Environmental Data Science (2026), 5: e3, 1–42
doi:10.1017/eds.2026.10031

CAMBRIDGE UNIVERSITY PRESS

APPLICATION PAPER

Which meteorological parameters influence extreme wind speed in a wind farm? A heterogeneous Granger causality approach

Kateřina Hlaváčková-Schindler[1], Rainer Wöss[1], Irene Schicker[2] and Claudia Plant[1]

[1]Research Group Data Mining and Machine Learning, Faculty of Computer Science and Data Science @ Uni Vienna, University of Vienna, Vienna, Austria
[2]GeoSphere Austria, Austria
Corresponding author: Kateřina Hlaváčková-Schindler; Email: katerina.schindlerova@univie.ac.at

Received: 12 July 2025; **Revised:** 14 November 2025; **Accepted:** 05 January 2026

Keywords: extreme wind speed; Granger causality; HMML; LiNGAM; wind energy

Abstract

For efficient wind farm management and optimized power generation under adverse weather conditions, understanding the causal meteorological drivers is essential. In this paper, we investigate the temporal causal influences of wind speed-related meteorological processes within a wind farm using the Heterogeneous Graphical Granger model (HMML). HMML is applied to synthetically generated wind power production data from Eastern Austria. To assess the plausibility of the identified causal processes, we compare the results with those obtained using the state-of-the-art LiNGAM method. Both methods are applied and evaluated across six different scenarios, each defined by distinct hydrological periods. The scenarios are defined by a set of time intervals characterized by either low/high extreme wind speeds or moderate wind speeds. We applied both methods across these scenarios and conducted causal reasoning to identify potential causes of extreme wind speeds within the wind farm. The sets of causal parameters obtained using HMML were found to be more realistic than those derived from LiNGAM. Combining the knowledge of causal variables affecting wind speed at the turbine hub, identified by HMML in each scenario, with weather forecasts can offer practical guidance for wind farm operators. Specifically, this knowledge can support more informed planning regarding when wind turbines should or should not be generating energy. For instance, the strong Granger-causal linkage identified between wind speed and temperature can inform curtailment strategies. In scenarios where rising temperatures are predictive of declining wind speeds, operators may preemptively adjust turbine output or schedule maintenance to optimize efficiency and reduce wear. Moreover, such predictive insights can feed into energy market models, where anticipated curtailment due to meteorological dependencies affects both generation forecasts and pricing strategies. By integrating these causal relationships into operational planning, the proposed tool offers a pathway toward more adaptive and economically efficient wind energy management.

Impact statement

Prediction focuses on the question of what might happen, while causal inference seeks to answer why it happens. In climate science, understanding the causal drivers, such as extreme wind events, enables us not only to explain phenomena but also to intervene, rather than merely forecast. In this context, identifying the most influential meteorological parameters that cause extreme wind speed events in wind farms is essential for effective wind energy

This research article was awarded Open Data badge for transparent practices. See the Data Availability Statement for details.

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

production management and planning. The results of our method provide causal indicators for six selected scenarios, defined by high, medium, and low wind speeds across two hydrological seasons. These insights can be used to support decision-making by wind farm operators.

1. Introduction

In recent years, renewable energy generation has increased rapidly, driven in part by the recent energy crisis and the global ambition to achieve fossil-free energy production by 2040 at the latest. However, as climate change continues to impact various aspects of daily life, including infrastructure and other critical sectors, it becomes increasingly important to investigate and understand the drivers and causal influences of adverse weather on sectors such as wind energy. In this context, machine learning methods have proven highly effective in uncovering causal relationships. Wind energy plays a pivotal role in global decarbonization efforts by providing a scalable, low-emission alternative to fossil fuel-based electricity generation. Wind power contributes not only to reducing greenhouse gas emissions but also to reshaping energy systems toward sustainability. Gazar et al. (2024) highlight the importance of causal inference tools in evaluating the environmental impacts of renewable energy projects, emphasizing the need for robust analytical frameworks to guide decision-making in complex socio-environmental contexts. Andersen and Geels (2023) argue that the speed and direction of net-zero transitions depend on the interplay between multiple systems—technological, institutional, and social—where wind energy serves as a key enabler of systemic change. However, as Sovacool et al., (2021) point out, the expansion of wind infrastructure must also account for issues of equity and justice, as decarbonization processes can lead to new forms of dispossession and socio-spatial inequality. In line with this, Köhler et al. (2019) call for a comprehensive research agenda that integrates technological innovation with societal transitions, highlighting the importance of interdisciplinary approaches to sustainability transformations in which wind energy is a central component.

An important direction in recent machine learning research is the focus on interpretability and explainability of methods. Graphical causal approaches inherently offer both of these properties. Recently, Runge et al. (2023) introduced a taxonomy of causal research questions in climatology. This taxonomy categorizes types of expert assumptions and the properties of available time series data. This framework provides a causal language through which researchers can formally define their study questions. Graphical Granger causality, and its nonlinear extensions, have played a sustained role in climatological research due to its ability to capture relationships among temporal variables. This is evidenced by a growing body of work over the past decades, see, e.g. Smirnov and Mokhov, (2009), Attanasio et al. (2013), Runge et al. (2015), Runge et al. (2023).

Focusing on the task of power generation in a wind farm and the associated airflow dynamics, the main objective of this work was to determine how relevant meteorological parameters influence selected low, medium, and high wind speed scenarios in a wind farm located in Eastern Austria, and to compare the differences among them. To achieve this, we employed the Heterogeneous Granger Causal Model HMML (Hlaváčková-Schindler and Plant, 2020), which enables the identification of Granger-causal relationships among processes that follow exponential distributions—distributions that more accurately reflect wind speed behavior in most of the scenarios considered.

Traditional correlation analyses cannot distinguish whether cloud coverage causes low wind speeds or whether both are consequences of synoptic-scale weather patterns. For wind farm operators, this distinction is critical: if cloud coverage is merely associated with low wind speeds, monitoring it provides no actionable insight. However, if cloud coverage causally influences boundary layer dynamics that subsequently affect hub-height wind speeds, it becomes a valuable leading indicator for operational decisions such as maintenance scheduling or reserve power allocation. Causal inference enables intervention: understanding that X causes Y allows prediction of how changes in X will affect Y, whereas correlation between X and Y provides no such predictive power for interventions Pearl (2009).

The objective is to provide the wind farm operators with more insights and a better understanding of the conditions leading to extreme wind speed events. This knowledge can support more informed

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

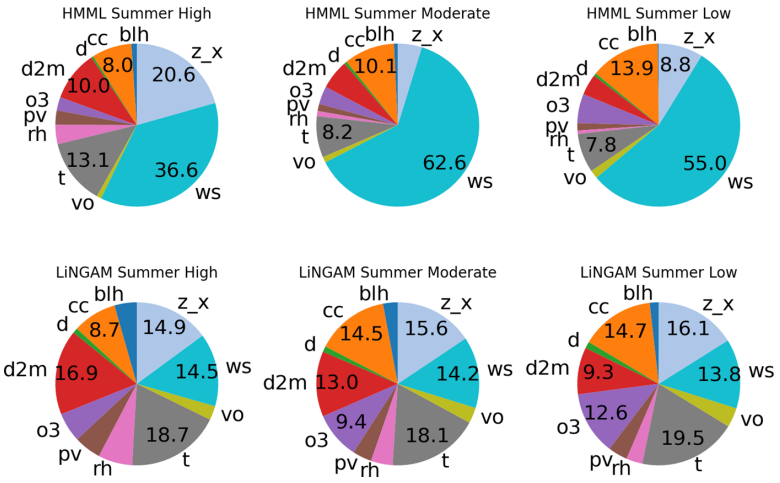


Figure 1. *Summary of the key results: The relative mean causal value of each variable to wind speed at the top of 135m high turbines for the summer in the wind speed scenarios, calculated by HMML and LiNGAM; The abbreviations denote: z—geopotential, blh—boundary layer height, d2m—dew point temperature at 2m, rel-h—relative humidity, ws—wind speed at 135 m, d—divergence, cc—cloud coverage, o3—ozone mixing ratio, pv—potential vorticity, t—temperature at 135, vo—relative vorticity.*

decision-making, enabling operators to optimize performance under prevailing conditions through enhanced wind farm management and power generation strategies.

The most related methods for causality among $p \geq 3$ non-Gaussian processes are LiNGAM (Shimizu et al., 2006) and the PCMCI method from Runge et al. (2019) developed for causal inference in multivariate time series in climate and meteorological time series. To the best of our knowledge, no non-Gaussian multivariate causal inference method has yet been applied to the analysis of wind farms and associated meteorological measurements.

To summarize the main contribution of our work, Figure 1 presents the results of our method in comparison to the benchmark method LiNGAM. Concretely, it shows the relative mean causal influence of each variable on wind speed measured at a height of 135 meters in summer across the scenarios of high, moderate, and low wind, as calculated by both HMML and LiNGAM. The values are average values at the locations of 38 turbines in a wind park in Eastern Austria. The average is over 100 scenarios of the same type and over all turbines. The spatial distribution of 38 turbines in the wind farm can be seen in Figure 2. One can see that the average distance between turbines is approximately 250 meters. The wind farm is situated on flat terrain, with elevation variations of less than 13 m across the approximately $\sim$2 km^2 area. All turbines have hub heights of 135 m, except turbine 38, which has a height of 99 m. The mean spacing among turbines is approximately 250 m. Topography is visualized using a hillshade overlay (azimuth 315°, altitude 45°) to highlight subtle elevation features. The flatness of the Pannonian Basin is evident in the narrow elevation range (115–130 m a.s.l.) in contrast to the Austrian Alps shown in panel (a), which reach elevations above >3000 m. Background topography is based on the EU Digital Elevation Model (EU-DEM) v1.1 in WGS84 projection. Turbine locations are sourced from open-access wind farm datasets, such as the Austrian Aviation Obstacle Database and the IG Windkraft website.

The following sections present the methodology, results, and conclusions of this work.

2. Method

2.1. Heterogeneous graphical Granger model and its estimation by method HMML

Granger causality, introduced by Granger (1969) to distinguish between a cause and effect, can be extended to the multivariate case, i.e. for $p > 2$ time series and model order $d \geq 1$, which is a time lag of past

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

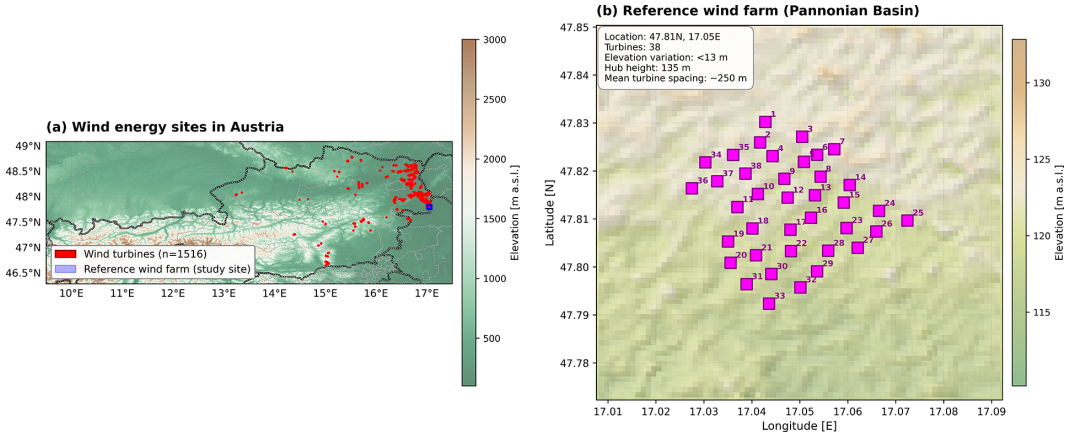


Figure 2. *Spatial distribution of wind energy infrastructure in Austria and a detailed view of the reference wind farm study site. (a) Wind turbines across Austria (red circles, n = 1516) overlaid on topography (EU-DEM v1.1, 100 m resolution). The blue box highlights the location of the reference wind farm (47.81°N, 17.05°E) in the eastern Pannonian Basin. Country borders (solid black curves) and federal state boundaries (dotted gray lines) are shown for reference. (b) Wind farm with spatial arrangement of 38 turbines (magenta squares) with turbine identification numbers…(1–38).*

lagged observations included in the model. The model order can be determined via an information theoretic criteria such as the Bayesian or Akaike information criterion. For p time-series $\mathbf{x}_1,..,\mathbf{x}_p$ the vector auto-regressive (VAR) model is:

$$x_i^t = \boldsymbol{X}_{t,d}^{Lag}\boldsymbol{\beta}_{i'} + \epsilon_i^t \tag{2.1}$$

where $\boldsymbol{X}_{t,d}^{Lag} = \left(x_1^{t-d},..,x_1^{t-1},..,x_p^{t-d},..,x_p^{t-1}\right)$. $\boldsymbol{\beta}_{i'}$ is the transposition of the matrix $\boldsymbol{\beta}_i$ of the regression coefficients and ϵ^t the error (Behzadi et al., 2019).

One can state that *the time-series* $\boldsymbol{x}_j$ *Granger-causes the time-series* $\boldsymbol{x}_i$ *for lag d and denote* $\boldsymbol{x}_j \rightarrow \boldsymbol{x}_i$, *for* $i,j = 1,\ldots,p$ *if and only if at least one of the d coefficients in row j of* $\boldsymbol{\beta}_i$ *is non-zero.*

Thus, to detect causal relations, the coefficients of the VAR model must be estimated. The estimation of $\boldsymbol{\beta}_i$ can be done by a penalized least squares or a maximum likelihood error of the order d, together with the lasso penalty (Tibshirani, 1996).

Multivariate Granger causality among time series, as defined in Eq. (2.1), is a special case of graphical causal models (Glymour et al., 2019). It assumes that the random error time series follow Gaussian distribution with zero mean and constant variance.

This assumption is often violated in practical applications, which can lead graphical Granger models to infer inaccurate or spurious causal relationships. To address this limitation, (Behzadi et al., 2019), building on the generalized linear model (GLM) framework introduced in Nelder and Wedderburn (1972), proposed a more general model for detecting Granger-causal relationships among $p \geq 3$ number of time series that follow a distribution from the exponential family. In this framework, the relationship between the response variable and the covariates in the regression is no longer linear but is defined through a so-called link function $\boldsymbol{\eta}$. This function is a monotonic, twice-differentiable and depends on a specific distribution from the exponential family.

The heterogeneous graphical Granger model (HGGM), (Behzadi et al., 2019), considers time series $\mathbf{x}_i$ that follow a distribution from the exponential family, characterized by a canonical parameter $\boldsymbol{\theta}_i$. The generic density form for each $\mathbf{x}_i$ can be written as:

$$p\left(\mathbf{x}_i|\mathbf{X}_{t,d}^{Lag},\boldsymbol{\theta}_i\right) = h(\mathbf{x}_i)\exp(\mathbf{x}_i\boldsymbol{\theta}_i - \eta_i(\boldsymbol{\theta}_i)) \tag{2.2}$$

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

where $\boldsymbol{\theta}_i = \mathbf{X}_{t,d}^{Lag}\left(\boldsymbol{\beta}_i^*\right)'$, with $\boldsymbol{\beta}_i^*$ being the optimum, and η_i is a link function corresponding to a time series $\mathbf{x}_i$. The HGGM applies the idea of generalized linear models and to time series in the following form

$$x_i^t \approx \mu_i^t = \eta_i^t\left(\mathbf{X}_{t,d}^{Lag}\boldsymbol{\beta}_i'\right) = \eta_i^t\left(\sum_{j=1}^{p}\sum_{l=1}^{d} x_j^{t-l}\beta_j^l\right) \tag{2.3}$$

for x_i^t, $i = 1,\ldots,p, t = d+1,\ldots,n$, each having a probability density from the exponential family; $\boldsymbol{\mu}_i$ denotes the mean of $\mathbf{x}_i$ and $var(\mathbf{x}_i|\boldsymbol{\mu}_i,\phi_i) = \phi_i v_i(\boldsymbol{\mu}_i)$ where ϕ_i is a dispersion parameter and v_i is a variance function depending only on $\boldsymbol{\mu}_i$; η_i^t is the t-th coordinate of $\boldsymbol{\eta}_i$.

The causal inference in (2.3) can be solved as a maximum likelihood estimate for $\boldsymbol{\beta}_i$ for a given lag $d > 0$, $\lambda > 0$, and all $t = d+1,\ldots,n$ by adding an adaptive lasso penalty function (Behzadi et al., 2019). Similarly, one can say that *the time series* $\mathbf{x}_j$ Granger–causes time series $\mathbf{x}_i$ *for a given lag d*, *and denote* $\mathbf{x}_j \rightarrow \mathbf{x}_i$, *for* $i,j = 1,\ldots,p$ if and only if at least one of the d coefficients in $j-th$ row of $\hat{\boldsymbol{\beta}}_i$ *of the penalized solution is non-zero*, see (Behzadi et al., 2019).

Hlaváčková-Schindler and Plant (2020) developed the HMML method, which outperformed HGGM in terms of causal inference precision. The idea of the HMML method for the estimation of $\boldsymbol{\beta}_i$ coefficients is the following: it replaces the solution via p penalized linear regression problems by formulating the feature selection as a combinatorial optimization problem, as was done in Hlaváčková-Schindler and Plant (2020) for the multivariate Granger causal model with time series from the exponential family. It uses the information theoretic criterion "minimum message length" (MML), introduced by (Wallace, 2005) for general inference problems, to determine causal connections in the model. As demonstrated in Hlaváčková-Schindler and Plant (2020), applying this principle significantly improves the precision of causal inference, especially for "short" time-series (The length of a short time series is of the order of at most hundreds of the number of involved time series.). The MML principle is an information-theoretical formulation of Occam's razor: even when models have a comparable goodness-of-fit to the observed data, the one generating the shortest overall message is more likely to be correct (where the message consists of a statement of the model, followed by a statement of data encoded concisely using that model). The statistical version of the MML principle constructs a description in terms of probability functions and some prior knowledge of the parameter vector. MML seeks the model that minimizes this trade-off between model complexity and model capability. In the type of MML considered in Hlaváčková-Schindler and Plant (2020) and in this study and application, the parameter space $\boldsymbol{\theta}$ for the statistical model $p(.|\boldsymbol{\theta})$ is decomposed into a countable number of subsets and associated code words for members of these subsets. The parameter $\boldsymbol{\theta}$ in the MML criterion corresponds to the maximum likelihood estimates of the regression coefficients $\boldsymbol{\beta}_i$ and the dispersion coefficient of the target time series.

Each regression problem for $i = 1,\ldots,p$ is expressed via the incorporation of a subset of indices of regressor variables, denoted by $\boldsymbol{\gamma}_i \subseteq \{1,..,p\}$ and $k_i = |\boldsymbol{\gamma}_i|$ into (2.3)

$$x_i^t = \eta_i^t\left(\mathbf{X}_{t,d}^{Lag}(\boldsymbol{\gamma}_i)\boldsymbol{\beta}_i'(\boldsymbol{\gamma}_i)\right) = \eta_i^t\left(\sum_{j=1}^{k_i}\sum_{l=1}^{d} x_j^{t-l}\beta_j^l\right) \tag{2.4}$$

where $\mathbf{X}_{t,d}^{Lag}(\boldsymbol{\gamma}_i)$ is the design matrix with regressors only from $\boldsymbol{\gamma}_i$ and $\boldsymbol{\beta}_i'(\boldsymbol{\gamma}_i)$ are their regression coefficients. The best structure of $\boldsymbol{\gamma}_i$ in the sense of the MML principle is determined either by a genetic algorithm, version HMMLGA, or an exhaustive search algorithm, version exHMML. For more details, see Hlaváčková-Schindler and Plant (2020). Similarly to the Gaussian case (Eq. 2.1), the time lag (i.e. the model order) of the target variable in HMML can be determined by expert knowledge or by the information-theoretic criteria. Since HMML is an instance of GLM models, the consequences about collinear or almost collinear time series also hold for HMML. Collinearity does not violate any assumptions of GLMs, unless there is perfect collinearity.

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

2.1.1. Performance of HMML on synthetic data with known ground truth causal graphs

Hlaváčková-Schindler and Plant (2020) examined the precision of the output causal graphs obtained by HMML in comparison to benchmark methods on synthetic data, where the ground truth, i.e. the target causal graph, was known. Randomly generated processes having an exponential distribution of Gaussian and gamma were examined together with the correspondingly generated target causal graphs. The performance of of both versions of HMML, namely HMMLGA, exHMML, as well as of the benchmark methods HGGM (Behzadi et al., 2019), SFGC (Kim et al., 2011) and LINGAM (Shimizu et al., 2006), depends on various parameters including the number of time series (features), the number of causal relations in Granger causal graph (dependencies), the length of time series and finally on the lag parameter. Behzadi et al. (2019) observed that varying the lag parameter from 3 to 50 did not influence either the performance of HGGM nor SFGC significantly. Based on that, the considered lags were 3 and 4 for all methods in the synthetic experiments. Causal graphs with mixed types of time series for $p = 5$ and $p = 8$ number of features were examined. For each case, causal graphs with higher edge density (dense case) and lower edge density (sparse case) were considered, which corresponds to the parameter "dependency" in the code, where the full graph has for p time series $p(p-1)$ possible directed edges. The length of the generated time series varied from 100 to 1000. In the experiments with causal networks with 5 and 8 time series, 5 time series with 2 gamma, 2 Gaussian, and 1 Poisson distributions were considered, which we generated randomly together with the corresponding network. For the denser case with 5 time series, random graphs were generated with 18 edges and for the sparser case, random graphs with 8 edges. The results of the experiments on causal graphs with 5 time series are presented in Table 1. Each value in the rows represents the mean value of all F-measures over 10 random generations of causal graphs for length n and lag d. For dependency 8, strength = 0.9 of causal connections was taken, for dependency 18, strength = 0.5.

Table 1 shows that HMMLGA and exHMML gave considerably higher precision in terms of F-measure than the other three comparison methods for all considered n up to 1000. In the second network, 8 time series with 7 gamma and 1 Gaussian distributions were considered, which were generated randomly together with a corresponding network. Dense and sparse graphs were considered. Similarly, in the experiments with $p = 5$, both exHMML and HMMLGA gave considerably higher F-measure than the comparison methods for the considered n up to 1000. The results of the experiments and more details can be found in Table 2 in Hlaváčková-Schindler and Plant (2020). The same paper presents the application of HMML (version HMMLGA) and LiNGAM to infer causal graph dynamics among selected climatological processes during rainy and dry period with the result of more realistic causal graph for the HMML

Table 1. *Hlaváčková-Schindler and Plant (2020): time series, average F-measure for methods HMML, HGGM, LiNGAM, and SFGC. The first subtable is for $d = 3$, the second one for $d = 4$*

dense g. 18, $n =$	100	300	500	1000;	sparse g. 8, $n =$	100	300	500	1000
exHMML	0.69	0.83	0.82	0.88	**exHMML**	0.70	0.72	0.72	**0.67**
HMMLGA	**0.73**	**0.90**	**0.89**	**0.90**	**HMMLGA**	**0.73**	**0.76**	**0.74**	**0.67**
HGGM	0.5	0.48	0.54	0.52	HGGM	0.52	0.36	0.66	0.36
LiNGAM	0.57	0.58	0.62	0.58	LiNGAM	0.58	0.54	0.69	0.45
SFGC	0.33	0.26	0.26	0.33	SFGC	0.14	0.35	0.44	0.31
dense g. 18, $n =$	100	300	500	1000;	sparse g. 8, $n =$	100	300	500	1000
exHMML	0.71	0.73	0.83	0.83	**exHMML**	**0.67**	**0.80**	**0.80**	0.68
HMMLGA	**0.82**	**0.79**	**0.87**	**0.92**	**HMMLGA**	**0.67**	0.73	0.77	**0.70**
HGGM	0.44	0.37	0.40	0.39	HGGM	0.53	0.47	0.65	0.36
LiNGAM	0.71	0.58	0.58	0.65	LiNGAM	0.33	0.52	0.74	0.46
SFGC	0.43	0.55	0.42	0.63	SFGC	0.35	0.59	0.42	0.38

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

Table 2. *An example of an excerpt of measurements of selected variables for turbine "1" on May 18, 2017*

	turb_ index	time	z_x (m^2s^{-2})	blh (m)	d2m (K)	rel_humidity (%)	ws (ms^{-1})	d (s^{-1})	cc (%)	o3 ($kgkg^{-1}$)	pv (kg^{-1})	t (K)	vo ($m^2s^{-1}kg^{-1}$)
5666822	1	2017–05–18 14:00:00	0.770995	2.800336	0.013221	–2.731491	0.869545	–0.034922	–0.36543	0.046724	–0.575555	1.152448	–0.135121
5666823	1	2017–05–18 15:00:00	0.770995	2.789344	0.006970	–2.723470	0.809865	–0.290214	–0.36543	0.024580	–0.570066	1.173738	–0.078755
5666824	1	2017–05–18 16:00:00	0.770995	2.737826	–0.015868	–2.700197	0.782642	–0.241364	–0.36543	0.008387	–0.559012	1.168723	–0.049297
5666825	1	2017–05–18 17:00:00	0.770995	2.236300	–0.015443	–2.617868	0.696706	–0.017124	–0.36543	–0.004526	–0.547915	1.137549	0.017904
5666826	1	2017–05–18 18:00:00	0.770995	0.521569	0.104885	–2.347876	0.730746	0.221854	–0.36543	0.004648	–0.534643	1.095379	–0.012785
5666827	1	2017–05–18 19:00:00	0.770995	–0.122236	0.238547	–1.888304	0.892757	0.707102	–0.36543	0.037969	–0.523183	1.056456	–0.217848
5666828	1	2017–05–18 20:00:00	0.770995	–0.247371	0.343639	–1.510327	0.936069	0.551278	–0.36543	0.076749	–0.533447	1.029846	–0.515331
5666829	1	2017–05–18 21:00:00	0.770995	–0.322509	0.385495	–1.241618	0.919796	0.285232	–0.36543	0.119750	–0.534713	1.005269	–0.625763
5666830	1	2017–05–18 22:00:00	0.770995	–0.359932	0.383213	–1.039513	0.944726	0.732300	–0.36543	–0.015654	–0.535805	0.997749	–0.767186
5666831	1	2017–05–18 23:00:00	0.770995	–0.379351	0.369368	–0.894722	0.953857	1.030358	–0.36543	0.018195	–0.522176	0.973753	–1.001889

method. The highest precision of HMML in comparison to other methods in synthetic experiments, as well as more plausible results in the precipitation, we consider HMML to be a good choice for causal inference among wind-related processes in the wind farm. For the equations and criteria for computing the causal values explicitly, we refer to Hlaváčková-Schindler and Plant (2020). In this paper, we use the MML criterion only for the target variable wind speed (one i).

2.1.2. Limitations and advantages of HMML

Granger causality has faced criticism since its introduction. It does not account for counterfactuals, see e.g. Mannino and Bressler (2015), Maziarz (2015), nor does it satisfy causal sufficiency, meaning it may overlook hidden common causes. Defending his method, Granger (1988) noted that causation is considered only among variables where prior theoretical belief suggests a causal link. Thus, conclusions about causal direction require domain knowledge of the mechanisms involved. While Granger causality does not capture all aspects of causality, it remains a valuable tool for empirical analysis (see Singh and Hawati, 2019). Besides other works, Runge et al. (2019) demonstrated that graphical Granger methods are empirically effective in large-scale climate and economic datasets.

Like most Granger causality methods, HMML assumes that the underlying time series are stationary. While HMML is designed for heterogeneous data in the sense of their probability distributions, it still requires that the heterogeneity be captured in a way that the encoding scheme can exploit, namely, for exponential probabilities. Noise can lead to spurious dependencies being encoded as causal links or true causal links being missed, however, this also holds for the other methods we used for comparison. The robustness of HMML to noise depends on the effectiveness of the message length criterion in penalizing such spurious patterns. Granger causality depends on the choice of lag length. HMML must either assume or estimate appropriate lags. Regarding the computation time, the MML-based search over possible graph structures can be computationally intensive, especially for large numbers of variables. This may limit scalability to a higher number p of processes.

Method HMML, in general, has the following advantages. It supports mixed data: HMML can work with both continuous and discrete variables, which is a major advantage over classical Granger causality methods that typically assume all variables are continuous and Gaussian. HMML handles nonlinear and non-Gaussian data. Further, it allows for modeling interactions between variables of different types without needing to transform or homogenize the data. The MML criterion balances model complexity and goodness of fit, helping to avoid overfitting and underfitting. HMML uses MML to infer the optimal graphical structure (i.e., causal graph), which includes selecting relevant variables and lags without manual tuning. The graphical output of HMML produces interpretable causal graphs. HMML is suitable for moderately high-dimensional data (for p up to 20).

Since some climatological processes are better fitted by exponential distributions than by a Gaussian one, using HMML can be beneficial for inference on our data set. As discussed in Section 2.1.1, HMML demonstrated significantly higher precision of causal inference regarding the compared methods on synthetic time series, particularly for short time series. This is relevant to our case, as we work with short time series. Specifically, we analyze measurements consisting of 96 temporal samples for wind and related meteorological data.

2.2. HMML compared conventional techniques

To compare the HMML approach with conventional techniques like rank correlation and linear correlation analysis, especially in the context of identifying drivers of extreme wind speeds, one needs to consider several dimensions: causality, data heterogeneity, temporal dynamics, and robustness. HMML identifies which variables cause changes in wind speed, not just those that co-vary. This is crucial for understanding drivers of extreme events. HMML incorporates lagged relationships, capturing how past values of predictors (i.e. selected climatological variables in the presented work) influence future wind speeds. Correlation methods ignore time ordering and missing key dynamics. HMML is designed to handle the data heterogeneity inherently, unlike correlation methods, which require preprocessing or

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

transformation. It does not assume linearity or Gaussian distributions, making it suitable for complex meteorological phenomena. Correlation methods often fail when relationships are nonlinear or involve threshold effects (common in extreme weather). The Minimum Message Length (MML) principle penalizes overly complex models, helping HMML avoid overfitting and spurious causal links. This is especially valuable in noisy environmental data. The output of HMML is a causal graph showing which variables influence wind speed and how strongly. This is an interpretable graphical output. This aids in visualizing and communicating the structure of dependencies, which is not possible with correlation matrices.

2.3. Wind farm data and HMML

The data used in this study consist of synthetically generated wind power production data, along with accompanying meteorological parameters derived from ERA5 reanalysis data (Hersbach et al., 2020). This data set was generated for a wind farm located in Eastern Austria, consisting of 24 meteorological parameters for the 38 individual wind turbines. (The full ERA5 dataset is available from Copernicus Climate Data Store: https://cds.climate.copernicus.eu/datasets.) The data set is available with hourly resolution for the past 21 years (2000–2020). The following parameters are used: geopotential in m^2s^{-2} (z), boundary layer height in m (blh), dew point temperature at 2m in K (d2m), relative humidity in $\%$ (rel-h), wind speed at 135 m in ms^{-1} (ws), divergence in s^{-1} (d), cloud coverage in $\%$ (cc), ozone mixing ratio in $kgkg^{-1}$ (o3), potential vorticity in $m^2s^{-1}kg^{-1}$ (pv), temperature at 135 in K (t), relative vorticity in $m^2s^{-1}kg^{-1}$ (vo). A short excerpt of the data measurements for the turbine labeled "1" is shown in Table 2.

The target variable is wind speed at 135m, corresponding to the hub height of the turbines. Pressure (surface and at 1000 hPa) was not considered as a feature as the idea was to look into not-so-obvious causal relationships. Pressure is a confounding parameter rather than a causal one to the target and other variables. The aim of our study was to identify by HMML, which meteorological variables have a causal effect on wind speed at 135 meters, and consequently, on power production.

Remark 2.1. Since both HMML and LiNGAM rely on autoregressive modeling, the target variable, wind speed at 135 meters in our application, is included among the set of causal variables in both methods. However, our primary interest lies in identifying which other variables, aside from wind speed at 135 meters, have a causal influence on this target variable.

2.3.1. Validation of synthetic data

To assess the realism of the ERA5-derived synthetic dataset, we compared synthetic wind speeds and power estimates from ERA5 with anonymized turbine-level measurements from the same wind farm over the period 2016–2021. Figure 3 summarizes the results.

In terms of power production, Figure 3 shows daily-aggregated power production comparing ERA5-synthetic values to measured operational data. The validation reveals strong agreement with Pearson correlation $r = 0.920$ and normalized mean absolute error (NMAE) of 27.1%. The monthly-averaged time series (Figure 3a) demonstrates ERA5 captures seasonal and inter-annual variability. The scatter plot (Figure 3b) confirms good linearity ($R^2 = 0.807$), though with a systematic bias of +1569 kWh/day (9.6%), attributable to ERA5's 30km grid not resolving microscale turbulence and wake effects. The wind speed validation (Figure 3) compares ERA5 wind speeds at 135m to measured hub-height observations. Results confirm that ERA5 adequately captures wind speed variability at this site. Seasonal distributions show ERA5 slightly overestimates high-wind events but captures the overall distribution shape. The strong temporal correlation confirms ERA5 captures the meteorological variability patterns that drive wind speed changes. While absolute power production differs due to turbine-specific power curves not modeled in ERA5, the day-to-day and month-to-month variations, which are the basis for causal inference, are well represented. Therefore, ERA5 is suitable for identifying which meteorological

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

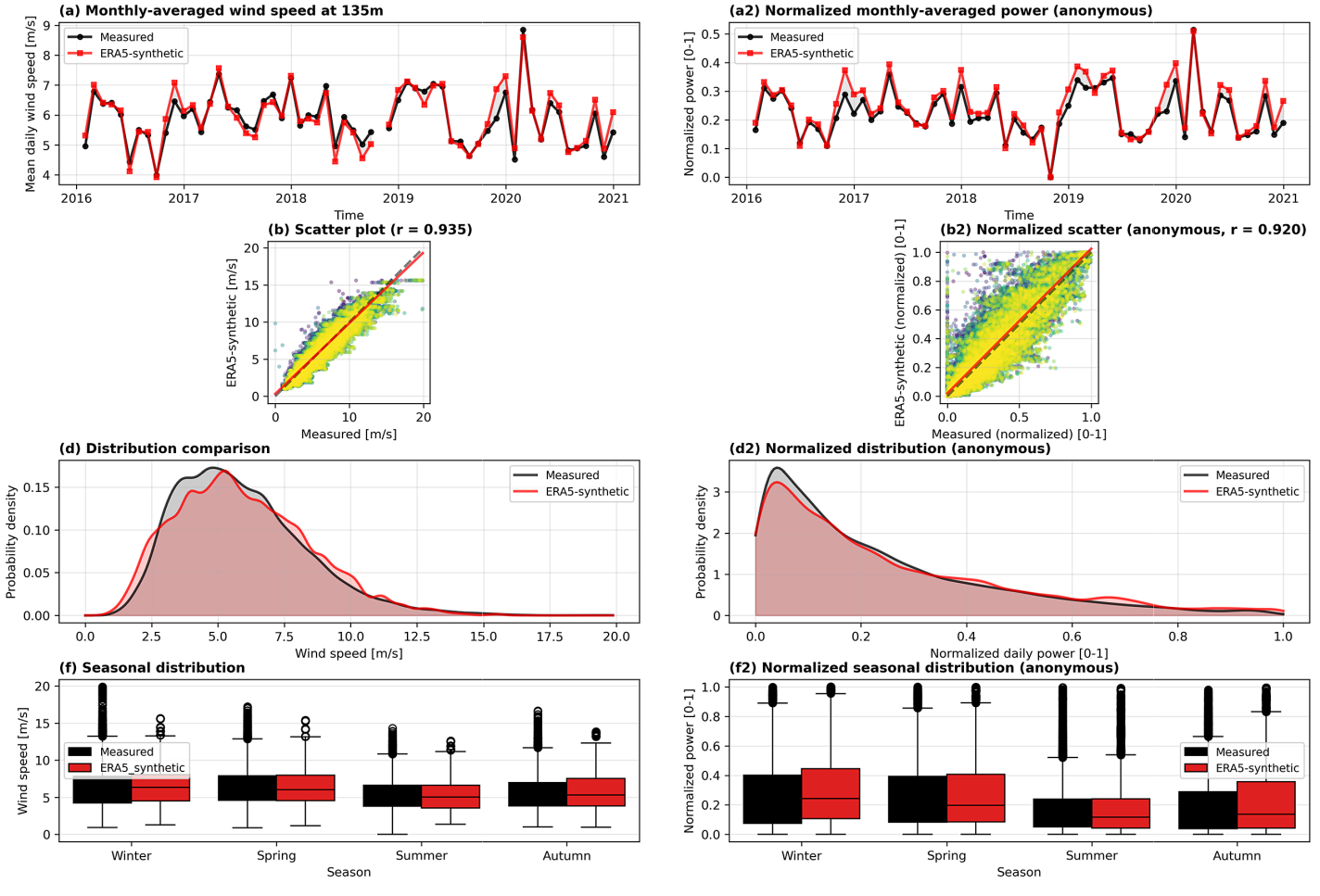


Figure 3. *Validation of the* ***ERA5-based synthetic dataset*** *against anonymized wind-farm observations. Left: Comparison of measured and ERA5-synthetic wind speeds at 135 m. Panels (a–f) show monthly averages, scatter relation, probability density, and seasonal distributions. Right: Normalized power comparison between measured and ERA5-synthetic data (anonymized). Panels (a2–f2) show monthly averages, scatter relation, distributions, and seasonal patterns. The synthetic series reproduces the temporal variability and magnitude of both wind speed and power well (correlation* $r = 0.94$ *for wind speed and* $r = 0.92$ *for normalized power). Probability-density and seasonal distributions confirm that the synthetic data capture the observed spread and seasonal cycle, supporting their suitability for the causal-analysis framework.*

variables drive wind speed variability. However, absolute thresholds should be interpreted with caution and adapted when applying the method to other wind turbine types.

All time series were standardized. Due to the nature of wind speed dynamics, only short time series are relevant. Therefore, we used the previous 96 hours of data, including wind speed (for the selected scenario) and a time series of equal length for the other meteorological variables. The HMML method was applied separately to each turbine, as well as across six defined scenarios at 135 meters height: for both the winter and summer hydrological half-years, we considered periods of low extreme wind, moderate wind, and high extreme wind.

The summer and winter half-years are defined as follows: the summer half-year includes all samples observed between March 20 and September 22, while the winter half-year includes samples from September 23 to March 19, respectively.

Our research objectives were as follows: First, we identified intervals of extreme wind events and moderate wind periods for each turbine in the wind farm over the past 21 years (2000–2020). Second, we applied HMML and LiNGAM to estimate the corresponding β-values for each meteorological parameter, indicating the strength of its causal relationship to wind speed in a given scenario and turbine. To ensure the reliability of the β-values over the 21-year period for each half-year, we averaged the results across 100 scenarios distributed throughout these years. Dividing the analysis into six scenarios enabled us to perform causal reasoning based on the values obtained by HMML and LiNGAM. Specifically, we compared the β-values of the meteorological variables across scenarios, evaluated the differences among them, and interpreted the plausibility of the results.

To make the findings from HMML and LiNGAM more accessible to wind power producers, we visualized the individual causal variables for each of the 38 turbines in the wind farm.

3. Related work

The related work in this paper focuses on the graphical causal models using multivariate time series, namely the graphical Granger causality and its non-linear extensions, the method LiNGAM from Shimizu et al. (2006) and the PCMCI method from Runge et al. (2019).

Causal inference methods, i.e. the bivariate Granger causality, Granger (1969), and its multivariate or non-linear graphical extensions, have been used to address causal questions in observational time-series data relying on different types of assumptions, see e.g. Zhu et al. (2015, Papagiannopoulou et al. (2017, Behzadi et al. (2019, Hlaváčková-Schindler and Plant (2020, Silva et al. (2021), Hlaváčková-Schindler et al. (2022). Since these methods are designed for time series, they can provide insights into complex dynamical systems, such as the Earth's climate, where conducting experiments or randomized trials is infeasible or impossible. Zhu et al. (2015) proposed the so-called spatio-temporal extended Granger causality model to analyze causalities among urban dynamics for air quality estimation using geographically sparse time-series data. Papagiannopoulou et al. (2017) emphasize the necessity of non-linear extensions of Granger Causality and propose a non-linear framework improving the predictive power of GC. Behzadi et al. (2019) introduced the method HGGM (Heterogeneous Graphical Granger Model), especially suited to time-series generated by distributions from the exponential family. The method uses adaptive lasso as a variable selection method. HGGM was applied to investigate spatio-temporal relationships in German and Austrian climatological data sets in Behzadi et al. (2019). The same model but with an algorithm based on minimum message length (HMML) was applied to the causal analysis of wind speed extreme events from Hersbach et al. (2020) data of hourly meteorological parameters in Hlaváčková-Schindler et al. (2022).

LiNGAM (Shimizu et al., 2006) was also proposed for multivariate causal modeling among time series. The method estimates causal structures in Bayesian networks among non-Gaussian time series using structural equation models and independent component analysis.

Runge et al. (2019) proposed method PCMCI for causal inference in time series of observational data. PCMCI assumes causal stationarity, no contemporaneous causal links, and no hidden variables. It outputs

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

directed lagged links and undirected contemporaneous links. More recent adaptation of PCMCI is PCMCI+ (Runge, 2020) and allows contemporary links with outputs of directed lagged links.

Since the LiNGAM method is based on a vector autoregressive model and is therefore conceptually similar to HMML, we will compare our results with those obtained using LiNGAM in our experiments.

4. Experiments

To apply both the HMML and LiNGAM methods, we identified scenarios based on the mean value wind speed over a 96-time-step window. The considered intervals were: high-wind extreme [9.5,18] m/s, low-wind extreme [1,4] m/s and moderate wind [6,8] m/s.

The thresholds for defining "high-wind extreme" ([9.5, 18] m/s) and "low-wind extreme" ([1, 4] m/s) are based on conditions approaching turbine rated capacity, typically 12-15 m/s for modern turbines, where curtailment decisions become critical (Manwell et al., 2010; Pryor and Barthelmie, 2021). Low extreme [1-4 m/s] captures near cut-in conditions, typically 3-4 m/s, where forecasting challenges are greatest and correspond to sub-optimal generation conditions near the cut-in speed (Manwell et al., 2010). Moderate [6-8 m/s] represents the optimal operating range. These thresholds correspond to the 15th, 40th-60th, and 85th percentiles, respectively of the 21-year wind speed distribution at hub height, and align with IEC 61400-1 wind turbine class considerations for the region.

We first filtered the data from the years 2000 to 2020 according to hydrological half years, summer and winter. For each half-year, we calculated the mean wind speed within each 96-sample window. If the mean fell within one of the predefined intervals, the timestamp of the window's onset was stored in the corresponding set of appropriate windows. To avoid overlap between the used windows, a minimum gap of two time steps from previously selected windows was enforced. Then, for each scenario and turbine, we randomly sampled 100 of these appropriate windows uniformly to perform causal inference and averaged the results.

The dominance of geopotential (z_x) in HMML's summer high-wind scenario (beta value 20.6%) aligns with the established understanding that extreme wind events in this region are driven by synoptic-scale pressure gradients, see Ban et al. (2014). The elevated influence of temperature (beta value $t = 13.1\%$) and dewpoint (beta value $d2m = 10.0\%$) reflects thermal instability that promotes vertical mixing. In contrast, winter high-wind events show a stronger dew point influence (beta value $= 11.1\%$), indicating frontal passages where moisture gradients are key indicators.

4.1. Maximum lag in time series and best-fitting exponential distribution

Both HMML and LiNGAM require determining the maximum lag of the target wind speed time series. Additionally, HMML requires identifying the best-fitting exponential distribution for each of the 100 wind speed time series.

In HMML, the maximum lag d was determined using the Akaike Information Criterion (AIC), defined as $AIC = 2k - 2l$ where l is the log likelihood, k is the number of parameters and n number of samples used for fitting. We constructed several $AR(d)$ autoregressive models with varying lag values d and selected that one with the lowest AIC value. Theoretically, one could use the BIC criterion. However, under the assumption time series follows an exponential distribution (which is light-tailed and non-Gaussian), the AIC is preferred over BIC for these reasons: It is asymptotically efficient for prediction, meaning it tends to select models that generalize better to new data. For non-Gaussian data (like exponential), where the true model may not be in the candidate set, AIC is more robust in selecting a model that approximates the data-generating process well. BIC is derived under the assumption that the true model is in the candidate set and that the likelihood is well-behaved, often assuming Gaussian errors. Simulation studies and empirical analyses, e.g. Hurvich and Tsai (1989) show that AIC tends to perform better in selecting lag lengths when the data is non-normal, heteroscedastic, or light-tailed, as is the case with exponential distributions.

The best-fitting distribution of wind speed within each interval was identified using the Residual Sum of Squares (RSS) and the Kolmogorov–Smirnov (K–S) test. The candidate distributions tested included

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

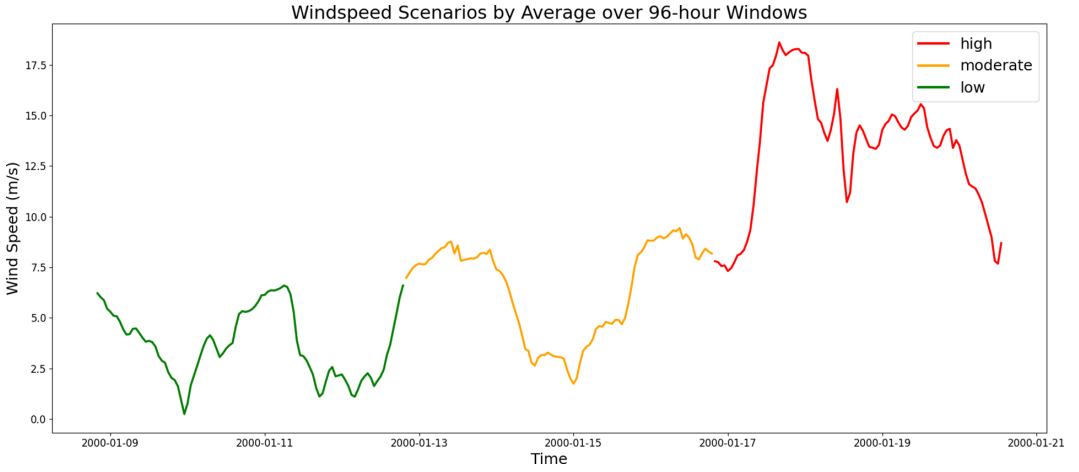


Figure 4. *The graph shows the wind speed in m/s over a time frame of 12 days (three 96-hour windows) in January 2000, given at the wind turbine with index zero. It showcases that we identify scenarios for low (green color), moderate (yellow), and high wind speed (red) based on it s average in the given time window.*

the Gaussian distribution, the inverse Gaussian distribution, and the gamma distribution. Different scenarios may exhibit different best-fitting distributions, as illustrated in Figure 4, which shows wind speed over time for the year 2007. In this example, wind speed in both extreme scenarios followed a Gaussian distribution, whereas the moderate wind scenario was best described by a gamma distribution.

4.2. β_i-Proportionality measure and statistical validation of causal values

We applied both methods to the six scenarios and provided lists of β values. The lists correspond to the p variables ($p = 11$ variables in our case) that were considered to find a causal connection to the target variable, wind speed measured at a height of 135 meters. For HMML, each list contains d entries, where d is the lag value determined for each scenario. For LiNGAM, we obtain $d + 1$ values, as it accounts for contemporaneous contributions. The lag d varied between 2 and 10 across different scenarios. From the β values for each scenario, we compute the

$$\beta_i\text{-proportionality} = \sum_{l=1}^{d} \frac{|\beta_i^l|}{\sum_{j=1}^{p}\sum_{k=1}^{d}|\beta_j^k|} \tag{4.1}$$

that can be understood as the relative causal strength of the i^{th} variable with respect to all p variables.

To ensure statistical validity of the causal values, we compute the arithmetic mean of the β_i-proportionality across all 100 windows per turbine and scenario. This results in a total of 3,800 windows per scenario.

4.3. Visualization of the causal values in the farm

The β_i-proportionality can be interpreted as the percentage contribution of each individual variable to the prediction of the target variable wind speed. These percentages are visualized as pie charts (see Figure 5 for HMML and Figure 6 for LiNGAM), where each pie represents one of the 38 turbines in the wind park, identified by the index at the center of the pie). The pie charts are spatially arranged according to the geographic coordinates of the respective turbines.

Table 3 presents the average β_i-proportionality values across all 38 turbines for each scenario for both HMML and LiNGAM. The three strongest causal variables are indicated in bold font, with the largest

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

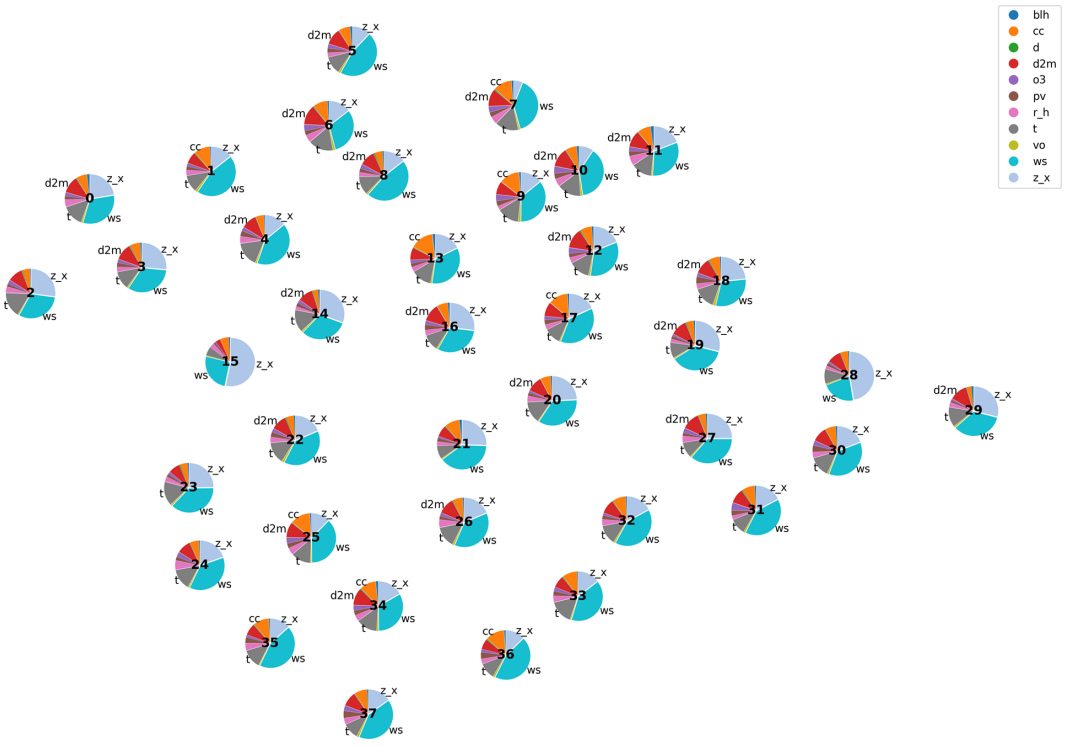


Figure 5. *β_i-proportionality of all variables in a pie diagram for each turbine selected by HMML for the summer high-wind scenario (averaged over 100 time windows). Variables with up to 5 the most significant percentage values are denoted at each turbine. The concrete values for the rest of the variables are in Table 4.*

value in blue. Moreover, we provide also the list of β_i-proportionality values for individual turbines for HMML in Table A1 and LiNGAM in Table A2 in Appendix A.1.

In contrast, most turbines analyzed with HMML share the same three causal variables with the highest magnitudes. Notably, HMML also shows a much higherproportionality value for turbines 15 and 28 compared to the rest. In the averaged results shown in Table 2, HMML consistently identifies wind speed as the dominant contributor, whereas LiNGAM selects a broader range of variables with more evenly distributed causal strengths. Wind speed appears among the top three contributors in only 3 out of 6 scenarios for LiNGAM, while it is the top contributor in all six scenarios for HMML.

The following observations can be made from Figures 5 and 6: with LiNGAM, a clear distinction is visible in the causal variables for turbines 15 and 28, where *z_x* being selected as the strongest causal variable compared to the others. In contrast, most turbines analyzed with HMML share the same three causal variables with the highest magnitudes. Also here, *z_x* has a much larger β_i proportionality value in turbines 15 and 28, compared to other turbines. In the averaged results shown in Table 3, HMML consistently identifies wind speed as the dominant contributor, whereas LiNGAM selects a broader range of variables with more evenly distributed causal strengths. Wind speed appears among the top three contributors in only 3 out of 6 scenarios for LiNGAM, while it is the top contributor in all six scenarios for HMML.

4.4. Stability of the β_i-proportionality values

To examine the stability of the causal values obtained by HMML with respect to the lag parameter *d* and the time-series length *n*, additional experiments were conducted using the approach described in the previous sections.

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

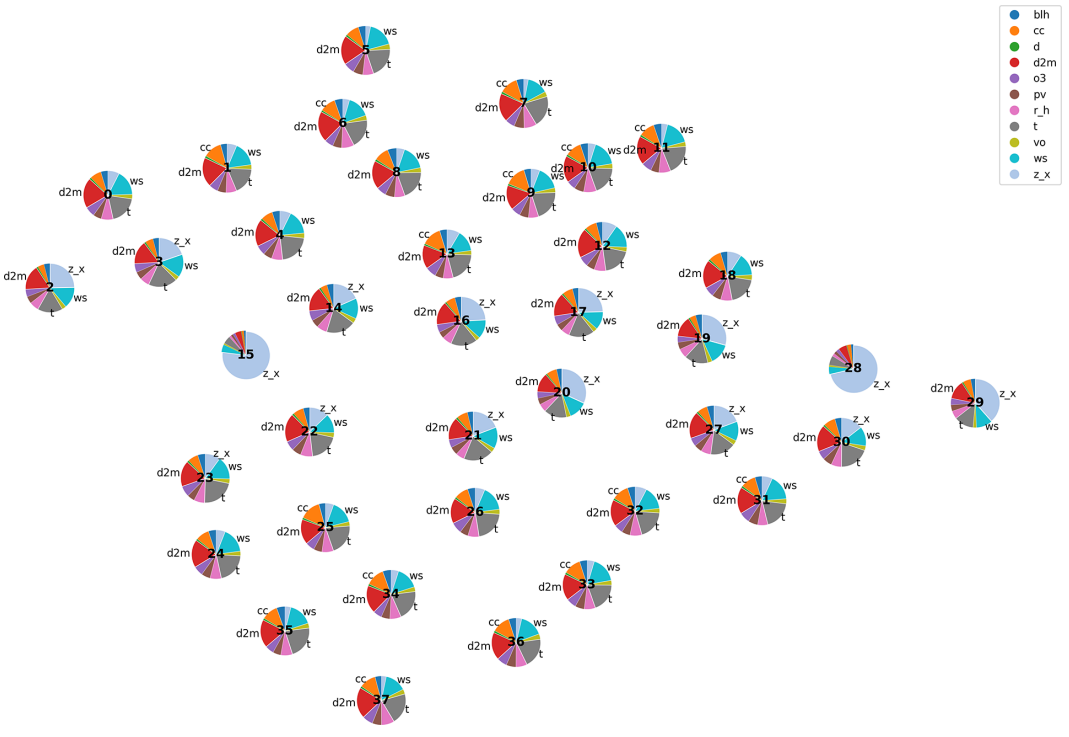


Figure 6. *β_i-proportionality of all variables in a pie diagram for each turbine selected by LiNGAM for the summer high-wind scenario (averaged over 100 time windows). Variables with up to 5 the most significant percentage values are denoted at each turbine. The concrete values for the rest of the variables are in Table 5.*

For the lag parameter, where d denotes the lag chosen by AIC—we test these variations: $[d-1, d, d+1]$. For the time-series length, we use the following numbers of samples n per window: [94,96,98]. We conducted experiments with HMML and LiNGAM again for each combination of these parameters and compared the results to the baseline approach (with the lag selected by AIC and 96 samples time-series length).

We investigated the stability of the achieved β_i-proportionality variables with respect to the variable d and n from two perspectives: (1) the stability of both methods in terms of the percentage of deviation of the obtained causal values and (2) the stability o with respect to the ordering of the causal variables contributing most to the wind speed.

4.4.1. Percentage of deviation

From the first test, we can conclude that the percentage deviations of the β_i-proportionality values—relative to their mean— are higher for HMML than for LiNGAM across all six weather scenarios. Detailed results for each scenario are provided in the Appendix B.1.

4.4.2. Ordering of causal variables according to magnitude

In the second test, we examined the stability of the ordering of the β_i-proportionality values—based on their magnitude—in relation to wind speed at 135 meters.

For HMML, in the summer high scenario, the ordering of the strongest three variables is constant with respect to these parameters (see Table 4). For LiNGAM, the ordering is stable with respect to the window-length n but not for the lag d (see Table 5).

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

Table 3. *Average β_i-proportionality over all 38 turbines per scenario for methods HMML and LiNGAM. The three strongest causal variables are indicated bold, out of which the largest value is blue*

	blh (m)	cc (%)	d (s^{-1})	d2m (K)	o3 ($kgkg^{-1}$)	$pv(kg^{-1})$	relative_ humidity (%)	t (K)	vo ($m^2s^{-1}kg^{-1}$)	ws (ms^{-1})	z_x (m^2s^{-2})
hmml_summer_high	0.010718	0.079535	0.004296	0.099815	0.026698	0.028379	0.038402	**0.130577**	0.009313	**0.366104**	**0.206165**
hmml_summer_moderate	0.007511	**0.100917**	0.005569	0.058434	0.035654	0.014051	0.010381	**0.081753**	0.011433	**0.626433**	0.047865
hmml_summer_low	0.002231	**0.138644**	0.005513	0.041688	0.058175	0.014398	0.006758	0.077774	0.016632	**0.550005**	**0.088181**
hmml_winter_high	0.030218	0.029386	0.004598	0.111286	0.073426	0.015348	0.033999	**0.148024**	0.011094	**0.429314**	**0.113308**
hmml_winter_moderate	0.019419	0.029173	0.002366	**0.072296**	0.053059	0.007739	0.020270	**0.091549**	0.009450	**0.653804**	0.040875
hmml_winter_low	0.033402	0.026969	0.003894	0.060693	**0.093625**	0.007804	0.020339	**0.089240**	0.011791	**0.567959**	0.084284
LiNGAM_summer_high	0.044372	0.087343	0.010182	**0.169025**	0.058528	0.053601	0.067977	**0.186625**	0.028091	0.145060	**0.149195**
LiNGAM_summer_moderate	0.029283	0.145048	0.011271	0.129643	0.093561	0.037885	0.043297	**0.181143**	0.031034	**0.141622**	**0.156212**
LiNGAM_summer_low	0.017905	**0.146604**	0.013205	0.093012	0.126015	0.038562	0.032256	**0.195150**	0.038639	0.137788	**0.160862**
LiNGAM_winter_high	0.065327	0.052218	0.010815	**0.167298**	0.111132	0.037054	0.064306	**0.224315**	0.026572	0.118339	**0.122624**
LiNGAM_winter_moderate	0.069029	0.060717	0.012725	**0.138700**	0.109793	0.031899	0.057071	**0.203016**	0.031271	**0.158153**	0.127627
LiNGAM_winter_low	0.059501	0.051235	0.016041	0.104872	0.140458	0.031327	0.045804	**0.180356**	0.044170	**0.179412**	**0.146822**

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

Table 4. *Ordering and β_i-proportionality values of variables by* ***HMML****, with respect to the chosen parameters in the* ***summer*** *half year and* ***high wind*** *scenario. The window-size is represented by wsize, dmod indicates the modifier added to the lag d chosen by AIC, and var/val represent the variables/values respectively*

	1	2	3	4	5	6	7	8	9	10	11
wsize=94,dmod=−1 var	ws	z_x	t	d2m	cc	pv	relative_humidity	o3	Vo	blh	d
wsize=94,dmod=−1 val	0.489	0.155	0.1	0.077	0.067	0.035	0.022	0.021	0.017	0.01	0.007
wsize=94,dmod=0 var	ws	z_x	t	d2m	cc	relative_humidity	pv	o3	Blh	vo	d
wsize=94,dmod=0 val	0.351	0.214	0.134	0.102	0.073	0.04	0.03	0.027	0.012	0.012	0.005
wsize=94,dmod=1 var	ws	z_x	t	d2m	cc	relative_humidity	o3	pv	blh	vo	d
wsize=94,dmod=1 val	0.247	0.21	0.178	0.123	0.09	0.046	0.045	0.033	0.014	0.01	0.003
wsize=96,dmod=−1 var	ws	z_x	t	d2m	cc	pv	o3	relative_humidity	vo	blh	d
wsize=96,dmod=−1 val	0.495	0.149	0.101	0.071	0.068	0.036	0.023	0.021	0.017	0.01	0.007
wsize=96,dmod=0 var	ws	z_x	t	d2m	cc	relative_humidity	pv	o3	blh	vo	d
wsize=96,dmod=0 val	0.366	0.206	0.131	0.1	0.08	0.038	0.028	0.027	0.011	0.009	0.004
wsize=96,dmod=1 var	ws	z_x	t	d2m	cc	relative_humidity	o3	pv	blh	vo	d
wsize=96,dmod=1 val	0.254	0.208	0.181	0.12	0.092	0.045	0.041	0.032	0.014	0.01	0.004
wsize=98,dmod=−1 var	ws	z_x	t	cc	d2m	pv	o3	relative_humidity	vo	d	blh
wsize=98,dmod=−1 val	0.513	0.142	0.098	0.07	0.068	0.038	0.023	0.018	0.015	0.008	0.007
wsize=98,dmod=0 var	ws	z_x	t	d2m	cc	relative_humidity	pv	o3	blh	vo	d
wsize=98,dmod=0 val	0.372	0.206	0.134	0.094	0.08	0.035	0.029	0.027	0.01	0.009	0.005
wsize=98,dmod=1 var	ws	z_x	t	d2m	cc	relative_humidity	o3	pv	blh	vo	d
wsize=98,dmod=1 val	0.269	0.204	0.18	0.111	0.094	0.043	0.039	0.035	0.012	0.01	0.004

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

Table 5. *Ordering and β_i-proportionality values of variables by* ***LiNGAM****, with respect to the chosen parameters in the* ***summer*** *half year and* ***high wind*** *scenario*

	1	2	3	4	5	6	7	8	9	10	11
wsize=94,dmod=−1 var	t	ws	d2m	z_x	cc	relative_humidity	blh	o3	pv	vo	d
wsize=94,dmod=−1 val	0.172	0.154	0.14	0.127	0.112	0.069	0.066	0.058	0.052	0.037	0.013
wsize=94,dmod=0 var	t	d2m	z_x	ws	cc	relative_humidity	o3	pv	blh	vo	d
wsize=94,dmod=0 val	0.185	0.168	0.151	0.143	0.089	0.068	0.058	0.055	0.045	0.028	0.01
wsize=94,dmod=1 var	t	d2m	z_x	ws	cc	relative_humidity	o3	pv	blh	vo	d
wsize=94,dmod=1 val	0.226	0.174	0.155	0.088	0.081	0.072	0.068	0.053	0.044	0.029	0.011
wsize=96,dmod=−1 var	t	ws	d2m	z_x	cc	relative_humidity	blh	o3	pv	vo	d
wsize=96,dmod=−1 val	0.165	0.154	0.153	0.126	0.103	0.074	0.071	0.055	0.051	0.034	0.012
wsize=96,dmod=0 var	t	d2m	z_x	ws	cc	relative_humidity	o3	pv	blh	vo	d
wsize=96,dmod=0 val	0.187	0.169	0.149	0.145	0.087	0.068	0.059	0.054	0.044	0.028	0.01
wsize=96,dmod=1 var	t	d2m	z_x	ws	cc	relative_humidity	o3	pv	blh	vo	d
wsize=96,dmod=1 val	0.228	0.173	0.153	0.089	0.082	0.073	0.068	0.053	0.042	0.028	0.011
wsize=98,dmod=−1 var	t	ws	d2m	z_x	cc	relative_humidity	blh	o3	pv	vo	d
wsize=98,dmod=−1 val	0.169	0.155	0.154	0.121	0.101	0.073	0.069	0.058	0.052	0.032	0.014
wsize=98,dmod=0 var	t	d2m	z_x	ws	cc	relative_humidity	o3	pv	blh	vo	d
wsize=98,dmod=0 val	0.185	0.168	0.15	0.15	0.085	0.067	0.06	0.053	0.043	0.028	0.01
wsize=98,dmod=1 var	t	d2m	z_x	ws	cc	relative_humidity	o3	pv	blh	vo	d
wsize=98,dmod=1 val	0.231	0.173	0.147	0.092	0.083	0.073	0.067	0.052	0.042	0.029	0.011

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

Table 6. *Ordering and β_i-proportionality values of variables by* ***HMML****, with respect to the chosen parameters in the* ***summer*** *half year and* ***low wind*** *scenario*

	1	2	3	4	5	6	7	8	9	10	11
wsize=94,dmod=–1 var	ws	cc	z_x	t	o3	d2m	pv	vo	relative_humidity	d	blh
wsize=94,dmod=–1 val	0.664	0.082	0.065	0.057	0.042	0.036	0.02	0.016	0.009	0.006	0.004
wsize=94,dmod=0 var	ws	cc	z_x	t	o3	d2m	vo	pv	relative_humidity	d	blh
wsize=94,dmod=0 val	0.543	0.129	0.091	0.081	0.064	0.044	0.018	0.016	0.007	0.005	0.003
wsize=94,dmod=1 var	ws	cc	z_x	t	o3	d2m	pv	vo	relative_humidity	d	blh
wsize=94,dmod=1 val	0.49	0.158	0.1	0.093	0.067	0.045	0.016	0.016	0.008	0.003	0.002
wsize=96,dmod=–1 var	ws	cc	z_x	t	o3	d2m	pv	vo	relative_humidity	d	blh
wsize=96,dmod=–1 val	0.66	0.087	0.066	0.061	0.039	0.036	0.018	0.014	0.009	0.005	0.004
wsize=96,dmod=0 var	ws	cc	z_x	t	o3	d2m	vo	pv	relative_humidity	d	blh
wsize=96,dmod=0 val	0.55	0.139	0.088	0.078	0.058	0.042	0.017	0.014	0.007	0.006	0.002
wsize=96,dmod=1 var	ws	cc	z_x	t	o3	d2m	pv	vo	relative_humidity	d	blh
wsize=96,dmod=1 val	0.481	0.171	0.107	0.089	0.069	0.039	0.016	0.015	0.008	0.003	0.002
wsize=98,dmod=–1 var	ws	cc	z_x	t	o3	d2m	pv	vo	relative_humidity	d	blh
wsize=98,dmod=–1 val	0.664	0.079	0.069	0.061	0.041	0.036	0.017	0.014	0.009	0.006	0.004
wsize=98,dmod=0 var	ws	cc	z_x	t	o3	d2m	vo	pv	relative_humidity	d	blh
wsize=98,dmod=0 val	0.554	0.13	0.09	0.079	0.059	0.041	0.018	0.016	0.006	0.005	0.002
wsize=98,dmod=1 var	ws	cc	z_x	t	o3	d2m	pv	vo	relative_humidity	d	blh
wsize=98,dmod=1 val	0.495	0.173	0.107	0.083	0.064	0.037	0.015	0.014	0.007	0.003	0.002

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

Table 7. *Ordering and β_i-proportionality values of variables by* ***LiNGAM****, with respect to the chosen parameters in the* ***summer*** *half-year and* ***low wind*** *scenario*

	1	2	3	4	5	6	7	8	9	10	11
wsize=94,dmod=–1 var	ws	z_x	t	cc	o3	d2m	pv	vo	relative_humidity	blh	d
wsize=94,dmod=–1 val	0.172	0.168	0.159	0.15	0.111	0.09	0.042	0.037	0.035	0.022	0.014
wsize=94,dmod=0 var	t	z_x	cc	ws	o3	d2m	vo	pv	relative_humidity	blh	d
wsize=94,dmod=0 val	0.197	0.165	0.143	0.137	0.125	0.093	0.039	0.038	0.032	0.018	0.013
wsize=94,dmod=1 var	t	z_x	o3	cc	d2m	ws	pv	vo	relative_humidity	blh	d
wsize=94,dmod=1 val	0.202	0.195	0.127	0.124	0.1	0.097	0.043	0.04	0.039	0.018	0.014
wsize=96,dmod=–1 var	ws	z_x	t	cc	o3	d2m	pv	vo	relative_humidity	blh	d
wsize=96,dmod=–1 val	0.172	0.167	0.157	0.152	0.108	0.093	0.04	0.037	0.036	0.023	0.014
wsize=96,dmod=0 var	t	z_x	cc	ws	o3	d2m	vo	pv	relative_humidity	blh	d
wsize=96,dmod=0 val	0.195	0.161	0.147	0.138	0.126	0.093	0.039	0.039	0.032	0.018	0.013
wsize=96,dmod=1 var	t	z_x	o3	cc	d2m	ws	pv	vo	relative_humidity	blh	d
wsize=96,dmod=1 val	0.2	0.198	0.126	0.125	0.1	0.098	0.042	0.04	0.039	0.018	0.014
wsize=98,dmod=–1 var	ws	z_x	cc	t	o3	d2m	pv	vo	relative_humidity	blh	d
wsize=98,dmod=–1 val	0.176	0.168	0.156	0.154	0.106	0.092	0.04	0.036	0.034	0.024	0.013
wsize=98,dmod=0 var	t	z_x	cc	ws	o3	d2m	vo	pv	relative_humidity	blh	d
wsize=98,dmod=0 val	0.195	0.161	0.147	0.14	0.125	0.093	0.038	0.038	0.032	0.017	0.013
wsize=98,dmod=1 var	t	z_x	cc	o3	ws	d2m	pv	vo	relative_humidity	blh	d
wsize=98,dmod=1 val	0.2	0.198	0.127	0.125	0.1	0.099	0.041	0.04	0.039	0.018	0.014

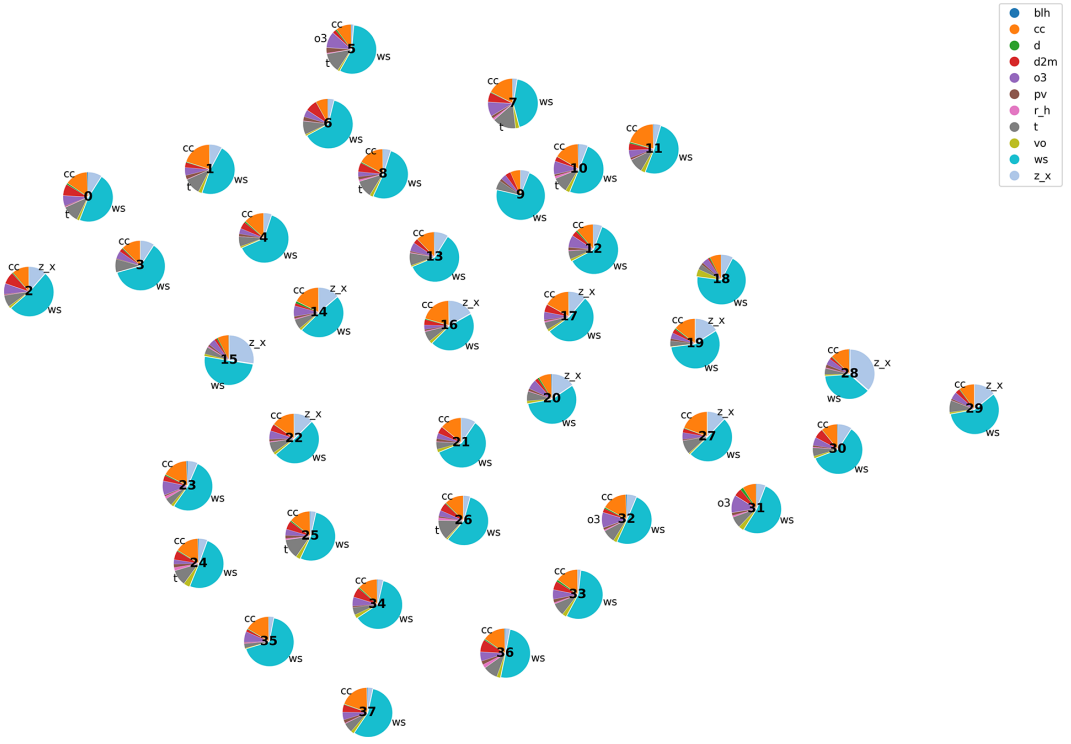


Figure 7. *Causal parameters by HMML in the summer low wind scenario.*

In the winter high scenario, neither method is stable for the three strongest causal variables (see Tables B8 and B12 in Appendix B.2). In the summer low scenario, HMML is stable for the six strongest causal variables, but LiNGAM does not exhibit stability with respect to any parameter, see Tables 6 and 7 and Figures 6 and 7. In the winter low scenario, no method is stable (see Tables B9 and B13 in Appendix B.2). In the moderate wind speed scenario, HMML shows stable results in both half years, whereas LiNGAM's three strongest causal variables and their orderings change with respect to both parameters d and n (see Tables B11 and B14 in Appendix B.2). Based on all these experiments, we conclude that the ordering of causal variables is more stable in both summer extreme scenarios when using HMML, compared to LiNGAM, see Figures 7 and 8. The results of both types of robustness tests suggest that HMML is more stable than LiNGAM in terms of the obtained causal values across all six weather scenarios. Furthermore, the ordering (ranking) of causal variables contributing most to wind speed is more stable in summer scenarios for HMML than for LiNGAM.

5. Conclusions

The objective of this work was to identify the most causally influential meteorological parameters affecting wind speed at 135 meters above ground level (approximately the hub height of the wind turbines) in the Andau wind park in Eastern Austria, across various seasonal scenarios, and to compare the resulting parameter sets between seasons. Using the HMML method, which is based on Granger causality, we gained insights into the causal variables influencing wind speed. Our findings indicate that the set of causal parameters for wind speed remains stable in at least two summer extreme scenarios. Concretely, in the summer high-wind scenario, HMML consistently identifies the following three parameters as having the highest causal strength: wind speed windspeed (wspeed135m/ws), geopotential (z_x) and temperature (t), with a stable order of influence. In the summer low-wind scenario, the order of causal strength is stable for the six most influential variables: windspeed, cloud-coverage (cc), geopotential, temperature, ozone mass mixing ratio (o3) and dewpoint temperature measured at 2 meters height (d2m).

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

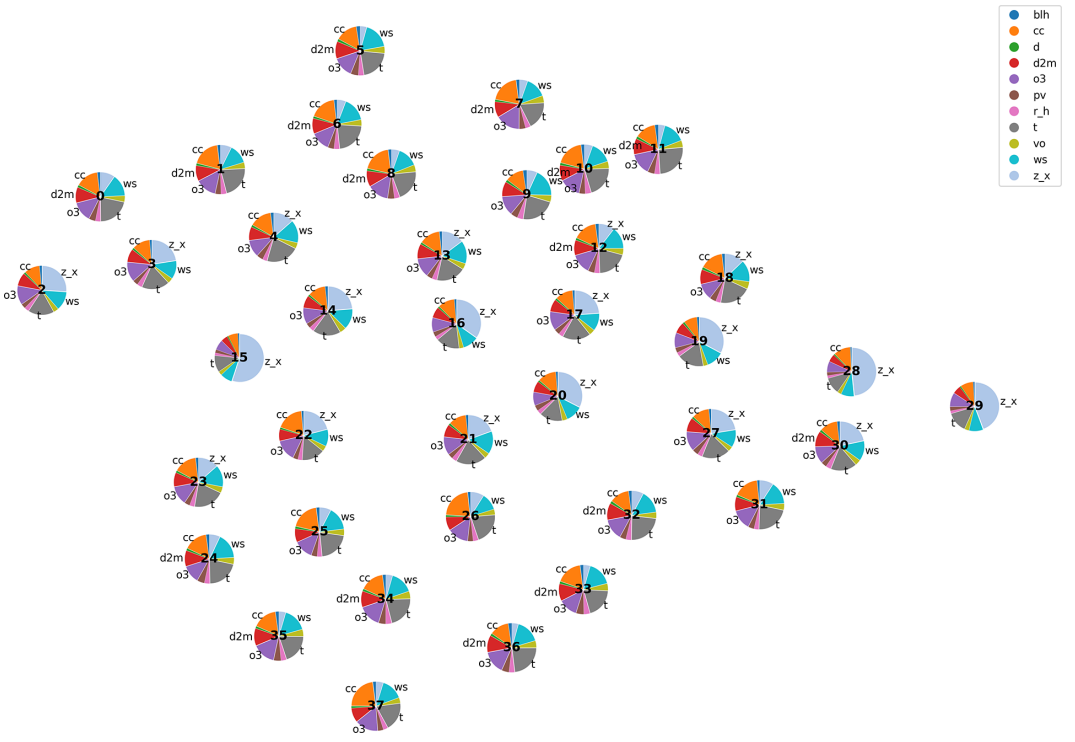


Figure 8. *Causal parameters by LiNGAM in the summer low wind scenario.*

In both high-wind scenarios (summer and winter), the three parameters with the highest causal influence on wind speed - aside from wind speed itself - are geopotential, temperature, and dew point temperature at 2 meters. Wind speed consistently exhibits the strongest causal influence by a significant margin, while the other three variables typically show similar levels of influence. In contrast, for low-wind scenarios during the summer half-year, cloud coverage emerges as the most influential variable, whereas in winter, the ozone mass mixing ratio shows the highest causal influence on wind speed.

The order of causal influence among parameters identified by HMML varies between summer and winter, as well as across different wind speed regimes. For instance, the ozone mass mixing ratio shows a stronger influence in summer than in winter. This can be attributed to ozone-related processes having a greater impact on wind turbine performance during summer, due to their role in radiative forcing and boundary layer mixing. In contrast, humidity and dew point temperature are more relevant for phenomena such as cloud formation, fog, or icing - factors that are more critical in winter or colder climates.

Among the eleven variables considered, the following were not found to have a significant causal influence on wind speed: relative humidity, boundary layer height (blh), potential vorticity (pv), relative vorticity (vo), and divergence (d).

The practical application of our results lies in combining the scenario-specific causal insights provided by HMML with weather forecasts. This integration can offer valuable guidance for wind farm operators. Specifically, understanding which meteorological variables most strongly influence wind speed at hub height allows for more informed planning of turbine operation—helping to determine when turbines should or should not be producing energy.

While the results are specific to the Pannonian Basin's flat terrain and the continental climate, the applicability to more complex terrain (mountains, coastlines) or other flat regions requires site-specific analysis due to different boundary layer physics. However, this does not imply that the HMML methodology is not transferable. On the contrary, adaptations to specific locations (size independent) are always needed, as local conditions like land-use (agriculture, forest) in the vicinity of a wind farm

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

always influences the site. Future work should validate in: (1) coastal wind farms with sea-breeze dynamics, (2) mountainous regions with complex flow, (3) offshore farms with different stability regimes.

In practice, the following added values can be drawn from the HMML results mainly for curtailment forecasting, maintenance scheduling, seasonal strategies and turbine-specific optimization.

Curtailment forecasting for 6–12 hours ahead involves monitoring numerical weather prediction (NWP) forecasts, specifically focusing on geopotential anomalies and hub-height temperature gradients. When the zonal geopotential anomaly (z_x) exceeds +1.5 standard deviations and the temperature shows a warming greater than $2K$, there is a 73% probability of high-wind events (greater than 12 m/s) occurring within that time frame. This indicates the need to prepare for curtailment. For maintenance scheduling, low-wind periods —which are optimal for turbine maintenance—are typically associated with increased cloud coverage and decreased ozone levels. Scheduling maintenance during times when cloud coverage exceeds 60% and ozone levels are trending downward can reduce lost production by approximately 15% compared to random scheduling. Seasonal strategies should also be considered. In winter, events show stronger causality with dew point temperature at 2 meters (d2m), with a beta value of 11.1% compared to 10.0% in summer, suggesting the influence of frontal passages. In contrast, summer events are more strongly linked to temperature and zonal geopotential anomalies (with beta values of 13.1% and 20.6%, respectively), indicating convective dynamics. Therefore, differentiated forecasting strategies are recommended for different seasons. Finally, turbine-specific optimization reveals that turbines 15 and 28 exhibit higher sensitivity to z_x (with beta values of 0.53 and 0.25, respectively, compared to the farm average of 0.21). This suggests these turbines are more exposed to synoptic-scale forcing, and thus may benefit from differentiated curtailment thresholds.

Open peer review. To view the open peer review materials for this article, please visit http://doi.org/10.1017/eds.2026.10031.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/eds.2026.10031.

Acknowledgements. This work was funded in part by the Austrian Climate and Research Programme (ACRP) project KR19AC0K17614.

Author contribution. Conceptualization: K.H-S., R.W.; Data curation: I.S.; Data visualization: R.W., K.H-S.; Formal analysis: K.H-S., R.W.; Methodology: K.H-S.; Software: R.W., K.H.; Writing - original draft: K.H-S., R.W.; Writing - review and editing: K.H-S., R.W., I.S.; Funding acquisition, resources: C.P., I.S. All authors approved the final submitted draft.

Competing interests. The authors declare no competing interests exist.

Data availability statement. The code for this project, including all experiments, is hosted on Gitlab: https://git01lab.cs.univie.ac.at/rainerwoess/hmml-windfarm. The dataset, which is a subset of ERA5, and code version as of 24.10.2024 are published on Zenodo: https://www.doi.org/10.5281/zenodo.13986329. The full ERA5 dataset is available from Copernicus Climate Data Store: https://cds.climate.copernicus.eu/datasets.

Ethics statement. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

References

Andersen AD and Geels FW (2023) Multi-system dynamics and the speed of net-zero transitions: Identifying causal processes related to technologies, actors, and institutions. *Energy Research & Social Science 102*, 103178.

Attanasio A, **Pasini A and Triacca U** (2013) Granger causality analyses for climatic attribution. *Atmospheric and Climate Sciences 3*, 515–522.

Ban N, **Schmidli J and Schär C** (2014) Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations. *Journal of Geophysical Research: Atmospheres 119*(13), 7889–7907. https://doi.org/10.1002/2014JD021478.

Behzadi S, **Hlaváčková-Schindler K and Plant C** (2019) Granger causality for heterogeneous processes. *Advances in Knowledge Discovery and Data Mining: PAKDD 2019, Proceedings, Part III 23*, 463–475.

Gazar AM, **Borsuk ME and Calder RS** (2024) Causal inference to scope environmental impact assessment of renewable energy projects and test competing mental models of decarbonization. *Environmental Research: Infrastructure and Sustainability 4*(4), 045005.

Glymour C, **Zhang K and Spirtes P** (2019) Review of causal discovery methods based on graphical models. *Frontiers in Genetics 10*, 524.

Granger CW (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society 37*, 424–438.

Granger CW (1988) Some recent development in a concept of causality. *Journal of Econometrics 39*(1–2), 199–211.

Hersbach H, **Bell B**, **Berrisford P**, **Hirahara S**, **Horányi A**, **Muñoz-Sabater J**, **Nicolas J**, **Peubey C**, **Radu R**, **Schepers D**, **Simmons A**, **Soci C**, **Abdalla S**, **Abellan X**, **Balsamo G**, **Bechtold P**, **Biavati G**, **Bidlot J**, **Bonavita M**, **De Chiara G**, **Dahlgren P**, **Dee D**, **Diamantakis M**, **Dragani R**, **Flemming J**, **Forbes R**, **Fuentes M**, **Geer A**, **Haimberger L**, **Healy S**, **Hogan RJ**, **Holm E**, **Janiskova M**, **Keeley S**, **Laloyaux P**, **Lopez P**, **Lupu C**, **Radnoti G**, **de Rosnay P**, **Rozum I**, **Vamborg F**, **Villaume S and Thepaut JN** (2020) The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society 146*(730), 1999–2049. https://doi.org/10.1002/qj.3803.

Hlaváčková-Schindler K, **Fuchs A**, **Plant C**, **Schicker I and DeWit R** (2022) The influence of meteorological parameters on wind speed extreme events: A causal inference approach. *EGU General Assembly Conference Abstracts*, EGU22–5756.

Hlaváčková-Schindler K and Plant C (2020) Heterogeneous graphical Granger causality by minimum message length. *Entropy 22*(12), 1400.

Hurvich CM and Tsai C-L (1989) Regression and time series model selection in small samples. *Biometrika 76*(2), 297–307.

Kim S, **Putrino D**, **Ghosh S and Brown EN** (2011) A granger causality measure for point process models of ensemble neural spiking activity. *PLoS Computational Biology 7*(3), e1001110.

Köhler J, **Geels FW**, **Kern F**, **Markard J**, **Onsongo E**, **Wieczorek A**, **Alkemade F**, **Avelino F**, **Bergek A**, **Boons F**, et al. (2019) An agenda for sustainability transitions research: State of the art and future directions. *Environmental Innovation and Societal Transitions 31*, 1–32.

Mannino M and Bressler SL (2015) Foundational perspectives on causality in large-scale brain networks. *Physics of Life Reviews 15*, 107–123.

Manwell JF, **McGowan JG and Rogers AL** (2010) *Wind Energy Explained: Theory, Design and Application*. Chichester: John Wiley & Sons.

Maziarz M (2015) A review of the Granger-causality fallacy. *The Journal of Philosophical Economics: Reflections on Economic and Social Issues 8*(2), 86–105.

Nelder JA and Wedderburn RW (1972) Generalized linear models. *Journal of the Royal Statistical Society: Series A (General) 135*(3), 370–384.

Papagiannopoulou C, **Miralles DG**, **Decubber S**, **Demuzere M**, **Verhoest NEC**, **Dorigo WA and Waegeman W** (2017) A non-linear Granger-causality framework to investigate climate–vegetation dynamics. *Geoscientific Model Development 10*(5), 1945–1960. https://doi.org/10.5194/gmd-10-1945-2017.

Pearl J (2009) *Causality: Models, Reasoning and Inference*, 2nd Edn. Cambridge: Cambridge University Press.

Pryor SC and Barthelmie RJ (2021) A global assessment of extreme wind speeds for wind energy applications. *Nature Energy 6*(3), 268–276.

Runge J (2020) Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In *Conference on Uncertainty in Artificial Intelligence*. Runge: Proceedings of Machine Learning Research. Vol *124*, pp. 1388–1397.

Runge J, **Gerhardus A**, **Varando G**, **Eyring V and Camps-Valls G** (2023) Causal inference for time series. *Nature Reviews Earth & Environment 4*, 487–505.

Runge J, **Nowack P**, **Kretschmer M**, **Flaxman S and Sejdinovic D** (2019) Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances 5*(11), eaau4996.

Runge J, **Petoukhov V**, **Donges JF**, **Hlinka J**, **Jajcay N**, **Vejmelka M**, **Hartman D**, **Marwan N**, **Paluš M and Kurths J** (2015) Identifying causal gateways and mediators in complex spatio-temporal systems. *Nature Communications 6*(1), 8502.

Shimizu S, **Hoyer PO**, **Hyvärinen A**, **Kerminen A and Jordan M** (2006) A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research 7*(10), 2003–2030.

Silva F, **Vega-Oliveros D**, **Yan X**, **Flammini A**, **Menczer F**, **Radicchi F**, et al. (2021) Detecting climate teleconnections with Granger causality. *Geophysical Research Letters 48*, e2021GL094707.

Singh S and Hawati J. (2019) Study on the causality nexus between macroeconomic variables using vector error correction modeling. *International Economics/Economia Internazionale 72*.1.

Smirnov DA and Mokhov II (2009) From Granger causality to long-term causality: Application to climatic data. *Physical Review E 80*(1), 016208.

Sovacool BK, **Turnheim B**, **Hook A**, **Brock A and Martiskainen M** (2021) Dispossessed by decarbonisation: Reducing vulnerability, injustice, and inequality in the lived experience of low-carbon pathways. *World Development 137*, 105116.

Tibshirani R (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology 58*(1), 267–288.

Wallace CS (2005) *Statistical and Inductive Inference by Minimum Message Length*. New York, NY: Springer Science & Business Media.

Zhu JY, **Sun C and Li VO** (2015) *Granger-causality-baesd air quality estimation with spatiotemporal (s-t) heterogeneous big data. In The First International Workshop on Smart Cities and Urban Informatics 2015.*

Cite this article: Hlaváčková-Schindler K, Wöss R, Schicker I and Plant C (2026). Which meteorological parameters influence extreme wind speed in a wind farm? A heterogeneous Granger causality approach. *Environmental Data Science*, 5: e3. doi:10.1017/eds.2026.10031

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

A. Appendix

A.1. β_i -Proportionality values for summer extreme scenario for all turbines and both methods

In the moderate wind scenario during the summer half-year, HMML attributes most of the causal strength to wind speed, with additional influence assigned to cloud coverage, temperature, and geopotential. Compared to the summer high-wind scenario, dew point temperature (d2m) and temperature (t) have a lower influence. For LiNGAM, the contribution of wind speed is notably lower, and a broader range of variables is selected. As observed in the high-wind scenario, both methods show a significantly increased influence of geopotential (z_x) for turbines 15 and 28 compared to the others. This pattern is consistently present across all considered scenarios.

Table A1. *β_i-proportionality values of all 38 turbines by HMML in the summer high-wind scenario as displayed in Figure 5*

turb_index	z_x (m^2s^{-2})	blh (m)	d2m (K)	rel_humidity (%)	ws (ms^{-1})	d (s^{-1})	cc (%)	o3 ($kgkg^{-1}$)	pv (kg^{-1})	t (K)	vo ($m^2s^{-1}kg^{-1}$)
1	**0.223825**	0.017277	0.115204	0.048840	**0.321946**	0.004020	0.069998	0.025440	0.020880	**0.143093**	0.009478
2	**0.143758**	0.006751	0.077016	0.036892	**0.451875**	0.004723	0.106889	0.022441	0.022428	**0.113444**	0.013782
3	**0.271543**	0.005897	0.098679	0.041237	**0.308723**	0.006128	0.048323	0.027130	0.017205	**0.169991**	0.005144
4	**0.264667**	0.009320	0.111965	0.031956	**0.328830**	0.006386	0.067458	0.023190	0.028223	**0.119786**	0.008218
5	**0.140626**	0.006297	0.102040	0.045324	**0.412718**	0.002421	0.053606	0.025241	0.037002	**0.164919**	0.009805
6	**0.121895**	0.014248	0.108868	0.028959	**0.460668**	0.002511	0.075672	0.027867	0.029583	**0.118593**	0.011138
7	**0.145133**	0.010503	0.133008	0.048532	**0.314511**	0.004821	0.093660	0.041543	0.028646	**0.166202**	0.013442
8	0.059375	0.012639	0.114830	0.049770	**0.397038**	0.006904	**0.121064**	0.040137	0.032122	**0.155693**	0.010429
9	**0.144827**	0.008809	0.104118	0.028994	**0.470313**	0.007400	0.050974	0.029597	0.024477	**0.121073**	0.009418
10	**0.142088**	0.011363	0.087033	0.026433	**0.359706**	0.000961	0.133646	0.043539	0.033354	**0.147380**	0.014498
11	0.095238	0.014972	**0.127418**	0.050008	**0.381545**	0.004116	0.072659	0.045707	0.037622	**0.159171**	0.011545
12	**0.191915**	0.021191	0.110000	0.062744	**0.317236**	0.002243	0.087952	0.030140	0.031217	**0.137725**	0.007638
13	**0.188941**	0.012150	0.133346	0.036483	**0.333549**	0.006596	0.074729	0.038061	0.026504	**0.139263**	0.010376
14	**0.176408**	0.017640	0.081005	0.030964	**0.346262**	0.004845	**0.152822**	0.023462	0.023888	0.135597	0.007106
15	**0.305915**	0.010450	0.100515	0.022766	**0.316114**	0.003974	0.035499	0.023456	0.023338	**0.148128**	0.009846
16	**0.533113**	0.009476	0.031411	0.031325	**0.254766**	0.000351	0.057358	0.007183	0.010791	**0.058846**	0.005380
17	**0.269279**	0.012367	**0.118811**	0.035792	**0.313035**	0.001714	0.072396	0.026863	0.034229	0.106959	0.008556
18	**0.181565**	0.012140	0.094821	0.033662	**0.377989**	0.000725	**0.127340**	0.023326	0.031321	0.110571	0.006541
19	**0.233350**	0.009081	0.107661	0.036233	**0.304126**	0.006108	0.071336	0.023752	0.045925	**0.146154**	0.016274
20	**0.290031**	0.009452	**0.110095**	0.023941	**0.370046**	0.006241	0.047821	0.014462	0.018264	0.102587	0.007059
21	**0.242779**	0.009361	0.103136	0.039927	**0.349998**	0.000700	0.068752	0.014911	0.020301	**0.145180**	0.004955
22	**0.255143**	0.016087	0.075818	0.028335	**0.393566**	0.004368	**0.098483**	0.012217	0.022416	0.086780	0.006788

Continued

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

turb_index	z_x (m^2s^{-2})	blh (m)	d2m (K)	rel_humidity (%)	ws (ms^{-1})	d (s^{-1})	cc (%)	o3 ($kgkg^{-1}$)	pv (kg^{-1})	t (K)	vo ($m^2s^{-1}kg^{-1}$)
23	**0.186000**	0.009267	0.103775	0.037385	**0.390965**	0.006854	0.049850	0.024452	0.033013	**0.145125**	0.013314
24	**0.248338**	0.008890	0.075883	0.032825	**0.372795**	0.007041	0.047889	0.018289	0.018140	**0.160935**	0.008975
25	**0.196187**	0.006757	0.091165	0.061989	**0.375081**	0.003745	0.058007	0.034132	0.023847	**0.140711**	0.008379
26	0.122103	0.009476	0.100389	0.045407	**0.376881**	0.003392	**0.131864**	0.039925	0.032493	**0.131193**	0.006878
27	**0.186189**	0.008326	0.108669	0.052983	**0.378479**	0.005254	0.064057	0.022866	0.021615	**0.140914**	0.010649
28	**0.249735**	0.009008	**0.120658**	0.044134	**0.364062**	0.003601	0.047769	0.031467	0.020703	0.101374	0.007489
29	**0.472557**	0.007449	0.092402	0.022068	**0.218173**	0.003572	0.048572	0.011588	0.017985	**0.097649**	0.007985
30	**0.289758**	0.014113	0.119014	0.032271	**0.344784**	0.003581	0.030102	0.010314	0.013500	**0.133396**	0.009167
31	**0.191313**	0.010964	0.099776	0.038294	**0.366396**	0.003001	0.067318	0.031191	0.044381	**0.138608**	0.008756
32	**0.173232**	0.009841	0.097764	0.025959	**0.401645**	0.006650	0.083130	0.051573	0.034791	**0.105849**	0.009567
33	**0.171545**	0.007882	0.096112	0.042197	**0.408886**	0.004229	0.087505	0.016514	0.024865	**0.130294**	0.009971
34	**0.145061**	0.005236	0.080387	0.042212	**0.401251**	0.004378	0.099529	0.027001	0.030618	**0.158591**	0.005736
35	**0.167228**	0.016578	0.115822	0.039421	**0.332111**	0.005754	0.105397	0.035904	0.031033	**0.140696**	0.010056
36	**0.134074**	0.007804	0.078047	0.048298	**0.437425**	0.004153	0.103704	0.016515	0.038155	**0.126574**	0.005252
37	**0.127728**	0.010278	0.069403	0.033652	**0.446294**	0.006269	**0.125299**	0.019329	0.044299	0.106763	0.010686
38	**0.151791**	0.007933	0.096889	0.041061	**0.412175**	0.003521	0.083899	0.033784	0.049221	**0.106117**	0.013608

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

Table A2. *β_i-proportionality values of all 38 turbines by LiNGAM in the summer high-wind scenario, as displayed in Figure 6*

turb_index	z_x (m^2s^{-2})	blh (m)	d2m (K)	rel_humidity (%)	ws (ms^{-1})	d (s^{-1})	cc (%)	o3 ($kgkg^{-1}$)	pv (kg^{-1})	t (K)	vo ($m^2s^{-1}kg^{-1}$)
1	0.075595	0.046144	**0.201551**	0.078772	**0.169424**	0.010557	0.076882	0.065802	0.054992	**0.192171**	0.028110
2	0.064144	0.042107	**0.196386**	0.070278	**0.163259**	0.012471	0.124745	0.060682	0.056826	**0.180558**	0.028543
3	**0.246661**	0.040304	**0.166673**	0.061456	0.142051	0.009561	0.043275	0.048802	0.047094	**0.168419**	0.025704
4	**0.197032**	0.040517	**0.156113**	0.062126	0.152875	0.009234	0.052922	0.057550	0.051391	**0.193822**	0.026419
5	0.072770	0.051156	**0.180897**	0.072821	**0.162637**	0.011035	0.079985	0.064060	0.057603	**0.215195**	0.031843
6	0.031924	0.047974	**0.194603**	0.071289	**0.175720**	0.011378	0.091166	0.071258	0.065011	**0.205485**	0.034192
7	0.044807	0.053801	**0.205724**	0.082533	**0.152202**	0.009838	0.106219	0.057838	0.059443	**0.197103**	0.030493
8	0.030248	0.047755	**0.190533**	0.083042	**0.140606**	0.012607	0.122906	0.062107	0.067173	**0.211139**	0.031884
9	0.052618	0.059708	**0.190095**	0.068682	**0.164984**	0.012210	0.098833	0.063297	0.062130	**0.197277**	0.030166
10	0.058091	0.046539	**0.164285**	0.066372	**0.156670**	0.010635	0.140434	0.061552	0.059969	**0.206030**	0.029421
11	0.049413	0.050766	**0.177780**	0.084339	**0.174537**	0.011160	0.108985	0.058069	0.064445	**0.190048**	0.030457
12	0.039443	0.052318	**0.191365**	0.077624	**0.168920**	0.011335	0.106046	0.067022	0.054580	**0.202459**	0.028888
13	0.096828	0.040512	**0.196487**	0.077318	**0.158614**	0.009579	0.078229	0.065577	0.054615	**0.194060**	0.028180
14	0.085330	0.047945	**0.157425**	0.066457	**0.142271**	0.010127	0.133381	0.071350	0.054133	**0.204451**	0.027131
15	**0.183141**	0.044726	**0.166512**	0.068836	0.136186	0.010246	0.045829	0.065019	0.051864	**0.196969**	0.030673
16	**0.769193**	0.012793	0.041521	0.013948	**0.051264**	0.002992	0.014721	0.016382	0.014705	**0.052597**	0.009883
17	**0.241857**	0.042638	**0.156599**	0.062821	0.125577	0.008757	0.063965	0.053036	0.042725	**0.177660**	0.024365
18	**0.245877**	0.039901	**0.156659**	0.055863	0.120888	0.009140	0.068943	0.058651	0.046315	**0.170924**	0.026839
19	0.092005	0.046128	**0.188978**	0.076314	**0.149470**	0.011291	0.084360	0.063985	0.059131	**0.195264**	0.033074
20	**0.292686**	0.042715	0.138118	0.061439	**0.139808**	0.009440	0.041749	0.041014	0.047179	**0.157858**	0.027994
21	**0.318720**	0.034250	**0.128283**	0.053411	0.122701	0.007920	0.070595	0.043021	0.042233	**0.152420**	0.026447
22	**0.191707**	0.037635	**0.151363**	0.057473	0.141683	0.011209	0.074042	0.053227	0.053073	**0.200308**	0.028280
23	**0.128284**	0.042358	**0.190629**	0.079842	0.126565	0.011444	0.071610	0.062864	0.060372	**0.196035**	0.029996
24	0.100037	0.049737	**0.173343**	0.070365	**0.153631**	0.010477	0.070533	0.071143	0.049376	**0.223172**	0.028185
25	0.059612	0.050496	**0.186395**	0.074037	**0.169785**	0.010137	0.090362	0.063452	0.059405	**0.207623**	0.028697
26	0.054490	0.042220	**0.167580**	0.076124	**0.156030**	0.010784	0.141365	0.059300	0.055876	**0.209435**	0.026796
27	0.063604	0.049908	**0.169024**	0.073776	**0.166704**	0.010364	0.094094	0.073428	0.054404	**0.214060**	0.030632
28	**0.193234**	0.046379	**0.180156**	0.066030	0.128636	0.010211	0.068308	0.055570	0.051611	**0.170467**	0.029397
29	**0.713706**	0.014947	**0.055869**	0.019989	0.048934	0.003660	0.026658	0.022066	0.019283	**0.063303**	0.011586
30	**0.384460**	0.030061	**0.126059**	0.052594	0.107809	0.008106	0.053366	0.043384	0.041975	**0.130489**	0.021696

Continued

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

turb_index	z_x (m^2s^{-2})	blh (m)	d2m (K)	rel_humidity (%)	ws (ms^{-1})	d (s^{-1})	cc (%)	o3 ($kgkg^{-1}$)	pv (kg^{-1})	t (K)	vo ($m^2s^{-1}kg^{-1}$)
31	**0.145032**	0.044135	**0.177959**	0.068423	0.129045	0.009894	0.079546	0.057552	0.061520	**0.198235**	0.028659
32	0.068653	0.046148	**0.177636**	0.067364	**0.169291**	0.011494	0.101184	0.071267	0.063440	**0.192462**	0.031060
33	0.075992	0.050183	**0.178506**	0.078369	**0.155802**	0.010903	0.111976	0.057789	0.056667	**0.195547**	0.028266
34	0.046655	0.049668	**0.175806**	0.073608	**0.175033**	0.011936	0.116782	0.065557	0.060795	**0.194718**	0.029441
35	0.049097	0.055148	**0.186449**	0.074380	**0.151380**	0.011271	0.124786	0.057062	0.057272	**0.203656**	0.029499
36	0.038738	0.055270	**0.187609**	0.076980	**0.160157**	0.010748	0.108693	0.057354	0.052928	**0.221249**	0.030274
37	0.036057	0.049620	**0.184548**	0.073155	**0.157861**	0.012121	0.120975	0.067453	0.064148	**0.200377**	0.033687
38	0.031677	0.041538	**0.207431**	0.084866	**0.143270**	0.010663	0.110592	0.069521	0.061122	**0.208701**	0.030619

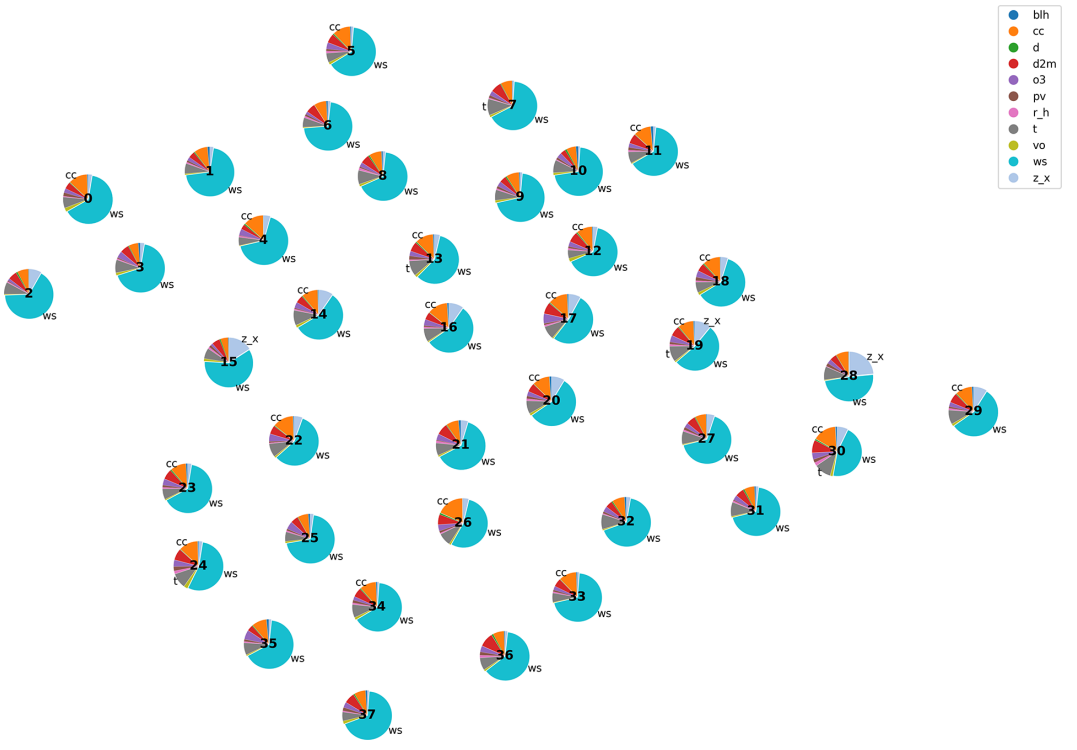


Figure A1. *Causal parameters by HMML in the summer moderate wind scenario.*

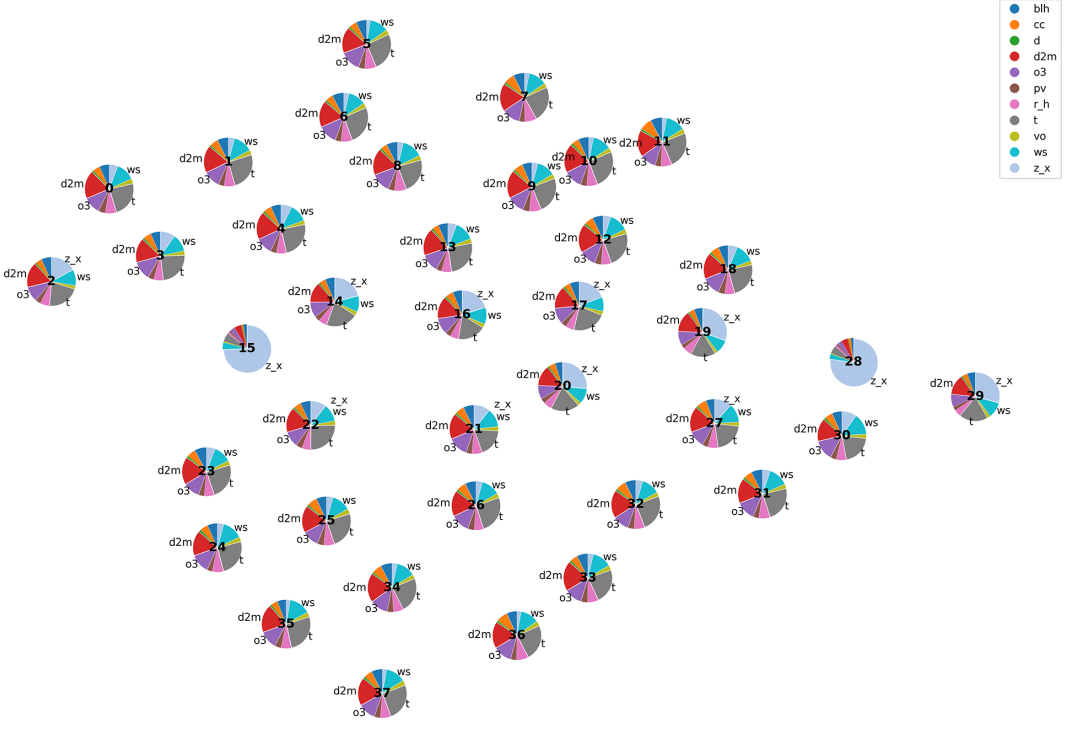


Figure A2. *Causal parameters by LiNGAM in the summer moderate wind scenario.*

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

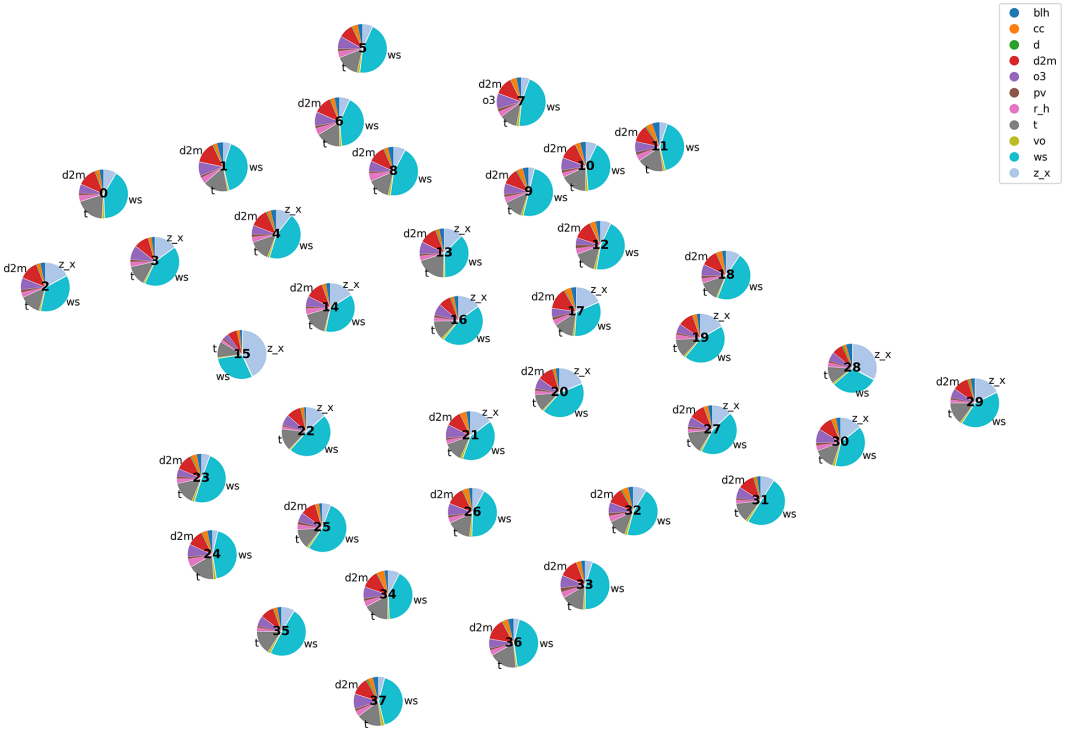


Figure A3. *Causal parameters by HMML in the winter high wind scenario.*

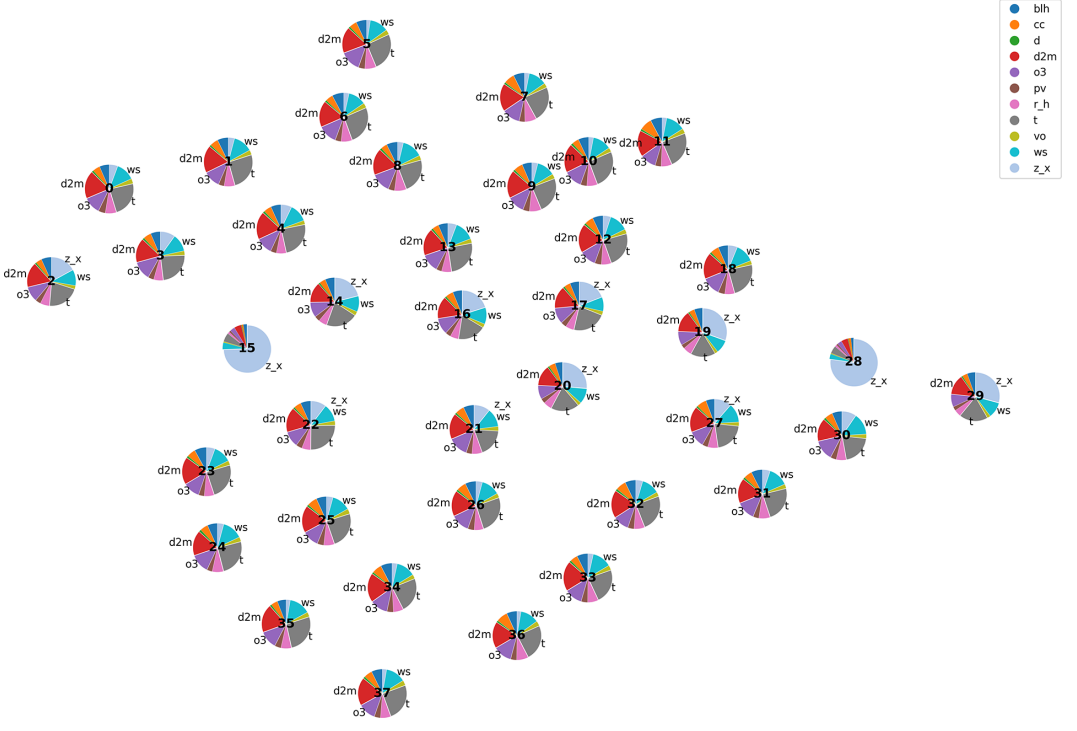


Figure A4. *Causal parameters by LiNGAM in the winter high wind scenario.*

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

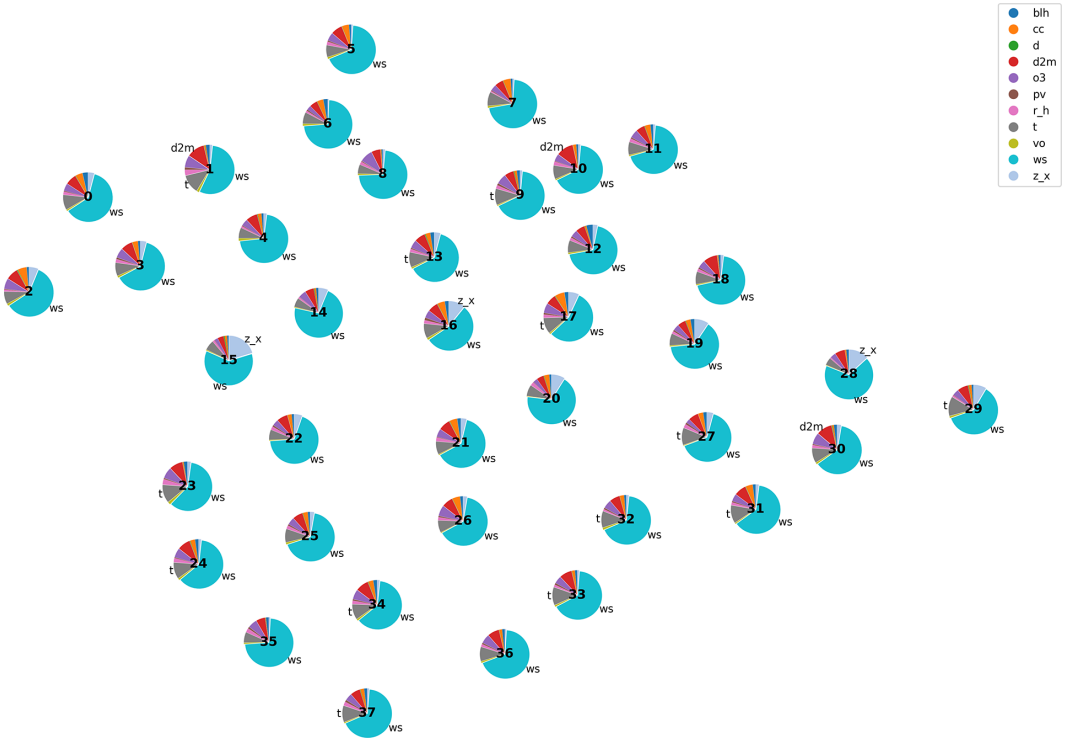


Figure A5. *Causal parameters by HMML in the winter moderate wind scenario.*

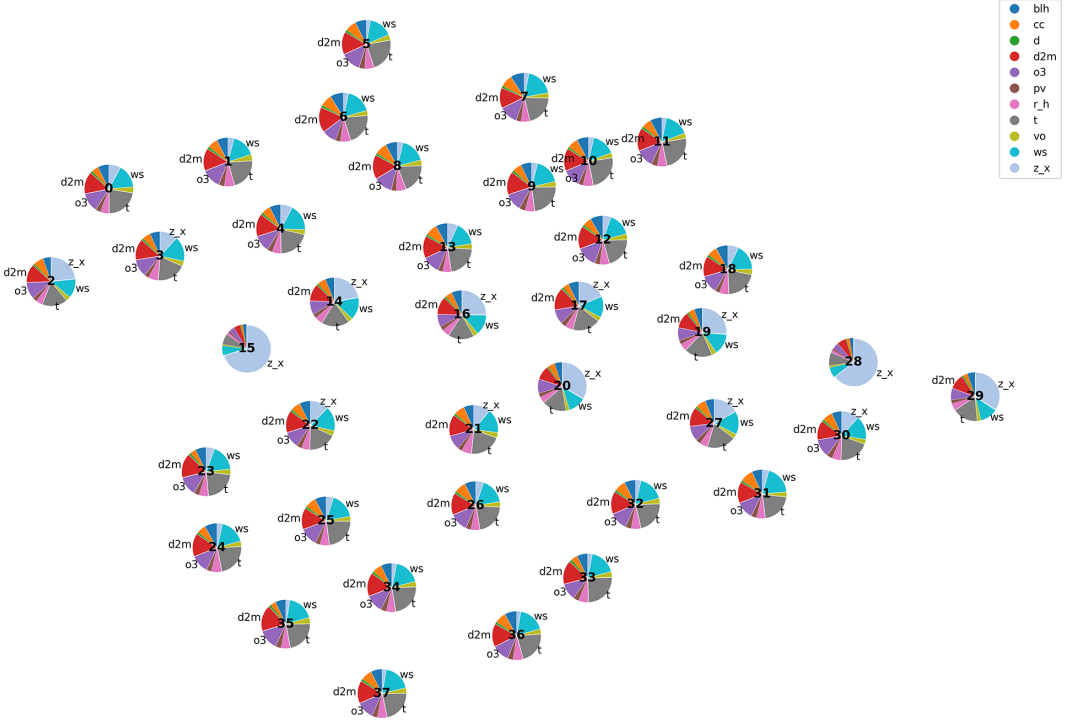


Figure A6. *Causal parameters by LiNGAM in the winter moderate wind scenario.*

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

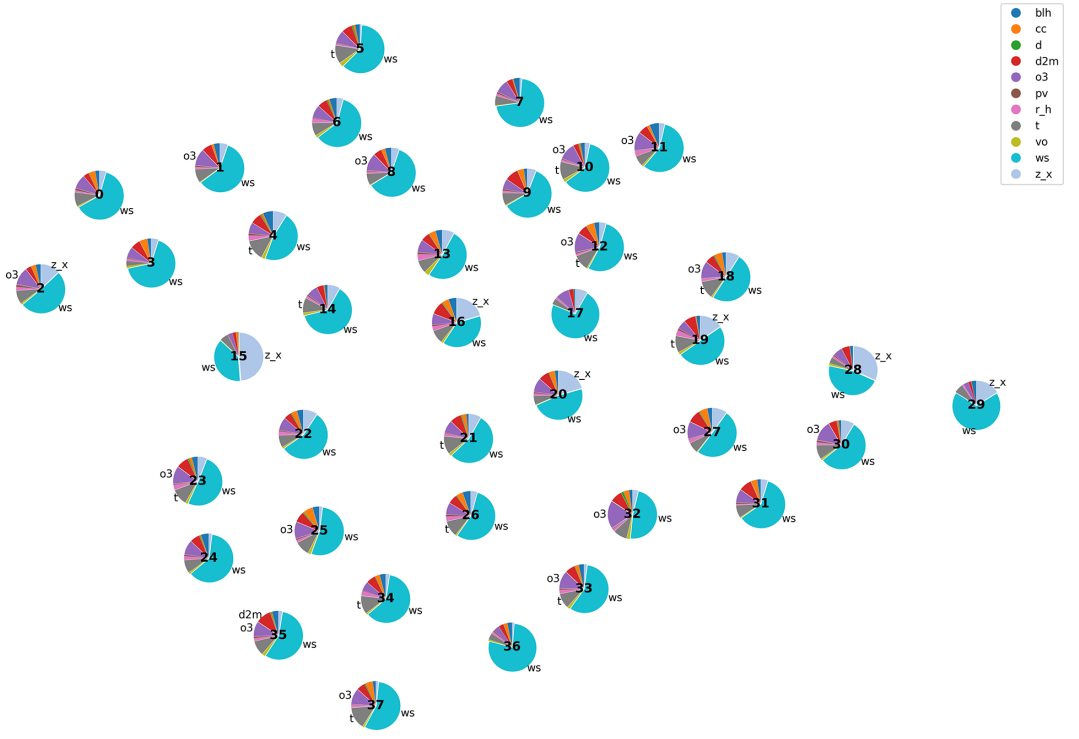


Figure A7. *Causal parameters by HMML in the winter low wind scenario.*

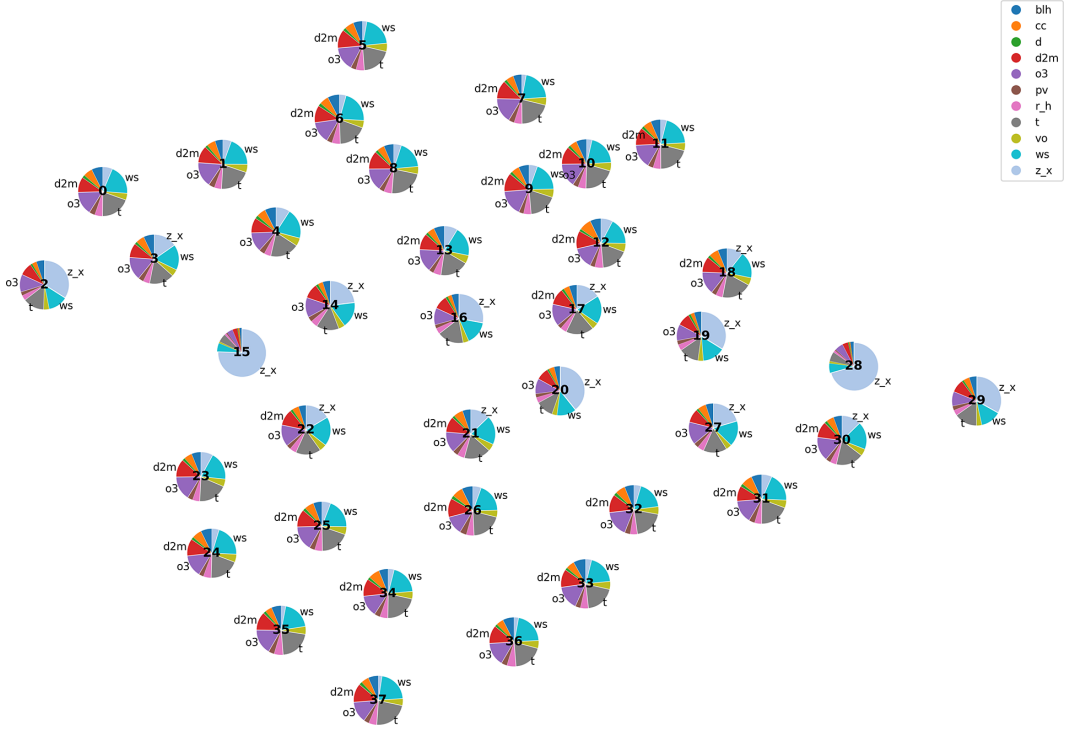


Figure A8. *Causal parameters by LiNGAM in the winter low wind scenario.*

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

B. Appendix

B.1. Stability Test 1

We investigated the deviations of the β_i-proportionality values, expressed as percentages, for HMML and LiNGAM. Tables A2–B6 present two rows per method: the first row ("mean ± std") reports the mean β-proportionality and standard deviation for each variable, while the second row ("std %") shows the standard deviation as a percentage of the mean.

Table B1. *β_i-proportionality values and their standard deviations in summer high scenarios*

	blh (m)	cc (%)	d (s^{-1})	d2m (K)	o3 ($kgkg^{-1}$)	pv (kg^{-1})	relative_humidity (%)	t (K)	vo ($m^2s^{-1}kg^{-1}$)	ws (ms^{-1})	z_x (m^2s^{-2})
HMML mean (+- std)	0.011(+– 0.002)	0.079(+– 0.011)	0.005(+– 0.002)	0.096(+– 0.02)	0.03(+– 0.009)	0.033(+– 0.003)	0.034(+– 0.011)	0.137(+– 0.035)	0.012(+– 0.003)	0.373(+– 0.106)	0.188(+– 0.03)
HMML std %	18.2	13.9	40.0	20.8	30.0	9.1	32.4	25.5	25.0	28.4	16.0
LiNGAM mean (+- std)	0.052(+–0.013)	0.092(+–0.011)	0.012(+–0.001)	0.164(+–0.012)	0.061(+–0.005)	0.053(+–0.001)	0.071(+–0.003)	0.194(+–0.027)	0.03(+–0.003)	0.13(+–0.031)	0.142(+–0.013)
LiNGAM std%	25.0	12.0	8.3	7.3	8.2	1.9	4.2	13.9	10.0	23.8	9.2

Table B2. *β_i-proportionality values and their standard deviations in summer moderate scenarios*

	blh (m)	cc (%)	d (s^{-1})	d2m (K)	o3 ($kgkg^{-1}$)	pv (kg^{-1})	relative_humidity (%)	t (K)	vo ($m^2s^{-1}kg^{-1}$)	ws (ms^{-1})	z_x (m^2s^{-2})
HMML mean (+- std)	0.008(+– 0.001)	0.116(+– 0.054)	0.007(+– 0.004)	0.062(+– 0.02)	0.042(+– 0.015)	0.016(+– 0.002)	0.013(+– 0.003)	0.085(+– 0.029)	0.012(+– 0.0)	0.585(+– 0.142)	0.055(+– 0.023)
HMML std %	12.5	46.6	57.1	32.3	35.7	12.5	23.1	34.1	0.0	24.3	41.8
LiNGAM mean (+- std)	0.028(+– 0.001)	0.144(+– 0.009)	0.012(+– 0.001)	0.128(+– 0.006)	0.089(+– 0.007)	0.038(+– 0.001)	0.045(+– 0.002)	0.17(+– 0.016)	0.031(+– 0.0)	0.14(+– 0.039)	0.175(+– 0.025)
LiNGAM std %	3.6	6.2	8.3	4.7	7.9	2.6	4.4	9.4	0.0	27.9	14.3

Table B3. *β_i-proportionality values and their standard deviations in summer low scenarios*

	blh (m)	cc (%)	d (s^{-1})	d2m (K)	o3 ($kgkg^{-1}$)	pv (kg^{-1})	relative_humidity (%)	t (K)	vo ($m^2s^{-1}kg^{-1}$)	ws (ms^{-1})	z_x (m^2s^{-2})
HMML mean (+- std)	0.003(+– 0.001)	0.127(+– 0.037)	0.005(+– 0.001)	0.04(+– 0.004)	0.056(+– 0.012)	0.017(+– 0.002)	0.008(+– 0.001)	0.076(+– 0.013)	0.016(+– 0.002)	0.567(+– 0.077)	0.087(+– 0.017)
HMML std %	33.3	29.1	20.0	10.0	21.4	11.8	12.5	17.1	12.5	13.6	19.5
LiNGAM mean (+- std)	0.019(+– 0.003)	0.141(+– 0.013)	0.014(+– 0.0)	0.095(+– 0.004)	0.12(+– 0.009)	0.04(+– 0.002)	0.036(+– 0.003)	0.184(+– 0.021)	0.038(+– 0.001)	0.137(+– 0.033)	0.176(+– 0.016)
LiNGAM std %	15.8	9.2	0.0	4.2	7.5	5.0	8.3	11.4	2.6	24.1	9.1

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

Table B4. *β_i-proportionality values and their standard deviations in winter high scenarios*

	blh (m)	cc (%)	d (s^{-1})	d2m (K)	o3 ($kgkg^{-1}$)	pv (kg^{-1})	relative_humidity (%)	t (K)	vo ($m^2s^{-1}kg^{-1}$)	ws (ms^{-1})	z_x (m^2s^{-2})
HMML mean (+- std)	0.031(+– 0.001)	0.033(+– 0.011)	0.005(+– 0.001)	0.122(+– 0.032)	0.083(+– 0.028)	0.019(+– 0.003)	0.038(+– 0.007)	0.144(+– 0.036)	0.012(+– 0.001)	0.392(+– 0.124)	0.121(+– 0.008)
HMML std %	3.2	33.3	20.0	26.2	33.7	15.8	18.4	25.0	8.3	31.6	6.6
LiNGAM mean (+- std)	0.067(+– 0.009)	0.053(+– 0.001)	0.011(+– 0.002)	0.164(+– 0.012)	0.109(+– 0.012)	0.038(+– 0.002)	0.063(+– 0.002)	0.21(+– 0.012)	0.028(+– 0.002)	0.11(+– 0.028)	0.145(+– 0.016)
LiNGAM std %	13.4	1.9	18.2	7.3	11.0	5.3	3.2	5.7	7.1	25.5	11.0

Table B5. *β_i-proportionality values and their standard deviations in winter moderate scenarios*

	blh (m)	cc (%)	d (s^{-1})	d2m (K)	o3 ($kgkg^{-1}$)	pv (kg^{-1})	relative_humidity (%)	t (K)	vo ($m^2s^{-1}kg^{-1}$)	ws (ms^{-1})	z_x (m^2s^{-2})
HMML mean (+- std)	0.025(+– 0.004)	0.039(+– 0.016)	0.004(+– 0.001)	0.084(+– 0.029)	0.06(+– 0.019)	0.011(+– 0.002)	0.028(+– 0.007)	0.101(+– 0.032)	0.013(+– 0.002)	0.586(+– 0.123)	0.05(+– 0.015)
HMML std %	16.0	41.0	25.0	34.5	31.7	18.2	25.0	31.7	15.4	21.0	30.0
LiNGAM mean (+- std)	0.067(+– 0.005)	0.06(+– 0.003)	0.013(+– 0.0)	0.135(+– 0.01)	0.106(+– 0.006)	0.033(+– 0.001)	0.057(+– 0.002)	0.189(+– 0.015)	0.032(+– 0.001)	0.151(+– 0.038)	0.157(+– 0.025)
LiNGAM std %	7.5	5.0	0.0	7.4	5.7	3.0	3.5	7.9	3.1	25.2	15.9

B.2. Stability Test 2

We investigated the stability of the ordering of the β_i-proportionality values with respect to wind speed at 135m, based on their magnitudes. The ordering and corresponding β-proportionality values for HMML and LiNGAM are presented in Tables B7–B14.

Table B6. *β_i-proportionality values and their standard deviations in winter low scenarios*

	blh (m)	cc (%)	d (s^{-1})	d2m (K)	o3 ($kgkg^{-1}$)	pv (kg^{-1})	relative_humidity (%)	t (K)	vo ($m^2s^{-1}kg^{-1}$)	ws (ms^{-1})	z_x (m^2s^{-2})
HMML mean (+- std)	0.029(+– 0.007)	0.033(+– 0.012)	0.006(+– 0.001)	0.071(+– 0.023)	0.098(+– 0.031)	0.011(+– 0.005)	0.027(+– 0.007)	0.105(+– 0.047)	0.017(+– 0.008)	0.52(+– 0.154)	0.081(+– 0.017)
HMML std %	24.1	36.4	16.7	32.4	31.6	45.5	25.9	44.8	47.1	29.6	21.0
LiNGAM mean (+- std)	0.057(+– 0.001)	0.054(+– 0.003)	0.016(+– 0.001)	0.103(+– 0.004)	0.129(+– 0.009)	0.031(+– 0.003)	0.046(+– 0.001)	0.172(+– 0.017)	0.042(+– 0.004)	0.17(+– 0.04)	0.18(+– 0.024)
LiNGAM std %	1.8	5.6	6.2	3.9	7.0	9.7	2.2	9.9	9.5	23.5	13.3

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

Table B7. *Ordering and β_i-proportionality values obtained by* ***HMML****, with respect to the chosen parameters in the* ***summer*** *half year and* ***moderate wind*** *scenario*

	1	2	3	4	5	6	7	8	9	10	11
wsize=94,dmod=−1 var	ws	cc	t	d2m	z_x	o3	pv	relative_humidity	vo	d	blh
wsize=94,dmod=−1 val	0.721	0.064	0.054	0.041	0.032	0.028	0.015	0.013	0.012	0.012	0.007
wsize=94,dmod=0 var	ws	cc	t	d2m	z_x	o3	pv	vo	relative_humidity	blh	d
wsize=94,dmod=0 val	0.614	0.103	0.085	0.059	0.049	0.037	0.016	0.011	0.011	0.008	0.005
wsize=94,dmod=1 var	ws	cc	t	d2m	z_x	o3	pv	relative_humidity	vo	blh	d
wsize=94,dmod=1 val	0.39	0.185	0.126	0.09	0.084	0.063	0.019	0.019	0.012	0.009	0.003
wsize=96,dmod=−1 var	ws	cc	t	d2m	z_x	o3	pv	vo	d	relative_humidity	blh
wsize=96,dmod=−1 val	0.731	0.059	0.052	0.042	0.033	0.028	0.015	0.012	0.011	0.011	0.007
wsize=96,dmod=0 var	ws	cc	t	d2m	z_x	o3	pv	vo	relative_humidity	blh	d
wsize=96,dmod=0 val	0.626	0.101	0.082	0.058	0.048	0.036	0.014	0.011	0.01	0.008	0.006
wsize=96,dmod=1 var	ws	cc	t	d2m	z_x	o3	pv	relative_humidity	vo	blh	d
wsize=96,dmod=1 val	0.409	0.183	0.118	0.086	0.084	0.061	0.019	0.018	0.012	0.008	0.003
wsize=98,dmod=−1 var	ws	cc	t	d2m	z_x	o3	pv	vo	d	relative_humidity	blh
wsize=98,dmod=−1 val	0.729	0.061	0.055	0.042	0.033	0.027	0.015	0.012	0.011	0.011	0.006
wsize=98,dmod=0 var	ws	cc	t	d2m	z_x	o3	pv	vo	relative_humidity	blh	d
wsize=98,dmod=0 val	0.624	0.104	0.081	0.057	0.05	0.037	0.014	0.011	0.01	0.008	0.005
wsize=98,dmod=1 var	ws	cc	t	d2m	z_x	o3	pv	relative_humidity	vo	blh	d
wsize=98,dmod=1 val	0.419	0.182	0.116	0.086	0.082	0.06	0.018	0.016	0.012	0.007	0.003

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

Table B8. *Ordering and β-proportionality values obtained by* ***HMML****, with respect to the chosen parameters in the* ***winter*** *half year and* ***high wind*** *scenario*

	1	2	3	4	5	6	7	8	9	10	11
wsize=94,dmod=−1 var	ws	z_x	t	d2m	o3	relative_humidity	blh	cc	pv	vo	d
wsize=94,dmod=−1 val	0.519	0.115	0.101	0.091	0.055	0.033	0.031	0.021	0.017	0.011	0.006
wsize=94,dmod=0 var	ws	t	z_x	d2m	o3	relative_humidity	cc	blh	pv	vo	d
wsize=94,dmod=0 val	0.388	0.147	0.129	0.118	0.079	0.037	0.035	0.032	0.019	0.012	0.004
wsize=94,dmod=1 var	ws	t	d2m	z_x	o3	relative_humidity	cc	blh	pv	vo	d
wsize=94,dmod=1 val	0.227	0.187	0.163	0.132	0.123	0.048	0.048	0.032	0.023	0.013	0.003
wsize=96,dmod=−1 var	ws	z_x	t	d2m	o3	relative_humidity	blh	cc	pv	vo	d
wsize=96,dmod=−1 val	0.522	0.114	0.102	0.091	0.055	0.033	0.029	0.021	0.017	0.01	0.006
wsize=96,dmod=0 var	ws	t	z_x	d2m	o3	relative_humidity	blh	cc	pv	vo	d
wsize=96,dmod=0 val	0.429	0.148	0.113	0.111	0.073	0.034	0.03	0.029	0.015	0.011	0.005
wsize=96,dmod=1 var	ws	t	d2m	z_x	o3	relative_humidity	cc	blh	pv	vo	d
wsize=96,dmod=1 val	0.24	0.184	0.162	0.133	0.12	0.047	0.046	0.03	0.023	0.012	0.003
wsize=98,dmod=−1 var	ws	z_x	t	d2m	o3	relative_humidity	blh	cc	pv	vo	d
wsize=98,dmod=−1 val	0.528	0.113	0.1	0.089	0.056	0.031	0.029	0.019	0.016	0.01	0.006
wsize=98,dmod=0 var	ws	t	z_x	d2m	o3	relative_humidity	blh	cc	pv	vo	d
wsize=98,dmod=0 val	0.418	0.148	0.118	0.113	0.074	0.035	0.031	0.03	0.016	0.012	0.005
wsize=98,dmod=1 var	ws	t	d2m	z_x	o3	relative_humidity	cc	blh	pv	vo	d
wsize=98,dmod=1 val	0.257	0.181	0.161	0.127	0.116	0.046	0.044	0.031	0.022	0.012	0.003

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

Table B9. *Ordering and β-proportionality values obtained by* ***HMML****, with respect to the chosen parameters in the* ***winter*** *half year and* ***low wind*** *scenario*

	1	2	3	4	5	6	7	8	9	10	11
wsize=94,dmod=–1 var	ws	o3	z_x	t	d2m	relative_humidity	cc	blh	vo	pv	d
wsize=94,dmod=–1 val	0.673	0.065	0.063	0.056	0.051	0.024	0.021	0.021	0.01	0.008	0.007
wsize=94,dmod=0 var	ws	t	o3	z_x	d2m	blh	cc	relative_humidity	vo	pv	d
wsize=94,dmod=0 val	0.527	0.101	0.097	0.093	0.062	0.034	0.031	0.024	0.016	0.009	0.005
wsize=94,dmod=1 var	ws	t	o3	d2m	z_x	cc	relative_humidity	blh	vo	pv	d
wsize=94,dmod=1 val	0.311	0.169	0.138	0.102	0.102	0.052	0.036	0.034	0.029	0.018	0.007
wsize=96,dmod=–1 var	ws	o3	t	z_x	d2m	relative_humidity	cc	blh	vo	pv	d
wsize=96,dmod=–1 val	0.676	0.063	0.058	0.057	0.052	0.024	0.024	0.021	0.011	0.008	0.007
wsize=96,dmod=0 var	ws	o3	t	z_x	d2m	blh	cc	relative_humidity	vo	pv	d
wsize=96,dmod=0 val	0.568	0.094	0.089	0.084	0.061	0.033	0.027	0.02	0.012	0.008	0.004
wsize=96,dmod=1 var	ws	t	o3	d2m	z_x	cc	relative_humidity	blh	vo	pv	d
wsize=96,dmod=1 val	0.329	0.162	0.138	0.102	0.096	0.048	0.037	0.036	0.027	0.018	0.007
wsize=98,dmod=–1 var	ws	o3	z_x	t	d2m	cc	relative_humidity	blh	vo	pv	d
wsize=98,dmod=–1 val	0.69	0.062	0.059	0.055	0.048	0.023	0.021	0.018	0.01	0.009	0.005
wsize=98,dmod=0 var	ws	o3	t	z_x	d2m	blh	cc	relative_humidity	vo	pv	d
wsize=98,dmod=0 val	0.559	0.097	0.094	0.084	0.063	0.032	0.028	0.02	0.011	0.007	0.004
wsize=98,dmod=1 var	ws	t	o3	d2m	z_x	cc	blh	relative_humidity	vo	pv	d
wsize=98,dmod=1 val	0.35	0.159	0.131	0.101	0.093	0.047	0.037	0.034	0.026	0.016	0.007

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

Table B10. *Ordering and β-proportionality values obtained by* ***HMML****, with respect to the chosen parameters in the* ***winter*** *half year and* ***moderate wind*** *scenario*

	1	2	3	4	5	6	7	8	9	10	11
wsize=94,dmod=−1 var	ws	t	d2m	o3	z_x	cc	relative_humidity	blh	vo	pv	d
wsize=94,dmod=−1 val	0.684	0.071	0.059	0.042	0.039	0.027	0.026	0.023	0.014	0.012	0.004
wsize=94,dmod=0 var	ws	t	d2m	o3	z_x	cc	blh	relative_humidity	vo	pv	d
wsize=94,dmod=0 val	0.634	0.091	0.074	0.056	0.044	0.031	0.024	0.023	0.011	0.009	0.003
wsize=94,dmod=1 var	ws	t	d2m	o3	z_x	cc	relative_humidity	blh	vo	pv	d
wsize=94,dmod=1 val	0.411	0.143	0.122	0.087	0.072	0.063	0.037	0.032	0.015	0.013	0.004
wsize=96,dmod=−1 var	ws	t	d2m	o3	z_x	cc	relative_humidity	blh	vo	pv	d
wsize=96,dmod=−1 val	0.695	0.07	0.055	0.041	0.037	0.026	0.025	0.023	0.013	0.011	0.004
wsize=96,dmod=0 var	ws	t	d2m	o3	z_x	cc	relative_humidity	blh	vo	pv	d
wsize=96,dmod=0 val	0.654	0.092	0.072	0.053	0.041	0.029	0.02	0.019	0.009	0.008	0.002
wsize=96,dmod=1 var	ws	t	d2m	o3	z_x	cc	relative_humidity	blh	vo	pv	d
wsize=96,dmod=1 val	0.425	0.141	0.121	0.086	0.068	0.06	0.037	0.03	0.015	0.013	0.004
wsize=98,dmod=−1 var	ws	t	d2m	o3	z_x	cc	relative_humidity	blh	vo	pv	d
wsize=98,dmod=−1 val	0.693	0.072	0.056	0.041	0.036	0.028	0.024	0.022	0.014	0.01	0.005
wsize=98,dmod=0 var	ws	t	d2m	o3	z_x	cc	relative_humidity	blh	vo	pv	d
wsize=98,dmod=0 val	0.641	0.09	0.074	0.056	0.043	0.031	0.022	0.022	0.01	0.008	0.002
wsize=98,dmod=1 var	ws	t	d2m	o3	z_x	cc	relative_humidity	blh	vo	pv	d
wsize=98,dmod=1 val	0.437	0.142	0.118	0.082	0.067	0.06	0.034	0.029	0.014	0.012	0.004

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

Table B11. *Ordering and β-proportionality values obtained by* ***LiNGAM****, with respect to the chosen parameters in the* ***summer*** *half year and* ***moderate wind*** *scenario*

	1	2	3	4	5	6	7	8	9	10	11
wsize=94,dmod=−1 var	ws	z_x	cc	t	d2m	o3	relative_humidity	pv	vo	blh	d
wsize=94,dmod=−1 val	0.182	0.16	0.155	0.148	0.12	0.081	0.043	0.038	0.031	0.029	0.013
wsize=94,dmod=0 var	t	z_x	cc	ws	d2m	o3	relative_humidity	pv	vo	blh	d
wsize=94,dmod=0 val	0.181	0.157	0.148	0.138	0.129	0.093	0.044	0.038	0.031	0.029	0.011
wsize=94,dmod=1 var	z_x	t	cc	d2m	o3	ws	relative_humidity	pv	vo	blh	d
wsize=94,dmod=1 val	0.212	0.18	0.133	0.133	0.094	0.093	0.047	0.038	0.031	0.027	0.011
wsize=96,dmod=−1 var	ws	z_x	cc	t	d2m	o3	relative_humidity	pv	vo	blh	d
wsize=96,dmod=−1 val	0.185	0.16	0.153	0.148	0.122	0.079	0.044	0.037	0.031	0.029	0.013
wsize=96,dmod=0 var	t	z_x	cc	ws	d2m	o3	relative_humidity	pv	vo	blh	d
wsize=96,dmod=0 val	0.181	0.156	0.145	0.142	0.13	0.094	0.043	0.038	0.031	0.029	0.011
wsize=96,dmod=1 var	z_x	t	d2m	cc	ws	o3	relative_humidity	pv	vo	blh	d
wsize=96,dmod=1 val	0.208	0.182	0.134	0.133	0.096	0.094	0.047	0.038	0.031	0.027	0.011
wsize=98,dmod=−1 var	ws	z_x	cc	t	d2m	o3	relative_humidity	pv	vo	blh	d
wsize=98,dmod=−1 val	0.187	0.158	0.154	0.15	0.12	0.079	0.044	0.037	0.031	0.029	0.013
wsize=98,dmod=0 var	t	z_x	cc	ws	d2m	o3	relative_humidity	pv	vo	blh	d
wsize=98,dmod=0 val	0.181	0.157	0.145	0.144	0.128	0.094	0.043	0.038	0.031	0.029	0.011
wsize=98,dmod=1 var	z_x	t	cc	d2m	ws	o3	relative_humidity	pv	vo	blh	d
wsize=98,dmod=1 val	0.204	0.18	0.134	0.134	0.098	0.094	0.048	0.038	0.032	0.027	0.011

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

Table B12. *Ordering and β-proportionality values obtained by* ***LiNGAM****, with respect to the chosen parameters in the* ***winter*** *half year and* ***high wind*** *scenario*

	1	2	3	4	5	6	7	8	9	10	11
wsize=94,dmod=–1 var	t	z_x	d2m	ws	o3	blh	relative_humidity	cc	pv	vo	d
wsize=94,dmod=–1 val	0.195	0.148	0.147	0.138	0.094	0.077	0.059	0.055	0.039	0.032	0.014
wsize=94,dmod=0 var	t	d2m	z_x	o3	ws	blh	relative_humidity	cc	pv	vo	d
wsize=94,dmod=0 val	0.22	0.168	0.13	0.114	0.111	0.064	0.064	0.055	0.037	0.026	0.011
wsize=94,dmod=1 var	t	d2m	z_x	o3	ws	relative_humidity	blh	cc	pv	vo	d
wsize=94,dmod=1 val	0.208	0.176	0.169	0.122	0.074	0.064	0.058	0.053	0.04	0.027	0.01
wsize=96,dmod=–1 var	t	d2m	z_x	ws	o3	blh	relative_humidity	cc	pv	vo	d
wsize=96,dmod=–1 val	0.196	0.151	0.148	0.14	0.093	0.08	0.061	0.053	0.036	0.03	0.014
wsize=96,dmod=0 var	t	d2m	z_x	ws	o3	blh	relative_humidity	cc	pv	vo	d
wsize=96,dmod=0 val	0.224	0.167	0.123	0.118	0.111	0.065	0.064	0.052	0.037	0.027	0.011
wsize=96,dmod=1 var	t	d2m	z_x	o3	ws	relative_humidity	blh	cc	pv	vo	d
wsize=96,dmod=1 val	0.213	0.175	0.163	0.122	0.076	0.064	0.058	0.052	0.04	0.027	0.01
wsize=98,dmod=–1 var	t	d2m	z_x	ws	o3	blh	relative_humidity	cc	pv	vo	d
wsize=98,dmod=–1 val	0.199	0.15	0.145	0.142	0.094	0.079	0.061	0.053	0.036	0.03	0.012
wsize=98,dmod=0 var	t	d2m	z_x	ws	o3	blh	relative_humidity	cc	pv	vo	d
wsize=98,dmod=0 val	0.224	0.167	0.126	0.116	0.112	0.063	0.063	0.053	0.037	0.027	0.011
wsize=98,dmod=1 var	t	d2m	z_x	o3	ws	relative_humidity	blh	cc	pv	vo	d
wsize=98,dmod=1 val	0.213	0.178	0.157	0.122	0.078	0.065	0.058	0.051	0.04	0.027	0.01

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

Table B13. *Ordering and β-proportionality values obtained by* ***LiNGAM****, with respect to the chosen parameters in the* ***winter*** *half year and* ***low wind*** *scenario*

	1	2	3	4	5	6	7	8	9	10	11
wsize=94,dmod=–1 var	ws	z_x	t	o3	d2m	blh	cc	relative_humidity	vo	pv	d
wsize=94,dmod=–1 val	0.212	0.183	0.148	0.124	0.099	0.057	0.052	0.047	0.037	0.027	0.016
wsize=94,dmod=0 var	t	ws	z_x	o3	d2m	blh	cc	relative_humidity	vo	pv	d
wsize=94,dmod=0 val	0.184	0.171	0.155	0.14	0.103	0.058	0.053	0.045	0.044	0.031	0.016
wsize=94,dmod=1 var	z_x	t	o3	ws	d2m	cc	blh	vo	relative_humidity	pv	d
wsize=94,dmod=1 val	0.211	0.182	0.125	0.118	0.107	0.058	0.057	0.046	0.044	0.034	0.017
wsize=96,dmod=–1 var	ws	z_x	t	o3	d2m	blh	cc	relative_humidity	vo	pv	d
wsize=96,dmod=–1 val	0.211	0.186	0.15	0.121	0.098	0.056	0.052	0.047	0.036	0.027	0.016
wsize=96,dmod=0 var	t	ws	z_x	o3	d2m	blh	cc	relative_humidity	vo	pv	d
wsize=96,dmod=0 val	0.18	0.179	0.147	0.14	0.105	0.06	0.051	0.046	0.044	0.031	0.016
wsize=96,dmod=1 var	z_x	t	o3	ws	d2m	cc	blh	vo	relative_humidity	pv	d
wsize=96,dmod=1 val	0.204	0.185	0.126	0.12	0.108	0.058	0.057	0.046	0.045	0.034	0.017
wsize=98,dmod=–1 var	ws	z_x	t	o3	d2m	blh	cc	relative_humidity	vo	pv	d
wsize=98,dmod=–1 val	0.216	0.181	0.15	0.12	0.099	0.056	0.051	0.047	0.037	0.028	0.016
wsize=98,dmod=0 var	t	ws	z_x	o3	d2m	blh	cc	relative_humidity	vo	pv	d
wsize=98,dmod=0 val	0.182	0.177	0.149	0.14	0.104	0.058	0.052	0.046	0.044	0.031	0.016
wsize=98,dmod=1 var	z_x	t	o3	ws	d2m	blh	cc	vo	relative_humidity	pv	d
wsize=98,dmod=1 val	0.202	0.185	0.125	0.123	0.108	0.058	0.056	0.046	0.045	0.034	0.018

https://doi.org/10.1017/eds.2026.10031 Published online by Cambridge University Press

Table B14. *Ordering and β-proportionality values obtained by* ***LiNGAM****, with respect to the chosen parameters in the* ***winter*** *half year and* ***moderate wind*** *scenario*

	1	2	3	4	5	6	7	8	9	10	11
wsize=94,dmod=−1 var	ws	t	z_x	d2m	o3	blh	cc	relative_humidity	pv	vo	d
wsize=94,dmod=−1 val	0.189	0.169	0.156	0.123	0.099	0.071	0.063	0.054	0.032	0.032	0.013
wsize=94,dmod=0 var	t	ws	d2m	z_x	o3	blh	cc	relative_humidity	pv	vo	d
wsize=94,dmod=0 val	0.2	0.152	0.137	0.135	0.111	0.069	0.063	0.057	0.033	0.032	0.013
wsize=94,dmod=1 var	t	z_x	d2m	o3	ws	blh	relative_humidity	cc	pv	vo	d
wsize=94,dmod=1 val	0.197	0.192	0.145	0.109	0.102	0.061	0.059	0.057	0.034	0.033	0.013
wsize=96,dmod=−1 var	ws	t	z_x	d2m	o3	blh	cc	relative_humidity	vo	pv	d
wsize=96,dmod=−1 val	0.192	0.169	0.154	0.123	0.098	0.073	0.061	0.054	0.031	0.031	0.013
wsize=96,dmod=0 var	t	ws	d2m	z_x	o3	blh	cc	relative_humidity	pv	vo	d
wsize=96,dmod=0 val	0.203	0.158	0.139	0.128	0.11	0.069	0.061	0.057	0.032	0.031	0.013
wsize=96,dmod=1 var	t	z_x	d2m	o3	ws	blh	relative_humidity	cc	pv	vo	d
wsize=96,dmod=1 val	0.196	0.19	0.144	0.109	0.104	0.062	0.059	0.056	0.034	0.033	0.013
wsize=98,dmod=−1 var	ws	t	z_x	d2m	o3	blh	cc	relative_humidity	vo	pv	d
wsize=98,dmod=−1 val	0.195	0.171	0.15	0.122	0.097	0.072	0.062	0.054	0.032	0.032	0.013
wsize=98,dmod=0 var	t	ws	d2m	z_x	o3	blh	cc	relative_humidity	pv	vo	d
wsize=98,dmod=0 val	0.2	0.157	0.138	0.13	0.11	0.07	0.062	0.057	0.032	0.032	0.013
wsize=98,dmod=1 var	t	z_x	d2m	o3	ws	blh	relative_humidity	cc	pv	vo	d
wsize=98,dmod=1 val	0.198	0.181	0.145	0.11	0.107	0.062	0.059	0.057	0.034	0.033	0.013