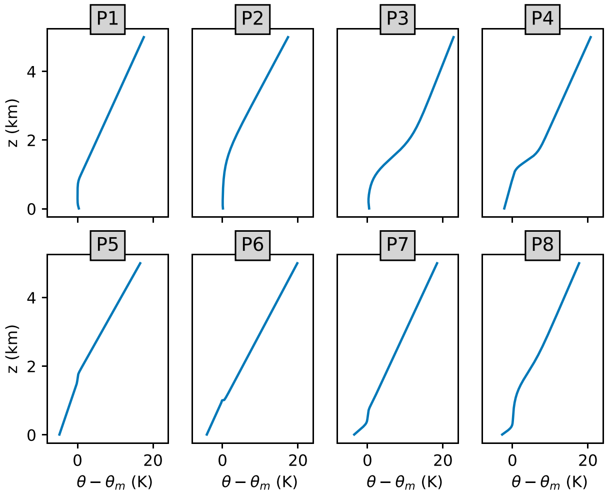

FIG. 6. Representative profiles from the fingerprinting analysis, ordered on the basis of their value for c. Since the variables θ_m were not included in the clustering, the representatives were set with $\theta_m = 0$.

values. The optimal representative profiles show the variety of profiles that are known to appear during the daily cycle (Stull 1988): Profile 1 (P1) and profile 2 (P2) have a constant ABL temperature without a significant CI. Representative 3 (P3) corresponds to a profile with a constant ABL temperature and a deep CI. Representative 4 (P4) has a strong and more shallow CI and its ABL has a weak, positive temperature gradient. Representative 5 (P5) and 6 (P6) are fully stable profiles, where the difference between the two lies in the height and significance of the nick in the temperature gradient. Representatives 7 (P7) and 8 (P8) correspond to profiles that have a stable surface layer and a neutral residual layer, where the CI for P7 is lower than for P8. All representatives have occurrence rates of over 10% in the entire dataset, except for P2. P2 only occurs in 3.5% of the cases, implying that these profiles are likely outliers that are clustered together into one representative, which could be connected to the large value of Δh_1 without the presence of a significant CI. Nevertheless, as the DBI indicated the optimal number of profiles, it is still included in the further analysis, though it will be ignored for large parts of the discussion.

The clusters and their corresponding aggregated fingerprints, shown in Fig. 7 and summarized in Table 2, give insight into the temperature profile behavior. Furthermore, Figs. 8 and 9 show the clusters and their aggregated fingerprints when limiting the data to day/night and winter/summer, respectively.

Overall, it is remarkable that the pixels within each cluster are grouped together geographically and are not randomly distributed in space. This indicates that the behavior of the temperature profiles is strongly correlated in space and that this correlation is captured by the algorithm. Moreover, the clusters mostly follow the surface type and locations of known physical systems, such as the East Icelandic Current in the northwest and the Azores high in the southwest of the domain. The majority of continental Europe is also classified in the same cluster (C), and coastal areas are grouped together in the same cluster (B).

In what follows, we will discuss each cluster depicted in Fig. 7, its geographical spread, and possibly diurnal and seasonal differences in its drivers that emerge from the analyses in Figs. 8 and 9, respectively. Furthermore, to quantify each cluster and the differences between clusters, Tables 2 and 3 are included that contain the median frequencies of each representative within each cluster, and the pairwise Euclidean distances between these median frequencies.

Cluster F is situated over the southern part of the Atlantic Ocean (with respect to our domain) and accounts for 9.3% of the pixels. The most prevalent profile for this cluster is P4 (median value of 34.6%), which has a clear CI and a weak, positive temperature gradient in the ABL itself. Furthermore, representatives P7 and P8 are mostly absent in this area with a combined occurrence rate of about 5%–6%. This is likely

TABLE 1. Physical quantities for each representative profile. The last column shows the overall occurrence rate of each representative.

	γ (K km^{-1})	z_{abl} (m)	Δh_1 (m)	$\Delta\theta_1$ (K)	Δh_2 (m)	$\Delta\theta_2$ (K)	Frequency (%)
P1	4.15	893	517	1.57	0	−0.380	12.8
P2	4.86	1437	2113	5.17	0	−0.107	3.5
P3	3.60	1417	1814	13.17	0	−0.358	12.2
P4	4.10	1359	737	7.30	692	2.18	22.0
P5	5.11	1763	250	0.71	1384	4.82	13.6
P6	4.83	1009	10	0.57	889	4.20	13.3
P7	4.23	807	537	1.83	315	3.61	10.1
P8	3.95	1623	1889	8.06	210	2.87	12.5

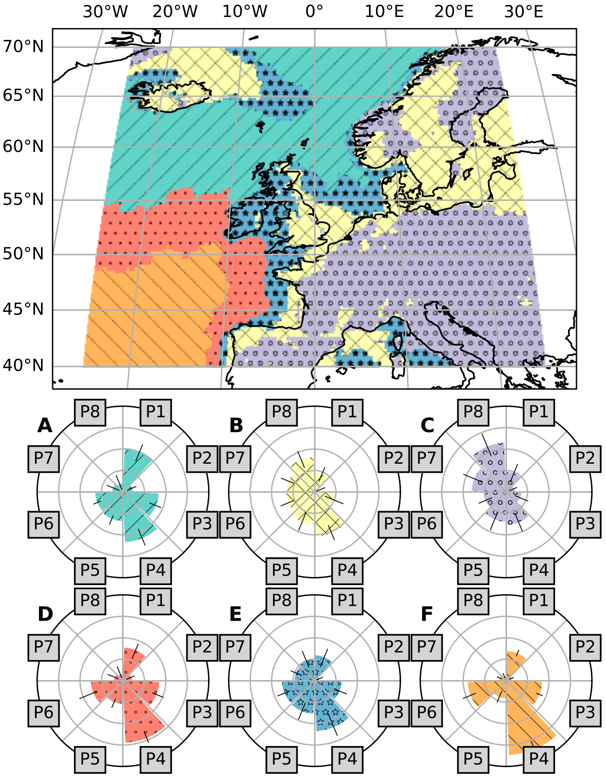

FIG. 7. Fingerprints of the ABL. (top) The clusters of pixels with similar stratification behavior. (bottom) the aggregated fingerprints of the cluster, where the median value of the occurrence rate of each representative is shown as a bar, with an error bar of the 90% interval to indicate the spread. The concentric circles in the bar chart indicate multiples of 10%. The median statistics of this map are also summarized in Table 2.

because cluster F is completely situated over the open ocean, where, due to the slow sea surface temperature variation because of the ocean's larger thermal capacity, the ABL tends to be in equilibrium with the sea surface temperature. This is also confirmed by looking at the diurnal analysis of this cluster (Fig. 8), where we see only marginal differences between the aggregated fingerprints and geographical locations of the clusters situated at the same location as cluster F in Fig. 7.

Geographically, cluster F coincides with the Azores high, which is a large high pressure system situated around 30°N in the Atlantic Ocean close to western Africa and Spain. This permanent high pressure system causes subsidence of air masses that leads to stabilization of the temperature profile (Wallace and Hobbs 2006), which strengthens the CI and causes a weak, positive temperature gradient in the ABL. The link between the Azores high and P4 is also confirmed when looking at the seasonal fingerprint analysis (Fig. 9), where we see that during summer, when the Azores high moves northward, the cluster containing the dominant P4 profile also moves northward (cluster F in both summer and winter fingerprints).

Cluster A is situated in the North Atlantic Ocean and the Norwegian Sea and covers 22.1% of the domain. It has two profiles that have an occurrence rate of over 20%, P1 and P2, and, similarly to cluster F, it has low prevalence rates for P7 and P8. Relative to cluster F covering the Azores high, cluster A has a substantially smaller occurrence rate for the P4 profile. Furthermore, we see that the P1 profile is much more prevalent in winter than in summer (Fig. 9) for this cluster and, except for part of the Norwegian Sea, there does not seem to be large diurnal differences for the North Atlantic. The reason for the change in the spatial distribution of the clusters over the Norwegian Sea when limiting the clustering to either day or night is not entirely clear, however, increasing the number of clusters for the daily fingerprints showed that the change in distribution for the cluster containing the Norwegian Sea was mainly caused by variation in continental profiles contained in the same cluster.

Cluster D is the smallest cluster, accounting for only 8.1% of all pixels. It is the cluster that contains the transition area between the Azores high cluster (F), the North Atlantic cluster (A), and cluster E to the east. The dominant profile in this cluster is still P4, though it is less extreme than in cluster F. Furthermore, the Euclidean distance between cluster D and F is lower than all other pairwise distances computed in Table 3, implying that clustering with only 5 clusters would merge clusters D and F with respect to the clustering based on 6 spatial clusters in Fig. 7. This was also confirmed in the clustering with 4 spatial clusters (appendix C), as well as the fact this area is not classified as a separate cluster in the diurnal and seasonal analysis, when there is more variation elsewhere in the domain.

Cluster C contains the majority of continental Europe, along with the Adriatic Sea. This cluster contains 27% of all pixels in our domain and has the highest prevalence rates for P7, and P8. These two representatives are characterized by a significant temperature gradient in the surface layer, which is caused by the diurnal variation of the surface temperature over land (Pal and Haeffelin 2015). This is also confirmed by examining the

TABLE 2. Results from the fingerprinting analysis shown in Fig. 7. The main body of the table shows the median frequency rate for each representative within each cluster. The bottom row corresponds to the overall frequency of each representative. The last column corresponds to the fraction of pixels classified to each cluster.

Median frequency (%)	P1	P2	P3	P4	P5	P6	P7	P8	Cluster frequency
A	20.5	4.5	16.6	23.3	13.5	12.9	5.0	4.0	22.1
B	8.9	2.4	9.3	20.8	15.2	13.2	13.1	16.3	21.9
C	9.4	4.0	7.1	15.3	13.8	10.3	16.0	22.9	27.0
D	15.4	4.1	16.9	28.9	10.8	14.9	4.1	3.1	8.1
E	11.8	2.2	12.6	23.4	15.4	15.1	9.3	10.3	11.6
F	13.8	2.6	16.8	34.6	8.2	17.9	3.2	2.2	9.3
Representative frequency	12.8	3.5	12.2	22.0	13.6	13.3	10.1	12.5	

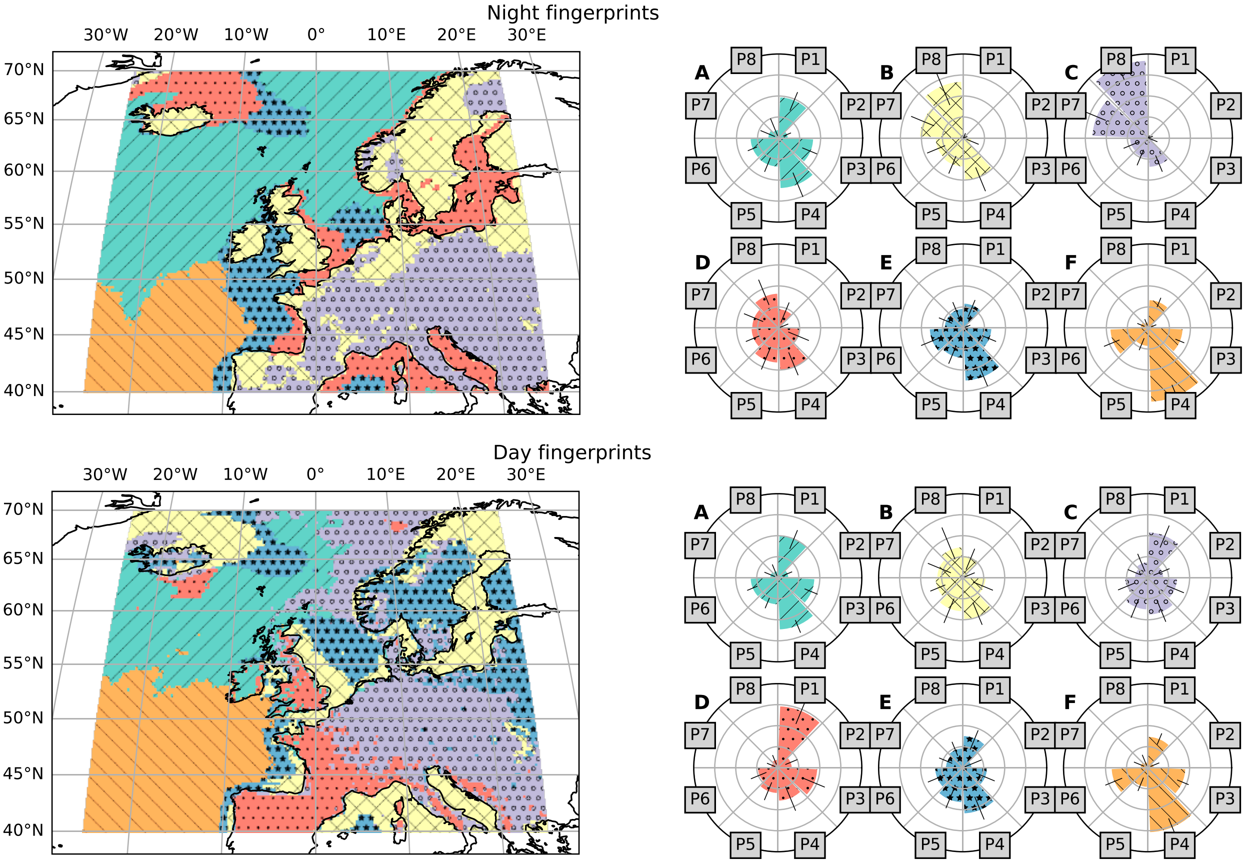

FIG. 8. Fingerprinting analysis applied to the (bottom) day hours (1200–1700 UTC) and (top) night hours (0000–0500 UTC) separately. Note that the representatives were still determined using the full year but the second step (agglomerative clustering) was computed for each time window separately.

clusters covering continental Europe in the diurnal analysis (Fig. 8): During the day, profiles over land are generally more unstable (P1), while during the night, there is a clear peak in profiles with a stable surface layer (P7 and P8). Moreover, the diurnal analysis also shows that the time of the day has a larger influence on the spatial distribution of the clustering over land than it has on the clustering over the open ocean, confirming the strong continental diurnal forcing.

Furthermore, cluster C also covers the Adriatic Sea. This is likely due to the surrounding land, which makes advection of continental air mass possible for all wind directions. However, the Adriatic Sea is only classified in the same cluster when including the entire dataset. When limiting the clustering to summer/winter or day/night, it is classified in clusters covering coastal seas and shores. Therefore, a more extensive analysis of the Mediterranean is recommended when dealing with the Adriatic Sea (and the Mediterranean Sea in general).

Clusters B (21.9%) and E (11.6%) contain the Baltic Sea, most coastal areas and shallow seas, the United Kingdom, Iceland, part of Scandinavia, and the seas to the north and east of Iceland. The fact that coastal areas are classified in a different cluster than the European continent or the Atlantic Ocean clearly shows that these areas have substantially different temperature profiles. Moreover, there is overall more variation in temperature profiles for clusters B and E than there is for the clusters covering the open ocean. This larger variation can be explained by the advection of air from the nearby land, which, because of its strong diurnal forcing, has a preference for P7 and P8, and the advection of air from the Atlantic Ocean, which contains more P1 and P4. Hence, coastal seas and the coast itself are transition areas where both characteristics of the continental land and open ocean are present, which can also be seen by comparing their respective fingerprints. This is clear from the Euclidean distances between the different clusters (Table 3), where the cluster of continental Europe (C) is closest to the cluster for the North Sea and Baltic Sea (B), which in turn is closest to cluster E, which, finally, is close to both cluster A (North Atlantic) and cluster D (Azores high transition area).

The influence of the land on the coastal seas and the seas surrounded by land is also present in the seasonal clustering (Fig. 9) where during summer, the prevalence of profiles with a stable surface layer (P7, P8) is higher for coastal seas such as the Baltic, Adriatic, and North Sea than it is for the Atlantic Ocean and Norwegian Sea.

Last, profiles with a stable internal boundary layer (P7, P8) are mostly absent over the open ocean. However, a maritime system that contains more of these profiles emerges close to Iceland (cluster B and E in Fig. 7). This area coincides with the

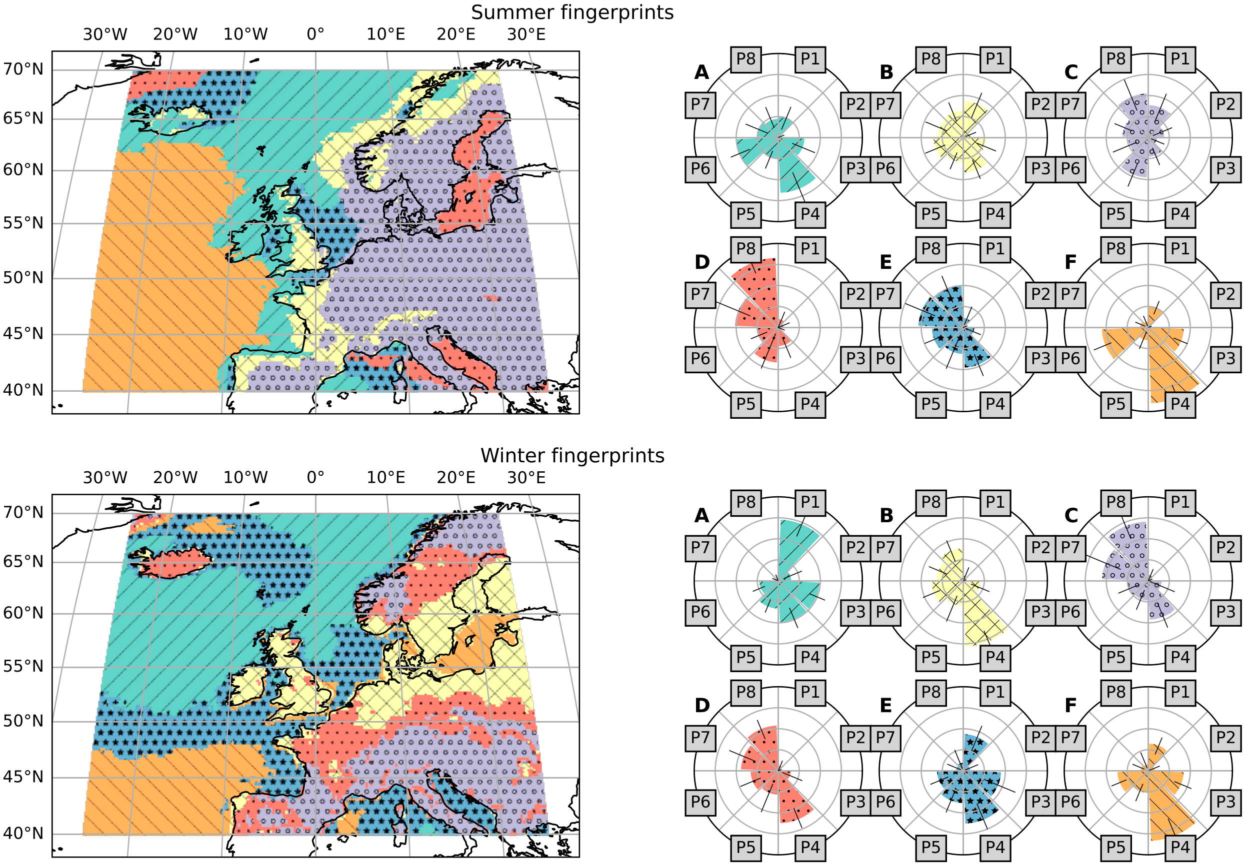

FIG. 9. Fingerprinting analysis applied to the summer months and winter months separately. Note that the representatives were still determined using the full year but the second step (agglomerative clustering) was computed for each season separately.

East Icelandic Current (EIC), which transports cold, Arctic water southeastward, which could cause this change in temperature behavior of the lower atmosphere (Casanova-Masjoan et al. 2020). This is again confirmed by looking at its seasonality, which is known to strengthen this current during winter (Casanova-Masjoan et al. 2020).

c. *Limitations and tests of robustness*

1) SERZ

The SERZ model is a compromise between the best fit of the temperature and the physical interpretability of the results. Therefore, it has drawbacks that should be taken into account:

TABLE 3. Relative Euclidean distances between the clusters based on their median values. The average distance to each other cluster was added in the last column. Note that only the relative differences in distances are meaningful.

	A	B	C	D	E	F	Avg distance
A	0	20	28	8	13	15	16.9
B	20	0	10	21	9	25	17.2
C	28	10	0	30	18	34	24.0
D	8	21	30	0	13	7	15.9
E	13	9	18	13	0	18	14.1
F	15	25	34	7	18	0	20.0

First of all, the SERZ model has two model parameters (C_1 and C_2) influencing the overall curvature of the profile that are fixed before the optimization. To assess the sensitivity toward C_2, local experiments were carried out in which C_2 was changed to both 50 and 200 m (at the Belgian–Dutch wind farms at 51°44′N, 2°58′E). Both cases yielded similar results as the $C_2 = 100$ m case, and the extracted physical parameters were (highly) correlated to the physical parameters computed with $C_2 = 100$ m, with a correlation of 0.88 between the results with $C_2 = 50$ m and $C_2 = 100$ m, and a correlation of 0.89 for $C_2 = 200$ m and $C_2 = 100$ m. Although these local experiments showed a relatively high correlation, an extensive and nonlocal study on the influence of these parameters could still be valuable.

Second, the possibility of physically interpreting the fit of a profile depends on the kind of profile, and the constraints or regularization used in the optimization. This is both an advantage and a disadvantage, as it allows the user to create user-specific constraints, but it also can create discrepancies between interpretations depending on the constraints set. For example, some wind farm simulations require the wind turbines to be completely below the CI. Therefore, they might consider adding an extra constraint that ensures this. However, since each of the parameters is intrinsically linked in the optimization, this might also result in different values for other parameters such as the ABL height and CI strength.

2) FINGERPRINTING ANALYSIS

A first drawback of the fingerprinting algorithm is the fact that there are hyperparameters that need to be chosen by the user. The main hyperparameters are the number of clusters in the final clustering, and the type of normalization used on the data before the start of the clustering. The former is a parameter that determines a trade-off between finding interesting systems and keeping the results interpretable. The latter determines the distribution of the data before the first clustering stage, and, therefore, can have an impact on the final result. It is worth noting, however, that the kind of normalization did not significantly change the results in our application. Since the fingerprinting analysis is relatively cheap, we recommend experimenting with the hyperparameter settings to assess the sensibility for the domain and application of interest.

A second drawback is that the results from our fingerprint analysis are unique to the domain under consideration and the data used. Hence, changing the domain or resolution of the data will result in different representatives and possibly different clusters. Therefore, it is recommended to users that the second clustering is repeated for their domain and data of interest.

Because of computational constraints, the analysis was limited to only one year of ERA5 data. This implies that interannual variability (IAV) of the temperature profiles and the ABL depth (Pal and Haeffelin 2015) is absent in the final atlas. Nevertheless, the main conclusions that are drawn from one year of data should remain valid since they are coupled to large-scale physical systems that are present every year, even though their exact location and intensity may vary.

3) TESTS OF ROBUSTNESS

To test the temperature model dependency of our fingerprint algorithm, the fingerprinting analysis was also carried out on parameters extracted from a linear fit (see appendix C). This showed that, although the interpretation of the representatives and aggregated fingerprints is less clear, the cluster regions mostly coincide with the results of the SERZ model, implying that the clusters and their aggregated fingerprints are relatively model independent.

To test the effect of the number of clusters on the results, the fingerprint analysis was carried out with 2, 4, and 8 clusters in the final clustering of the fingerprints (see appendix C). This analysis did not show substantially different results overall. However, increasing the number of clusters revealed that highly mountainous regions, which are known to behave atypically (Frei 2014), emerged as separate clusters. However, it is not clear if this is an artifact of the SERZ model fitted to atypical (mountainous) profiles or because of a currently unknown reason. Therefore, subsequent studies are necessary to assess the SERZ model performance for these areas. Furthermore, care should be taken with clusters that are not specifically located in any geographically known system and have similar aggregated fingerprints: Seemingly random splits between two clusters can be a consequence of the choice of the number of clusters, which can cause some clusters to be split even though they are sufficiently similar, or can cause two clusters to merge even though they differ significantly (especially for datasets of limited size).

The previous analysis was also carried out on other years of ERA5 data that yielded similar results, both in the different representatives extracted from the data as in the clusters found over the grid. The analysis presented here is tailored to the European domain but can be applied elsewhere. Moreover, subregions in Europe can be analyzed in more detail. For example, in our application the Azores high system is found to have one dominant profile (P4), yet if the analysis was done for the domain only covering the area of the Azores high this one representative might be split into multiple, similar representatives. Nevertheless, this new set of representatives will be physically close to the P4 profile, allowing direct comparison with the findings in this work.

4. Conclusions and future work

This work introduces a modified vertical potential temperature model, the surface extended Rampanelli and Zardi model, for the atmospheric boundary layer and lower troposphere, and presents a temperature profile atlas for Europe, excluding the Mediterranean. To this end, a new spatiotemporal analysis framework based on pixel fingerprints was developed to deal with the spatiotemporal data and create generally applicable representative profiles.

The new SERZ model is successful in modeling different atmospheric boundary layer stratifications and can be fine-tuned to user-specific applications by including extra constraints or regularization. It significantly reduces (with a substantial margin) the fitting error for stable profiles by 43.2% while it only increased the mean error for unstable profiles with 1.6%. In doing so, SERZ performs equally well for stable and unstable ABLs, allowing it to be used in a general setting where the stability of the ABL is unknown. The general applicability allowed us to reduce the complexity of the data, with originally around 40 varying height levels, to only 6 parameters, which allowed further analysis.

The fingerprint analysis is a unique and flexible way to visualize and analyze spatiotemporal data successfully. In this analysis, optimally chosen representatives are used to remove the temporal component from the data, which allows the resulting pixel fingerprints to represent the temporal variation of the signal. Furthermore, the number of representatives can be chosen independently from the number of parameters in the data, which is useful when the number of parameters becomes too large to manually interpret and cross correlate. In addition, the fingerprint analysis results in a number of universal representatives and clusters that provide an easy way to visualize and interpret complex spatiotemporal data.

The resulting clusters with their agglomerated fingerprints show that the ABL stratification behaves fundamentally differently between land, ocean, and coastal seas. Over land and coastal areas, profile type variation is larger than over the ocean. This is likely due to the diurnal variation in the surface temperature, which is higher over land due to its lower heat capacity, which is advected toward the sea. Over the open ocean, the behavior of the ABL is limited to only two or three profiles and the diurnal variation is mostly absent. Furthermore, the

Azores high pressure system and the East Icelandic Current behaved distinctively different from the surrounding ocean, showing that differences in synoptic-scale systems or ocean currents substantially influence the ABL profiles.

The stratification of the ABL can differ significantly based on geographical location, time of day, and time of the year, and these differences in stratification should be taken into account when selecting and comparing atmospheric conditions. Especially for offshore wind energy applications that require subselections of the atmospheric state, good selections of the temperature profiles are important, as the choice of background profiles for these locations influences the results of these simulations and is often based on onshore experience. The universal applicability of our new SERZ model, together with the framework of the fingerprint analysis, provides an easy way of determining these background profiles. We applied this to Europe to generate the ideal representatives for this area resulting in a stratification atlas that can be consulted for various applications. By implementing and including the SERZ model, fingerprinting analysis, and the data in a user-friendly way, it is also readily available to apply the complete framework to other areas.

Future work could include wind profiles and surface variables in this atlas, as it only consists of temperature profiles at this stage. Wind resource assessments would then be able to use the same atlas for all atmospheric conditions in their simulations, which would also remove the need to interpret correlations between these variables. Furthermore, a comparison of the resulting fingerprints with local temperature profile measurements (Pandolfi et al. 2013; Granados-Muñoz et al. 2012; De Tomasi et al. 2011) would be interesting to study the uncertainties of these results caused by using ERA5 in combination with the fingerprinting analysis.

Acknowledgments. The authors acknowledge support from the project FREEWIND, funded by the Energy Transition Fund of the Belgian Federal Public Service for Economy, SMEs, and Energy (FOD Economie, K.M.O., Middenstand en Energie), from the Department of Earth and Environmental Sciences of the KU Leuven, and the Department of Mechanical Engineering of the KU Leuven. The computational resources and services in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation Flanders (FWO) and the Flemish government department EWI. The ERA5 data from Hersbach et al. (2020) on model levels were downloaded with the help of the German Climate Compute Centre (DKRZ) infrastructure. Michiel Dhont (ORCID: 0000-0002-5679-0991), Jan Schween (University of Cologne), and Dino Zardi (University of Trento) are thanked for their input on the visualization, the physical interpretation of the Azores high pressure system, and the discussion on the equation of the strength of the capping inversion, respectively.

Data availability statement. The Python scripts for the SERZ model and fingerprint analysis, visualization scripts, the representative profiles, and the individual pixel fingerprints are available at the git repository (https://github.com/SebastiaanJamaer/Jamaer_etal_2022.git) without access restrictions. The ERA5 data (Hersbach et al. 2020) are a public dataset that is available at the Copernicus Climate Data Store.

APPENDIX A

Numerical Implementation

In general, the implementation is straightforward, but some pitfalls should be avoided.

a. Functions f, g, and h

The functions f, g and h are defined as

$$f(\eta) = \frac{\tanh(\eta) + 1}{2} = \frac{1}{1 + e^{-2\eta}}, \tag{A1}$$

$$g(\eta) = \frac{\ln[2\cosh(\eta)] + \eta}{2} = \frac{\ln(e^{\eta} + e^{-\eta}) + \eta}{2}, \quad \text{and} \tag{A2}$$

$$h(\eta') = -\frac{\ln[2\cosh(\eta')] - \eta'}{2} = \frac{\eta' - \ln(e^{\eta'} + e^{-\eta'})}{2} \tag{A3}$$

and are implemented using internal functions of the NumPy module. The hyperbolic tangent and hyperbolic cosine are defined by NumPy. The direct implementation for f is always numerically stable (in absolute error). However, for g and h, overflow can occur because, for larger $|x|$, e^x can become either zero or infinity. This then causes the ln function to return either $-\infty$ or ∞. Therefore, for these cases, the values are replaced by

$$g(\eta) \approx \frac{|\eta| + \eta}{2} \quad \text{and}$$

$$h(\eta') \approx \frac{\eta' - |\eta'|}{2}.$$

b. Including regularization in the optimization problem

The optimization problem in Eq. (3) was implemented such that regularization on individual parameters to improve physical interpretability was possible through the Tikhonov regularization.

$$\underset{\Theta}{\text{minimize}}$$

$$S = \frac{1}{N}\sum_i [\theta_i - \hat{\theta}(z_i; \Theta)]^2 + \lambda_r R(\Theta)$$

subject to

$$0 \le \theta_m, a, b, z_{\text{abl}}, \Delta h_1,$$

$$0 \le \alpha \le 1,$$

$$\alpha z_{\text{abl}} < z_{\text{abl}} - \frac{\Delta h_1}{2}, \quad \text{and}$$

$$z_{\text{abl}} + \frac{\Delta h_1}{2} < 4000\ \text{m},$$

with

$$R(\Theta) = \sum_{p \in \Theta} (p/w_p)^2, \tag{A4}$$

which is the regularization term, where w_p is the weight assigned to parameter p and λ_r is a factor determining the overall weight of the regularization. Note that for the results in this work, we set $\lambda_r = 0$.

c. Ill conditioning

To see why an unconstrained version of the optimization problem can be ill conditioned, we construct the linear sub-optimization problem over the parameters a, b, c, and θ_m. We start from the fact that minimizing S without constraints will require that all of its partial derivatives are zero: Minimizing the function S will, for all of its unconstrained values, require that its partial derivative is zero, so

$$\begin{aligned}
&\sum_i[\theta_i - \hat{\theta}(z_i;\Theta)]\frac{\partial\hat{\theta}}{\partial\theta_m}(z_i) + \lambda_r\frac{\partial}{\partial\theta_m}R(\Theta) = 0,\\
&\sum_i[\theta_i - \hat{\theta}(z_i;\Theta)]\frac{\partial\hat{\theta}}{\partial a}(z_i) + \lambda_r\frac{\partial}{\partial a}R(\Theta) = 0,\\
&\sum_i[\theta_i - \hat{\theta}(z_i;\Theta)]\frac{\partial\hat{\theta}}{\partial b}(z_i) + \lambda_r\frac{\partial}{\partial b}R(\Theta) = 0,\\
&\sum_i[\theta_i - \hat{\theta}(z_i;\Theta)]\frac{\partial\hat{\theta}}{\partial c}(z_i) + \lambda_r\frac{\partial}{\partial c}R(\Theta) = 0,\\
&\sum_i[\theta_i - \hat{\theta}(z_i;\Theta)]\frac{\partial\hat{\theta}}{\partial\Delta h_1}(z_i) + \lambda_r\frac{\partial}{\partial\Delta h_1}R(\Theta) = 0,\\
&\sum_i[\theta_i - \hat{\theta}(z_i;\Theta)]\frac{\partial\hat{\theta}}{\partial z_{\mathrm{abl}}}(z_i) + \lambda_r\frac{\partial}{\partial z_{\mathrm{abl}}}R(\Theta) = 0, \quad \text{and}\\
&\sum_i[\theta_i - \hat{\theta}(z_i;\Theta)]\frac{\partial\hat{\theta}}{\partial\alpha}(z_i) + \lambda_r\frac{\partial}{\partial\alpha}R(\Theta) = 0,
\end{aligned}$$

Now, by definition,

$$\begin{aligned}
&\frac{\partial\hat{\theta}}{\partial\theta_m}(z_i) = 1,\\
&\frac{\partial\hat{\theta}}{\partial a}(z_i) = f(\eta_i),\\
&\frac{\partial\hat{\theta}}{\partial b}(z_i) = g(\eta_i), \quad \text{and}\\
&\frac{\partial\hat{\theta}}{\partial c}(z_i) = h(\eta_i')
\end{aligned}$$

with the corresponding definitions of η and η'. Then, using the definition of $\hat{\theta}$, we can rewrite the first four equations to

$$\begin{aligned}
&\theta_m N + a\sum f(\eta_i) + b\sum g(\eta_i) + s\sum h(\eta_i') + 2\lambda_r\frac{\theta_m}{w_{\theta_m}} = \sum\theta_i,\\
&\theta_m\sum f(\eta_i) + a\sum f^2(\eta_i) + b\sum f(\eta_i)g(\eta_i) + c\sum f(\eta_i)h(\eta_i') + 2\lambda_r\frac{a}{w_a} = \sum f(\eta_i)\theta_i,\\
&\theta_m\sum g(\eta_i) + a\sum g(\eta_i)f(\eta_i) + b\sum g^2(\eta_i) + c\sum g(\eta_i)h(\eta_i') + 2\lambda_r\frac{b}{w_b} = \sum g(\eta_i)\theta_i, \quad \text{and}\\
&\theta_m\sum h(\eta_i') + a\sum h(\eta_i')f(\eta_i) + b\sum h(\eta_i')g(\eta_i) + c\sum h^2(\eta_i') + 2\lambda_r\frac{c}{w_c} = \sum h(\eta_i')\theta_i,
\end{aligned} \tag{A5}$$

which is a linear system in the form of $\mathbf{Ax} = \mathbf{b}$ for the parameters θ_m, a, b, and c, where $\mathbf{A}$ is a square 4 × 4 matrix, $\mathbf{x}$ is the vector of those four parameters, and $\mathbf{b}$ is a vector containing the right-hand side of the equations in Eq. (A5). This linear system in Eq. (A5) can be reformulated in the elegant shape $\mathbf{A} = \mathbf{M}^{\mathrm{T}}\mathbf{M} + 2\lambda_r\mathbf{W}$ with

$$\mathbf{M} = \begin{bmatrix} 1 & f(\eta_1) & g(\eta_1) & h(\eta_1') \\ 1 & f(\eta_2) & g(\eta_2) & h(\eta_2') \\ \vdots & \vdots & \vdots & \vdots \\ 1 & f(\eta_N) & g(\eta_N) & h(\eta_N') \end{bmatrix} \quad \text{and} \quad \mathbf{W} = \begin{pmatrix} w_{\theta_m} & 0 & 0 & 0 \\ 0 & w_a & 0 & 0 \\ 0 & 0 & w_b & 0 \\ 0 & 0 & 0 & w_c \end{pmatrix}. \tag{A6}$$

If z_{abl} becomes very large, both f and g approach a constant in the lower atmosphere, and therefore the matrix $\mathbf{M}$ becomes singular. Therefore, it also becomes clear why for large values for z_{abl}, the optimization problem is ill conditioned if no regularization is applied.

APPENDIX B

Extra Material and Tables

a. Summary tables: Seasonal and diurnal clustering

Day clustering is shown in Table B1, night clustering is shown in Table B2, summer clustering is shown in Table B3, and winter clustering is shown in Table B4.

b. Map domain and height distribution

Figure B1 is a map of the study area, and Fig. B2 shows the mean height levels of the ERA5 data.

TABLE B1. Results from the fingerprinting analysis shown in Fig. 8, constrained to the noon hours. The main body of the table shows the median frequency rate for each representative within each cluster. The bottom row corresponds to the overall frequency of each representative. The last column corresponds to the fraction of pixels classified to each cluster. Note that the cluster labels (A, B, …) correspond to the clusters for the clustering shown in Fig. 8, day panel.

Mean frequency (%)	P1	P2	P3	P4	P5	P6	P7	P8	Cluster frequency
A	21.5	7.4	14.0	17.5	15.0	10.7	6.0	7.3	23.0
B	10.4	2.6	10.9	19.7	15.9	12.9	12.1	14.5	19.5
C	15.1	4.1	11.6	21.7	16.4	13.1	7.9	9.1	15.9
D	20.1	5.0	17.4	24.6	12.6	12.9	3.6	3.1	17.7
E	29.2	6.6	18.8	15.7	12.0	9.4	3.1	3.7	7.8
F	14.7	2.9	17.9	30.7	9.1	17.1	3.2	2.5	16.1
Representative frequency	17.5	4.9	14.6	21.9	14.0	12.9	6.6	7.6	

TABLE B2. Results from the fingerprinting analysis shown in Fig. 8, constrained to the nightly hours. The main body of the table shows the median frequency rate for each representative within each cluster. The bottom row corresponds to the overall frequency of each representative. The last column corresponds to the fraction of pixels classified to each cluster. Note that the cluster labels (A, B, …) correspond to the clusters for the clustering shown in Fig. 8, night panel.

Mean frequency (%)	P1	P2	P3	P4	P5	P6	P7	P8	Cluster frequency
A	0.7	0.7	3.8	19.8	13.5	13.2	20.7	26.8	19.7
B	0.4	0.9	2.4	13.7	9.4	8.7	26.6	36.7	18.8
C	19.4	4.3	16.4	23.8	13.6	13.3	4.9	3.8	29.4
D	8.9	2.2	10.3	20.0	16.4	12.8	12.5	16.3	13.7
E	13.3	2.6	16.3	35.1	8.5	18.0	3.6	2.6	10.2
F	11.6	2.3	13.3	25.0	14.2	16.0	8.9	9.1	8.2
Representative frequency	9.3	2.4	10.2	21.8	12.8	13.2	13.5	16.7	

TABLE B3. Results from the fingerprinting analysis shown in Fig. 9, constrained to the summer months. The main body of the table shows the median frequency rate for each representative within each cluster. The bottom row corresponds to the overall frequency of each representative. The last column corresponds to the fraction of pixels classified to each cluster. Note that the cluster labels (A, B, …) correspond to the clusters for the clustering shown in Fig. 9, summer panel.

Mean frequency (%)	P1	P2	P3	P4	P5	P6	P7	P8	Cluster frequency
A	15.4	7.6	4.0	9.1	19.0	9.4	12.2	21.0	31.4
B	4.0	2.5	1.2	8.8	16.7	10.1	20.6	33.2	5.4
C	9.9	0.8	16.9	36.1	6.1	22.1	3.8	2.9	24.4
D	10.0	1.0	12.5	26.3	10.2	19.8	9.8	8.9	21.9
E	17.3	3.4	11.0	16.6	11.7	14.1	12.0	12.7	9.7
F	4.5	0.5	5.3	19.1	12.6	14.9	21.3	19.7	7.2
Representative frequency	11.7	3.4	9.8	21.0	12.8	16.1	11.2	14.0	

TABLE B4. Results from the fingerprinting analysis shown in Fig. 9, constrained to the winter months. The main body of the table shows the median frequency rate for each representative within each cluster. The bottom row corresponds to the overall frequency of each representative. The last column corresponds to the fraction of pixels classified to each cluster. Note that the cluster labels (A, B, …) correspond to the clusters for the clustering shown in Fig. 9, winter panel.

Mean frequency (%)	P1	P2	P3	P4	P5	P6	P7	P8	Cluster frequency
A	2.7	0.9	6.4	18.5	9.3	10.1	22.0	28.2	14.4
B	17.4	3.2	17.9	25.2	15.3	12.5	3.3	3.8	21.1
C	29.1	6.1	19.6	19.7	13.2	9.0	1.2	1.0	24.7
D	2.4	0.5	8.1	31.3	11.9	15.2	12.8	15.4	14.7
E	13.0	3.0	16.8	33.3	10.5	14.7	3.6	3.1	12.7
F	2.4	0.7	6.0	25.0	11.4	13.2	17.8	21.5	12.4
Representative frequency	13.4	3.2	13.9	24.5	12.7	12.3	9.2	10.8	

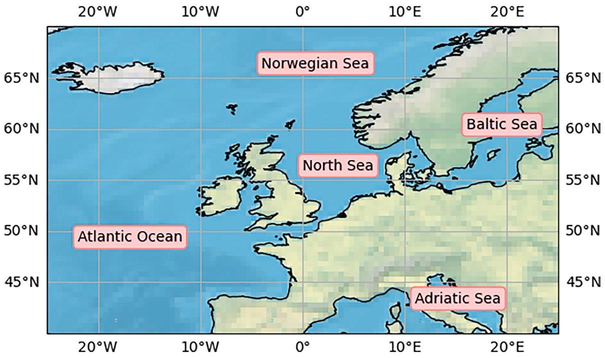

FIG. B1. Map of the study area.

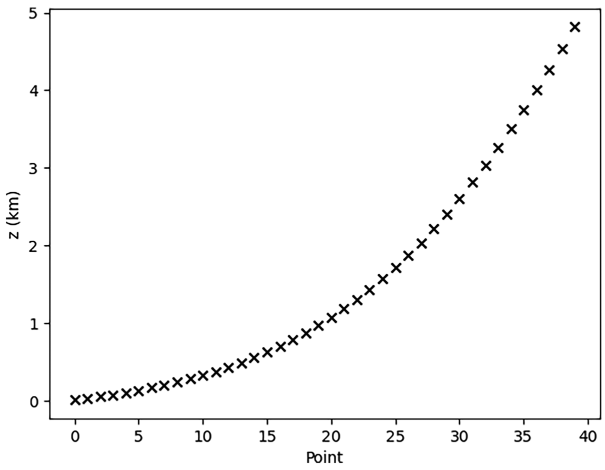

FIG. B2. Mean height levels of the ERA5 data.

APPENDIX C

Extra Fingerprint Analyses

a. Linear fingerprint analysis

Figure C1 shows the representatives for the year 2020 fitted with three-line piecewise linear profiles. The representatives were ordered based on their gradient of the lowest line fragment. The resulting representatives are slightly different from the representatives found with the SERZ fits, although this is mainly due to the less straightforward ordering of these profiles. In general, we do see similar representatives: The representative of the SERZ model (P4) is very similar to the fourth representative with the linear profile (L4). Furthermore, P8 is relatively similar to L7, P7 is similar to L8, P2 is similar to L1 and L5, and P1 is similar to P3.

The clusters and aggregated fingerprints based on the fits with the linear profile (Fig. C2) also show similar spatial distributions, especially over sea. In these clusters, we again recognize the Azores high pressure system and the East Icelandic Current. Furthermore, coastal waters behave differently from the open ocean, although this is less clear for the linear fits.

b. Varying number of clusters

To assess the influence of the number of clusters on our results, we repeated our analysis with a varying number of clusters. The number of clusters varied between 2 and 8, in steps of 2. For all setups, the same representatives were used (which are independent of the number of spatial clusters). Since the difference between the different setups is small, only the clustering based on the full dataset is included in this appendix. Furthermore, since the spatial clustering uses agglomerative clustering, increasing the number of clusters by 1 will split the cluster with the most variance into two new clusters without

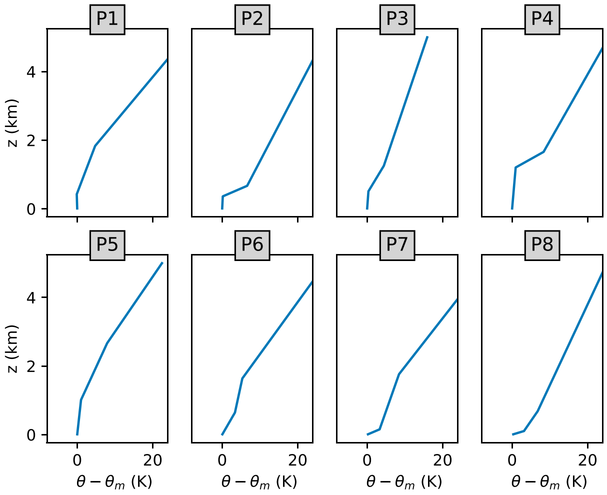

FIG. C1. Representative profiles based on piecewise linear fits.

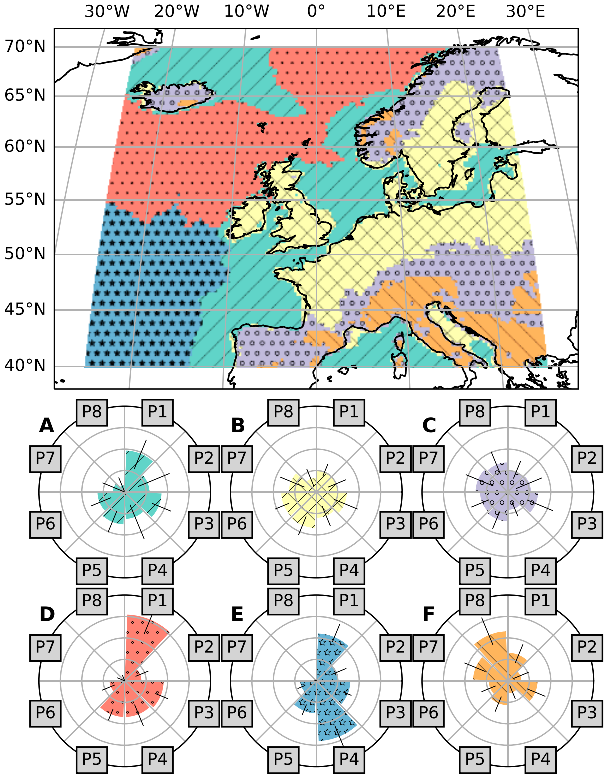

FIG. C2. Fingerprint analysis based on piecewise linear profiles.

redistributing the other points. Hence, increasing the number of clusters by 2 will either split two existing clusters into two or will split an existing cluster in three.

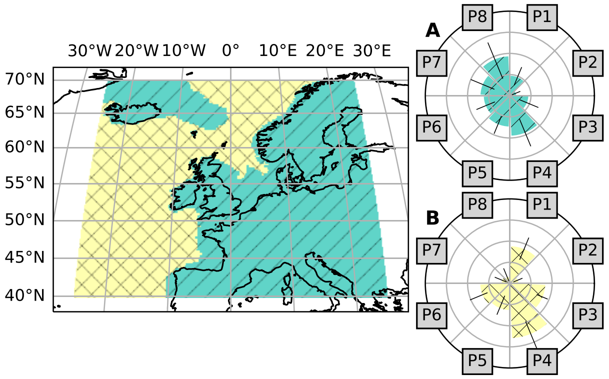

FIG. C3. Fingerprint analysis with only two clusters.

Figure C3 shows that the fingerprinting algorithm splits the domain into ocean and land/coastal areas when using only two clusters. Interesting to see here is that the East Icelandic Current behaves more similar to the land then it does to the open ocean. Furthermore, we can already see with two clusters that P7 and P8 mainly occur over land/coast, while P1 and P4 are much more prevalent over open ocean.

Figure C4 shows the fingerprinting algorithm with four clusters. Here we see that both clusters from Fig. C3 are split: the ocean cluster is split along the borders of the Azores high pressure system and the rest of the Atlantic Ocean and open North Sea, and the land and coastal cluster is split in a cluster that mainly contains continental Europe and a pixel that encompasses coastal seas, the EIC, Iceland, and the United Kingdom. This figure already shows the main characteristics of the atlas with six spatial clusters.

Figure C5 shows the analysis with eight spatial clusters. In comparison with the map with six clusters (Fig. 7) we see that cluster A_6 is split into two and the Europe mainland cluster is split into two. However, for each of these splits, the difference in the agglomerated fingerprints for the two new clusters is relatively small. Yet, an interesting pattern around the higher and more mountainous regions seems to appear. However, it was beyond the scope of this work to investigate this further.

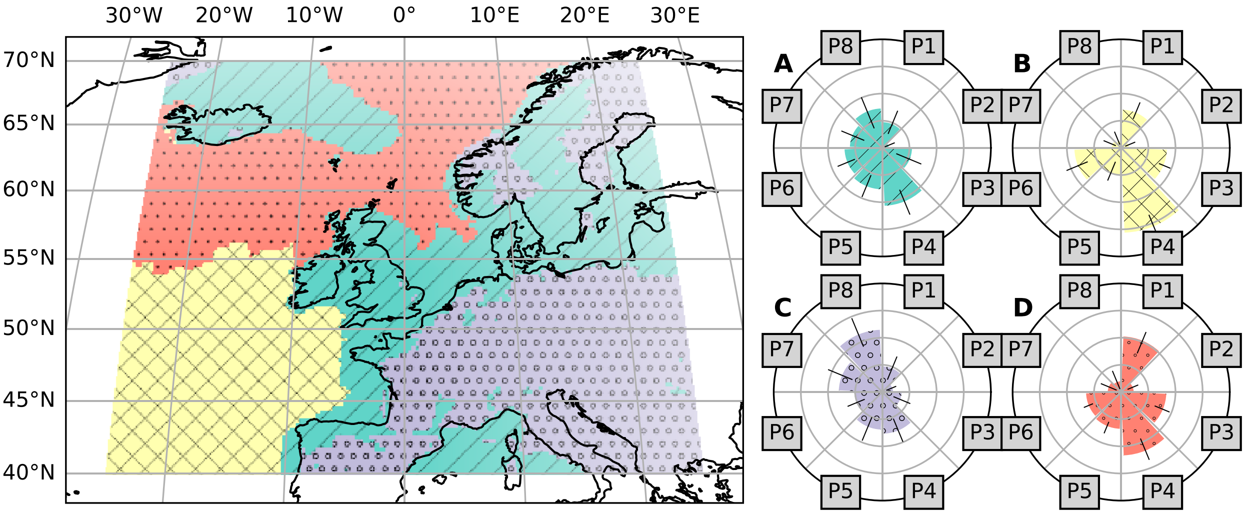

FIG. C4. Fingerprint analysis with four clusters.

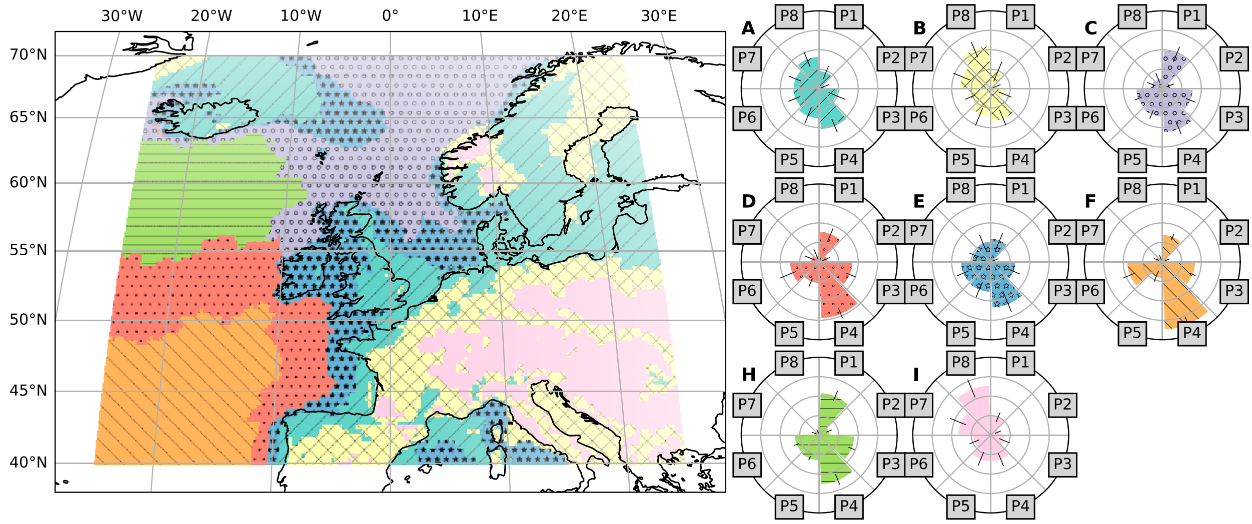

FIG. C5. Fingerprint analysis with eight clusters.

REFERENCES

Allaerts, D., and J. Meyers, 2015: Large eddy simulation of a large wind-turbine array in a conventionally neutral atmospheric boundary layer. *Phys. Fluids*, **27**, 065108, https://doi.org/10.1063/1.4922339.

——, and ——, 2017: Boundary-layer development and gravity waves in conventionally neutral wind farms. *J. Fluid Mech.*, **814**, 95–130, https://doi.org/10.1017/jfm.2017.11.

——, S. Vanden Broucke, N. van Lipzig, and J. Meyers, 2018: Annual impact of wind-farm gravity waves on the Belgian–Dutch offshore wind-farm cluster. *J. Phys. Conf. Ser.*, **1037**, 072006, https://doi.org/10.1088/1742-6596/1037/7/072006.

Andersen, S. J., J. N. Sørensen, and R. F. Mikkelsen, 2017: Performance and equivalent loads of wind turbines in large wind farms. *J. Phys. Conf. Ser.*, **854**, 012001, https://doi.org/10.1088/1742-6596/854/1/012001.

Argyle, P., and S. J. Watson, 2014: Assessing the dependence of surface layer atmospheric stability on measurement height at offshore locations. *J. Wind Eng. Ind. Aerodyn.*, **131**, 88–99, https://doi.org/10.1016/j.jweia.2014.06.002.

Baas, P., F. C. Bosveld, and G. Burgers, 2016: The impact of atmospheric stability on the near-surface wind over sea in storm conditions. *Wind Energy*, **19**, 187–198, https://doi.org/10.1002/we.1825.

Batchvarova, E., and S.-E. Gryning, 1994: An applied model for the height of the daytime mixed layer and the entrainment zone. *Bound.-Layer Meteor.*, **71**, 311–323, https://doi.org/10.1007/BF00713744.

Bleeg, J., M. Purcell, R. Ruisi, and E. Traiger, 2018: Wind farm blockage and the consequences of neglecting its impact on energy production. *Energies*, **11**, 1609, https://doi.org/10.3390/en11061609.

Branlard, E., E. Quon, A. R. M. Forsting, J. King, and P. Moriarty, 2020: Wind farm blockage effects: Comparison of different engineering models. *J. Phys. Conf. Ser.*, **1618**, 062036, https://doi.org/10.1088/1742-6596/1618/6/062036.

Brisson, E., M. Demuzere, B. Kwakernaak, and N. P. M. van Lipzig, 2011: Relations between atmospheric circulation and precipitation in Belgium. *Meteor. Atmos. Phys.*, **111**, 27–39, https://doi.org/10.1007/s00703-010-0103-y.

Casanova-Masjoan, M., and Coauthors, 2020: Along-stream, seasonal, and interannual variability of the North Icelandic Irminger Current and East Icelandic Current around Iceland. *J. Geophys. Res. Oceans*, **125**, e2020JC016283, https://doi.org/10.1029/2020JC016283.

Chatterjee, F., D. Allaerts, U. Blahak, J. Meyers, and N. P. M. van Lipzig, 2016: Evaluation of a wind-farm parametrization in a regional climate model using large eddy simulations. *Quart. J. Roy. Meteor. Soc.*, **142**, 3152–3161, https://doi.org/10.1002/qj.2896.

Dang, R., Y. Yang, H. Li, X.-M. Hu, Z. Wang, Z. Huang, T. Zhou, and T. Zhang, 2019: Atmosphere Boundary Layer Height (ABLH) determination under multiple-layer conditions using micro-pulse lidar. *Remote Sens.*, **11**, 263, https://doi.org/10.3390/rs11030263.

Davies, D. L., and D. W. Bouldin, 1979: A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.*, **PAMI-1**, 224–227, https://doi.org/10.1109/TPAMI.1979.4766909.

Deardorff, J. W., 1979: Prediction of convective mixed-layer entrainment for realistic capping inversion structure. *J. Atmos. Sci.*, **36**, 424–436, https://doi.org/10.1175/1520-0469(1979)036<0424:POCMLE>2.0.CO;2.

Demuzere, M., M. Werner, N. P. M. van Lipzig, and E. Roeckner, 2009: An analysis of present and future ECHAM5 pressure fields using a classification of circulation patterns. *Int. J. Climatol.*, **29**, 1796–1810, https://doi.org/10.1002/joc.1821.

De Tomasi, F., M. M. Miglietta, and M. R. Perrone, 2011: The growth of the planetary boundary layer at a coastal site: A case study. *Bound.-Layer Meteor.*, **139**, 521–541, https://doi.org/10.1007/s10546-011-9592-6.

Devesse, K., L. Lanzilao, S. Jamaer, N. van Lipzig, and J. Meyers, 2022: Including realistic upper atmospheres in a wind-farm gravity-wave model. *Wind Energy Sci.*, **7**, 1367–1382, https://doi.org/10.5194/wes-7-1367-2022.

Docquier, D., W. Thiery, S. Lhermitte, and N. van Lipzig, 2016: Multi-year wind dynamics around Lake Tanganyika. *Climate Dyn.*, **47**, 3191–3202, https://doi.org/10.1007/s00382-016-3020-z.

Emeis, S., S. Siedersleben, A. Lampert, A. Platis, J. Bange, B. Djath, J. Schulz-Stellenfleth, and T. Neumann, 2016: Exploring the wakes of large offshore wind farms. *J. Phys. Conf.*

Ser., **753**, 092014, https://doi.org/10.1088/1742-6596/753/9/092014.

Fedorovich, E. E., and D. V. Mironov, 1995: A model for a shear-free convective boundary layer with parameterized capping inversion structure. *J. Atmos. Sci.*, **52**, 83–96, https://doi.org/10.1175/1520-0469(1995)052<0083:AMFASF>2.0.CO;2.

Fischereit, J., K. S. Hansen, X. G. Larsén, M. P. V. D. Laan, P. E. Réthoré, and J. P. M. Leon, 2022: Comparing and validating intra-farm and farm-to-farm wakes across different mesoscale and high-resolution wake models. *Wind Energy Sci.*, **7**, 1069–1091, https://doi.org/10.5194/wes-7-1069-2022.

Fitch, A. C., J. B. Olson, and J. K. Lundquist, 2013: Parameterization of wind farms in climate models. *J. Climate*, **26**, 6439–6458, https://doi.org/10.1175/JCLI-D-12-00376.1.

Frei, C., 2014: Interpolation of temperature in a mountainous region using nonlinear profiles and non-Euclidean distances. *Int. J. Climatol.*, **34**, 1585–1605, https://doi.org/10.1002/joc.3786.

Granados-Muñoz, M. J., F. Navas-Guzmán, J. A. Bravo-Aranda, J. L. Guerrero-Rascado, H. Lyamani, J. Fernández-Gálvez, and L. Alados-Arboledas, 2012: Automatic determination of the planetary boundary layer height using lidar: One-year analysis over southeastern Spain. *J. Geophys. Res.*, **117**, D18208, https://doi.org/10.1029/2012JD017524.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

IEA, 2021: World energy outlook 2021. Tech Rep., 386 pp., https://iea.blob.core.windows.net/assets/4ed140c1-c3f3-4fd9-acae-789a4e14a23c/WorldEnergyOutlook2021.pdf.

Kelly, M., R. A. Cersosimo, and J. Berg, 2019: A universal wind profile for the inversion-capped neutral atmospheric boundary layer. *Quart. J. Roy. Meteor. Soc.*, **145**, 982–992, https://doi.org/10.1002/qj.3472.

Lamb, H. H., 1972: *British Isles Weather Types and a Register of the Daily Sequence of Circulation Patterns 1861–1971*. Geophysical Memoirs, No. 116, H.M. Stationery Office, 85 pp., https://digital.nmla.metoffice.gov.uk/download/file/digitalFile_19cef4a2-3070-414c-b628-12308f5c978c.

Lang, F., Y. Huang, S. T. Siems, and M. J. Manton, 2018: Characteristics of the marine atmospheric boundary layer over the Southern Ocean in response to the synoptic forcing. *J. Geophys. Res. Atmos.*, **123**, 7799–7820, https://doi.org/10.1029/2018JD028700.

Lanzilao, L., and J. Meyers, 2022: Effects of self-induced gravity waves on finite wind-farm operations using a large-eddy simulation framework. *J. Phys. Conf. Ser.*, **2265**, 022043, https://doi.org/10.1088/1742-6596/2265/2/022043.

Liu, L., and R. J. A. M. Stevens, 2022: Vertical structure of conventionally neutral atmospheric boundary layers. *Proc. Natl. Aca. Sci. USA*, **119**, e2119369119, https://doi.org/10.1073/pnas.21193x69119.

Mahrt, L., and O. Acevedo, 2023: Types of vertical structure of the nocturnal boundary layer. *Bound.-Layer Meteor.*, **187**, 141–161, https://doi.org/10.1007/s10546-022-00716-7.

Ollier, S. J., S. J. Watson, and C. Montavon, 2018: Atmospheric gravity wave impacts on an offshore wind farm. *J. Phys. Conf. Ser.*, **1037**, 072050, https://doi.org/10.1088/1742-6596/1037/7/072050.

Pal, S., and M. Haeffelin, 2015: Forcing mechanisms governing diurnal, seasonal, and interannual variability in the boundary layer depths: Five years of continuous lidar observations over a suburban site near Paris. *J. Geophys. Res. Atmos.*, **120**, 11 936–11 956, https://doi.org/10.1002/2015JD023268.

Pandolfi, M., G. Martucci, X. Querol, A. Alastuey, F. Wilsenack, S. Frey, C. D. O'Dowd, and M. Dall'Osto, 2013: Continuous atmospheric boundary layer observations in the coastal urban area of Barcelona during SAPUSS. *Atmos. Chem. Phys.*, **13**, 4983–4996, https://doi.org/10.5194/acp-13-4983-2013.

Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.*, **12**, 2825–2830.

Platis, A., and Coauthors, 2018: First in *situ* evidence of wakes in the far field behind offshore wind farms. *Sci. Rep.*, **8**, 2163, https://doi.org/10.1038/s41598-018-20389-y.

——, and Coauthors, 2021: Evaluation of a simple analytical model for offshore wind farm wake recovery by in situ data and weather research and forecasting simulations. *Wind Energy*, **24**, 212–228, https://doi.org/10.1002/we.2568.

——, M. Hundhausen, A. Lampert, S. Emeis, and J. Bange, 2022: The role of atmospheric stability and turbulence in offshore wind-farm wakes in the German bight. *Bound.-Layer Meteor.*, **182**, 441–469, https://doi.org/10.1007/s10546-021-00668-4.

Porchetta, S., D. Muñoz-Esparza, W. Munters, J. van Beeck, and N. van Lipzig, 2021: Impact of ocean waves on offshore wind farm power production. *Renew. Energy*, **180**, 1179–1193, https://doi.org/10.1016/j.renene.2021.08.111.

Porté-Agel, F., M. Bastankhah, and S. Shamsoddin, 2020: Wind-turbine and wind-farm flows: A review. *Bound.-Layer Meteor.*, **174** (1), 1–59, https://doi.org/10.1007/s10546-019-00473-0.

Rampanelli, G., and D. Zardi, 2004: A method to determine the capping inversion of the convective boundary layer. *J. Appl. Meteor.*, **43**, 925–933, https://doi.org/10.1175/1520-0450(2004)043<0925:AMTDTC>2.0.CO;2.

Reyers, M., J. G. Pinto, and J. Moemken, 2015: Statistical–dynamical downscaling for wind energy potentials: Evaluation and applications to decadal hindcasts and climate change projections. *Int. J. Climatol.*, **35**, 229–244, https://doi.org/10.1002/joc.3975.

Schelbergen, M., P. C. Kalverla, R. Schmehl, and S. J. Watson, 2020: Clustering wind profile shapes to estimate airborne wind energy production. *Wind Energy Sci.*, **5**, 1097–1120, https://doi.org/10.5194/wes-5-1097-2020.

Schneemann, J., F. Theuer, A. Rott, M. Dörenkämper, and M. Kühn, 2021: Offshore wind farm global blockage measured with scanning lidar. *Wind Energy Sci.*, **6**, 521–538, https://doi.org/10.5194/wes-6-521-2021.

Schneider, T., 2007: The thermal stratification of the extratropical troposphere. *The Global Circulation of the Atmosphere*, Princeton University Press, 47–77, https://doi.org/10.1515/9780691236919-005.

Sculley, D., 2010: Web-scale *k*-means clustering. *WWW'10: Proc. 19th Int. Conf. on World Wide Web*, Raleigh, NC, Association for Computing Machinery, 1177–1178, https://doi.org/10.1145/1772690.1772862.

Seibert, P., F. Beyrich, S.-E. Gryning, S. Joffre, A. Rasmussen, and P. Tercier, 2000: Review and intercomparison of operational methods for the determination of the mixing height. *Atmos. Environ.*, **34**, 1001–1027, https://doi.org/10.1016/S1352-2310(99)00349-0.

Souverijns, N., W. Thiery, M. Demuzere, and N. P. M. Van Lipzig, 2016: Drivers of future changes in East African precipitation. *Environ. Res. Lett.*, **11**, 114011, https://doi.org/10.1088/1748-9326/11/11/114011.

Stull, R. B., 1988: *An Introduction to Boundary Layer Meteorology*. Atmospheric Sciences Library, Vol. 13, Kluwer Academic, 666 pp., https://doi.org/10.1007/978-94-009-3027-8.

Wallace, J. M., and P. V. Hobbs, 2006: *Atmospheric Science: An Introductory Survey*. 2nd ed. Elsevier Inc., 488 pp., https://doi.org/10.1016/C2009-0-00034-8.