Environmental Data Science (2025), 4: e48, 1–10
doi:10.1017/eds.2025.10008

CAMBRIDGE UNIVERSITY PRESS

APPLICATION PAPER

From winter storm thermodynamics to wind gust extremes: discovering interpretable equations from data

Frederick Iat-Hin Tam[1,2], Fabien Augsburger[1] and Tom Beucler[1,2]

[1]Faculty of Geosciences and Environment, University of Lausanne, Lausanne, VD, Switzerland
[2]Expertise Center for Climate Extremes, University of Lausanne, Lausanne, VD, Switzerland
Corresponding author: Frederick Iat-Hin Tam; Email: ft21894@gmail.com

Received: 14 June 2025; **Revised:** 14 June 2025; **Accepted:** 16 June 2025

Keywords: dimensionality reduction; equation discovery; interpretable statistical learning; pareto optimization

Abstract

Reliably identifying and understanding temporal precursors to extreme wind gusts is crucial for early warning and mitigation. This study proposes a simple data-driven approach to extract key predictors from a dataset of historical extreme European winter windstorms and derive simple equations linking these precursors to extreme gusts over land. A major challenge is the limited training data for extreme events, increasing the risk of model overfitting. Testing various mitigation strategies, we find that combining dimensionality reduction, careful cross-validation, feature selection, and a nonlinear transformation of maximum wind gusts informed by Generalized Extreme Value distributions successfully reduces overfitting. These measures yield interpretable equations that generalize across regions while maintaining satisfactory predictive skill. The discovered equations reveal the association between a steady drying low-troposphere before landfall and wind gust intensity in Northwestern Europe.

Impact Statement

Data-driven equation discovery is a growing field in scientific machine learning, yet its application to extreme events remains underexplored, particularly for small, noisy, multidimensional datasets. Using European extreme winter storm data, we demonstrate that careful cross-validation, nonlinear transformations informed by extreme value theory, and multi-objective optimization yield interpretable equations that generalize well across regions and unseen storms. This paves the way for objective knowledge discovery for weather extremes.

1. Introduction

Damaging windstorms associated with extratropical cyclones are among the leading weather-related disasters in the mid-latitudes winters, especially in Europe (Schwierz et al., 2010). From a weather forecasting standpoint, early detection of reliable precursors leading to extreme windstorms helps with timely disaster mitigation. Beyond horizontal temperature differences strengthening storms (Laurila et al., 2021), studies have highlighted how lower-atmosphere humidity fuels wind jets

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

© The Author(s), 2025. 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.2025.10008 Published online by Cambridge University Press

(Martínez-Alvarado et al., 2022) and transports high-wind air toward the surface (Browning et al., 2015), yet simple models linking winterstorm characteristics to wind gusts are lacking.

Interpretable machine learning has enabled trustworthy data-driven models of thunderstorms (Hilburn, 2023) and tropical cyclones (Tam et al., 2024), improving process understanding by reducing the attribution uncertainty of post-hoc explainable AI methods (Mamalakis et al., 2023). This motivates us to apply data-driven methods to uncover novel temporal patterns in environmental precursors to extreme European windstorms. A key challenge is the limited availability of extreme storm data, which restricts the use of complex nonlinear models prone to overfitting in small-sample conditions. However, this limitation also presents an opportunity to derive a simple, interpretable equation approximating complex physical processes, paving the way for data-driven discovery.

Data-driven equation discovery distills simple laws from empirical data (Schmidt and Lipson, 2009) with methods ranging from sparse selection algorithms to symbolic regression (Song et al., 2024). While data-driven approaches have successfully identified physical laws from controlled fluid dynamics experiments (Brunton et al., 2016; Cheng and Alkhalifah, 2024), finding interpretable equations from noisy, multidimensional weather and climate data remains under-explored. Balancing model performance and algorithmic complexity, Grundner et al. (2024) uncovered simple analytical cloud cover equations from high-resolution model outputs. We similarly implement a hierarchical modeling framework to track the trade-off between the model complexity and generalization capability, identifying a set of sparse yet descriptive equations that achieves satisfactory prediction skills for unseen storm cases across European regions. Additionally, we explore the best strategy for dimensionality reduction and feature selection for reliable discovery of simple equations from small yet high-dimensional datasets.

2. Data

2.1. Raw datasets description

Winter windstorm tracks for Europe. The winter storm data analyzed in this study is based on the ECMWF Winter Windstorm Indicator dataset (C3S CDS, 2022). This dataset documents historical severe windstorms between 1979 and 2021, and their footprints. The tracking of storms is performed by applying a tracking algorithm on the ERA5 reanalysis data from October through March. The tracking algorithm tracks local maxima of 850 hPa relative vorticity, with a minimal 10 m wind speed threshold of 25 m s^{-1}. A total of 118 storms is identified in the 42-year period. This study analyzes storms in the last 30 years (96 storms). Since we focus on the impact of severe windstorms over land in this study, short-lived or non-landfalling storms are filtered out, which leaves us with 63 storms.

ERA5 meteorological reanalysis. We use the hourly ERA5 reanalysis dataset with a horizontal resolution of 0.25°–0.25° to characterize the environmental conditions for all tracked storms. Our analysis considers 28 potential environmental drivers for wind gust formation and intensity (Schulz and Lerch, 2022) listed in Supplementary Information Table 6. To find the driver conditions local to each storm, we adopt a system-following 4°–4° grid box centered around the storm track of all 28 environmental drivers. These drivers include measurements of the winterstorm kinematics (e.g., 10 m wind speeds), precipitation characteristics (e.g., mean precipitation rates), and the thermodynamic characteristics (e.g., latent heat flux; K-index) near the windstorms. A full list of the ERA5 variables and their units is provided in Supplementary Information Table 6.

2.2. Data preprocessing: dimensionality reduction and partitioning

Focusing on extreme events limits the training data for data-driven models. To ensure reliable equation discovery, we reduce the dimensionality of storm history and post-landfall gust data, turning our dataset into one that can be analyzed with simple data-driven models without overfitting.

Features: prescribed spatial dimensionality reduction. To encode the 4D storm history fields as time series, we compute four spatial statistics: maximum (max), minimum (min), mean, and standard deviation (std). Using predefined statistics instead of learned filters facilitates physical interpretability. For example,

https://doi.org/10.1017/eds.2025.10008 Published online by Cambridge University Press

a higher temperature standard deviation may indicate the presence of weather fronts (Lagerquist et al., 2022). To further reduce dimensionality, we apply Principal Component Analysis (PCA) to compress time series into orthogonal temporal modes ($\Pi_X(t)$) and PC loadings (PC_X), which are scalars denoting the projection of the time series onto $\Pi_X(t)$:

$$X(t) - \overline{X(t)} = \sum_{i=1}^{N} \mathrm{PC}_{X,i}\Pi_{X,i}(t), \quad \text{where} \quad \langle \Pi_{X,i}(t) | X(t) \rangle_X = \mathrm{PC}_{X,i}. \tag{1}$$

Here, N is the number of retained PC modes indexed by i, and $\langle \cdot | \cdot \rangle_X$ denotes the inner product associated with the PCA decomposition of X. The modes $\Pi_{X,i}(t)$ are normalized such that $\langle \Pi_{X,i}(t) | \Pi_{X,j}(t) \rangle_X = \delta_{ij}$ where δ_{ij} is the Kronecker delta, ensuring orthonormality of the PC modes.

Targets: geographical clustering. In a low-sample-size regime, predicting wind gusts for every pixel across Europe is impractical and difficult to interpret. We use K-means++ clustering (Arthur and Vassilvitskii, 2007) to partition the landmass into K subregions, incorporating elevation, latitude, longitude, and maximum gusts from 63 severe windstorms. This ensures clusters reflect windstorm gust characteristics while remaining spatially coherent. To enforce geographical contiguity, we apply k-nearest neighbors (kNN; Enas and Choi, 1986), reassigning edge pixels based on the most frequent cluster label. Although silhouette scores suggest an optimal $K = 4$–5 (Supplementary Figure S1), these clusters group disjointed regions with distinct wind gust characteristics (Supplementary Figure S2). We therefore consider three additional clustering metrics (e.g., Xie and Beni, 1991), which identify local optima at $K = 9$, 14, and 15 (Supplementary Figure S3). Among these, $K = 15$ best preserves geographical compactness and aligns with established European climatic zones (Jylhä et al., 2010) (Supplementary Figure S4). With clusters defined, we extract, for each storm, the highest instantaneous 10 m wind gust within each cluster during the 15 hours post-landfall. This window ensures short-lived but intense storms like Lothar (Wernli et al., 2002) are included. Our target is the maximum 15-hour wind gust within each cluster ($U_{\mathrm{gust},i}$), where i indexes the 15 clusters.

Cross-validation procedure. To ensure data-driven models select gust predictors applicable to severe windstorms, we always leave out the same seven storms for testing. The test set is randomly chosen, except for the Lothar storm, which is explicitly withheld to assess performance on an extreme, unseen case. The remaining 56 storms are split into 7 training-validation folds by randomly selecting 7 sets of 8 storms without replacement for validation ("Random split"). Alternatively, as a step toward invariant causal prediction (Peters et al., 2016), we identify features robust to distributional shifts by clustering the output vector $\mathbf{U}_{\mathbf{gust}}$ into seven groups via K-means++ (Supplementary Table S2). We then train on six clusters and validate on the remaining one, repeating this process seven times ("ICP split"), following Häfner et al. (2023). Sweet et al. (2023) show that such an "ICP split" outperforms naive k-fold cross-validation for regression.

2.3. Extreme value theory for wind gust distributions

Assessing wind gust severity relative to local climatology and distinguishing moderate from extreme gusts are key to risk assessment. We describe extreme gusts using the Generalized Extreme Value (GEV) distribution, a family of continuous distributions designed for block maxima such as extreme sea levels (Méndez et al., 2022). The cumulative distribution function (CDF) and its functional inverse (quantile function) are

$$G_{\mu,\sigma,\xi}(x) = \begin{cases} \exp\left[-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right], & \xi \neq 0, \\ \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right], & \xi = 0, \end{cases} \quad G^{-1}_{\mu,\sigma,\xi}(p) = \begin{cases} \mu + \frac{\sigma}{\xi}\left[(-\ln p)^{-\xi} - 1\right], & \xi \neq 0, \\ \mu - \sigma\ln(-\ln p), & \xi = 0. \end{cases} \tag{2}$$

where $\mu \in \mathbb{R}$ (location), $\sigma \in \mathbb{R}^*_+$ (scale), and $\xi \in \mathbb{R}$ (shape). Unlike empirical CDFs, the GEV enables extrapolation beyond observed data to estimate return periods for extreme events. Due to the high spatial variability in land surface and topography across Europe, we fit a separate GEV for each cluster

https://doi.org/10.1017/eds.2025.10008 Published online by Cambridge University Press

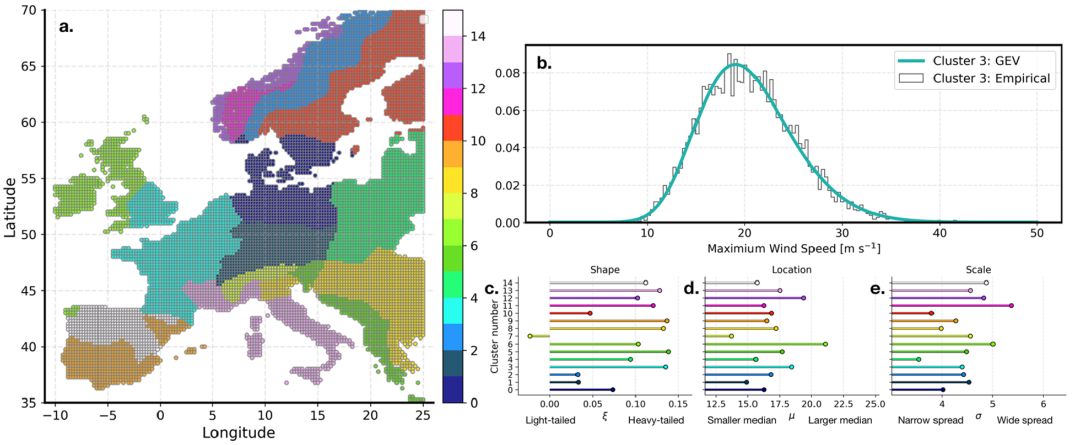


Figure 1. *(a) We derive geographically contiguous regions of the European continent via K-means++ and kNN clustering to define maximum wind gust targets. (b) Extreme gust distributions are modeled using the GEV distribution, with shape (c), location (d), and scale (e) parameters fitted for each region.*

(Figure 1c–e; Supplementary Table S1). After testing multiple definitions and durations for block maxima sampling, we opted for *daily* maximum gusts over all cluster grid points during winter months (Supplementary Figure S5). This approach differentiates regions frequently impacted by strong gusts from those less affected (Figure 1a,d) and allows modeling of unseen extremes, which empirical distributions cannot capture (Figure 1b).

3. Methodology

Feature selection. We aim to derive *interpretable* equations linking storm characteristics to post-landfall gusts. Even with prescribed spatial statistics and PCA for dimensionality reduction, the number of features (28 predictors × 4 spatial statistics × 1–12 PC loadings) still far exceeds the number of training and validation samples (56). To address this underdetermined problem, we apply sequential forward feature selection, minimizing the mean or maximum validation root-mean-squared error (RMSE) across our seven cross-validation splits. Features are added until the RMSE objective no longer decreases.

Model hierarchy. We construct a hierarchy of data-driven models expressible as simple equations by training multiple linear regressions (MLR) with PC loadings as features while varying three hyperparameters: feature smoothing, retained feature variance, and target transformation. For feature smoothing, we apply a low-pass rectangular filter to the PC loading time series, testing cutoff frequencies from 0.52 rad/hr to 1.92 rad/hr. For retained variance, we set four thresholds to determine the number of retained PC components per predictor: [75%, 80%, 85%, 90%]. The number of retained PC components per predictor acts as a regularization hyperparameter controlling the predictor variance fed to the ML framework. Finally, we explore a nonlinear, data-adaptive transformation of maximum wind gusts to encourage out-of-distribution generalization for extremes (Buriticá and Engelke, 2024). For each sample, we first obtain its percentile value from the CDF fitted in each region, then apply a nonlinear transformation to better distinguish extreme values in the GEV distribution tails: $Z_i = -\ln\left[1 - \mathrm{CDF}_i\left(U_{\mathrm{gust,i}}\right)\right]$, where Z is the transformed target, and CDF_i is the cumulative distribution function of the GEV fitted in region i. All MLRs are trained using the least squares algorithms from the Scikit-Learn Python package (Pedregosa et al., 2011).

Pareto optimality for equation discovery. Balancing performance and complexity is a multi-objective optimization task that can be addressed through Pareto optimization (Jin and Sendhoff, 2008). Following sequential feature selection, we select optimal models from the empirical "Pareto front," corresponding to

https://doi.org/10.1017/eds.2025.10008 Published online by Cambridge University Press

the lowest error achieved for a given complexity before the mean (or max) validation RMSE across seven cross-validation splits plateaus. We define complexity as the number of unique features (predictor + spatial statistic) selected, as this better reflects equation simplicity than, for instance, the number of trainable parameters.

Model evaluation. Pareto optimization seeks models with the lowest error while maintaining simplicity. We evaluate model performance using the mean validation RMSE across seven folds and measure complexity by the number of selected unique features (predictor + spatial statistic). Using the max validation RMSE yields different Pareto-optimal models, which are also evaluated. For models predicting **Z**, we convert predictions back to wind gust speed in physical units using the per-cluster GEV quantile function ($G^{-1}_{\mu,\sigma,\xi}$) before computing RMSE, which ensures consistency across model hierarchies.

4. Results

4.1. Pareto-optimal models for wind gust prediction

We explore the (complexity, error) space of trained MLRs by varying hyperparameters—feature smoothing, retained variance, and target transformation—as well as the cross-validation strategy (mean or max validation RMSE). This allows us to identify Pareto-optimal maximum wind gust equations in low sample size conditions, forming the empirical *Pareto front* (dashed black line, Figure 2b). When features are selected by minimizing the max validation RMSE, the minimum achievable RMSE stabilizes after five selected unique variables, suggesting that additional variables may not improve generalization to unseen cases. The steady decline in validation RMSE beyond five features sets an upper complexity limit, beyond which models lose their ability to generalize across regions and storms with different gust spatial patterns. Additionally, the Pareto front (Figure 2b) indicates that models perform best when retaining 90% of PC variance and applying moderate smoothing (removing oscillations with frequency >1.5 rad/hr). These hyperparameters yield features most likely to generalize well across all storm cases.

As expected, the performance table for the best *sparse* models from different feature selection methods (Table 1) shows that models targeting the transformed gusts ($\boldsymbol{Z}$) generally underperform models directly targeting $\boldsymbol{U}_{\text{gust}}$ on the validation set when RMSE is measured in $\boldsymbol{U}_{\text{gust}}$ units. Additionally, cross-validation using the "ICP split" yields no validation RMSE improvement over naive random splitting. However, generalizable equations must perform robustly on unseen storms. While the "ICP split" sacrifices some

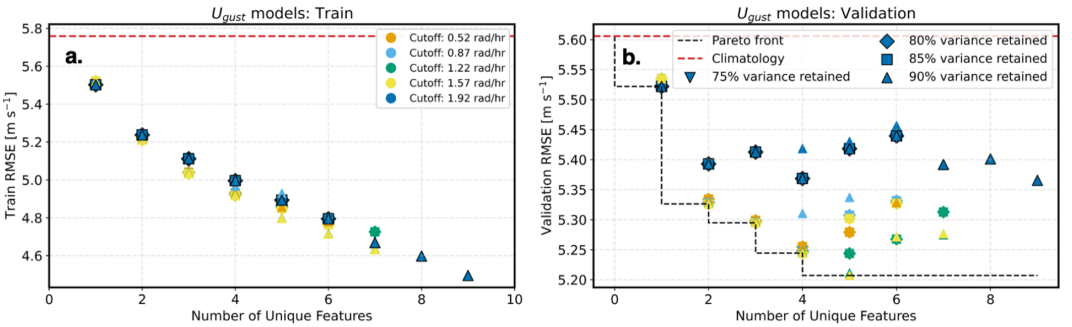


Figure 2. *(a) Training and (b) validation (complexity, error) plots for MLR models predicting* U_{gust}*, with features selected to minimize the maximum validation RMSE across seven randomly split cross-validation folds. Each colored symbol represents a model with different hyperparameters: cutoff frequency for smoothing is indicated by transparency levels, and retained feature variance by shape and color. The Pareto front (black dashed line in b) shows improved validation skills up to five unique features, suggesting that sparser linear models generalize better. In contrast, training RMSE decreases linearly with additional features, suggesting overfitting beyond 5 features. All models outperform the climatological baseline, which predicts the training mean* U_{gust} *and is shown as a red dashed line.*

https://doi.org/10.1017/eds.2025.10008 Published online by Cambridge University Press

Table 1. *RMSE for the best models in the hierarchy, along with their unique feature count. We report the mean and the standard deviation (in parentheses) across feature selection methods and use bold font for the best trained, validation and test RMSE for all trained models (smallest mean RMSE and spread). Model performance is compared to the climatology mean baseline to assess skill.*

Feature selection methods	Feature count	Training	Validation	Test
Max RMSE; Random split; U	4	4.92 (**0.06**)	5.24 (**0.30**)	6.57 (0.02)
Mean RMSE; Random split; U	4	**4.86** (0.13)	**5.16** (0.75)	7.42 (0.01)
Max RMSE; Random split; Z	4	5.22 (0.09)	5.62 (0.42)	7.17 (0.02)
Mean RMSE; Random split; Z	4	5.04 (0.12)	5.39 (0.61)	7.17 (0.05)
Mean RMSE; ICP split; U	4	5.05 (0.14)	5.76 (1.14)	6.20 (**0.01**)
Max RMSE; ICP split; U	3	5.20 (0.13)	5.86 (0.93)	6.68 (0.01)
Mean RMSE; ICP split; Z	4	5.12 (0.14)	5.90 (1.15)	**6.08** (0.03)
Max RMSE; ICP split; Z	3	5.40 (0.18)	6.00 (0.79)	6.47 (0.02)
Climatology; U; Random split	0	5.75	5.61	6.23
Climatology; U; ICP split	0	5.76	5.71	6.23

training and validation skill, it successfully identifies sparse equations that generalize well to extreme cases like Storm Lothar, as evidenced by its low test RMSE. In contrast, equations discovered using random splits consistently perform worse on test sets, suggesting implicit overfitting to validation sets by selecting variables optimized for specific gust spatial distributions. The smaller error spread of $\boldsymbol{U}_{\text{gust}}$ models confirms their tendency to select sparse equations suboptimal for unseen cases, regardless of hyperparameters. Thus, splitting storms by their wind gust spatial distribution for cross-validation ("ICP split"), transforming the target using extreme value theory, and leveraging this information for feature selection leads to more reliable models, revealing features that better anticipate unseen cases.

4.2. Reducing geographical bias through GEV-informed wind gust transformations

Why do the $\boldsymbol{Z}$ models outperform the $\boldsymbol{U}_{\text{gust}}$ models on the test set while underperforming on training and validation sets? To address this, we examine whether this performance gap varies across geographical regions. We compare the best $\boldsymbol{U}_{\text{gust}}$ and $\boldsymbol{Z}$ models trained with the "ICP split" feature selection method to assess regional differences in model skill. The $\boldsymbol{U}_{\text{gust}}$ model exhibits a distinct west–east RMSE gradient (Figure 3a), with higher errors in Northwestern Europe, where severe windstorms frequently make landfall. This suggests that the model prioritizes inland clusters with weaker winds, potentially at the expense of distinguishing differences in strong coastal gusts (Figure 3a). It may also struggle to anticipate rare inland windstorms, such as Storm *Aila* (2020) (Rantanen et al., 2021). Therefore, the main difference between the $\boldsymbol{U}_{\text{gust}}$ and $\boldsymbol{Z}$ error maps (Figure 3) is that the nonlinear gust transformation imposes stronger spatial regularization. The $\boldsymbol{Z}$ model shows decreased skill in Eastern Europe and the Balkans but improved performance in Northwestern Europe, suggesting reduced geographical bias. These results indicate that the GEV-informed gust transformation mitigates regional biases in MLR predictions, leading to a fairer model despite its slightly higher overall RMSE compared to $\boldsymbol{U}_{\text{gust}}$ MLRs.

4.3. Leveraging interpretable, pareto-optimal equations for scientific insight

Both the best 4-feature $\boldsymbol{U}_{\textbf{gust}}$ and $\boldsymbol{Z}$ models are linear functions of the *standardized* PC loadings, defined as $\widetilde{\text{PC}}_{X,j} = \frac{\text{PC}_{X,j} - \text{PC}_{X,j}^{\text{mean}}}{\text{PC}_{X,j}^{\text{std}}}$, where the mean and standard deviation are calculated over the training set and j indexes the selected features. After fitting, these standardized PC loadings are multiplied by region-specific weights $\boldsymbol{a}_{\boldsymbol{X}} \in \mathbb{R}^{15\times 4}$ and then added to biases $\boldsymbol{b} \in \mathbb{R}^{15}$ for each of the 15 regions of interest. Using the inner product from Eq. 1, we express the models as scaled projections of winterstorm fields onto data-driven temporal patterns $\Pi_{X,j}(t)$. The equation for the $\boldsymbol{U}_{\text{gust}}$ model is:

https://doi.org/10.1017/eds.2025.10008 Published online by Cambridge University Press

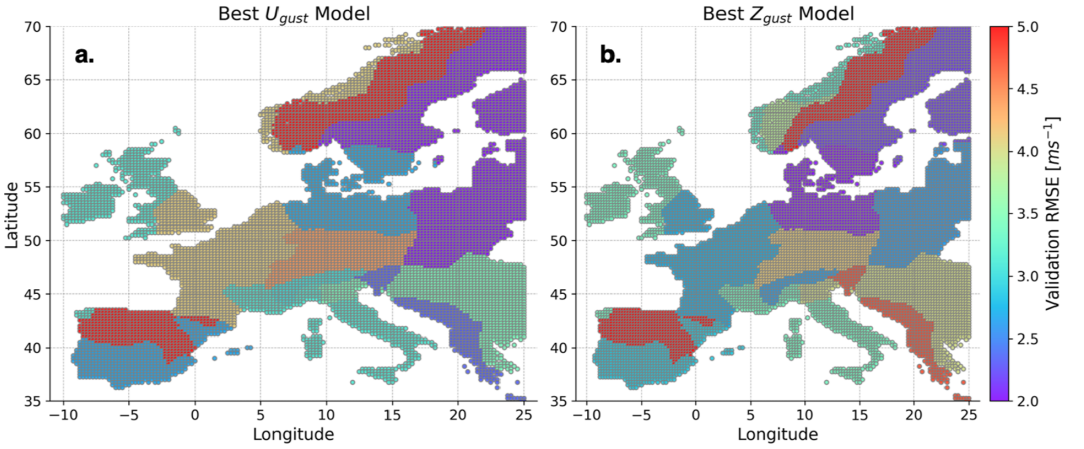


Figure 3. *Validation error maps for (a) the best* $\boldsymbol{U}_{\text{gust}}$ *model and (b) the best* $\boldsymbol{Z}$ *model in physical units (m* s^{-1}*). The nonlinear, GEV-informed transformation reduces the west–east error gradient in* $\boldsymbol{U}_{gust}$ *and improves gust predictions in windstorm-prone Northwestern Europe.*

$$\boldsymbol{U}^{\boldsymbol{U}}_{\textbf{gust}} = \boldsymbol{\beta}^{\boldsymbol{U}} + \sum_{j=1}^{4} \boldsymbol{\alpha}^{\boldsymbol{U}}_{\boldsymbol{X},\boldsymbol{j}} \langle \Pi_{X,j}(t) | X_j(t) \rangle_{X_j} \text{ where: } \quad \boldsymbol{\alpha}^{\boldsymbol{U}}_{\boldsymbol{X},\boldsymbol{j}} = \frac{\boldsymbol{a}^{\boldsymbol{U}}_{\boldsymbol{X},\boldsymbol{j}}}{\text{PC}^{\text{std}}_{X,j}}, \boldsymbol{\beta}^{\boldsymbol{U}} = \boldsymbol{b}^{\boldsymbol{U}} - \sum_{j=1}^{4} \frac{\text{PC}^{\text{mean}}_{X,j}}{\text{PC}^{\text{std}}_{X,j}} \mathbf{1}. \tag{3}$$

For the $\boldsymbol{Z}$ model, using the definition of Z and the GEV quantile function from Eq. 2, we obtain:

$$\boldsymbol{U}^{\boldsymbol{Z}}_{\textbf{gust}} = \mathbf{G}^{-1}_{\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\xi}} \left(1 - \exp \left[-\boldsymbol{\beta}^{\boldsymbol{Z}} - \sum_{j=1}^{4} \boldsymbol{\alpha}^{\boldsymbol{Z}}_{\boldsymbol{X},\boldsymbol{j}} \langle \Pi_{X,j}(t) | X_j(t) \rangle_{X_j} \right] \right) \text{ with } \left(\boldsymbol{\alpha}^{\boldsymbol{Z}}_{\boldsymbol{X},\boldsymbol{j}}, \boldsymbol{\beta}^{\boldsymbol{Z}} \right) \text{ similarly defined.} \tag{4}$$

$\boldsymbol{U}^{\boldsymbol{U}}_{\text{gust}}$ and $\boldsymbol{U}^{\boldsymbol{Z}}_{\text{gust}}$ are the predicted maximum wind gusts for each region using the direct and transformed models, respectively; $\boldsymbol{b}^{\boldsymbol{U}}, \boldsymbol{b}^{\boldsymbol{Z}} \in \mathbb{R}^{15}$ are region-specific biases; $\boldsymbol{a}^{\boldsymbol{U}}_{\boldsymbol{X},\boldsymbol{j}}, \boldsymbol{a}^{\boldsymbol{Z}}_{\boldsymbol{X},\boldsymbol{j}} \in \mathbb{R}^{15}$ are region-specific regression weights for the selected PC loadings; $\Pi_{X,j}(t)$ are the data-driven temporal patterns associated with each winterstorm feature; $\langle \Pi_{X,j}(t) | X_j(t) \rangle_{X_j}$ represents the projection of the storm field $X_j(t)$ onto the temporal mode $\Pi_{X,j}(t)$; $\boldsymbol{G}^{-1}_{\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\xi}}$ is the GEV quantile function, mapping transformed values back to physical wind gust units; and $\mathbf{1} \in \mathbb{R}^{15}$ is the identity vector (a column vector of ones).

For a region i and a feature j, models predict higher $U_{\text{gust},i}$ when the product of $a_{X,j,i}$ and $\widetilde{PC}_{X,j}$ is positive, and vice versa. Assuming $a_{X,j,i} > 0$, time series that project positively onto the PC eigenvectors lead to higher wind gust predictions. The regression coefficients $\boldsymbol{a}_{\boldsymbol{X},\boldsymbol{i}}$ in the best $\boldsymbol{Z}$ model exhibits significant geographical variability (Figure 4a-d), indicating that the influence of temporal patterns on gusts varies across regions. For example, two low-tropospheric humidity PCs in the $\boldsymbol{Z}$ model are positively correlated with gusts in region 4 (France and the Netherlands; Figure 4a-b), while the other two PC terms are negatively correlated (Figure 4c-d). The temporal evolution of the first PC mode of RH^{850}_{max} and the third PC mode of RH^{975}_{max} (Figure 4e) suggests that higher gusts in France occur when drying persists within or at the top of the boundary layer (850 hPa) before storm landfall. A drier lower troposphere promotes the descent of high-speed winds via evaporation (Browning et al., 2015) or through downward mixing (Pantillon et al., 2018). In contrast, a larger standard deviation of 500 hPa geopotential height (Φ) before landfall (Figure 4f) may indicate more perturbed west–east winds, which can intensify winterstorms. To summarize, Figure 4 demonstrates that we can extract important storm predictors leading to extreme wind gusts, such as low-tropospheric humidity, and quantify their relative importance. Finally, the temporal variability encoded in the PCs reveals additional insights into the timing of the drying or moistening tendency relative to landfall, which increases the likelihood of strong gusts.

https://doi.org/10.1017/eds.2025.10008 Published online by Cambridge University Press

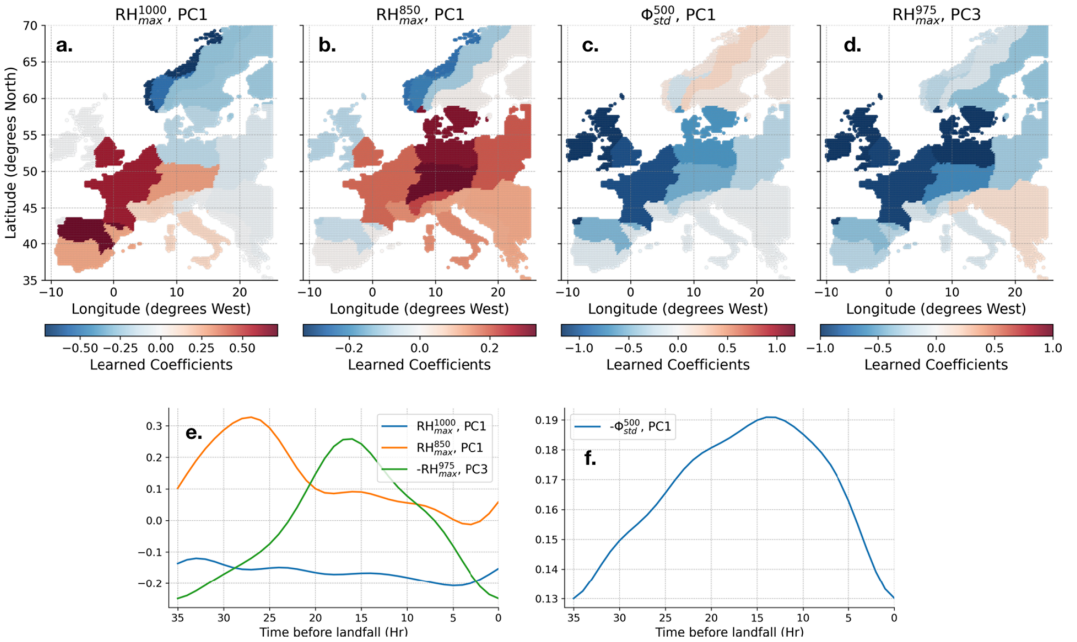


Figure 4. *A visual representation of the learned weights for the 4 features in equation of our best* $\boldsymbol{Z}$ *model: (a)* RH^{1000}_{max}*, (b)* RH^{850}_{max}*, (c)* RH^{975}_{max}*, and (d)* Φ^{500}_{std}*. Comparing these weights across clusters to the time series of the four selected PC modes (e-f) highlights key temporal patterns in storm history that promote or suppress extreme winds after landfall in different regions.*

5. Conclusion

We derived interpretable equations linking storm history to post-landfall extreme wind gusts through a PC-regression-based framework. By combining dimensionality reduction and temporal smoothing, we trained a hierarchy of equations from small datasets while mitigating overfitting via careful cross-validation, sequential feature selection, and Pareto optimization. Our results suggest that hierarchical modeling improves generalizability in small-sample settings, with the "Invariant Causal Prediction split" providing the greatest benefit for unseen test cases. Pareto optimization indicates that equations should contain at most 3–4 features to balance generalizability and predictive skill. While the best-performing $\boldsymbol{U}_{\text{gust}}$ model performs well globally (mean validation RMSE), spatial analysis reveals a west–east error structure, as the model prioritizes weak-wind regions at the expense of capturing gust variations in strong-wind regions. A GEV-informed nonlinear transformation reduces this spatial bias by accentuating distinctions in gust percentiles at the distribution tail. The discovered equations provide insights into how the timing of low-tropospheric drying is associated with wind gusts in Northwestern Europe, underscoring the importance of storm thermodynamic history in predicting post-landfall wind gusts. Future extensions could apply this framework to larger datasets and explore alternative interpretable modeling approaches, including nonlinear methods such as symbolic regression.

Supplementary material. The supplementary material for this article can be http://doi.org/10.1017/eds.2025.10008.

Open peer review. To view the open peer review materials for this article, please visit http://doi.org/10.1017/eds.2025.10008.

Acknowledgements. We are grateful for the technical assistance of M. Gomez and the advice of E. Koch.

Author contribution. All authors contributed to the conceptualization, methodology, writing, and approved the final submitted draft. Data curation and visualization: F.I.T., F.A.

Competing interests. The authors declare none.

Data availability statement. The released version of the Github repository containing code to reproduce the main figures of this work can be accessed through Zenodo and has been assigned the following DOI: 10.5281/zenodo.14766994. The ECMWF Windstorm Indicator Data can be downloaded from the Copernicus Windstorm Indicator data archive. ERA5 datasets were downloaded from the Copernicus website (multiple pressure levels as well as single pressure levels).

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

Financial support. This research was supported by the Canton of Vaud, Switzerland, through the base funding of UNIL.

References

Arthur D and Vassilvitskii S (2007) k-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. pp. 1027–1035.

Browning KA, **Smart DJ**, **Clark MR and Illingworth AJ** (2015) The role of evaporating showers in the transfer of sting-jet momentum to the surface. *Quarterly Journal of the Royal Meteorological Society 141*(693), 2956–2971.

Brunton SL, **Proctor JL and Kutz JN** (2016) Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences 113*(15), 3932–3937.

Buriticá G and Engelke S (2024) Progression: An extrapolation principle for regression. Preprint, arXiv:2410.23246.

Cheng S and Alkhalifah T (2024) Discovery of physically interpretable wave equations. Preprint, arXiv:2404.17971.

Copernicus Climate Change Service, Climate Data Store (2022) Winter windstorm indicators for Europe from 1979 to 2021 derived from reanalysis. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. https://doi.org/10.24381/cds.9b4ea013

Enas GG and Choi SC (1986) Choice of the smoothing parameter and efficiency of K-nearest neighbor classification. *Statistical Methods of Discrimination and Classification*, 235–244.

Grundner A, **Beucler T**, **Gentine P and Eyring V** (2024) Data-driven equation discovery of a cloud cover parameterization. *Journal of Advances in Modeling Earth Systems 16*(3), e2023MS003763.

Häfner D, **Gemmrich J and Jochum M** (2023) Machine-guided discovery of a real-world rogue wave model. *Proceedings of the National Academy of Sciences 120*(48), e2306275120. https://doi.org/10.1073/pnas.2306275120.

Hilburn KA (2023) Understanding spatial context in convolutional neural networks using explainable methods: Application to interpretable GREMLIN. *Artificial Intelligence Earth Systems 2*, 220093.

Jin Y and Sendhoff B (2008) Pareto-based multiobjective machine learning: An overview and case studies. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 38*(3), 397–415.

Jylhä K, **Tuomenvirta H**, **Ruosteenoja K**, **Niemi-Hugaerts H**, **Keisu K and Karhu JA** (2010) Observed and projected future shifts of climatic zones in Europe and their use to visualize climate change information. *Weather, Climate, and Society 2*, 148–167.

Lagerquist R, **Allen JT and McGovern A** (2022) Climatology and variability of warm and cold fronts over North America from 1979 to 2018. *Journal of Climate 33*, 6531–6554.

Laurila TK, **Gregow H**, **Cornér J and Sinclair VA** (2021) Characteristics of extratropical cyclones and precursors to windstorms in northern Europe. *Weather and Climate Dynamics 2*(4), 1111–1130.

Mamalakis A, **Barnes EA and Ebert-Uphoff I** (2023) Carefully choose the baseline: Lessons learned from applying XAI attribution methods for regression tasks in geoscience. *Artificial Intelligence for the Earth Systems 2*, e220058. https://doi.org/10.1175/AIES-D-22-0058.1.

Martínez-Alvarado O, **Gray SL**, **Clark PA and Baker LH** (2022) Objective detection of sting jets in low-resolution datasets. *Meteorological Applications 20*(1), 41–55.

Méndez FJ, **Menéndez M**, **Luceño A and Losada IJ** (2022) Analyzing monthly extreme sea levels with a time-dependent GEV model. *Journal of Atmospheric and Oceanic Technology 24*, 894–911.

Pantillon F, **Lerch S**, **Knippertz P and Corsmeier U** (2018) Forecasting wind gusts in winter storms using a calibrated convection-permitting ensemble. *Quarterly Journal of the Royal Meteorological Society 144*(715), 1864–1881.

Pedregosa F, **Varoquaux G**, **Gramfort A**, **Michel V**, **Thirion B**, **Grisel O and Duchesnay É** (2011) Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research 12*, 2825–2830.

Peters J, **Bühlmann P and Meinshausen N** (2016) Causal inference by using invariant prediction: Identification and confidence intervals. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78*(5), 947–1012. https://doi.org/10.1111/rssb.12167.

Rantanen M, **Laurila TK**, **Sinclair V and Gregow H** (2021) Storm Aila: An unusually strong autumn storm in Finland. *Weather 76*(9), 306–312.

Schmidt M and Lipson H (2009) Distilling free-form natural laws from experimental data. *Science 324*(5923), 81–85.

Schulz B and Lerch S (2022) Machine learning methods for postprocessing ensemble forecasts of wind gusts: A systematic comparison. *Monthly Weather Review 150*, 235–257.

Schwierz C, **Köllner-Heck P**, **Zenklusen Mutter E**, **Bresch DN**, **Vidale PL**, **Wild M and Schär C** (2010) Modelling European winter wind storm losses in current and future climate. *Climatic Change 101*, 485–514.

Song W, **Jiang S**, **Camps-Valls G**, **Williams M**, **Zhang L**, **Reichstein M**, **Vereecken H**, **He L**, **Hu X and Shi L** (2024) Towards data-driven discovery of governing equations in geosciences. *Communications Earth & Environment 5*(1), 589.

Sweet L, **Müller C**, **Anand M and Zscheischler J** (2023) Cross-validation strategy impacts the performance and interpretation of machine learning models. *Artificial Intelligence for the Earth Systems 2*, e230026.

Tam FIH, **Beucler T and Ruppert Jr JH** (2024) Identifying three-dimensional radiative patterns associated with early tropical cyclone intensification, *Journal of Advances in Modeling Earth Systems 16*(12), e2024MS004401.

Wernli H, **Dirren S**, **Liniger MA and Zillig M** (2002) Dynamical aspects of the life cycle of the winter storm 'Lothar' (24–26 December 1999), *Quarterly Journal of the Royal Meteorological Society 128*, 405–429.

Xie XL and Beni G (1991) A validity measure for fuzzy clustering. *Transactions on Pattern Analysis and Machine Intelligence 13*(08), 841–847.

Cite this article: Tam FI-H, Augsburger F and Beucler T (2025). From winter storm thermodynamics to wind gust extremes: discovering interpretable equations from data. *Environmental Data Science*, 4: e48. doi:10.1017/eds.2025.10008

https://doi.org/10.1017/eds.2025.10008 Published online by Cambridge University Press