Environmental Data Science (2020), 4: 1–1
doi:10.1017/eds.2020.xx

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Clustering of causal graphs to explore drivers of river discharge

Wiebke Günther[1*], Peter Miersch[2], Urmi Ninad[3] and Jakob Runge[1,3]

[1]German Aerospace Center, Institute of Data Science, Jena, Germany
[2]Helmholtz Centre for Environmental Research, Department of Computational Hydrosystems, Leipzig, Germany
[3]Technische Universität Berlin, Faculty of Electrical Engineering and Computer Science, Berlin, Germany
*Corresponding author. Email: wiebke.guenther@dlr.de

Received: 01 February 2023; **Revised:** ; **Accepted:**

Keywords: causal inference; causal effect estimation; clustering; catchment hydrology

Abstract

This work aims to classify catchments through the lens of causal inference and cluster analysis. In particular, it uses causal effects (CE) of meteorological variables on river discharge while only relying on easily obtainable observational data. The proposed method combines time series causal discovery with CE estimation to develop features for a subsequent clustering step. Several ways to customize and adapt the features to the problem at hand are discussed. In an application example, the method is evaluated on 358 European river catchments. The found clusters are analyzed using the causal mechanisms that drive them and their environmental attributes.

Impact Statement

This paper discusses how one can classify river catchments based on causal effects between temperature, precipitation and discharge, i.e. the volume of water that leaves the given area over a certain time. The proposed method is applied to 358 European catchments.

1. Motivation

In hydrological research, basic classification of catchments, that is an area where all water drains to a single outlet, is done through analyzing the response of discharge to precipitation input. Here, discharge refers to the volume of water flowing through a river channel per time unit (Turnipseed and Sauer 2010). However, discharge characteristics are highly heterogeneous, as they depend on catchment characteristics like area, slope, and land cover. Furthermore, catchment behaviour is also driven by regional climate and the interaction of hydrometeorological processes, for instance, snow melt, soil moisture, and precipitation events. Historically, such classification or clustering of catchments has been done using a carefully selected subset of their attributes (Wagener et al. 2007). However, many of these attributes are correlated, making it difficult to select a minimal predictive set. Hydrological signatures have also been used as a basis for clustering, leading to clusters that are partially close to ones based on climate behaviour (Kuentz et al. 2017; Jehn et al. 2020).

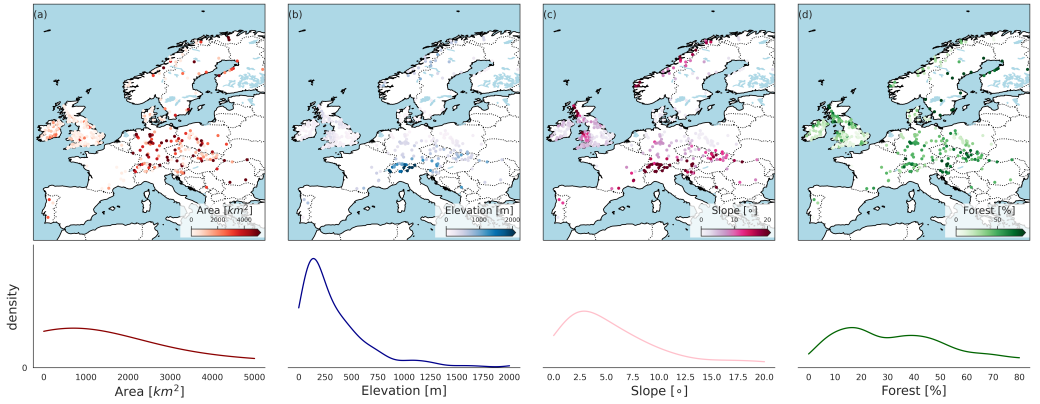

Figure 1. *Geographic overview and distribution of European catchments characteristics. (a) Area, (b) average elevation, (c) average slope, (d) forest cover.*.

In recent years, the importance of data driven analysis has been recognized (Peters-Lidard et al. 2017) and catchment classification has also been done using machine learning techniques (Jiang, Bevacqua, and Zscheischler 2022).

However, we note that the tools of causal inference seem to be under-explored to derive relationships between meteorological variables that can serve as a foundation for classifying river catchments. Causal inference algorithms allow, under specific assumptions, to discover and quantify causal relationships from observational data. Moreover, their outputs are inherently explainable by design. This makes them especially suitable for the domain of Earth sciences, since here it is often infeasible to conduct controlled experiments to arrive at causal conclusions (Runge, Bathiany, et al. 2019; Samarasinghe, Deng, and Ebert-Uphoff 2020; Gnecco et al. 2019).

In our analysis, we investigate the impact of temperature and precipitation on observed discharge in European catchments. See figure 1 for an overview of the considered catchments and their environmental characteristics. We assume linear relationships between the variables, and employ the PCMCI framework by Runge, Nowack, et al. 2019, in combination with expert knowledge, to identify causal relationships. Based on the found causal graphs, we also quantify the causal effect between the variables based on the path method (Wright 1934) and Pearl's causal framework (Pearl et al. 2000) following Runge, Petoukhov, et al. 2015. Subsequently, we use the estimated causal effects as features for clustering using the k-means algorithm (Lloyd 1982).

In doing so, we show how methods from causal inference can improve our understanding of discharge generating mechanisms.

2. Method

We want to explore differences in the causal structure of discharge and its drivers across Europe. Our method relies on observations from multiple data sets that have been collected at different locations, i.e. on data that is heterogeneous with respect to the environment by which it is influenced. In our application setting, the considered data sets correspond to the catchments, where we observed temperature, precipitation and discharge. Within each of the M data sets (i.e. catchments), we have N observational time series, denoted by the vector $\mathbf{X}^m$ where m is the data set index, for variables that are the same across data sets. An observation of variable i at time point t within data set m is then denoted by $X_t^{i,m}$. To ease notation, we suppress the data set index in the following. Our analysis comprises three main steps, i.e.

1. Estimate a causal graph for each data set (i.e. catchment) separately using the PCMCI-algorithm.
2. Extract features of the causal graphs that will be used in the subsequent clustering step. Here, one is rather free in the choice of features, and we will discuss a few different options below. In this work, we focus on features based on linear CE estimation. This step can also be seen as selecting a mapping from the space of graphs into $V \subset \mathbf{R}^d$ with d the dimension of the feature space. For instance, d could be the number of lagged CEs between specific variables X^i and X^j. In this space V we can use the Euclidean distance for clustering in the next step.
3. Finally, employ the k-means clustering algorithm to find graphs (that correspond to catchments) with similar causal patterns.

2.1. Causal discovery

Interdependencies within a system of variables can be conveniently represented by graphical models, where the variables are represented by the nodes of the graph and edges indicate dependence between the respective nodes. Of specific interest to us are causal graphs, where an edge indicates a causal relationship between two nodes. Here, we consider them to be directed acyclic graphs.

To be able to learn such causal graphs from observational data alone, we have to impose some assumptions. In particular, we assume that the causal Markov condition, that is, the independence of error terms driving each subprocess holds, as well as faithfulness which ensures that the graph includes all conditional independencies that hold in the true underlying process. We also assume that all relevant variables are included in the model. This is known as causal sufficiency. We also make the stationarity assumption, i.e. that the causal relationships do not change over time. Furthermore, we assume that there are no contemporaneous effects.

In this setting, we apply the PCMCI-algorithm to learn the time series graph within each data set. PCMCI is a two-stage causal discovery algorithm based on the framework of constraint-based causal discovery (Spirtes et al. 2000) that is suitable for time series. Its first stage, called PC_1 is based on the PC-stable algorithm that discovers relevant conditions for each of the N variables by iterative independence testing. These conditions are a superset of the true causal parents with high probability. In stage two, the momentary conditional independence (MCI) test uses the conditions found in stage one to infer a causal link $X^i_{t-\tau} \to X^j_t$, i.e. it tests $X^i_{t-\tau} \perp Y_t | pa(Y_t), pa(X^i_{t-\tau})$ where $pa(X^i_t)$ denotes the parent superset of X^i_t found in the first step. Conditioning on the parents of the target $pa(X^i_t)$ increases the effect size, and conditioning on $pa(X^i_{t-\tau})$ helps to control for false-positives in the highly autocorrelated time series case. For further details on PCMCI, refer to (Runge, Nowack, et al. 2019).

2.2. Feature extraction based on causal effect estimation

Now, we will give further detail on the feature extraction procedure within step 2. The CE of setting $X^i_{t-\tau}$ to x^* on X^j_t is given by $\Psi_{ji}(\tau) := \frac{\partial}{\partial_{x^*}} \mathbf{E}[X^j_t \,|\, \mathrm{do}(X^i_{t-\tau}) = x^*]$. Note that the do-operator refers to a hard intervention on the system of forcing the value of $X^i_{t-\tau}$ to be x^*. Following Runge, Petoukhov, et al. 2015, to estimate the CE from observational data, we fit a linear model of the following form, where we only estimate the coefficients Φ_{ij} that correspond to links in our causal graph (see step 1)

$$X^j_t = \sum_{i=0}^{N-1} \sum_{\tau=1}^{\tau_{\max}} \Phi_{ji}(\tau) X^i_{t-\tau} + \varepsilon_t \quad \text{with } \Phi_{ji}(\tau) \neq 0 \text{ only if } X^i_{t-\tau} \text{ is a parent of } X^j_t. \tag{1}$$

One straightforward way of using this approximation of the causal process for clustering, is to directly consider the vector of path coefficients $(\mathbf{\Phi}_{ij}(\tau))_{\tau=1,\ldots,\tau_{\max}, i\in V, j\in W}$ for subsets of variables $V, W \subset \mathbf{X}$ as features. The path coefficient $\mathbf{\Phi}_{ji}(\tau)$ is a standardized version of the corresponding $\Phi_{ji}(\tau)$ in (1) and represents the change in the expectation of X^j_t (in units of its standard deviation) induced by raising $X^i_{t-\tau}$ by 1 standard deviation, while keeping all other parents constant. It therefore quantifies the direct

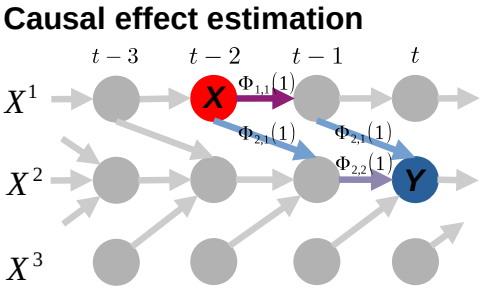

Figure 2. *Illustration of total causal effect estimation in a time series graph for* $\tau = 2$. *The CE* $\Psi_{ji}(2)$ *between* $X^1_{t-2} =: X$ *and* $X^2_t =: Y$ *is computed by summing over the products of path coefficients (link labels) along each path, i.e.* $\Psi_{21}(2) = \Phi_{1,1}(1)\Phi_{2,1}(1) + \Phi_{2,1}(1)\Phi_{2,2}(1)$.

effect of $X^i_{t-\tau}$ on X^j_t. One potential problem with this approach could be that the frequently (for every not directly linked pair of nodes) appearing zero-values could dominate or skew the clustering results. Graphs with the same non-parental relationships would be in the same cluster even though they exhibit very different behaviour in the present links.

Therefore, it might be preferable to use the lagged CEs of the variables in V onto the variables in W instead. The matrix $\mathbf{\Psi}(\tau)$ of the standardized CEs between all variables for time lag τ can be iteratively computed based on the standardized path coefficients $\mathbf{\Phi}(\tau)$, as the entry $\Psi_{ji}(\tau)$ corresponds to the sum over the products of path coefficients along all path between $X^i_{t-\tau}$ and X^j_t, see also figure 2. If we want to consider a large maximal time lag $\tau_{\max}$, using $(\mathbf{\Psi}(\tau))_{\tau=1,\dots,\tau_{\max}}$ will give us a high-dimensional feature space.

However, we can reduce its dimension by using aggregated measures. Here, there are various ways in which aggregation is possible. The lagged CEs could be aggregated over the number of variables, or by aggregating in the direction of the lags, e.g. by averaging

$$\frac{1}{\tau_{\max}} \sum_{\tau=1}^{\tau_{\max}} \mathbf{\Psi}(\tau),$$

or both, as is done in calculating the average causal effect (ACE) or average causal susceptibility (ACS). Please refer to Runge, Petoukhov, et al. 2015 for the formulas. Note that many different ways of aggregation besides averaging are possible and have to be carefully evaluated within each application context. Additional to averaging, we have included results for maximal lagged CEs, i.e. using the features $(max_\tau(\mathbf{\Psi_{ij}}(\tau)))_{i,j=1,\dots,N}$, in the figure 3.

3. Application

Now, we apply our method to the problem of classifying European catchments in terms of their causal interactions between temperature, precipitation and discharge.

3.1. Data and setup

In this study, we consider 358 gauged catchments across Europe selected based on data availability of daily observational discharge, meteorological observations, watershed boundaries, and morphological catchment characteristics for the study period from 1950 to 2021. Daily observational discharge and watershed boundaries were used from the Global Runoff Data Centre (GRDC) dataset

(https://www.bafg.de/GRDC, accessed: 21. July 2022). Using the observational 0.1° daily gridded E-OBS dataset (version 26e, Cornes et al. 2018), catchment averaged precipitation and mean surface temperature time series were derived using area-weighted averages of the gird cells within the catchments' boundary. For analysis of the clusters, catchment averaged slope and altitude were derived from the USGS digital elevation model (Earth Resources Observation And Science (EROS) Center 2017), land cover (impervious, forest, pervious) from the European Space Agency GlobCover (Arino et al. 2012). Deriving the morphological variables from gridded data products was done to allow for comparison to future studies including process based hydrological models, therefore we restricted our study to catchments where this data is available. We furthermore limit our study to catchments with a minimum of 30 years of continuous discharge records within the study period to ensure sufficient data for our causal inference approach. The catchment areas range between 16 and $10,000\ km^2$ - larger catchments, with increasing importance of spatial heterogeneity on discharge generation, were omitted. Overall, the selected catchments encompass a large variety of locations, areas, average altitudes and morphologies, with a cumulation in Great Britain and Ireland due to the availability of watershed boundaries in those regions (fig. 1).

The implementation of our method is based on the Tigramite-package (https://github.com/jakobrunge/tigramite/tree/master). For the causal discovery step (step 1 above), we use the PCMCI-algorithm. We assume a linear relationship between the variables and therefore use the partial correlation conditional independence test with a 0.05 significance level in the PCMCI-algorithm. We also provide the general knowledge that discharge *cannot* cause either temperature nor precipitation to the causal discovery method. We look for causal relationships up to 30 time steps in the past, i.e. $\tau_{\max} = 30$. Any time slices of samples with missing values are discarded. For step 2, we use the functionality provided in the LinearMediation class of Tigramite. Within step 3, we use the scipy-implementation of the k-means algorithm with 4 cluster centers. Results for 3 and 5 cluster centers are provided in the supplement. The choice of this hyperparameter was based on a combination of the Elbow method and the silhouette score (Kaufman and Rousseeuw 2009). The elbow method corresponds to finding the point after which the sudden drop in average distance from the centroid slows down, see Teoh and Rong 2022 for details. The silhouette score provides a more quantitative measure of how similar and separated the found clusters are. We plotted the average distance from the centroid, the silhouette score, and its average, respectively, over the number of cluster centers. These plots can be found in the supplement. The results vary slightly over the different features that we used as a basis for the clustering. However, generally speaking, 4 cluster centers seems to be a good choice that also leads to clusters that are informative and of roughly the same size.

3.2. Results

As one would expect, we find slightly different clusters depending on the causal aspect of the system we are focusing on during the feature extraction phase of our method. We present results for a few different feature choices in figure 3, more can be found in the supplement in figure 10. However, some metrics also yield very similar results. We find only minor variation between the clusters for ACE and ACS. Only considering the CE of precipitation on discharge gives us very similar clusters to considering the relationships between all variables. Moreover, clusters based on joint CE are almost identical to only considering the maximal CE over all lags for each variable pair.

Figure 10 in the supplement also illustrates the results for an increasing number of clusters. We observe that the clusters tend to become regionally very scattered as the number of clusters grows. This effect is especially apparent in middle and Western Europe. This is possibly due to the high spatial variability in the European climate and topography.

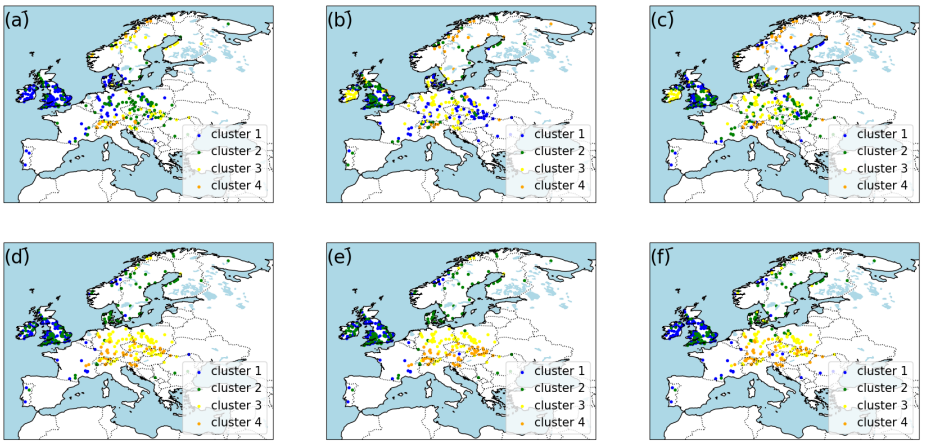

Figure 3. *Cluster results for different choices of features. For a brief explanation of these, and other, feature extraction methods see section 2.2. In the following X^1 denotes temperature, X^2 precipitation and X^3 discharge. The features correspond to (a) average lagged CE of X^1 on X^3, (b) average lagged CE of X^2 on X^3, (c) average lagged CE of X^i on X^j for all $i \neq j$, (d) maximal lagged CE of X^i on X^j for all $i \neq j$, (e) path coefficients, (f) ACE of X^i on the system for all i.*

3.3. Distinct causal behaviour

The time series graphs that are discovered in step 1 of our method do not differ much across catchments in terms of skeleton, i.e. the graph without orientations, and link direction, see figure 9 in the supplement. This is to be expected, since we provided the causal discovery algorithm with substantial expert knowledge. Therefore, we get more regionally varying clusters if we take the variations in strength of the links, or, in other words, of the value of the path coefficients, into account. This behaviour is visible in figure 10 in the supplement. In this section, we want to investigate the features that we used for clustering further by having a closer look at the distribution of each component of the feature vectors within each cluster. This will allow us to associate the observed clusters with a specific causal pattern. As an example, we will focus on the average joint CE and some variants of it.

In figure 4, we observe that for one dimensional feature-spaces, we can directly see that the clusters correspond to a specific range of the feature values. This is shown in the right column of figure 4 for the one-dimensional features average lagged CE of temperature on discharge, and average lagged effect of precipitation on discharge, respectively. For instance, in the yellow cluster, we see a strong average CE of temperature on discharge. As can be expected, this cluster has the lowest CE of precipitation on discharge. Interestingly, we find that this cluster is also associated with a specific European region, see figure 3.

Now, we look at similar plots for the case where we used the four dimensional feature vector of the average joint CE of temperature on precipitation, temperature on discharge, precipitation on temperature, and precipitation on discharge for clustering (left and middle column in figure 4). Here, we see that the fourth component almost exactly separates according to the clusters. We suspect that it is dominating the clustering, since the CEs between precipitation and temperature are very low in comparison. More plots can be found in the supplement.

3.4. Distribution of catchment attributes within clusters

We are also interested in analysing the distribution of environmental variables within each of the found clusters. We illustrate this using boxplots in figure 5. Also, note that the environmental attributes tend to be strongly correlated. For instance, steeper slope generally corresponds to higher altitude. In figures

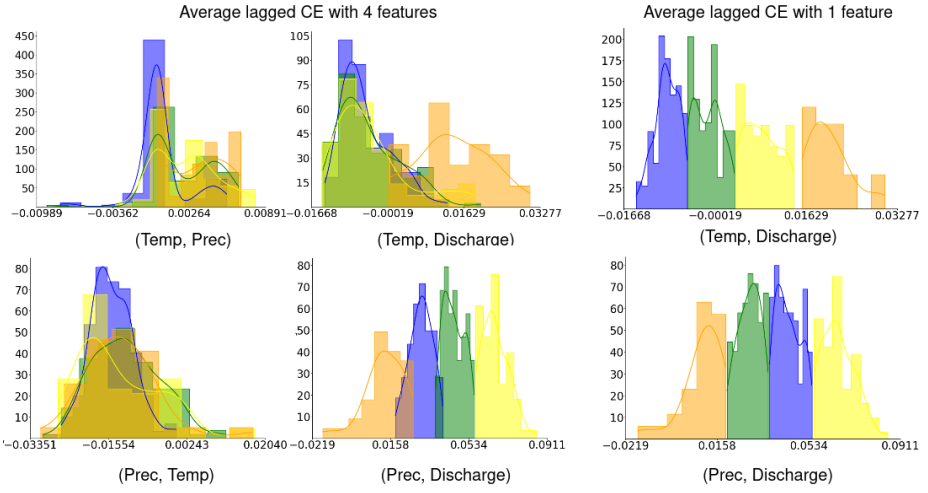

Figure 4. *Histogram of one-dimensional CE-based features (right column) in comparison to clustering based on feature vectors that include the average joint CE of temperature on precipitation, temperature on discharge, precipitation on temperature, and precipitation on discharge (left and middle column). For instance, the plot in the right upper corner shows the frequency of values of the average lagged CE of temperature on discharge within the different clusters. One can see that in the catchments of cluster 1 (blue) the average lagged CE of temperature on discharge is relatively low in comparison to the other clusters. Colors correspond to the clusters illustrated in figure 3.*

3 and 5, general patterns are also visible across all different choices of features, e.g. strong differences in mean altitude between one or two of the found clusters and the remaining ones. We can summarize the differences and similarities with respect to the attributes in the following way:

- 1. cluster: low altitude, small catchments, western/middle Europe
- 2. cluster: low to medium altitude, small, western/middle Europe
- 3. cluster: low to medium altitude, larger area, western/middle Europe
- 4. cluster: high altitude, with larger area: seems to be alps, Scandinavian mountains

Some of the clusters seem to be strongly related to a specific region, e.g. in terms of effect of precipitation on discharge we observe a very clear regional cluster in Ireland. However, it is hard to interpret without further domain knowledge. Moreover, the results could be skewed by the choice of catchments. A large proportion of the catchments are located in Great Britain and Ireland since here watershed boundaries are readily available. This was a limiting factor in our catchment selection because we wanted to be able to infer the catchment size. How this problem can be alleviated in future work is further discussed in the next section.

4. Discussion

In principle, our method is applicable in any situation where there are multiple data sets containing observations of the same variables. This makes our method applicable to a wide range of problems and research areas, not even limited to the domain of Earth science.

Of course, in practice the availability and quality of data are limiting factors. Therefore, an analogous analysis can be conducted in regions with good public data availability, like North America, whereas it would be more challenging in regions with a lower converage by weather and gauging stations.

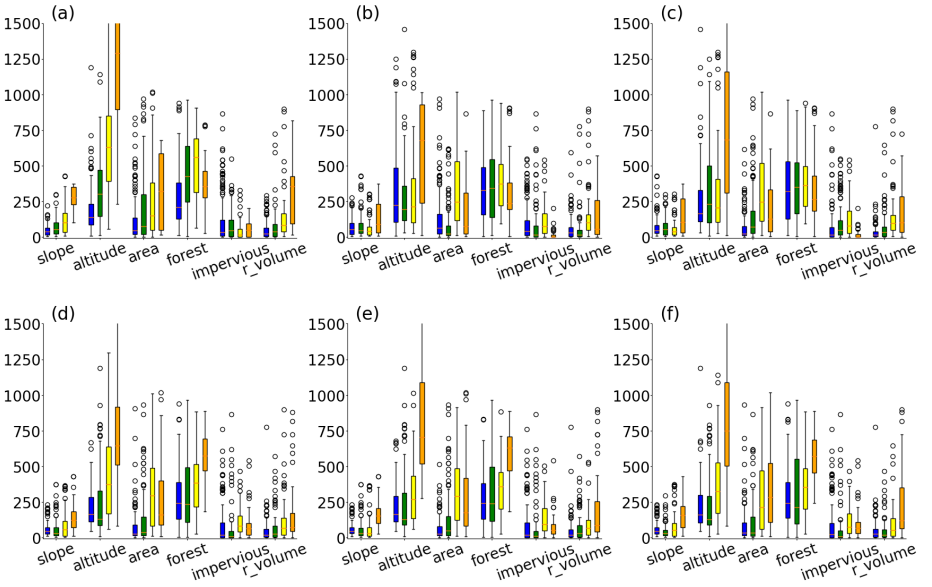

Figure 5. *Distribution of catchment attributes within each cluster. The attributes are slope in degrees times* 10*, average altitude in m above sea level, area in* km^2*, proportion of basin covered by forest, proportion of basin covered by impervious surfaces, mean annual rain volume in* $100km^3$*. Colors correspond to the clusters illustrated in figure 3. In the following* X^1 *denotes temperature,* X^2 *precipitation and* X^3 *discharge. The features correspond to (a) average lagged CE of* X^1 *on* X^3*, (b) average lagged CE of* X^2 *on* X^3*, (c) average lagged CE of* X^i *on* X^j *for all* $i \neq j$*, (d) maximal lagged CE of* X^i *on* X^j *for all* $i \neq j$*, (e) path coefficients, (f) ACE of* X^i *on the system for all* i*.*

Also, construction of geographically averaged timeseries and associated environmental attributes might not be as clean cut as in our application, where the catchment boundaries provide a natural distinction between different data regions. In other questions relevant to the field of Earth Science, the regions to average over might be more arbitrary. This essentially introduces more hyperparameters into the problem. Another potential problem is an insufficient amount of data. For instance, if one is interested in extreme events, like the flood generating process instead of the discharge generating process, then there are probably too few events to obtain stable causal discovery and clustering results.

There are still many avenues open for future work. The main next step that we want to take is to only focus on peak discharge events and to see how the drivers of extreme events differ from the baseline behaviour of normal fluctuations in discharge. Another major point that we want to explore further is the influence of climate change on the system. So far we have assumed stationarity of the time series but this assumption is violated in a warming climate.

Furthermore, the effect of including more variables into our model has to be investigated. In other words, effects of the possible violation of the underlying causal sufficiency assumption within the PCMCI-algorithm has to be explored.

Also, further analysis has to be done to distinguish the clusters more, especially cluster 1 and 2. In particular, the imbalance in the dataset of most clusters being located in Great Britain or Ireland has to be addressed. The reason for this is that here a lot of clusters with watershed boundaries are available. However, following the approach presented in Xie et al. 2022, it is possible to infer the watershed boundaries for many more clusters across Europe making them suitable for our analysis. Moreover, the European climate has a high spatial variation that depends on far more factors than have been analysed by us so far.

5. Conclusion

In this work, we presented how tools from causal discovery and effect estimation can be combined with clustering techniques. The presented approach is very customizable and thus suitable for a wide variety of problems and domains. In an application example, we have explored how causal drivers of discharge, specifically the strengths of their causal effects, differ across Europe. We saw how the causal inference methods allow us to find clusters of catchments that can guide domain experts in developing and evaluating hypothesis based on observational data, and how these clusters are affected by different choices in designing the features.

Acknowledgments. We thank Jakob Zscheischler and Shijie Jiang for their helpful comments.

Funding Statement. W.G. and P.M. were supported by the Helmholtz AI project *CausalFlood*. U.N. and J.R. were supported by grant no. 948112 CausalEarth of the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program.

Competing Interests. The authors declare no competing interests exist.

Data Availability Statement. Replication data can be found in the Global Runoff Data Centre (GRDC) data set: https://www.bafg.de/GRDC/EN/Home/homepage_node.html.

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

Author Contributions. Conceptualization: W.G.; P.M.; J.R. Methodology: W.G.; P.M.; J.R. Data curation: P.M. Data visualisation: W.G.; P.M. Writing original draft: W.G.; P.M. Writing review and editing: W.G.; P.M.; U.N.; J.R. All authors approved the final submitted draft.

Supplementary Material. Supplementary material has been provided with this submission.

References

Olivier Arino, Jose Julio Ramos Perez, Vasileios Kalogirou, Sophie Bontemps, Pierre Defourny, and Eric Van Bogaert (2012). *Global land cover map for 2009 (GlobCover 2009)*.

Richard Cornes, G. Schrier, Else Van den Besselaar, and P. Jones (Aug. 2018). "An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets". In: *Journal of Geophysical Research Atmospheres* 123. DOI: 10.1029/2017JD028200.

Earth Resources Observation And Science (EROS) Center (2017). *Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010)*. DOI: 10.5066/F7J38R2N. URL: https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-global-multi-resolution-terrain-elevation.

Nicola Gnecco, Nicolai Meinshausen, J. Peters, and Sebastian Engelke (2019). "Causal discovery in heavy-tailed models". In: *The Annals of Statistics*.

Florian U Jehn, Konrad Bestian, Lutz Breuer, Philipp Kraft, and Tobias Houska (2020). "Using hydrological and climatic catchment clusters to explore drivers of catchment behavior". In: *Hydrology and Earth System Sciences* 24.3, pp. 1081–1100.

S. Jiang, E. Bevacqua, and J. Zscheischler (2022). "River flooding mechanisms and their changes in Europe revealed by explainable machine learning". In: *Hydrology and Earth System Sciences* 26.24, pp. 6339–6359. DOI: 10.5194/hess-26-6339-2022. URL: https://hess.copernicus.org/articles/26/6339/2022/.

Leonard Kaufman and Peter J Rousseeuw (2009). *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons.

Anna Kuentz, Berit Arheimer, Yeshewatesfa Hundecha, and Thorsten Wagener (2017). "Understanding hydrologic variability across Europe through catchment classification". In: *Hydrology and Earth System Sciences* 21.6, pp. 2863–2879.

Stuart Lloyd (1982). "Least squares quantization in PCM". In: *IEEE transactions on information theory* 28.2, pp. 129–137.

Judea Pearl et al. (2000). "Models, reasoning and inference". In: *Cambridge, UK: CambridgeUniversityPress* 19.2.

Christa D. Peters-Lidard, Martyn Clark, Luis Samaniego, Niko E. C. Verhoest, Tim van Emmerik, Remko Uijlenhoet, Kevin Achieng, Trenton E. Franz, and Ross Woods (July 20, 2017). "Scaling, Similarity, and the Fourth Paradigm for Hydrology". In: *Hydrology and Earth System Sciences* 21.7, pp. 3701–3713. ISSN: 1027-5606. DOI: 10.5194/hess-21-3701-2017. URL: https://hess.copernicus.org/articles/21/3701/2017/ (visited on 01/27/2023).

Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel Mahecha, Jordi Muñoz, Egbert Nes, Jonas Peters, Rick Quax, Markus Reichstein, Marten Scheffer, Bernhard Schölkopf, Peter Spirtes, Jie Sun, and Jakob Zscheischler (Dec. 2019). "Inferring causation from time series in Earth system sciences". In: *Nature Communications* 10. DOI: 10.1038/s41467-019-10105-3.

Jakob Runge, Peer Nowack, Marlene Kretschmer, Seth Flaxman, and Dino Sejdinovic (2019). "Detecting and quantifying causal associations in large nonlinear time series datasets". In: *Science advances* 5.11, eaau4996.

Jakob Runge, Vladimir Petoukhov, Jonathan F Donges, Jaroslav Hlinka, Nikola Jajcay, Martin Vejmelka, David Hartman, Norbert Marwan, Milan Paluš, and Jürgen Kurths (2015). "Identifying causal gateways and mediators in complex spatio-temporal systems". In: *Nature communications* 6.1, pp. 1–10.

Savini M Samarasinghe, Yi Deng, and Imme Ebert-Uphoff (2020). "A causality-based view of the interaction between synoptic-and planetary-scale atmospheric disturbances". In: *Journal of the Atmospheric Sciences* 77.3, pp. 925–941.

Peter Spirtes, Clark N Glymour, Richard Scheines, and David Heckerman (2000). *Causation, prediction, and search*. MIT press.

Teik Toe Teoh and Zheng Rong (2022). "Clustering". In: *Artificial Intelligence with Python*. Springer, pp. 213–218.

D Phil Turnipseed and Vernon B Sauer (2010). *Discharge measurements at gaging stations*. Tech. rep. US Geological Survey.

Thorsten Wagener, Murugesu Sivapalan, Peter Troch, and Ross Woods (2007). "Catchment classification and hydrologic similarity". In: *Geography compass* 1.4, pp. 901–931.

Sewall Wright (1934). "The method of path coefficients". In: *The annals of mathematical statistics* 5.3, pp. 161–215.

Jiaxin Xie, Xiaomang Liu, Peng Bai, and Changming Liu (2022). "Rapid watershed delineation using an automatic outlet relocation algorithm". In: *Water Resources Research* 58.3, e2021WR031129.

A. Appendix. Additional Plots

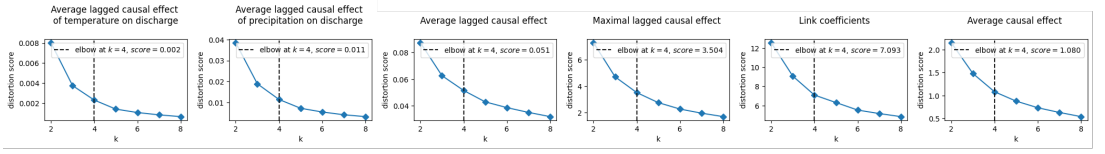

Figure 6. *Average distance to cluster center for different numbers of cluster centers, the optimal choice for the number of cluster centers is where the sudden drop starts to slow down. Here, this would be around* 4 *cluster centers.*

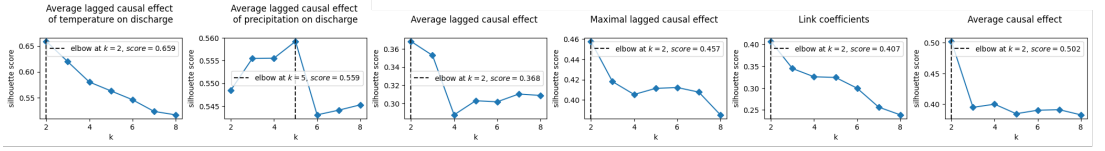

Figure 7. *Average silhouette score for different numbers of cluster centers, the optimal choice for the number of cluster centers is where the silhouette score is maximized.*

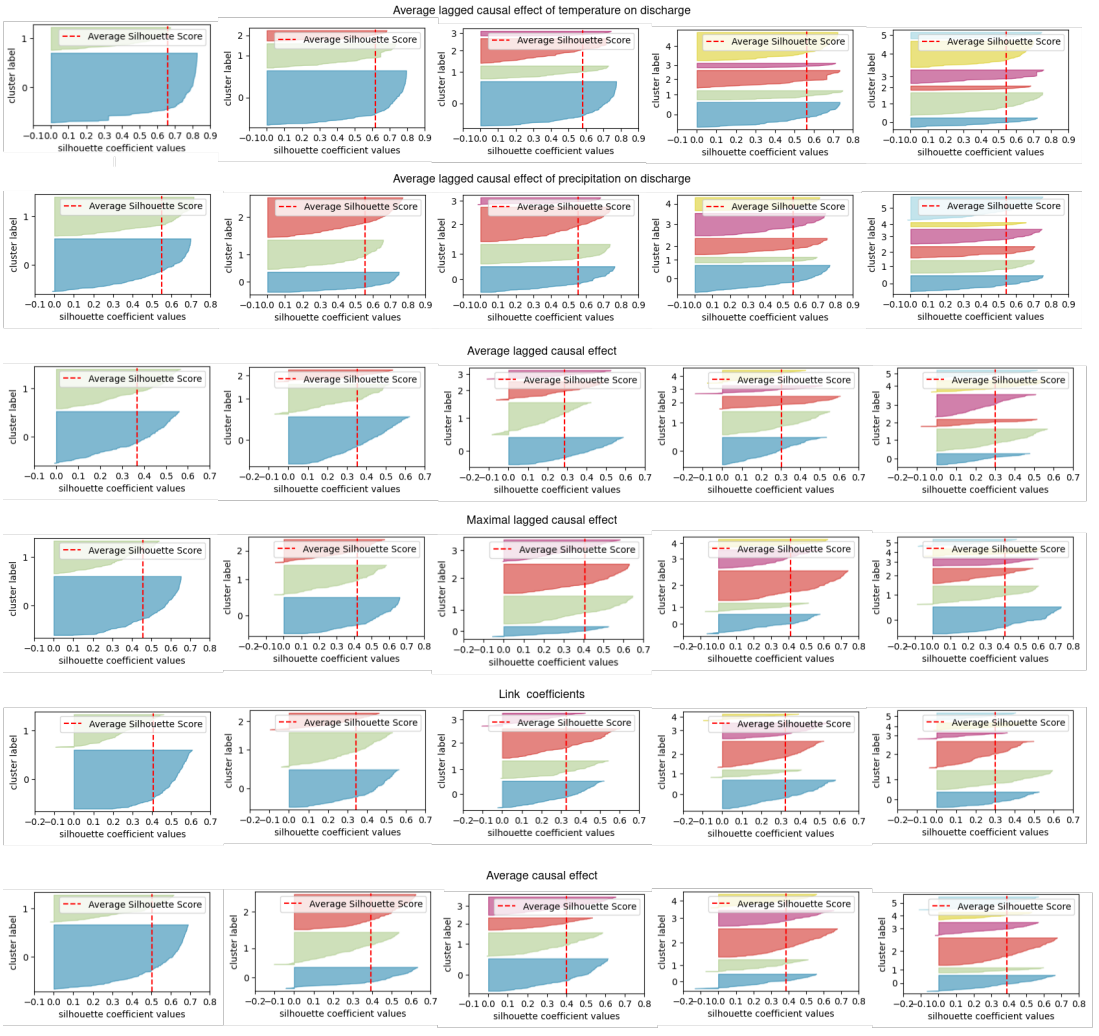

Figure 8. *Silhouette scores for different numbers of cluster centers, the optimal choice for the number of cluster centers is where the silhouette score of all the clusters is above the average score of the data set (indicated by dotted red line). Furthermore all the clusters should be roughly equal in size..*

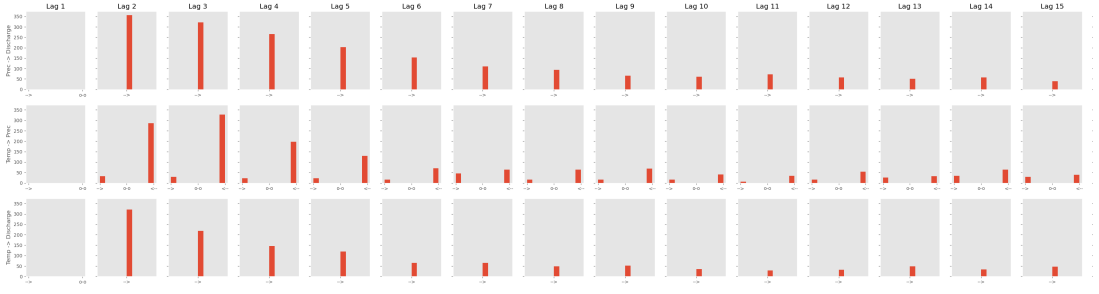

Figure 9. *Number of links across all graphs up to a maximal time lag of* 15.

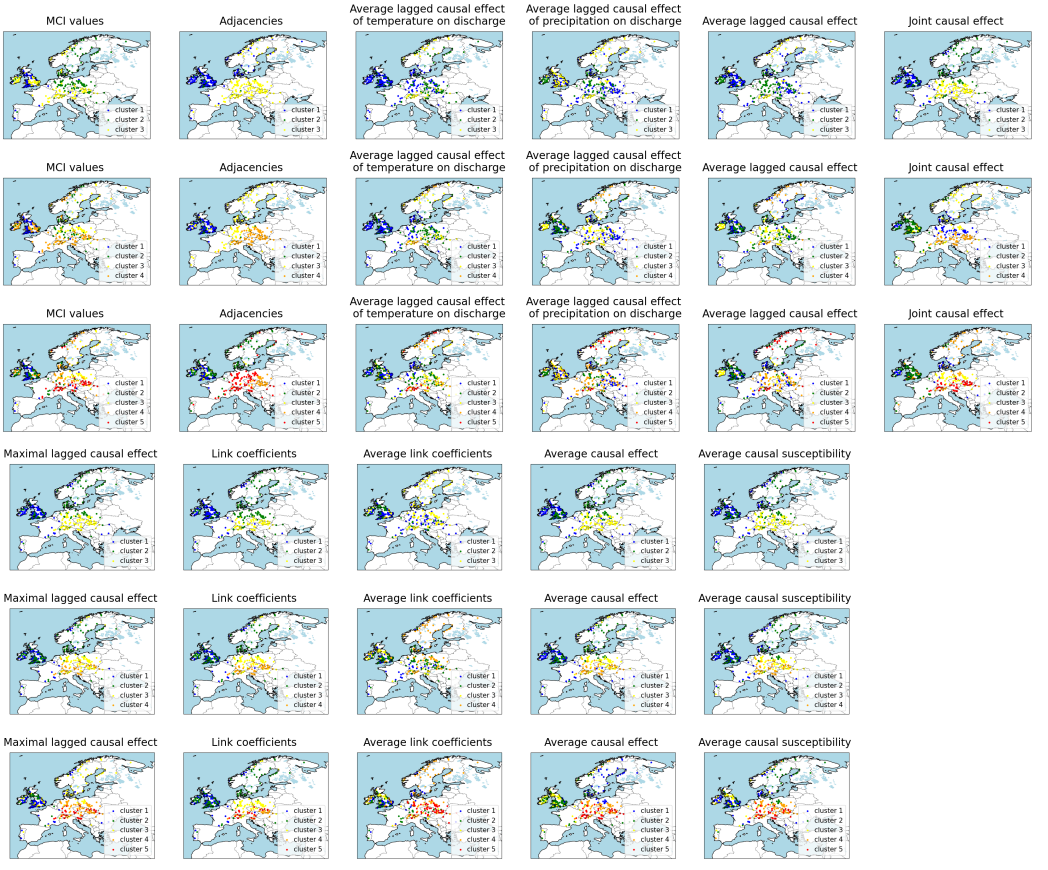

Figure 10. *Cluster results for various different choices of CE-based and other features for 3, 4, or 5 clusters.*

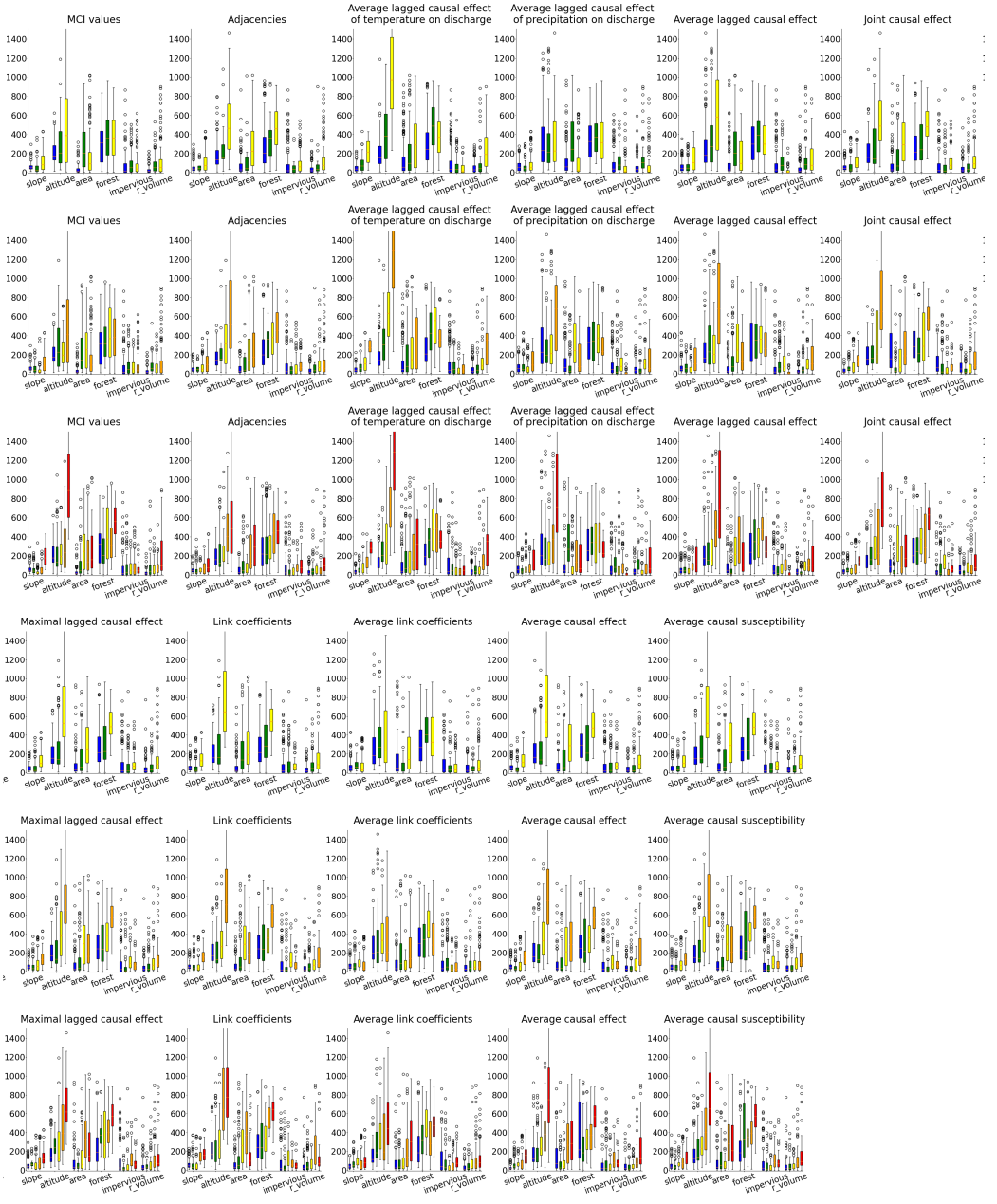

Figure 11. *Distribution of various attributes within each found cluster results for different choices of CE-based and other features for 3, 4, as well as 5 clusters.*

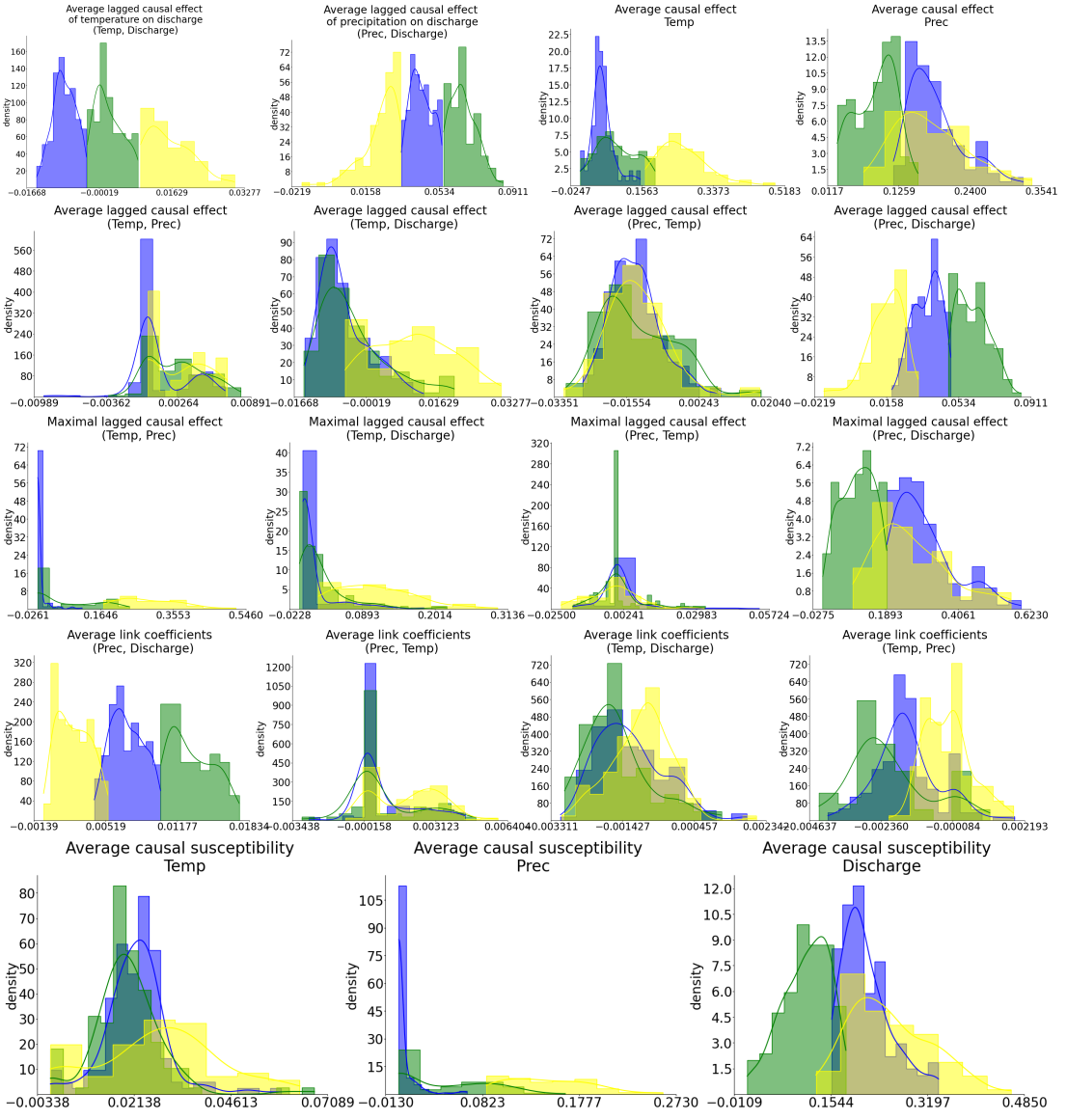

Figure 12. *Distribution of features within each cluster for* 3 *clusters.*

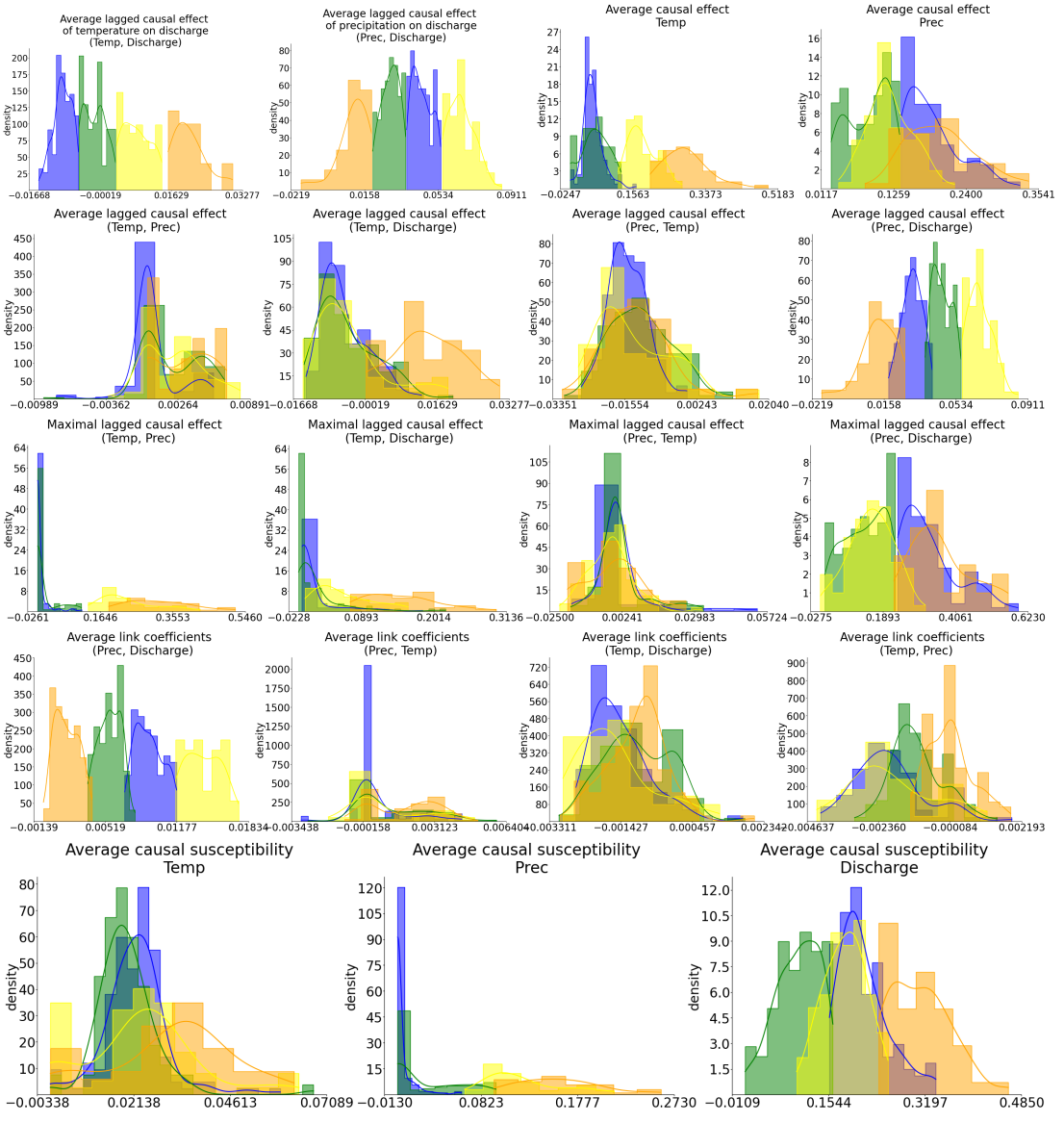

Figure 13. *Distribution of features within each cluster for* 4 *clusters.*

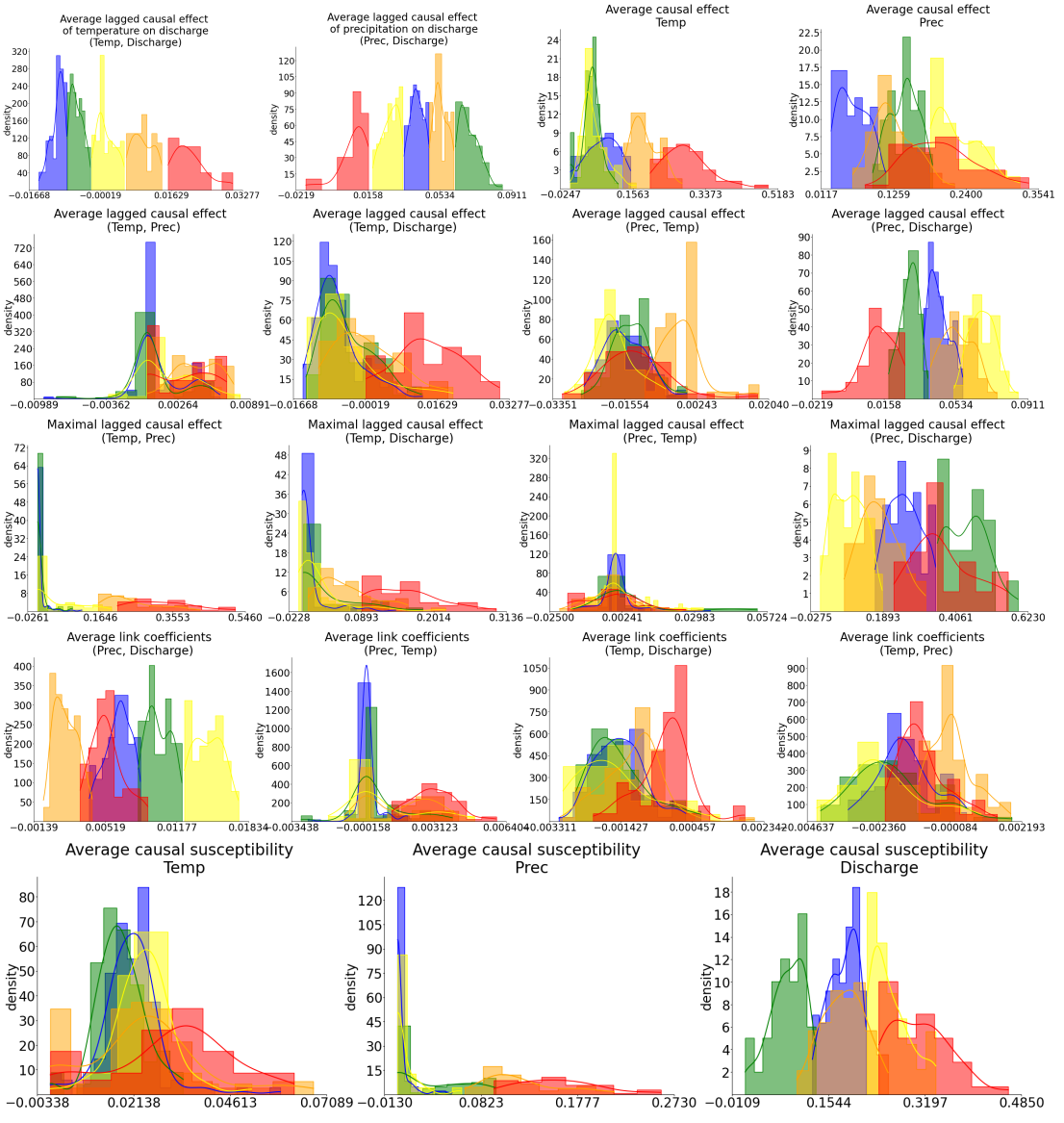

Figure 14. *Distribution of features within each cluster for* 5 *clusters.*