DETECTION OF METEOROLOGICAL VARIABLES IN A WIND FARM INFLUENCING THE EXTREME WIND SPEED BY HETEROGENEOUS GRANGER CAUSALITY

Kateřina Hlaváčková-Schindler
Faculty of Computer Science
University of Vienna
Währingerstrasse 29, Vienna, Austria
katerina.schindlerova@univie.ac.at

Irene Schicker
Geosphere Austria
Hohe Warte 38
Vienna, Austria
irene.schicker@geosphere.at

Kejsi Hoxhallari
Faculty of Computer Science
University of Vienna
Währingerstrasse 29, Vienna, Austria
a033521@unet.univie.ac.at

Claudia Plant
Faculty of Computer Science
University of Vienna
Währingerstrasse 29, Vienna, Austria
claudia.plant@univie.ac.at

ABSTRACT

For an efficiently managed wind farm and wind power generation under adverse weather, knowledge of meteorological parameters influencing wind speed is of crucial importance for optimized and improved forecasts. We investigate temporal effects of wind speed related processes such as wakes within the wind farm using the Heterogeneous Graphical Granger model. The ERA5 meteorological reanalysis was used to generate wind farm power production data in Eastern Austria. We evaluated six different scenarios for the hydrological half-year period, based on moderate wind speed and varying temporal intervals of low or high extreme wind speed as target variables. This splitting in scenarios allows us to carry out causal reasoning about possible causes of extreme wind speed in a wind farm. The discovered set of causal parameters for each of the scenarios provides information enabling future early warning as well as for taking management measures for wind farm power generation management under adverse weather conditions.

1 INTRODUCTION

Recent years have shown a rapid increase in renewable energy generation, especially with the energy crisis in the past years as well as with the aim of being fossil free in terms of energy generation by latest 2040. However, with climate change already affecting various parts of daily life including infrastructure and other sectors, the drivers and causal influence of weather to application sector such as wind energy need to be investigated. Here, machine learning methods have proven to be very useful. An important direction of recent ML research is on the interpretability and explainability of methods. Graphical causal methods possess both properties. Recently, [6] introduced a taxonomy of causal research questions in climatology. These are types of expert assumptions and properties of the available time series data to provide a causal language in which researchers can define their study questions. Graphical Granger causality and its non-linear modifications, due to its ability to express relationships among temporal data, has been playing an ongoing role in climatological research over decades [9], [1], [8], [6]. Focusing on the task of power generation in a wind farm and air flow within, the main question of this contribution was to determine, how the relevant meteorological parameters influence scenarios of a wind farm performance in Eastern Austria, under extremely strong, extreme low, and moderate wind scenarios and to compare the differences. For this we used the heterogeneous Granger causal model [5], allowing the determination of Granger-causal connections among processes which follow exponential distributions. These do fit better with the wind speed distriutions in some of the considered scenarios. The objective is to provide the wind farm operators with more inside knowledge and a better understanding of the conditions leading to

the extreme wind speed events and to take action to gain the most out of the prevailing conditions in combination with wind farm management and power generation. The most related work for causality among $p \geq 3$ non-Gaussian processes is the PCMCI method from [7] developed for causal inference in multivariate time series and applied to climatic data. To our best knowledge we are not aware of any non-Gaussian multivariate causal inference method applied to a wind farm.

2 METHOD

2.1 HETEROGENEOUS GRAPHICAL GRANGER MODEL

The Heterogeneous Graphical Granger model (HGGM), [2], is an extension of the original bivariate concept of causality for Gaussian processes defined by [3] to processes from exponential family. It considers $p \geq 3$ time series $\boldsymbol{x}_i$ which follow a distribution from the exponential family using a canonical parameter $\boldsymbol{\theta}_i$. The generic density form for each $\boldsymbol{x}_i$ can be written as $p(\boldsymbol{x}_i|\boldsymbol{X}_{t,d}^{Lag}, \boldsymbol{\theta}_i) = h(\boldsymbol{x}_i)\exp(\boldsymbol{x}_i\boldsymbol{\theta}_i - \eta_i(\boldsymbol{\theta}_i))$ where $\boldsymbol{X}_{t,d}^{Lag} = (x_1^{t-d}, .., x_1^{t-1}, .., x_p^{t-d}, .., x_p^{t-1})$, $\boldsymbol{\theta}_i = \boldsymbol{X}_{t,d}^{Lag}(\boldsymbol{\beta}_i^*)'$, with $\boldsymbol{\beta}_i^*$ being the optimum, and η_i a link function corresponding to time series $\boldsymbol{x}_i$. The HGGM applies the generalized linear models to time series in the following form

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

$\boldsymbol{\mu}_i$ denotes the mean of $\boldsymbol{x}_i$ and η_i^t is the t-th coordinate of $\boldsymbol{\eta}_i$. The causal inference in (1) can be solved as a maximum likelihood estimate for $\boldsymbol{\beta}_i$ for a given lag $d > 0$, $\lambda > 0$ (a regularization parameter), and all $t = d+1, \ldots, n$ with added adaptive lasso penalty function [2]. One can state that *time series $\boldsymbol{x}_j$ Granger–causes time series $\boldsymbol{x}_i$ for a given 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 j-th row of $\hat{\boldsymbol{\beta}}_i$ of the penalized solution is non-zero. ($\hat{\boldsymbol{\beta}}_i$ is an estimate of $\boldsymbol{\beta}_i$). Replacing p penalized linear regression problems by formulating (1) as a variable selection problem, [5] used the "minimum message length" (MML) to find causal connections in the HGGM. In this approach, the minimization of (1) with respect to β values is done by using the MML principle. The minimum message length principle chooses the model with the shortest message length as the best explanatory for the data [10]. For the equations and criterion for computing the causal values explicitly we refer to [5]. We use the MML criterion only for target variable wind speed (one i). As some climatological processes are better fitted by exponential distributions than by a Gaussian one, HMML can be beneficial to inference on our data set. As demonstrated in synthetic and real experiments in the same publication, HMML significantly improved the causal inference precision of those in [2] especially for short time-series. This is our case, where for wind speed and the wind related data, a short time series of 96 hour is relevant.

2.2 WIND FARM DATA AND HMML

Data used in this study are synthetically generated wind power production data plus accompanying meteorological parameters from the ERA5 data [4]. This data set was generated by Geosphere Austria for a wind farm located in Eastern Austria, consisting of 24 spatial and weather variables related to 38 individual wind turbine covering 21 years (2000-2020) measured hourly. The following variables are used: geopotential in m^2s^{-2} (z), boundary layer height measured in m (blh), dew point temperature at 2m in K (d2m), relative humidity in $\%$ (rel-h), wind speed at 135 m in ms^{-1} (wspeed135m), 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 (t135m), relative vorticity in $m^2s^{-1}kg^{-1}$ (vo). 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 feature as the idea was to look into not-so-obvious causal relationships. So one can see the pressure variable more as a confounding than causal variable to the target and other variables. The aim was to discover by HMML, which of the meteorological variables have a causal effect on wind speed at 135m, and consequently on power production. All time series were standardized. Due to the character of wind speed, only short time series are relevant, so the previous 96 hours including wind speed (the selected scenario) and the same length of time series for other variables. The HMML was used for each turbine separately as well as the following six scenarios (at 135m): for a winter and summer hydrological half-year we

consider low extreme wind, moderate wind and high extreme time periods. Our research objectives were: Firstly, we identified the intervals of extreme events and moderate wind periods for each turbine in the last 21 years (2000-2020) in the Andau wind farm. Then we applied HMML to find corresponding β values for each meteorological parameters indicating how strong (if any) is the causal relation of this parameter to wind speed in a concrete scenario and turbine. Further, to make the values of β reliable over a period of 21 years, we executed a statistical validation of these values by averaging the values over 21 years. Splitting experiments to six scenarios allows us to carry out causal reasoning about the causal values obtained by HMML in these scenarios. Concretely, we compared the β values of the variables over the scenarios, evaluated the differences among the β variables in different scenarios and interpreted the plausibility of the results. To make the results of HMML user-friendly for the wind power producers, we visualized the individual causal variables on each of the 38 turbines in the farm.

2.3 Scenarios and selection of their time intervals

As extreme events we consider a wind speed of $\geq 15\ m/s$ (for this value is a wind turbine producing rated power before being turned off at higher wind speeds) as well as a wind speed of $\leq 2\ m/s$ as minimum value of wind speed. Interval of $6\ m/s \leq ws \leq 8\ m/s$ is considered a moderate wind. For data pre-processing we first search for all of the occurrences of low/high wind or wind that falls within the 'moderate' category as denoted above. Then two suitable time intervals of 96 hours are selected for the three scenarios in all years, separately for the summer and winter half-years as illustrated in Table 1. The goal was to have at least 10% instances of extreme/moderate wind in each 96 hour intervals. Table 1 specifies wind-speed scenarios in time intervals in 2000.

half-years in 2000	Winter	Summer
High extreme wind	01-16 - 01-19	10-28 - 10-31
Low extreme wind	01-01 - 01-04	07-22 - 07-25
Moderate wind	12-02 - 12-06	06-16 - 06-20

Table 1: Scenarios and time intervals considered for year 2000

2.4 Best Fitting Exponential Distribution and maximum Lag of time series

The best fitting distribution of wind speed in each of the intervals is found by the residual sum of squares (RSS) and Kolmogorov-Smirnov test. The tested exponential distributions were Gaussian distribution, inverse Gaussian distribution and gamma distribution.

Different scenarios may have different best fitting distributions, as can be seen by the example in Figure 1 of wind speed w.r.t. time is shown for year 2007. Here, for both extreme scenarios, wind speed followed a Gaussian distribution whereas wind speed for the moderate wind scenario followed a gamma distribution. The maximum lag d in HMML was determined by the AIC information criterion. We created a number of $AR(d)$ autoregressive models with different d and chose that with minimal AIC value.

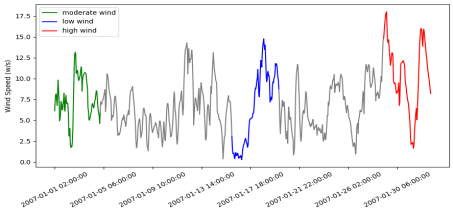

Figure 1: Wind speed in year 2007

2.5 Experiments and statistical validation of causal values

Our implementation in Python together together with implementation of HMML can be found under `https://anonymous.4open.science/r/Wind_farm_HMML--20BB/README.md` where also all figures of the wind farm for each scenario can be found. Method HMML was used with the genetic search algorithm `HMMLGA`. We run HMML on the six scenarios and provide a list of β values. The lists corresponds to the variables (11 variables in our case) which we consider to find causal connection to the target variable wind speed at 135m. Each of the lists contains d entries, where d corresponds to the lag value we have determined for each scenario. The lag d varied from 2 to 10 for various scenarios. After the β values for each scenario were found,

we computed β_i-proportionality $= \max_{k=1,\ldots,d}\{\beta_{ik}\}/(\sum_{j=1}^{11}(\max_{k=1,\ldots,d}\{\beta_{j,k}\})$ which can be understood as the relative strength of causality of i-th variable with respect to all 11 variables. To achieve a statistical validity of the causal values, for all scenarios, each turbine and each variable i we then calculate the corresponding arithmetical mean of β_i-proportionality over all 21 years.

3 Results and their interpretation

Each turbine in a scenario has been assigned a list of β_i proportionality values of the most significant causal variables for wind speed and a corresponding pie diagram. To visualize this information in a farm, we created 38 interactive pie charts for each season using matplotlib pie charts. The end output are six figures corresponding to the seasons, see an example for turbine 2 in Figure 2. As an interactive feature or our python code, when a computer mouse enters the pie chart, function `hover(event)` is triggered which shows a detailed pie chart with its respective beta values. When the mouse leaves the event, the function `leavePie(event)` is triggered, which sets the pie chart back to its original form. Different colors and a proportional size of chunk which corresponds to its beta value are used to discriminate against other variables. The visual outputs of the farm with the causal values for each scenario yield a lot of information to carry out causal reasoning. For example, one could pose a question, about spacial interactions, namely how the values of causal variables differ from one turbine to another one in a concrete scenario.

One can see from Figure 3 that the causal values in pies differ for turbines, e.g. cc (in pink) has different values for turbine 2 and 26. This can be inferred from the visual results. Additionally, the most influential variable is denoted at each turbine. Another question can be to compare how do the causal values of a particular turbine differ in various scenarios. For the objective of this work and due to the page limit we decided to compare average β_i-proportionality values for the whole farm of high extreme scenario to moderate wind speed scenario for summer hydrological year. The values of the three most significant variables for each scenario can be found in Table 2. Figure 3 illustrates the farm in high wind extreme scenario in summer (only a half of pies visualized for readability reasons), the moderate and low wind scenarios can be found in Appendix as well as the beta values for high wind (Table 3). All figures for all scenarios and turbines and tables of β_i-proportionality values are available under `https://anonymous.4open.science/r/Wind_farm_HMML--20BB/README.md`.

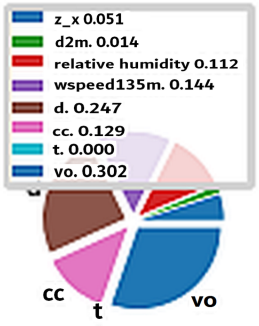

Figure 2: Interactive pie chart

Wind speed	Winter half-year	Summer half-year
High extreme wind	rh (0.184), vo (0.153), cc (0.127)	d (0.261), vo (0.231), rel-h (0.134)
Low extreme wind	cc (0.212), rel-h (0.206), d (0.154)	d (0.241), cc (0.212), rel-h (0.178)
Moderate wind	d (0.363), vo (0.216), rel-h (0.193)	cc (0.374), rel-h (0.183), d (0.141)

Table 2: Three most significant causal variables and their average values for ws135 for each scenario.

Table 2 indicates that mainly the changes in relative humidity (rel-h), in the winter season and the dew point (d) in summer combined with vorticity (vo) can be used as indicators beforehand for extreme/higher wind speed occurence whereas for the low wind scenarios the cloud cover (cc) is a higher indicator in winter and dew point for summer cases. In general, vorticity, cloud cover, and one of humidity indicators (dew point, relative humidity) are plausible indicators for extremes and cloud cover, dew point and relative humidity are can be used for below production wind speeds.

The variations within the wind farm can be explained, to a large extend, by wake effects caused by the turbines upfront the main wind direction and the position of the respective turbines within the wind farm. Additional effects include the land-use structure (e.g. lakes, agriculture, forest) and especially the build-area.

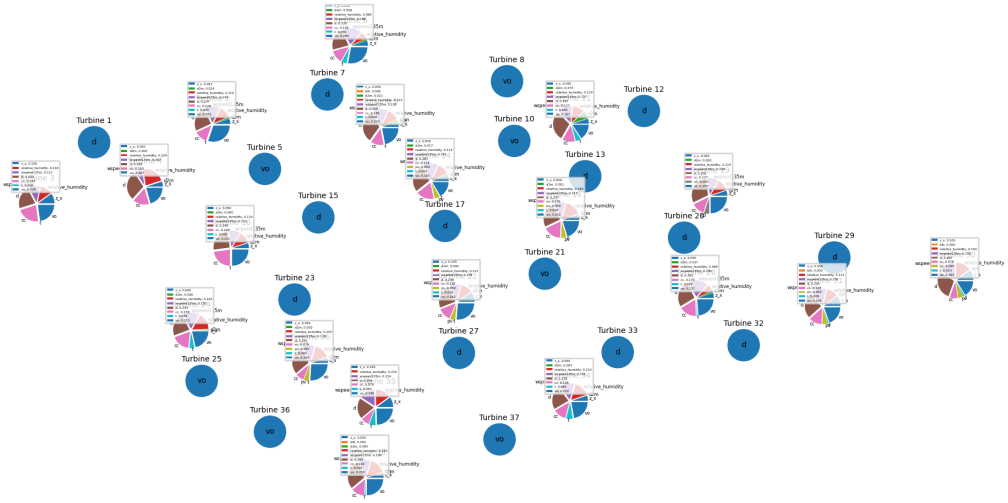

Figure 3: The farm in high wind extreme scenario in summer

4 Conclusions

In this work we demonstrate that HMML algorithm can detect causal meteorological variables for extreme wind speed using a wind farm in Eastern Austria as example. ERA5 reanalysis data was used to generate the synthetic wind farm data using real wind farm data as proxy. Results of the algorithm provide good indicators for the six selected scenarios, high, medium, and low wind in two seasons and can be used to advice wind farm operators. Using this knowledge combined with forecast planning of wind turbine non/production for those six scenarios can be made easier and early onwards. Furthermore, what-if scenarios for repowering of existing wind farms or planning of new wind farms can be calculated and used in the planning process. As next steps, the application to both operational weather forecasting for wind farm operators as well as to climate scenarios to detect future changes in operating conditions, is anticipated.

Acknowledgments

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

References

[1] Attanasio, A., Pasini, A., and Triacca, U. Granger causality analyses for climatic attribution. *Atmospheric and Climate Sciences 2013* (2013).

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

[3] Granger, C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society* (1969), 424–438.

[4] 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, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society 146*, 730 (2020), 1999–2049.

[5] HLAVÁČKOVÁ-SCHINDLER, K., AND PLANT, C. Heterogeneous graphical Granger causality by minimum message length. *Entropy 22*, 12 (2020), 1400.

[6] RUNGE, J., GERHARDUS, A., VARANDO, G., EYRING, V., AND CAMPS-VALLS, G. Causal inference for time series. *Nature Reviews Earth & Environment* (2023), 1–19.

[7] RUNGE, J., NOWACK, P., KRETSCHMER, M., FLAXMAN, S., AND SEJDINOVIC, D. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances 5*, 11 (2019), eaau4996.

[8] RUNGE, J., PETOUKHOV, V., DONGES, J. F., HLINKA, J., JAJCAY, N., VEJMELKA, M., HARTMAN, D., MARWAN, N., PALUŠ, M., AND KURTHS, J. Identifying causal gateways and mediators in complex spatio-temporal systems. *Nature Communications 6*, 1 (2015), 8502.

[9] SMIRNOV, D. A., AND MOKHOV, I. I. From Granger causality to long-term causality: Application to climatic data. *Physical Review E 80*, 1 (2009), 016208.

[10] WALLACE, C. S., AND BOULTON, D. M. An information measure for classification. *The Computer Journal 11*, 2 (1968), 185–194.

A APPENDIX

Table 3 contains the β_i- proportionality values for each variable under the high wind speed scenario in summer, corresponding to Figure 3.

Table 3: β_i-proportionality values for scenario winter half-year, high wind speed

β_i proportionalities of all variables											
Turbines	z	bhl	d2m	rel-h	ws135m	d	cc	o3	pv	t2m	vo
T0	0.051	0.000	0.010	0.162	0.148	0.230	0.129	0.000	0.050	0.000	0.220
T1	0.051	0.000	0.014	0.112	0.144	0.247	0.129	0.000	0.000	0.000	0.302
T2	0.100	0.000	0.000	0.116	0.112	0.222	0.223	0.000	0.000	0.000	0.228
T3	0.050	0.000	0.000	0.224	0.067	0.293	0.160	0.000	0.000	0.000	0.207
T4	0.050	0.000	0.000	0.212	0.143	0.216	0.128	0.000	0.000	0.016	0.234
T5	0.050	0.000	0.018	0.066	0.188	0.220	0.128	0.000	0.000	0.050	0.280
T6	0.051	0.000	0.020	0.113	0.138	0.318	0.129	0.000	0.050	0.000	0.181
T7	0.100	0.000	0.023	0.113	0.129	0.221	0.127	0.000	0.000	0.007	0.279
T8	0.050	0.000	0.021	0.113	0.138	0.329	0.129	0.000	0.000	0.000	0.220
T9	0.050	0.000	0.073	0.113	0.136	0.237	0.127	0.000	0.000	0.000	0.263
T10	0.050	0.000	0.073	0.114	0.137	0.293	0.127	0.000	0.000	0.050	0.157
T11	0.150	0.000	0.000	0.114	0.109	0.332	0.127	0.000	0.000	0.000	0.168
T12	0.100	0.000	0.020	0.114	0.139	0.292	0.127	0.000	0.050	0.000	0.158
T13	0.050	0.000	0.017	0.113	0.142	0.283	0.178	0.000	0.050	0.001	0.165
T14	0.050	0.000	0.000	0.120	0.109	0.277	0.171	0.000	0.050	0.000	0.223
T15	0.050	0.000	0.000	0.114	0.110	0.248	0.226	0.000	0.000	0.000	0.251
T16	0.100	0.000	0.015	0.113	0.194	0.249	0.128	0.000	0.000	0.000	0.200
T17	0.050	0.000	0.001	0.160	0.113	0.247	0.176	0.000	0.050	0.000	0.203
T18	0.050	0.000	0.000	0.114	0.109	0.300	0.127	0.050	0.000	0.000	0.250
T19	0.000	0.000	0.000	0.158	0.065	0.312	0.128	0.050	0.050	0.000	0.239
T20	0.050	0.000	0.000	0.113	0.109	0.293	0.128	0.000	0.000	0.000	0.307
T21	0.100	0.000	0.000	0.112	0.109	0.238	0.128	0.000	0.050	0.000	0.262
T22	0.050	0.000	0.006	0.162	0.153	0.235	0.129	0.000	0.000	0.050	0.214
T23	0.000	0.000	0.006	0.163	0.153	0.234	0.178	0.000	0.000	0.050	0.215
T24	0.050	0.000	0.000	0.202	0.118	0.226	0.080	0.000	0.000	0.000	0.324
T25	0.050	0.000	0.000	0.153	0.118	0.291	0.079	0.000	0.050	0.000	0.259
T26	0.050	0.000	0.000	0.112	0.109	0.291	0.129	0.000	0.050	0.050	0.208
T27	0.050	0.000	0.017	0.066	0.190	0.263	0.170	0.000	0.000	0.007	0.237
T28	0.050	0.000	0.014	0.113	0.145	0.254	0.128	0.000	0.050	0.000	0.246
T29	0.100	0.000	0.000	0.154	0.155	0.287	0.078	0.000	0.050	0.014	0.163
T30	0.100	0.000	0.000	0.113	0.150	0.255	0.128	0.000	0.050	0.009	0.195
T31	0.001	0.000	0.012	0.119	0.196	0.227	0.171	0.000	0.050	0.001	0.223
T32	0.050	0.000	0.050	0.117	0.109	0.283	0.174	0.000	0.000	0.000	0.217
T33	0.050	0.000	0.003	0.153	0.151	0.228	0.128	0.000	0.000	0.065	0.222
T34	0.100	0.000	0.000	0.153	0.154	0.204	0.079	0.000	0.000	0.064	0.246
T35	0.100	0.000	0.002	0.153	0.166	0.225	0.080	0.000	0.000	0.000	0.274
T36	0.050	0.000	0.001	0.203	0.167	0.230	0.080	0.000	0.000	0.000	0.269
T37	0.050	0.000	0.000	0.153	0.109	0.299	0.129	0.000	0.000	0.010	0.250

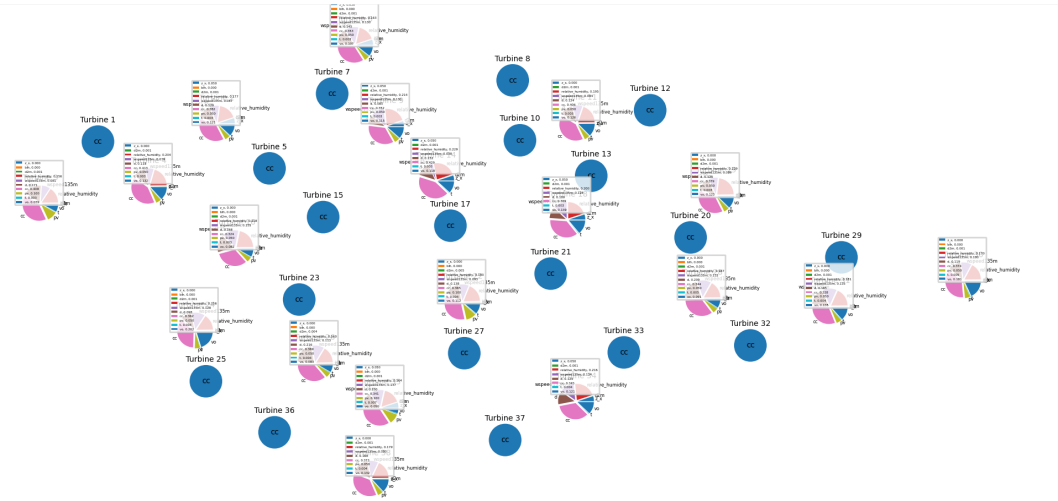

Figure 4: The farm in moderate wind speed scenario in summer

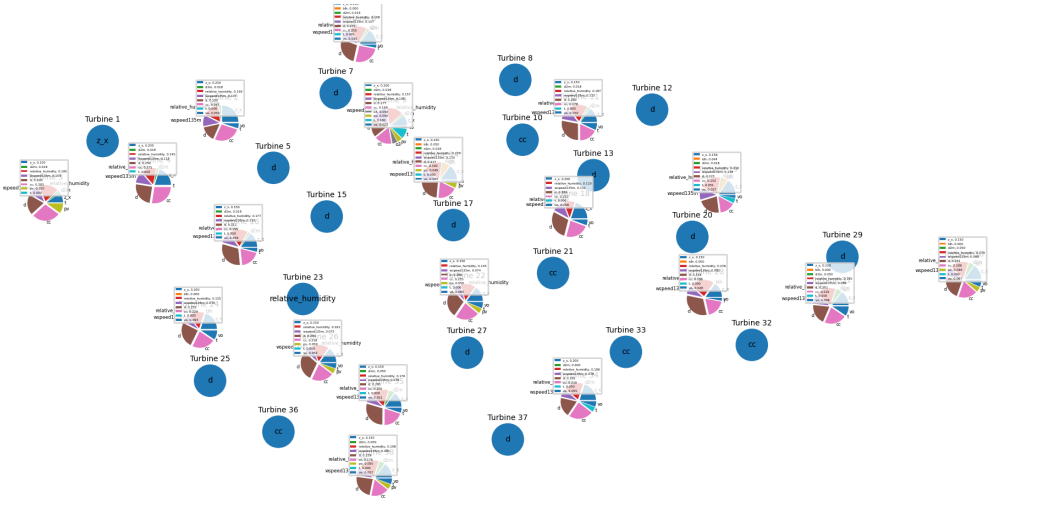

Figure 5: The farm in low wind extreme scenario in summer