Causal Discovery for Railway Health Condition Monitoring – A Case Study

Neumann, Thorsten[1], Rabel, Martin[2],
Popescu, Oana-Iuliana[2], Gerhardus, Andreas[3]

[1]German Aerospace Center (DLR),
Institute of Transportation Systems, Berlin
[2]University of Potsdam,
Institute of Computer Science, Potsdam
[3]German Aerospace Center (DLR),
Institute of Data Science, Jena

Abstract

Integrating knowledge of causal relationships into machine-learning models might help to overcome the problem of limited interpretability of these models. Causal methods can thus be expected to enable innovative data-driven approaches for transparent and reliable diagnostics in safety-critical domains such as railway health condition monitoring. Based on available interlocking data, this paper exemplarily demonstrates how causal discovery can be applied to learn the causal influences of certain weather parameters on the insulation resistance of the electric installation of electronic interlockings.

Keywords: Causal Discovery; Explainable AI; PHM; Interlocking

1 Introduction

Capacities and reliability of the railway infrastructure are key success factors for the rail sector in competition with other modes of transport. Innovative data-driven approaches using machine learning are expected to essentially support asset health monitoring including predictive maintenance. In this context, common machine learning relies on finding and utilizing *statistical associations* (e.g., correlations) based on large scale data.

But, typically, it is not capable of actually deciding about *causality* which however is crucial for transparent and holistic diagnostics. This is one of the main reasons why machine learning is still facing big challenges with regard to validation and verification in safety-critical domains such as rail, in particular.

Recent research in the field of *"Explainable AI"* – still without explicitly addressing causality – therefore attempts to make (black-box) machine learning models better accessible for human interpretation (cf. [1]). Besides that, the *"Causal Revolution"* – as it is called in [2] – has brought a profound new mathematical understanding of causality during the last decades with promising algorithmic tools for causal learning and reasoning (cf. [3], [4], [5], [6]) that are now ready to be widely applied and tested in practice. The remainder of this paper focusses on this second approach.

2 Causal graphs and causal discovery

It is an important and well-proven fact that "correlation is not causation" (see [2]). Indeed, there are numerous academic and real-world examples where correlation must not be misinterpreted in terms of causality. Though, given knowledge about the data-generating process or in case that certain assumptions such as *faithfulness* or *causal sufficiency* (cf. [7]) hold, associations in the data nevertheless carry important qualitative and quantitative information about the causal relations between the variables under consideration. Formalizing and utilizing this connection between statistical and causal associations in order to reason about causality in the given data is the topic of *causal inference* (cf. [4]).

Here, the task of *causal discovery* refers to learning cause-and-effect relationships between variables. Basically, constraint-based causal discovery algorithms retrieve the causal information from marginal and conditional independencies in the data distribution (see [5] and, for example, the review article [8]). In this paper, the PCMCI algorithm [9] (see Section 2.2) together with a partial correlation test (see Section 2.1) as implemented in the Python package Tigramite (see [10]) is employed. As a result, the found qualitative causal relations are described by *causal graphs* in which nodes stand for the variables under consideration and directed edges represent their causal interaction, respectively.

2.1 Partial correlation test

By definition, two random variables A and B are marginally independent if and only if their joint distribution $p(A,B)$ equals the product $p(A)\cdot p(B)$ of their marginal distributions. In this case, observing the value of B does not contain any information about the value of A. Similarly, A and B are conditionally independent given a third variable C if and only if the factorization $p(A, B|C) = p(A|C) \cdot p(B|C)$ holds, where $p(X|Y) = p(X, Y)/p(Y)$ denotes the conditional distribution of X given Y. In this case, if the value of C is already observed, then additionally observing the value of B does not yield any information about the value of A beyond the information already obtained by observing C. For more details see textbooks on probability theory or statistics (e.g., [11]).

In the case study presented here, the simplifying assumption is made that the data distribution is multivariate Gaussian. Then, A and B are marginally independent if and only if their correlation is zero. Similarly, A and B are conditionally independent given C under this assumption if and only if the so-called partial correlation of A and B given C is zero (cf. [12]). To calculate this partial correlation, one can proceed in three steps: First, linearly regress A on C and calculate the residuals r_A as the difference $A - \hat{A}$ where $\hat{A}$ is the value of A as predicted by the linear regression on C. Second, repeat the first step with B instead of A. Third, calculate the correlation between the residuals r_A and r_B (cf. [12]).

In practice, the correlation and partial correlation values are not given but rather need to be estimated from finite data. To build a statistical hypothesis test, with marginal respectively conditional independence being the null hypothesis, one then utilizes the fact that a certain transformation of the estimated correlation respectively estimated partial correlation follows a Student's t-distribution with a certain number of degrees of freedom (cf. [12]). The Python package Tigramite referenced above (see [10]) provides an implementation of this kind of partial correlation test.

2.2 The PCMCI algorithm

The constraint-based approach to causal discovery builds on the observation that the particular form of a causal graph typically imprints several marginal and conditional independencies into the data (cf. [13], [14]). For example, if there are three variables A, B, C and the causal graph is $A \rightarrow B \rightarrow C$, then A and C are conditionally independent given B. Intuitively speaking, this independence holds because all information about A that is contained in C must have "passed through" B and, hence, is already contained in B. Thus, if one has already observed B, then one does not learn more about C by additionally observing A.

Now suppose that, in addition to $A \rightarrow B \rightarrow C$, the causal graph also contains the directed edge $A \rightarrow C$. Then, A and C are no longer conditionally independent given B because there is a "pathway of information" from A to C, namely $A \rightarrow C$, that does not pass through B. The idea of constraint-based causal discovery is to find these marginal and conditional independencies based on data by statistical tests of independence, for instance by the partial correlation test described above, and to then reverse-engineer the causal graph. For example, if one finds that A and C are conditionally independent given B, then one can already conclude that the edge $A \rightarrow C$ is not part of the causal graph.

While such types of arguments appear intuitively plausible, they nevertheless do rely on certain assumptions that one has to be aware of. These are the causal Markov assumption and causal faithfulness assumptions (or variations thereof) as well as, depending on the particular algorithm that one decides to use, the causal sufficiency assumption and the assumption that there are no cyclic causal relationships (cf. [5]).

A prominent example for constraint-based causal discovery algorithms is the PC algorithm (cf. [5]). This algorithm utilizes the four assumptions listed in the previous paragraph and has originally been developed for non-time series data. The PCMCI algorithm (see [15]) is an adaptation of PC to time series data that is specifically designed to handle the statistical challenge of applying causal discovery to time series with a large number of variables. In addition to the assumptions of the PC algorithm, PCMCI assumes that the causal relationships remain constant in time.

The result of PCMCI is an estimated causal graph (see Figure 2 further below for an example), in which the curved directed edges signify causal relationships between variables. These edges are supplemented by their so-called "lag", written as white digits on top of the edges in the plot, which specifies how many time steps the respective causal relationships take to manifest themselves. For example, if A causally influences B at lag 2, then A_{t-2} has a causal influence on B_t. Here, the time index t is arbitrary due to the assumption of constant-in-time causal relationships. The same edge can have multiple lags. Causal relationships that act on time scales below the sampling interval, which thus have a time lag of zero, are drawn as straight-line edges with circles at their ends. This notation signifies that the algorithm does not decide about their direction; that is, if $A_t \circ\!\!-\!\!\circ B_t$ is in the estimated causal graph, then either A influences B or B influences A at time lag zero. An extension of PCMCI that attempts to orient this type of edge, if possible, is the PCMCI+ algorithm (see [9]).

3 Case Study: Interlocking

Properly monitoring the health condition of the control, communication and safety systems of the railway infrastructure is crucial for its maintenance as well as reliable train operations. Besides monitoring other components such as switches and crossings or the track itself, electronic interlockings are an example where continuously measuring meaningful health related parameters helps to detect failures and thus to avoid safety-critical situations.

Typically, the electrical systems of electronic interlockings in Germany are isolated against earth. The insulation resistance (IR) thus is a good measure for defects in the electric installation. While a single defect is not that problematic in general, parallel defects at more than one location may result in uncontrollable and safety-critical behavior of the interlocking system. Events of temporarily or suddenly reduced IR therefore require special attention. At the same time, it is known that weather has a major influence on the overall level of the IR (see [16]).

The goal of this case study is to analyze how and to what extent these known causal relations between relevant weather parameters (e.g., temperature or precipitation) and available IR measurements can be detected automatically from data by means of time-series based causal discovery (cf., for example, [17]). In practical applications, knowing such causal relations usually helps to better interpret the data which again can lead to better and faster maintenance decisions.

Note that the presentation in this paper is mostly to demonstrate the general potential of innovative causal methods for railway condition monitoring purposes rather than a deep analysis of the specific use case itself. Nonetheless, the selection of the use case is motivated by the fact that diagnostics of insulation faults as in context of interlockings, for instance, are an urgent and important topic related to railway infrastructure operation and safety (cf. [18]).

3.1 Data and Preprocessing

The IR data used in this case study comprise measurements for about 1.5 years between 2015 and 2016 collected from an electronic interlocking in southern Germany. Details on the data collection can be found in [19]. For the purpose of causal discovery, the data have then been aggregated to one sample per 10 minutes (see Figure 1a) with the option of further down-sampling, and synchronized with freely available weather data (see [16]) provided by Deutscher Wetterdienst (DWD). Missing values were interpolated linearly. Regarding the IR measurements, time intervals with missing (i.e., interpolated) data are indicated by the orange line in Figure 1a and were left out as input during the causal learning process (see Section 3.2).

Furthermore, in the middle (around time index 40,000) and at the end (around time index 80,000) it seems that there is a systematic drop in the overall level of the IR measurements where the true reason is not clear so far. One cause, of course, could be a drift or other sensor artefacts in the measurement system. But, also the connection to a temporary (untypical) configuration of the interlocking system depending on the combined operational states of all connected switches could be a conceivable explanation.

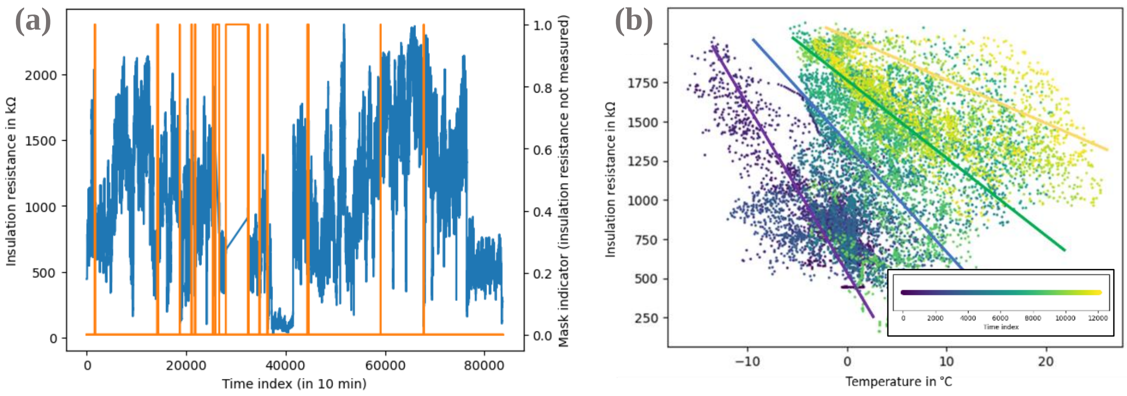

Figure 1: a) IR with a time-based resolution of 10 min (regions with missing data marked by the orange line) and b) negative correlation between temperature and IR based on a data subset covering the first three months of the available data.

Note that some relations between weather and IR variables have already been observed in previous studies and can thus perfectly be used for evaluating the plausibility of the causal discovery results in the remainder of this paper. For a more detailed discussion about known dependencies between IR and certain weather parameter (such as temperature and precipitation) see [16]. In addition, Figure 1b shows a (time-varying) negative correlation of IR and temperature which was found in this way during conducting the case study presented here.

3.2 Results and Discussion

Due to multiple periodicities on different time scales (e.g., daily, seasonal) overlapping with sudden effects, ensuring *(causal) stationarity* (cf. [7]) of the input time series for the learning process can be challenging which, by the way, is also supported by the previously mentioned finding of time-varying correlations as shown in Figure 1b. In the following, long-term effects are therefore addressed by learning separate causal graphs per season (e.g., summer or winter). At the same time, *hourly* aggregated samples are used to tolerably capture the short-to-medium-term effects while sufficiently limiting the number of backward time slices (i.e., four) as one important hyperparameter in the causal discovery process using PCMCI (cf. Section 2.2). Other relevant hyperparameters of the PCMCI algorithm were specified based on some common default values.

Figure 2 shows the derived causal graph based on data for summer seasons (June to August). Following this graph, both precipitation and ground temperature appear to have a negative causal association with the IR at small time lags which is in good agreement with previous findings (cf. Figure 1b and/or [16]). Interestingly, the (air) temperature is detected to only have an *indirect effect* on the IR via the ground temperature, while both temperature measures strongly correlate. Also, humidity seems to negatively affect the IR, but with a larger delay. Sunshine duration, on the other hand, has a positive impact on the IR, probably because it facilitates evaporation and thus a drier environment for the cables and field elements of the interlocking system opposed to the effect of humidity and precipitation. Lastly, temperature and sunshine duration are found to positively correlate while humidity and precipitation (at least in the short term) have a reverse trend as compared to temperature.

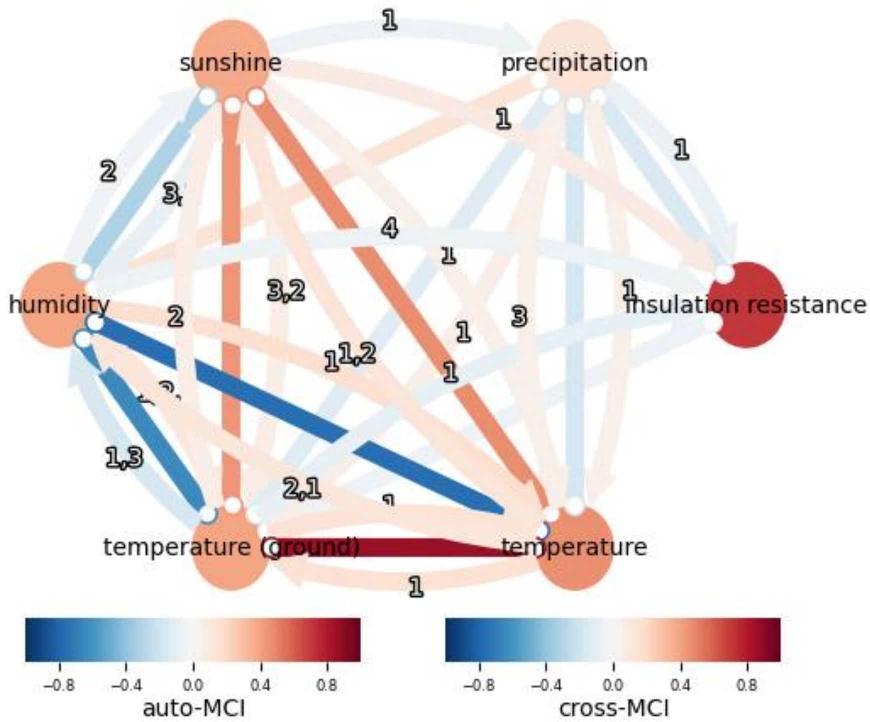

Figure 2: Causal graph exemplarily showing the results of causal discovery for summer based on hourly data. Undirected edges refer to causal relations whose direction is non-decidable via PCMCI (cf., last paragraph in Section 2.2), and numbers denote the time lag(s) per edge (in hours) related to the respective causal interactions. Further, red and blue colors respectively represent positive and negative correlations (see [10]).

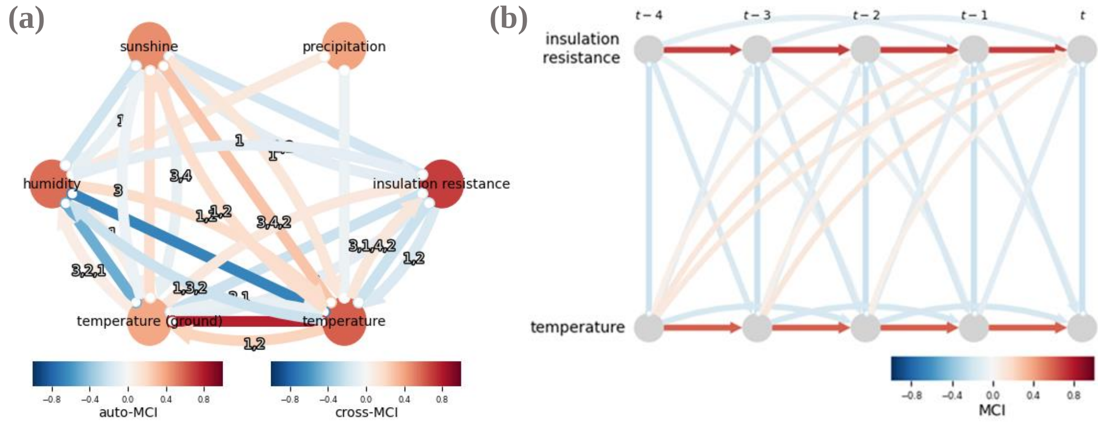

Figure 3: a) Causal graph exemplarily showing the results of causal discovery for winter based on hourly data and b) Unrolled graph for the relation between temperature and IR with explicit time slices. (Colors and labels as in Figure 2.)

In comparison to the summer graph, Figure 3a shows the respective causal discovery results for the winter season (December to February). As can be seen, there are at least three significant differences between both graphs.

First, the precipitation node now seems a bit more isolated from the IR measurements meaning that the former direct effect on IR vanishes. Whether or to what extent the latter is replaced by *indirect effects* (e.g., via the humidity node) however remains unclear due to the undirectedness of the zero-lag-edges that connect precipitation to the rest of the causal graph in Figure 3a. Yet unproven, a possible explanation for the divergent findings concerning the precipitation effect on IR could be the different forms of precipitation in summer and winter together with their distinct electric implications. That is, fluid water can be expected to penetrate the electric installations in the field much easier than snow which simply accumulates above with much less electrically relevant contact. In addition, water typically is a good conductor while snow is mostly non-conducting. Consequently, snow is indeed likely not to significantly affect the IR, but only when it is being reduced into its fluid state again later and thus not directly during the precipitation event as represented by the precipitation node in the causal graph.

Interestingly, if true, this might also explain a second observation, namely the inverted algebraic sign of the correlation between sunshine duration and IR in Figure 3a. That is, while summer sunshine facilitates evaporation and drying ground with the likely result of *higher* IR values as discussed earlier, in winter it may cause snow and ice getting melted which contrarily makes the direct environment of the cables and field elements of the interlocking system more wet and thus potentially *lowers* the IR level.

Finally, a third difference between Figure 2 and Figure 3a is the existence of a quite complex interaction of temperature and IR in the winter graph (cf. also the unrolled subgraph in Figure 3b). In particular, the causal relation from IR towards temperature looks very counterintuitive and has very likely to be considered as wrongly detected.

In this context, it has to be clearly stressed that the above results and discussions are subject to substantial uncertainty because of several reasons. Namely, as detailed above, the employed causal discovery method relies on various assumptions about the data-generating process, and some of these assumptions might be violated for the data at hand. For example, in potential violation of the causal sufficiency assumption, there might be unobserved variables that act as confounders which introduce additional dependencies that are not modelled by causal relationships between the observed variables. Moreover, in potential violation of the assumption behind the partial correlation test, the data distribution might be more complex than multivariate Gaussian.

In addition to that, the presented analysis has an explorative nature with an admittedly rather limited scope. To improve, one could repeat the causal discovery steps in which one systematically varies the hyperparameters of the analysis. This, for example, means modifying the hyperparameters of the PCMCI algorithm itself such as the maximal considered time lag, but also the choice of different conditional independence tests, the selection of data, or the details of the data preprocessing. Needless to say, relatively stable results across a proper range of hyperparameter settings in context of such an extended analysis would clearly improve the trustiness of the final results.

4 Conclusion and outlook

While further analyses are necessary, the presented case study nevertheless hints at the applicability and reasonableness of causal discovery in the field of railway health condition monitoring. Note that the general concept is not limited to the specific use case of monitoring electronic interlockings of course, but can also be applied to other critical components of the rail infrastructure. With regard to refining the results from Section 3.2, one could take an additional causal effect analysis based on the estimated causal graphs (see [6]) into account, and/or support the causal discovery step by integrating expert knowledge on the existence or non-existence of edges in the causal graphs (incl. their causal direction), for instance. Also, one might explore further causal discovery methods for time series and dynamical systems and analyze the consistency among the results.

Acknowledgement

This work was financed via institutional funding at the German Aerospace Center (DLR). All authors contributed in their role as scientific staff members of the DLR in the project “CausalAnomalies”.

Literature

[1] A. Sajid, T. Abuhmed, S. El-Sappagh, K. Muhammad, J. Alonso-Moral, R. Confalonieri, R. Guidotti, J. Del Ser, N. Díaz-Rodríguez and F. Herrera, "Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence," *Information Fusion,* vol. 99, no. 11, p. 101805, 2023.

[2] J. Pearl and D. Mackenzie, The Book of Why, First Trade Paperback ed., New York: Basic Books, 2020.

[3] A. R. Nogeira, A. Pugnana, S. Ruggieri, D. Pedreschi and J. Gama, "Methods and tools for causal discovery and causal inference," *WIREs Data Mining and Knowledge Discovery,* vol. 12, no. 2, 2022.

[4] J. Peters, D. Janzing and B. Schölkopf, Elements of causal inference, Cambridge, London: MIT Press, 2017.

[5] P. Spirtes, C. Glymour and R. Scheines, Causation, Prediction, and Search, 2nd ed., Cambridge, Massachussetts: MIT Press, 2001.

[6] J. Pearl, Causality: Models, Reasoning, and Inference, 2nd ed., Cambridge: Cambridge University Press, 2009.

[7] J. Runge, "Causal network reconstruction from time series: From theoretical assumptions to practical estimation," *Chaos,* vol. 28, no. 7, 2018.

[8] C. Glymour, K. Zhang and P. Spirtes, "Review of Causal Discovery Methods Based on Graphical Models," *Front. Genet.,* vol. 10, 2019.

[9] J. Runge, "Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets," in *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence*, Toronto, Canada, 2020.

[10] J. Runge, "Tigramite - Causal inference for time series datasets," [Online]. Available: https://github.com/jakobrunge/tigramite. [Accessed 26/03/2025].

[11] L. Wasserman, All of Statistics: A Concise Course in Statistical Inference, New York: Springer, 2004.

[12] T. W. Anderson, An Introduction to Multivariate Statistical Analysis, 3rd ed., Hoboken, New Jersey: John Wiley & Sons, 2003.

[13] T. Verma and J. Pearl, "Causal Networks: Semantics and Expressiveness," *Machine Intelligence and Pattern Recognition,* vol. 9, pp. 69-76, 1990.

[14] D. Geiger, T. Verma and J. Pearl, "Identifying independence in Bayesian networks," *Networks,* vol. 20, no. 5, pp. 507-534, 1990.

[15] J. Runge, M. Kretschmer, S. Flaxman and D. Sejdinovic, "Detecting and quantifying causal associations in large nonlinear time series datasets," *Sci. Adv.,* vol. 5, no. 11, 2019.

[16] J. Heusel and J. C. Groos, "Analysis of Systematic Influences on the Insulation Resistance of Electronic Railway Interlocking Systems," in *IAI 2021. Lecture Notes in Mechanical Engineering,* Cham, 2022.

[17] G. Camps-Valls, A. Gerhardus, U. Ninad, G. Varando, G. Martius, E. Balaguer-Ballester, R. Vinuesa, E. Diaz, L. Zanna and J. Runge, "Discovering causal relations and equations from data," *Physics Reports,* vol. 1044, pp. 1-68, 2023.

[18] C. Ziegler, "Condition-based maintenance and planting an apple tree," in *The Rise of IoT & Big Data in Rail*, Cologne, Germany, 2025.

[19] C. Linder and R. Schenkendorf, "Datengetriebene Diagnoseansätze für ESTW-Kabelanlagen," *SIGNAL + DRAHT,* vol. 107, no. 10, pp. 16-21, 2015.

Authors

	**Neumann, Thorsten (Dr.-Ing.)** Dipl.-Math. (Mathematics) in Münster, PhD (Engineering) in Braunschweig. Researcher at the DLR-Institute of Transportation Systems in Berlin as member of the working group on “Asset Monitoring and Management”. Research interests: Transparent and holistic diagnostics for railway infrastructures.
	**Rabel, Martin (Dr. rer. nat.)** M.Sc. (Physics), PhD (Mathematics) in Heidelberg. Formerly staff member at the DLR-Institute of Data Science in Jena; now Postdoc with focus on causal methods with Jakob Runge’s group, currently affiliated with the University of Potsdam. Research interests: Causal methods.
	**Popescu, Oana-Iuliana (M.Sc.)** M.Sc. (Computer Science). Formerly staff member at the DLR-Institute of Data Science in Jena; now PhD student affiliated with the Institute of Computer Science at the University of Potsdam. Research interests: Causal discovery for data with anomalies and distribution shifts.
	**Gerhardus, Andreas (Dr. rer. nat.)** M.Sc. in Physics (University of Bonn), Dr. rer. nat. (Faculty of Mathematics and Natural Sciences at the University of Bonn), now group leader at the DLR-Institute of Data Science in Jena. Research interests: Causal inference methods and their application