Granger Causality for Heterogeneous Processes

Sahar Behzadi[1], Kateřina Hlaváčková-Schindler[1], and Claudia Plant[1,2]

[1] Faculty of Computer Science, Data Mining, University of Vienna, Vienna, Austria
[2] ds:UniVie, University of Vienna, Vienna, Austria
{sahar.behzadi, katerina.schindlerova, claudia.plant}@univie.ac.at

Abstract. Discovery of temporal structures and finding causal interactions among time series have recently attracted attention of the data mining community. Among various causal notions graphical Granger causality is well-known due to its intuitive interpretation and computational simplicity. Most of the current graphical approaches are designed for homogeneous datasets i.e. the interacting processes are assumed to have the same data distribution. Since many applications generate heterogeneous time series, the question arises how to leverage graphical Granger models to detect temporal causal dependencies among them. Profiting from the generalized linear models, we propose an efficient **H**eterogeneous **G**raphical **G**ranger **M**odel (HGGM) for detecting causal relation among time series having a distribution from the exponential family which includes a wider common distributions e.g. Poisson, gamma. To guarantee the consistency of our algorithm we employ adaptive Lasso as a variable selection method. Extensive experiments on synthetic and real data confirm the effectiveness and efficiency of HGGM.

1 Introduction

Recently there is a significant interest in causal inference in various data mining tasks. Discovery of causal relations among different processes leads to characterize the evolution in time of regular instances. The regular pattern can be used to detect the deviated observations or outliers in anomaly detection [15]. A number of methods has been developed to infer causal relations from time series data by Granger causality [8] which is a popular method due to its computational simplicity. The presumption of this approach is that a cause helps to predict its effects in the future. Most of the existing methods in this area assume additive causal interactions among time series following a specific data type or a certain distribution. The well-know causality notion, Additive Noise Models (ANMs), have been proposed for either continuous [17] or discrete [14] time series. Moreover, most of the probabilistic approaches are designed for homogeneous datasets [5], [4]. However, in reality the interacting processes do not have to be homogeneous (having the same distribution). Such situations can occur, for example, in climatology when various measurements are provided for different meteorological stations. Figure 1 shows 10 weather stations and three major weather systems in Austria. The monthly amount of precipitation as well as the number of sunny days have been measured for every station, each of which

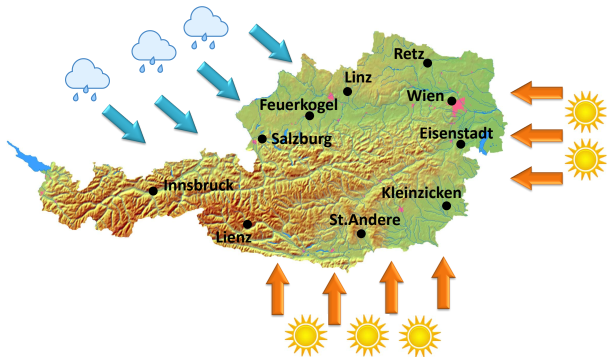

Fig. 1. Meteorological stations and three major weather systems influencing Austria.

with a non-Gaussian distribution. One can be interested in investigating how the number of sunny days in a station, influenced by one of the weather systems, can impact the amount of precipitation in the other locations.

Applying existing algorithms on such data sets can result an inaccurate Granger causal model since they have been designed for specific homogeneous data types. Moreover, the small set of algorithms, which are supposed to cope with the heterogeneity, mostly employ an exhaustive pairwise testing. This leads to inefficiency in a causal network discovery specially when the number of interacting proceses is increasing. In between, graphical Granger models are popular due to their efficiency and effectivness. They employ a penealized Vector Autoregression (VAR) to the Granger concept [1], [3], [7], [18]. However, to the best of our knowledge, so far they have been designed only for homogeneous data sets. Thus, in this paper we propose a penalized VAR-based algorithm to detect the **H**eterogeneous **G**raphical **G**ranger **M**odel (HGGM) by employing generalized linear models (GLMs). Similar to the other graphical models, we assume that the interactions among the involved processes are additive. Moreover, to ensure the convergence of HGGM to the true causal graph (i.e. consistency) we employ the well-know penalization approach, adaptive Lasso, with oracle properties [20]. The paper brings the following contributions:

- **Heterogeneity:** Applying the GLM methodology, we propose a heterogeneous graphical Granger model to discover the causal interactions among a wide variety of heterogeneous time series from the exponential family;
- **Consistency:** Assessing the causal relations via adaptive Lasso ensures consistency of our method;
- **Scalability:** Unlike other existing algorithms, HGGM avoids an exhaustive pairwise causality testing by penalized estimation of VAR models. Due to the computational simplicity of HGGM, it is convenient to be used in practice. Moreover, its reasonable runtime complexity makes our algorithm scalable for the large data sets consisting of long time series;
- **Effectiveness:** Following the result of our extensive experiments on synthetic and real datasets, HGGM is an effective algorithm even by detecting sparse causal graphs.

In the following we specify the problem and the theoretical background and propose our HGGM model. Section 2 presents the related work. In Section 3, we introduce the problem and our proposed framework to deal with heterogeneous data. In Section 4 we introduce our integrative algorithm HGGM and the theoretical considerations of it. Extensive experiments on synthetic and real data are demonstrated in Section 5. Our conclusion is in Section 6.

2 Related Work

Among various approaches to infer causality, Granger causality [8] is well-known due to its simplicity and computational efficiency. It states that a cause efficiently improves the predictability of its effect. There are various approaches depending on how to assess the predictability. Probabilistic approaches interpret it as the improvement in the likelihood (i.e. probability). However, several methods in this group are distinguished based on the way how they employ probability. Information-theoretic methods detect the causal direction by introducing some indicators. Among them, compression-based algorithms apply the Kolmogorov complexity and define a causal indicator by mean of the Minimum Description Length (MDL) [5], [4], [6]. Essentially, these algorithms are designed to infer the pairwise causal relations. Therefore, employing them for discovery of causal networks leads to inefficiency, especially when the number of processe increases. Moreover, to the best of our knowledge, almost all the algorithms in this category deal with homogeneous data sets except *Crack* [10], the most recent compression-based algorithm to deal with multivariate and heterogeneous processes. Beside the pairwise testing and its drawbacks, this algorithms lacks the accurate causal relations since there is no lag parameter considered in this approach. Transfer entropy, shortly TEN, is another approach among information-theoretic methods which is based on Shannon's Entropy [16]. In this approach it is more likely that the causal direction with the lower entropy corresponds to the true causal relation. Given a lag variable, TEN can detect both linear and non-linear causal relations. However, due to pairwise testing and its dependency on the lag variable, the computational complexity of this algorithm is exponential in the lag parameter. Moreover, similar to compression-based methods, TEN is not designed to deal with bidirectional causalities. As another method in this category, the authors in [9] employ the log-likelihood ratio to detect any causal relations among processes. They propose a statistical framework (SFGC) for mixed type data and assessing the causal relations between multiple time series is accomplished by the false discovery rate (FDR). The statistical power of the FDR based methods rapidly decreases with increasing number of hypotheses and these methods are computationally intensive. As the consequence, the statistical efficiency of SFGC decreases for the increasing number of investigated time series.

Another approach to assess the predictability is the graphical Granger method where a penalized VAR model is supposed to be estimated [1], [18]. Graphical Granger method is popular for its simplicity and efficiency since employing a penalized VAR model we avoid the pairwise testing. However most of the algorithms in this category are designed for Gaussian processes. Utilizing the advantages of this approach we introduced a graphical Granger algorithm for heterogeneous processes.

3 Theory

3.1 Granger Causality

Granger causality is a well-known notion of causality introduced by Granger in the area of econometrics [8]. Although the Granger causality is not meant to be equivalent to the true causality but it provides useful information capturing the temporal dependencies among time series. In a bivariate case let $x^{1:n} = \{x^t | t = 1, \ldots, n\}$ and $y^{1:n} = \{y^t | t = 1, \ldots, n\}$ denote two time series up to time n. Moreover, let the following two models represent two autoregressive models corresponding to time series y with and without taking past observations of x into consideration.

$$y^T = \alpha_1 y^1 + \cdots + \alpha_{T-1} y^{T-1} + \gamma_1 x^1 + \cdots + \gamma_{T-1} x^{T-1} + \varepsilon^T \tag{1}$$

$$y^T = \alpha_1 y^1 + \cdots + \alpha_{T-1} y^{T-1} + \varepsilon^T \tag{2}$$

Following the principle of Granger causality, x Granger-causes y if the Model 1 significantly improves the predictability of y comparing to the Model 2. The concept of Granger causality can be extended to more than two time series. Let $x_1^{1:n}, \ldots, x_p^{1:n}$ be p time series up to time n and X^T be the concatenated vector of all time series at time T, i.e. $X^T = (x_1^T, ..., x_p^T)$. The vector autoregressive (VAR) model is given by:

$$X^T = A_1 X^1 + \cdots + A_{T-1} X^{T-1} + \varepsilon^T \tag{3}$$

where A_t is a matrix of the regression coefficients at time $t = 1, \ldots, T-1$ and ε^t is a white noise. Thus, x_j Granger-causes x_i if at least one of the $(i, j)th$ elements in the coefficient matrices $A_1, \ldots, A_{T-1}$ is non-zero.

3.2 Causal Inference by Penalization

In order to detect the causal relations between several time series, one needs to estimate the coefficients of the VAR model introduced in the last section. Since this problem can be ill-posed, penalizing the VAR of order d (a time window) by means of a penalty function provides an efficient and sparse solution when the convergence to the true causal graph is ensured (e.g. [1], [18]). The penalization approach is referred to as variable selection as well. Thus, given the window size d for any time series $x_i, \quad i = 1, ..., p$, we consider the VAR model including all p time series. We slide the window over time series and get the corresponding VAR model. The fact is that the best regressors with the least squared error for any specific time series will have non-zero coefficients in the VAR model only for the dependent time series. More precisely, Let $X_{T,d}^{Lag} = \{{x_i}^{T-t} | i = 1, ..., p; \quad t = 1, ..., d\}$ denote the concatenated vector of all the lagged variables up to time T

for a given time window of length d. For simplicity we consider the same lag d for each time series. Applying the penalized optimization, the variable selection problem for the time series x_i is given by:

$$\hat{\beta}_i = \arg\min_{\beta_i} \sum_{T=d+1}^{n} (x_i^T - X_{T,d}^{Lag}\beta_i)^2 + \lambda R(\beta_i) \tag{4}$$

where $R(.)$ is the penalty function and λ is the regularization parameter. $\hat{\beta}_i = (\beta_1, ..., \beta_p)$ is a concatenated vector of the regression coefficients $\beta_1, ..., \beta_p$ corresponding to any time series $x_1, ..., x_p$. Back to the definition of Granger causality, x_j Granger-causes x_i if and only if at least one of the coefficients in β_j is non-zero.

3.3 Adaptive Lasso

One of the well-known variable selection methods is Lasso [19] where the penalty function considered in Equation 4 is the L_1 norm of the coefficients, i.e. $R(\beta_i) = ||\beta_i||_1$. Despite the efficiency of Lasso, the consistency [3] of this approach is not ensured. Therefore, we employ adaptive Lasso [20], a modification of Lasso, as the variable selection method in our model due to its consistency as well as its oracle properties. In this approach we assign adaptive weights for penalizing the L_1 norm of different coefficients. The penalty function is given by:

$$R(\beta_i) := \sum_{j=1}^{p} w_j |\beta_j| \quad where \quad w_j = \frac{1}{|\hat{\beta}_j^{(mle)}|^{\omega}} \tag{5}$$

In fact, w_j is the weight vector for some $\omega > 0$ and $\hat{\beta}_j^{(mle)}$ is the maximum likelihood estimate of the parameters. The consistency of adaptive Lasso is guaranteed under some mild regularity conditions in the following theorem [20]:

Theorem 1. *Let $\mathcal{A} = \{i : \hat{\beta}_i \neq 0\}$ be the set of all non-zero coefficient estimates. Suppose that $\lambda/\sqrt{n} \to 0$ and $\lambda n^{\frac{(\omega-1)}{2}} \to \infty$ then under some mild regularity conditions adaptive Lasso must be consistent for the variable selection.*

3.4 Heterogeneous Granger Causality

Most of the approaches to detect the Granger causality among time series have certain Gaussian assumptions for the interacting processes. However in many applications this assumption leads to an inaccurate causal model. Moreover, mostly the variable selection algorithms employed to penalize the VAR model are consistent under additional specific conditions on the Gaussian time series, see e.g. [1]. Profiting from the GLM framework, we propose a general integrative model to detect causal relations among a large number of heterogeneous time series. GLM, introduced by Nelder and Baker in [13], is a natural extension of linear regression to the cases when the regressed variables (time series) can

[3] I.e. the resulting sequence of estimates does not have to converge in probability to the optimal solution for variable selection under certain conditions (Section 2 in [20]).

have any distribution from the exponential family. In another word, the relation among the response variable and the covariates in a regression is not any more linear but defined by a link function g, a monotone twice differentiable function depending on concrete distribution functions from the exponential family.
In our model we assume the mean value of each time series at time T depends on its own history and the past values of the concurrent time series so that:

$$E(x_i^T) = g_i^{-1}(X_{T,d}^{Lag}.\beta_i). \tag{6}$$

Finally, our general objective function is defined as:

$$\hat{\beta}_i = \arg\min_{\beta_i} \sum_{T=d+1}^{n} \big[-x_i^T(X_{T,d}^{Lag}.\beta_i) + g_i^{-1}(X_{T,d}^{Lag}.\beta_i)\big] + \lambda.\sum_{j=1}^{p} w_j|\beta_j|. \tag{7}$$

The concrete form of our proposed objective function (7) concerning x_i to have binomial and Poisson distribution, respectively, is given by:

$$\hat{\beta}_i = \arg\min_{\beta_i} \sum_{T=d+1}^{n} \big[-x_i^T(X_{T,d}^{Lag}.\beta_i) + log(1+e^{(X_{T,d}^{Lag}.\beta_i)})\big] + \lambda.\sum_{j=1}^{p} w_j|\beta_j|, \tag{8}$$

$$\hat{\beta}_i = \arg\min_{\beta_i} \sum_{T=d+1}^{n} \big[-x_i^T(X_{T,d}^{Lag}.\beta_i) + \exp(X_{T,d}^{Lag}.\beta_i)\big] + \lambda.\sum_{j=1}^{p} w_j|\beta_j|. \tag{9}$$

4 HGGM Algorithm

Our method HGGM is summarized in Algorithm 1. At first it constructs the overall lagged matrix X^{Lag}, by sliding the window of size d over each time series. Then, HGGM solves the optimization problem (Equation 7) for each time series by calling $GLM - penalize()$, [11]. This procedure applies Fisher scoring algorithm to estimate the coefficients. We set the maximum λ as an input of $GLM - penalize()$ and the procedure employs the cross-validation to find the best regularization parameter.
Essentially one needs to know the distribution of every time series in order to specify an appropriate link function g. We utilize a statistical fitting procedure to find the most accurate distribution for every time series. We assign to any time series the distribution from the exponential family with the least Akaike Information Criterion (AIC). Finally, based on the definition of Granger causality we get pairwise Granger-causal relations among p time series out of which we construct the adjacency matrix corresponding to the final causal graph.

Consistency: The consistency of adaptive Lasso for the variable selection has been proven under some mild regularity conditions (Section 3). Thus, applying the adaptive Lasso for GLMs enables us to make the following statement about the consistency of HGGM.

Corollary 1. *Assume G be a true Granger causal graph corresponding to p time series, each of length n. Let the regularization parameter λ fullfils the conditions of Theorem 1. Then taking p time series as input, HGGM outputs a causal graph which converges to the true graph G with probability approaching 1 as $n \to \infty$.*

Algorithm 1 Causal Detection by HGGM

HGGM $(x_i, g_i, i = 1, \dots, p; d; \lambda_{max})$
Adj := adjacency matrix of the output graph
X^{lag} := lagged matrix of all temporal variables
// find Granger causalities for each feature
for all x_i **do**
  // solve the penalized optimization problem considering lagged variables
  $\beta_i = GLM - penalize(X^{Lag}, x_i, g_i, \lambda_{max}, d)$; // β_i := coefficients w.r.t x_i
  for all ${\beta_i}^j$ sub-vectors of β_i **do**
    $Adj(j, i) = 0$ //discover Granger-causalities
    if $(\exists t, 1 < t < d$ such that ${\beta_i}^j(t) > 0)$ **then**
      $Adj(j, i) = 1$
    end if
  end for
end for
return (Adj)

Proof. When $n \to \infty$ the conditions of Theorem 1 are fulfilled. Therefore it follows that the procedure *GLM–penalize(.)* in Algorithm 1 converges to the true Granger causal graph. Thus, HGGM is consistent as well.

Computational Complexity: Based on the proposed objective function (7), we investigate causal relationships for any time series $x_i, i = 1, \dots, p$ by fitting the most accurate VAR model. Therefore at any time we have p regression models each of which consists of d lagged variables corresponding to $x_1, \dots, x_p$. Applying Fisher scoring to estimate the parameters of VAR models, the number of computations required to solve a VAR of order d is $\mathcal{O}(d^2)$. Thus, the computational complexity of HGGM is in order of $\mathcal{O}(np^2d^2)$.

5 Experimental Results

In this section the performance of HGGM in comparison to other algorithms will be assessed in terms of *F-measure* which takes both precision and recall into account. Although there are many approaches to detect the Granger causality, only few of them are designed for heterogeneous time series. Therefore, we compare our algorithm to three methods which are applicable to mixed time series, i.e. transfer entropy, shortly TEN [16], Crack [10] and SFGC [9]. To evaluate HGGM we investigate the effectiveness and efficiency of HGGM by extensive experiments on synthetic and real-world data sets. HGGM is implemented in MATLAB and for the other comparison methods we use their publicly available implementations and recommended parameter settings. The source code and data sets are publicly available at: `https://bit.ly/2FkUB3Q`

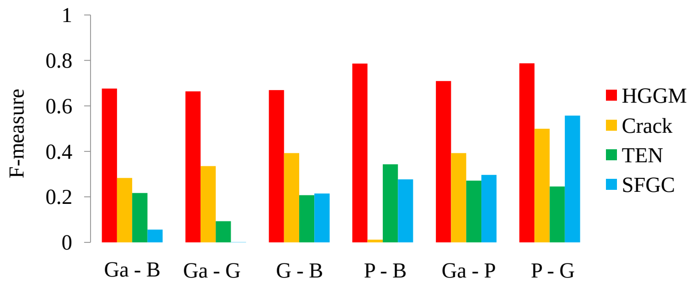

Fig. 2. Performance in various heterogeneous data sets. Ga: Gamma, G: Gaussian, B: Bernoulli, P: Poisson.

5.1 Synthetic Heterogeneous Data Sets

Firstly, we investigate the effectiveness of HGGM comparing to other algorithms in terms of *F-measure*. That is, we conduct various experiments each of which concerning a unique aspect. Then, we target the scalability of the algorithms varying the number of time series and the length of them. In any synthetic experiment, we report the average performance of 50 iterations performed on different data sets with the given characteristics. The length of generated time series n is always 1,000 except for the experiment on increasing the length. For any algorithm which requires to specify the lag variable we run the algorithm for various lags and take the average F-measure as the final result.

Effectiveness: HGGM is designed to deal with Gaussian as well as non- Gaussian time series having a distribution from the exponential family. In this experiment we generated time series with various combinations of Gaussian and non-Gaussian distributions in order to assess HGGM in various cases. Figure 2 shows that HGGM outperforms other algorithms in various combinations of Gaussian – non-Gaussian distributions and discrete - continuous time series. It confirms that our GLM-based objective function effectively copes with heterogeneity of time series comparing to the other methods. For the rest of the experiments we focus on Poisson - Gaussian combination as a representative for heterogeneous data sets.
Dependency: Figure 3 a illustrates how various algorithms perform when the dependency among time series, i.e. the coefficients in VAR model, is increasing ranging from 0.1 to 1. As one can expect, HGGM and SFGC have an ascending trend. However, the effectiveness of Crack and TEN is surprisingly decreasing. Although the performance of HGGM is smaller than SFGC and TEN in a very early stage, it outperforms other algorithm for the dependencies higher than 0.3 with a high margin.
Increasing the Number of Features: We increased the number of time series (features) iteratively in order to compare the performance of the algorithms when many time series are involved. Figure 3 b shows that the F-measure of any algorithm is descending while HGGM is still more efficient than others in any case. There is a big gap among the performance of two algorithms, Crack and TEN, comparing to HGGM in this figure. One of the reasons for this is that they are not able to deal with the bidirectional causality and by increasing the number of time series it effects the performance more and more.

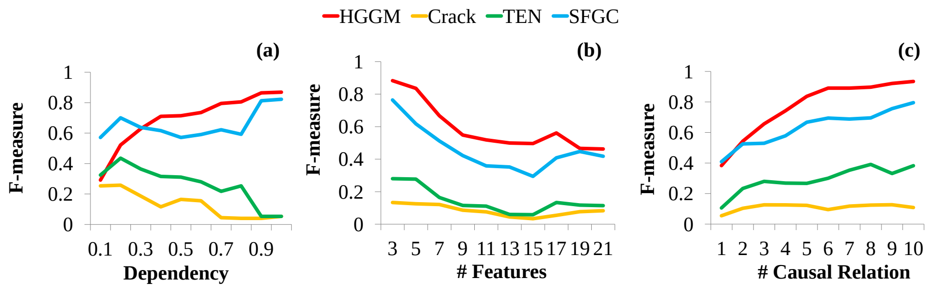

Fig. 3. Synthetic experiments.

Causal Relations: How will the various algorithms behave when the true causal graph is sparse? In this experiment we vary the number of causal relations among 5 mixed time series from Poisson - Gaussian combination. As expected, the effectiveness of any algorithm is increasing when the density of the true causal graph is increasing too. However Figure 3 c shows the superiority of our algorithm comparing to others even for sparse graphs.

Scalability: The scalability is investigated in two experiments. First, we increase the number of time series iteratively where the length is set to 1,000 i.e. $n = 1,000$. Then we vary n while every time four time series are generated. By the first experiment the efficiency of HGGM is shown (Figure 4 a) when the number of features is bigger than 6 comparing to Crack and TEN and bigger than 9 comparing to SFGC. However, considering the next experiment (Figure 4 b) the efficiency of our algorithm is confirmed. HGGM is the fastest algorithm almost always for the time series longer than 2,000.

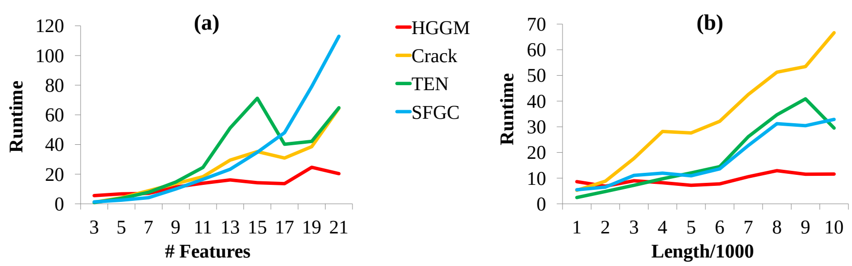

Fig. 4. Experiments on runtime in seconds

5.2 Real-world Applications

We conducted the experiments on publicly available real data sets considering two cases, whether a ground truth is given or not. In order to be fair in the real experiments we set $d = 15$ for all the algorithms which require a lag variable.

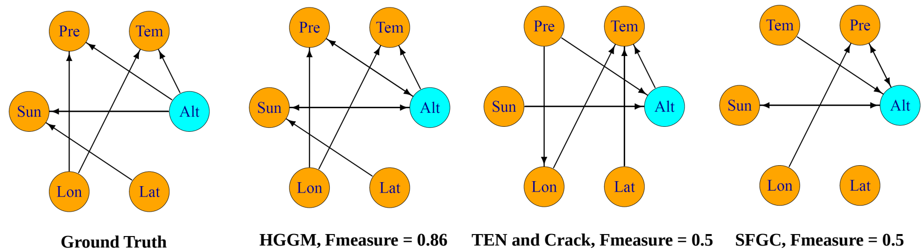

Fig. 5. Comparison on German weather data set.

Weather in Germany: The first data set *DWD* [4] is a climatological data consisting of 6 measurements, temperature, sunshine hours, altitude, precipitation, longitude and latitude for 394 weather stations all over Germany. The altitude measurement is already provided in a discrete time series while all other measurements are continuous. Applying the statistical fitting procedure (Section 4), we assign Gaussian distribution for all continuous time series and the Poisson distribution for the altitude. The ground truth is available in [12] which is provided by pairwise causal relations. In order to be fair by evaluating the results of the algorithms, we do not consider the causal interactions where no information is provided. Figure 5 shows the performance of HGGM comparing to other algorithms in terms of $F - measure$. HGGM ably finds all the existing causal relations. However, it detects causal relations where sunshine and temperature cause altitude.

Marks: The next two data sets together with the corresponding ground truth are publicly available [5]. *Marks* data set concerns the examination marks of 88 students on five different topics. The given true causal graph reveals any impacts the grades of a topic could have on the other topics. We assign Poisson distribution to any topic. In this experiment HGGM ($F - measure = 0.74$) was able to outperform TEN (0.55), Crack (0.6) and SFGC (0.71).
Gaussian: The Gaussian data set is a simulated data showing the causal interactions among 7 Gaussian time series. The time series are of the length 5,000. HGGM ($F - measure = 0.4$) performs more accurately comparing to other algorithms, TEN (0), Crack (0.14) and SFGC (0.14), although non of the algorithms was able to capture all the causal relations in the ground truth.

Austrian climatological data set: As a real world application we investigate causal spatio-temporal interactions among climatological phenomena for 10 sites uniformly distributed in Austria (Fig. 1). For any site we used the monthly measurements of precipitation and of the number of sunny days for 26 months. Employing the statistical fitting, we consider a Gamma distribution for the precipitation and a Poisson distribution for the number of sunny days. Because of the space limit we randomly focus on one of the stations, *Feuerkogel*, and the complete experiment is provided in the supplementary material.

[4] `http://www.dwd.de/DE/Home/home_node.html`
[5] `Http://www.bnlearn.com/documentation`

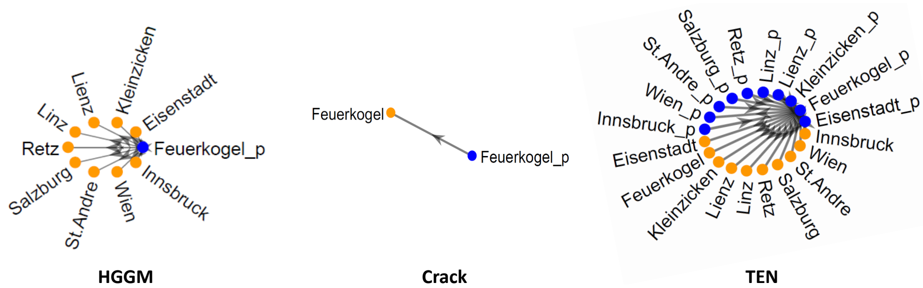

Fig. 6. Experiment on the Austrian climatological data. blue circles: amount of precipitation and orange circles: number of sunny days.

Moreover, the real meteorological data set is publicly available [6]. Essentially, Austrian weather is influenced by three climatic systems while any system has its own characteristics. Concerning the interpretation of results for the selected station, we concentrate on the Atlantic maritime climate from the north-west which is characterized by low-pressure fronts, mild air from the Gulf Stream, and precipitation [2]. The northern slopes of the Alps, the Northern Alpine Foreland, and the Danube valley are influenced by the Atlantic weather system.

Fig. 6 shows the causal graph discovered by HGGM, TEN and Crack. SFGC was not able to detect any causal relation therefore we exclude its result. Considering the impact of the Atlantic weather system, one expects the influence on the neighbour sites of *Feuerkogel* and the sites in eastern Austria. The sites in southern slope cannot be influenced by this system since the Alps are located in between. Comparing HGGM to other algorithms, HGGM is successful to detect more influenced sites by finding the correct causal direction among *Linz*, *Salzburg*, *Retz*, *Wien* and *Eisenstadt*. However it detects an interaction between *Feuerkogel* and *Lienz* which is not likely due to the large mountain area between the sites. Regarding Crack, although the only causal relation discovered by this algorithm sounds reasonable, there are other stations, e.g. *Linz* and *Salzburg*, where it is plausible to consider a causal interaction among them. On the other hand, TEN discovers a dense causal graph among all 20 time series and *Feuerkogel* which is hard to interpret. Moreover considering the Atlantic weather system, there is no interpretation for the causal direction from *Retz* to *Feuerkogel* detected by TEN since its direction is exactly in the opposite.

6 Conclusions and future work

In this paper we proposed HGGM, a graphical Granger model for discovery of causal relations among a number of heterogeneous processes. Profiting of a GLM framework our approach is generalized for time series having distributions from exponential family. Moreover to ensure the consistency of HGGM we employ adaptive Lasso with a proven consistency. We investigated the performance of HGGM in terms of effectiveness and efficiency comparing to state-of-the-art

[6] `Https://www.zamg.ac.at`

methods. Extensive experiments on synthetic and real data sets demonstrates the advantages of HGGM. As already mentioned, one of the interesting applications of our algorithm can be utilizing HGGM to detect anomalies among heterogeneous time series. To the best of our knowledge none of the current algorithms deal with heterogeneous anomalies by means of graphical Granger causality.

References

1. Arnold, A., Liu, Y., Abe, N.: Temporal causal modelling with graphical Granger methods. In: KDD (2007)
2. Bacsó, N.: Das Klima des Donauraumes. Geoforum (1971)
3. Bahadori, M.T., Liu, Y.: Granger causality analysis in irregular time series. In: SDM (2012)
4. Budhathoki, K., Vreeken, J.: Causal inference by compression. In: ICDM (2016)
5. Budhathoki, K., Vreeken, J.: Mdl for causal inference on discrete data. In: ICDM (2017)
6. Budhathoki, K., Vreeken, J.: Causal inference on event sequences. In: SDM (2018)
7. Cheng, D., Bahadori, M.T., Liu, Y.: Fblg: A simple and effective approach for temporal dependence discovery from time series data. In: KDD (2014)
8. Granger, C.W.: Investigating causal relations by econometric models and cross-spectral methods. Econometrica pp. 424–438 (1969)
9. Kim, S., Putrino, D., Ghosh, S., Brown, E.: A Granger causality measure for point process models of ensemble neural spiking activity. PLOS Comp. Biology pp. 1–13 (2011)
10. Marx, A., Vreeken, J.: Causal inference on multivariate mixed-type data by minimum description length. ECMLPKDD (2018)
11. McIlhagga, W.: penalized: A matlab toolbox for fitting generalized linear models with penalties. Journal of Statistical Software, Articles (2016)
12. Mooij, J.M., Peters, J., Janzing, D., Zscheischler, J., Schölkopf, B.: Distinguishing cause from effect using observational data: Methods and benchmarks. Journal of Machine Learning Research (2016)
13. Nelder, J.A., Baker, R.J.: Generalized linear models. Encyclopedia of statistical sciences (1972)
14. Peters, J., Janzing, D., Schölkopf, B.: Causal inference on discrete data using additive noise models. IEEE Trans. patt. an. and mach. intelligence p. 2436—2450 (2011)
15. Qiu, H., Liu, Y., Subrahmanya, N.A., Li, W.: Granger causality for time-series anomaly detection. In: ICDM (2012)
16. Schreiber, T.: Measuring information transfer. Physical review letters **85**(2), 461 (2000)
17. Shimizu, S., Hoyer, P.O., Hyvärinen, A., Kerminen, A.: A linear non-gaussian acyclic model for causal discovery. Journal of Machine Learning Research **7**(Oct), 2003–2030 (2006)
18. Shojaie, A., Michailidis, G.: Discovering graphical Granger causality using the truncating lasso penalty. Bioinformatics (2010)
19. Tibshirani, R.: Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodological) pp. 267–288 (1996)
20. Zou, H.: The adaptive lasso and its oracle property. Journal of the Am. Stat. Assoc. pp. 1418–1429 (2008)