HAL open science

Graphs in State-Space Models for Granger Causality in Climate Science

Víctor Elvira, Emilie Chouzenoux, Jordi Cerda, Gustau Camps-Valls

To cite this version:

Víctor Elvira, Emilie Chouzenoux, Jordi Cerda, Gustau Camps-Valls. Graphs in State-Space Models for Granger Causality in Climate Science. Colloquium 2023 - When Causal Inference Meets Statistical Analysis, Apr 2023, Paris, France. hal-04165347

HAL Id: hal-04165347

https://inria.hal.science/hal-04165347v1

Submitted on 18 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Graphs in State-Space Models for Granger Causality in Climate Science

VÍCTOR ELVIRA, University of Edinburgh, UK
ÉMILIE CHOUZENOUX, Inria Saclay, France
JORDI CERDÀ, Universitat de València, Spain
GUSTAU CAMPS-VALLS, Universitat de València, Spain

Granger causality (GC) is often considered not an actual form of causality. Still, it is arguably the most widely used method to assess the predictability of a time series from another one. Granger causality has been widely used in many applied disciplines, from neuroscience and econometrics to Earth sciences. We revisit GC under a graphical perspective of state-space models. For that, we use GraphEM, a recently presented expectation-maximisation algorithm for estimating the linear matrix operator in the state equation of a linear-Gaussian state-space model. Lasso regularisation is included in the M-step, which is solved using a proximal splitting Douglas-Rachford algorithm. Experiments in toy examples and challenging climate problems illustrate the benefits of the proposed model and inference technique over standard Granger causality methods.

CCS Concepts: • **Computing methodologies** → **Modeling**; *Machine learning algorithms.*

Additional Key Words and Phrases: State-space models, Granger causality, Sparsity, Expectation-Maximisation, Earth and climate sciences

ACM Reference Format:
Víctor Elvira, Émilie Chouzenoux, Jordi Cerdà, and Gustau Camps-Valls. 2023. Graphs in State-Space Models for Granger Causality in Climate Science. In *CausalStats'23: When Causal Inference meets Statistical Analysis, April 17-21, 2023, Paris, France.* ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Granger causality (GC) [14] is undoubtedly the most widely used method to infer not causal but *predictability* relations from observational time series, and is the common tool in Earth sciences [25], neurosciences [24], social sciences [21] and finance [15]. Granger causality tests whether one time series is predictive of another beyond what can be explained by the other series alone. GC implicitly tells us about the amount of *information* present in the first time series that is not in the second time series and is helpful for its *forecast*. Other methods rely on similar concepts of *information flow* and *predictability*: connections can be established between GC and transfer entropy [29], directed information [22], convergent cross-mapping [30], Liang's measure of information flow [19], and with the graphical causal model perspective [33]. Recent structural equation models (SEM) can also be seen as particular instantiations of GC: the time series Linear non-Gaussian acyclic model (TS-LiNGAM) [16] can be viewed as a vector autoregressive (VAR) model with the non-Gaussian variable, and the time series models with independent noise (TiMINo) [23] can be seen as a non-linear-extension of a VAR model. These models require autoregressive modelling to estimate the variance of residuals, which makes GC struggle when the process has been filtered, down-sampled or observed with additive noise [4], which are common situations in many fields of science and engineering. Also, finite-order VAR models require choosing the order and do not allow incorporating latent states that may not be observed.

This work focuses on state-space models (SSMs) to identify the causal structure. SSMs model a hidden state evolving in a Markovian manner over discrete time. A sequence of observations is acquired, each being a transformed, partial, and noisy version of the hidden state. A common task is estimating the hidden state conditioned on available observations (Bayesian filtering). The task of estimating each state using the whole sequence of observations is called Bayesian smoothing. The linear-Gaussian SSM is widely used, and exact computation of filtering and smoothing distributions is possible via the Kalman filter and the Rauch-Tung-Striebel (RTS) smoother [27, Chapter 8]. Static parameter estimation is possible using expectation-minimisation (EM) or optimisation-based methods [27, Chapter 12].

We here review the GraphEM method [12] for estimating the linear matrix of a linear-Gaussian state-space model's state equation and evaluating its performance for GC in synthetic and real-life experiments in the Earth and climate sciences. Estimating such a matrix allows us to obtain valuable information about the hidden process, not only for inference purposes but also for understanding the uncovered relations among the state dimensions, in the line of graphical modelling methods for time series [2, 3, 11]. GraphEM estimates directed graphs representing causal dependencies between states of a multi-dimensional state containing uni-dimensional time series, with a sparsity constraint in the linear matrix. It applies the EM framework for MAP estimation, with the E-step using Kalman filtering and RTS smoothing and the M-step solving a Lasso-like problem through Douglas-Rachford proximal splitting [9]. Experiments on toy and real climate problem simulations show fast, practical convergence, and better performance than state-of-the-art methods.

2 BACKGROUND

Let us consider the following Markovian SSM,

$$\begin{aligned} \mathbf{x}_k &= \mathbf{A}\mathbf{x}_{k-1} + \mathbf{q}_k, \\ \mathbf{y}_k &= \mathbf{H}\mathbf{x}_k + \mathbf{r}_k, \end{aligned} \tag{1}$$

where, for $k = 1, \ldots, K$, $\mathbf{x}_k \in \mathbb{R}^{N_x}$ is the hidden state at time k, $\mathbf{y}_k \in \mathbb{R}^{N_y}$ is the associated observation, $\mathbf{A} \in \mathbb{R}^{N_x \times N_x}$, $\mathbf{H} \in \mathbb{R}^{N_y \times N_x}$, $\{\mathbf{q}_k\}_{k=1}^K \sim \mathcal{N}(0, \mathbf{Q})$ is the i.i.d. state noise process and $\{\mathbf{r}_k\}_{k=1}^K \sim \mathcal{N}(0, \mathbf{R})$ is the i.i.d. observation noise process. The state process is initialized as $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{x}_0; \bar{\mathbf{x}}_0, \mathbf{P}_0)$ with known $\bar{\mathbf{x}}_0$ and $\mathbf{P}_0$.

CausalStats'23, April 17-21, 2023, Paris, France
© 2023 Association for Computing Machinery.
This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *CausalStats'23: When Causal Inference meets Statistical Analysis, April 17-21, 2023, Paris, France*, https://doi.org/XXXXXXX.XXXXXXX.

The interest in the linear-Gaussian model of (1) resides on computing the filtering distributions $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ and the smoothing distributions $p(\mathbf{x}_k|\mathbf{y}_{1:K})$, with $\mathbf{y}_{1:k} = \{\mathbf{y}_j\}_{j=1}^k$. The Kalman filter provides exact forms of these as $p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \mathcal{N}(\mathbf{x}_k|\mathbf{m}_k, \mathbf{P}_k)$ [18]. The RTS smoother also yields exact smoothing distributions, $p(\mathbf{x}_k|\mathbf{y}_{1:K}) = \mathcal{N}(\mathbf{x}_k|\mathbf{m}_k^s, \mathbf{P}_k^s)$. To use either algorithm, the exact model parameters $\mathbf{A}$, $\mathbf{H}$, $\mathbf{Q}$, and $\mathbf{R}$ must be known [27]. In the next section, we introduce GraphEM [12] for the joint estimation of the unknown $\mathbf{A}$ and the hidden states.

3 THE GRAPHEM APPROACH

3.1 Role of sparsity and connection to GC

GraphEM approach, introduced in [8, 12] provides a MAP estimation of matrix $\mathbf{A}$ using a sparsity prior knowledge. Matrix $\mathbf{A}$ describes the relations among the dimensions of the hidden state, that is, the interactions of the state dimensions between consecutive time steps. A sparse transition matrix $\mathbf{A}$ allows us to interpret the multi-variate process under Granger causality (also called predictive causality) on the process $\mathbf{x}_k$ (i.e., not in the observed time series, unlike in standard Granger causality). The (n, m) entry in the $\mathbf{A}$ encodes the weight in which the mth time series in the hidden state affects the nth time series in the next time step, being zero if it does not have any effect. A null entry can be understood as if the mth time series does not bring any information to predict the nth time series (given the others).

3.2 EM methodology

GraphEM seeks the maximum of the MAP function $p(\mathbf{A}|\mathbf{y}_{1:K}) \propto p(\mathbf{A})p(\mathbf{y}_{1:K}|\mathbf{A})$, i.e., at solving

$$\text{minimise}_{\mathbf{A}\in\mathbb{R}^{N_x\times N_x}} \left(\varphi_K(\mathbf{A}) = -\log p(\mathbf{A}) - \log p(\mathbf{y}_{1:K}|\mathbf{A})\right). \quad (2)$$

In order to promote sparsity in the estimation of the $\mathbf{A}$ matrix, the regularisation function $\varphi_0(\mathbf{A}) = -\log p(\mathbf{A})$ is set as

$$(\forall \mathbf{A} \in \mathbb{R}^{N_x\times N_x}) \quad \varphi_0(\mathbf{A}) = \gamma\|\mathbf{A}\|_1, \quad (3)$$

with $\gamma > 0$ a penalisation weight. The ℓ_1 penalty, also known as Lasso, is convex and, as such, can be efficiently handled through the proximal optimisation method.

The direct minimisation of φ_K is difficult due to the intricate form of the likelihood term [27]:

$$\log p(\mathbf{y}_{1:K}|\mathbf{A}) = \sum_{k=1}^{K} \frac{1}{2}\log|2\pi\mathbf{S}_k(\mathbf{A})| + \frac{1}{2}\mathbf{z}_k(\mathbf{A})^\top\mathbf{S}_k(\mathbf{A})^{-1}\mathbf{z}_k(\mathbf{A}), \quad (4)$$

with $\mathbf{z}_k(\mathbf{A}) = \mathbf{y}_k - \mathbf{H}\mathbf{A}\mathbf{m}_{k-1}(\mathbf{A})$ and $\mathbf{S}_k(\mathbf{A})$ the covariance matrix of the predictive distribution $p(\mathbf{y}_k|\mathbf{y}_{1:k-1}) = \mathcal{N}(\mathbf{y}_k; \mathbf{H}\mathbf{A}\mathbf{m}_{k-1}(\mathbf{A}), \mathbf{S}_k(\mathbf{A}))$, both being sided products of the Kalman filter ran for a given $\mathbf{A}$ (see [27, Section 4.3]).

The EM approach adopted in GraphEM allows deriving an upper bound that is more tractable and, thus, easier to minimise. As shown in [12], the following function, parameterised by $\mathbf{A}'$, majorises the MAP objective function φ_K for every $\mathbf{A} \in \mathbb{R}^{N_x\times N_x}$:

$$Q(\mathbf{A};\mathbf{A}') = \frac{K}{2}\text{tr}\left(\mathbf{Q}^{-1}(\Sigma - \mathbf{C}\mathbf{A}^\top - \mathbf{A}\mathbf{C}^\top + \mathbf{A}\Phi\mathbf{A}^\top)\right) + \varphi_0(\mathbf{A}) + C \quad (5)$$

where tr is the trace operator, and C is a constant term independent from $\mathbf{A}$. The GraphEM algorithm is then deduced, following a majorisation-minimisation principle [17] as described in Alg. 1. The inner problem involved in the M-step is a convex Lasso-like problem that is solved through Douglas-Rachford iterations [9] given in the next Section. The convergence of Alg. 1 is established in [12, Th. 1].

Algorithm 1 GraphEM algorithm.

Initialisation of $\mathbf{A}^{(0)}$
For $i = 1, 2, \ldots$
  (E-step) Run the Kalman filter and RTS smoother, $\mathbf{A}' := \mathbf{A}^{(i-1)}$
    Construct $Q(\mathbf{A};\mathbf{A}^{(i-1)})$ with (5).
  (M-step) Update $\mathbf{A}^{(i)} = \text{argmin}_{\mathbf{A}}\left(Q(\mathbf{A};\mathbf{A}^{(i-1)})\right)$ (see Sec. 3.3)

3.3 Computation in the M-step

At an iteration $i \in \mathbb{N}$, the M-step minimises the function $Q(\mathbf{A};\mathbf{A}^{(i-1)})$, obtained by plugging $\mathbf{A}' := \mathbf{A}^{(i-1)}$ in Eq.(5). This minimisation problem, sometimes called Lasso regression [31], has been much studied in the literature of optimisation [1, 28], and most of the methods proposed to solve it rely on the proximity operator [5, 10].[1]

Let us decompose, for every $\mathbf{A} \in \mathbb{R}^{N_x\times N_x}$, $Q(\mathbf{A};\mathbf{A}^{(i-1)}) = f_1(\mathbf{A}) + f_2(\mathbf{A})$, where $f_1(\mathbf{A}) = \frac{K}{2}\text{tr}\left(\mathbf{Q}^{-1}(\Sigma - \mathbf{C}\mathbf{A}^\top - \mathbf{A}\mathbf{C}^\top + \mathbf{A}\Phi\mathbf{A}^\top)\right)$, and $f_2 = \varphi_0$. GraphEM relies on the Douglas-Rachford (DR) algorithm to minimise $f_1 + f_2$; DR is a fixed-point proximal strategy for convex optimisation that benefits from sound convergence guarantees [9] and has demonstrated its great practical performance in matrix optimisation problems related to graphical inference applications [6].

DR iterations are summarised in Alg. 2, in the simplified case when $\mathbf{Q} = \sigma_{\mathbf{Q}}^2\mathbf{Id}$ (see general case in [8]). DR generates a sequence $\{\mathbf{A}_j\}_{j\in\mathbb{N}}$ guaranteed to converge to a minimiser of f_1+f_2. In practice, for the iteration i in GraphEM, we run the DR method with $\theta = 1$ and initialisation $\mathbf{Z}_0 = \mathbf{A}^{(i-1)}$, i.e. the majorant function tangency point. Moreover, we stop the DR loop as soon as $|(f_1 + f_2)(\mathbf{A}_{j+1}) - (f_1 + f_2)(\mathbf{A}_j)| \leq \varepsilon$ (typically, $\varepsilon = 10^{-3}$), the DR output $\mathbf{A}_j$ being used for defining the GraphEM iterate $\mathbf{A}^{(i)}$.

Algorithm 2 Douglas-Rachford algorithm for M-step.

Set $\mathbf{Z}_0 \in \mathbb{R}^{N_x\times N_x}$ and $\theta \in (0, 2)$
For $j = 1, 2, \ldots$
  $\mathbf{A}_j = \left(\text{sign}((\mathbf{Z}_j)_{nm}) \times \max(0, |(\mathbf{Z}_j)_{nm}| - \theta)\right)_{1\leq n,m\leq N_x}$
  $\mathbf{V}_j = \left(\frac{\theta K}{\sigma_{\mathbf{Q}}^2}\mathbf{C} + 2\mathbf{A}_j - \mathbf{Z}_j\right)\left(\frac{\theta K}{\sigma_{\mathbf{Q}}^2}\Phi + \mathbf{Id}\right)^{-1}$
  $\mathbf{Z}_{j+1} = \mathbf{Z}_j + \theta(\mathbf{V}_j - \mathbf{A}_j)$

4 EXPERIMENTAL EVALUATION

4.1 Synthetic examples

We generate synthetic data using (1) with $N_x = N_y$. $\mathbf{A}$ is composed of b blocks of dimensions $(b_j)_{1\leq j\leq b}$ such that $N_x = \sum_{j=1}^b b_j$.

[1] See also http://proximity-operator.net/

Each diagonal block is randomly generated, and a projection with a maximum singular value of less than one is applied for stability. $\mathbf{H}, \mathbf{Q}, \mathbf{R}, \mathbf{P}_0$ are set to $\mathbf{Id}$, $K = 10^3$, $\sigma_{\mathbf{Q}}^2\mathbf{Id}$, $\sigma_{\mathbf{R}}^2\mathbf{Id}$, $\sigma_{\mathbf{P}}^2\mathbf{Id}$ respectively. Four datasets were created (Table 1), and performance was assessed using RMSE on $\mathbf{A}$ and precision, recall, specificity, accuracy, and F1 score for graph edge detection (the elements of the estimated matrix are clipped if smaller than 10^{-10} in absolute value).

Dataset	$(b_j)_{1\leq j\leq b}$	$(\sigma_{\mathbf{Q}}, \sigma_{\mathbf{R}}, \sigma_{\mathbf{P}})$
A	$(3, 3, 3)$	$(10^{-1}, 10^{-1}, 10^{-4})$
B	$(3, 3, 3)$	$(1, 1, 10^{-4})$
C	$(3, 5, 5, 3)$	$(10^{-1}, 10^{-1}, 10^{-4})$
D	$(3, 5, 5, 3)$	$(1, 1, 10^{-4})$

Table 1. Description of datasets

For each dataset, we ran the GraphEM algorithm with the stopping rule $|\varphi_K(\mathbf{A}^{(i)}) - \varphi_K(\mathbf{A}^{(i-1)})| \leq 10^{-3}$. A maximum number of 50 iterations was set, and it was always sufficient to reach this criterion. The algorithm is initialized by setting a random and dense matrix $\mathbf{A}^{(0)}$, ensuring the stability of the process. Parameter γ, balancing the weight of the sparse prior, is optimized on a single realisation thanks to a manual grid search to maximize the accuracy score. We also provide the ML estimator's results, computed using the EM algorithm (MLEM). In this case, the M-step has a closed-form solution [27, Th.12.5]. In addition, we compare two Granger-causality approaches [7] for graphical modelling. The first algorithm, pairwise Granger Causality (PGC) explores the $N_x(N_x\text{-}1)$ possible dependencies among two nodes, at each time independently from the rest. The second approach, the conditional Granger Causality (CGC), operates similarly. Still, for each pair of nodes, it also considers the information of the other $N_x - 2$ signals to evaluate whether one node provides information to the other when the rest of the signals are observed. We also do a manual grid search for fine-tuning the parameters of both PGC and CGC (more information can be found in [20]). PGC and CGC do not estimate a weighted graph but a binary one, so RMSE is not calculated in those cases.

The results averaged on 50 realisations are presented in Table 3. MLEM does not promote sparsity in the graph, which explains the poor results in terms of edge detectability. Moreover, GraphEM provides a better RMSE score on all examples. Regarding the graph structure, we can observe that GraphEM has also better detection scores when compared with both PGC and CGC. We also display an example of graph reconstruction for dataset C in Fig. 1, illustrating the ability of GraphEM to recover the graph shape and weights.

4.2 Climate modelling and understanding

Studying the complex inter-dependencies in climate processes is a critical challenge. We use climate models and analytical data to gain insight through observational causal discovery [25]. We evaluate graphEM and compare it to other methods on climate-related problems, such as linking ENSO and the North Atlantic Oscillation.

We used pre-industrial control runs from [13], followed the methodology in [26], and selected 15 climate variables (hfls, hfss, huss, rlds, rlus, rlut, ta, tas, tasmax, tasmin, uas, va, vas, wap, zg). We

	method	RMSE	accur.	prec.	recall	spec.	F1
A	GraphEM	0.081	0.9104	0.9880	0.7407	0.9952	0.8463
	MLEM	0.149	0.3333	0.3333	1	0	0.5
	PGC	-	0.8765	0.9474	0.6667	0.9815	0.7826
	CGC	-	0.8765	1	0.6293	1	0.7727
B	GraphEM	0.082	0.9113	0.9914	0.7407	0.9967	0.8477
	MLEM	0.148	0.3333	0.3333	1	0	0.5
	PGC	-	0.8889	1	0.6667	1	0.8
	CGC	-	0.8889	1	0.6667	1	0.8
C	GraphEM	0.120	0.9231	0.9401	0.77	0.9785	0.8427
	MLEM	0.238	0.2656	0.2656	1	0	0.4198
	PGC	-	0.9023	0.9778	0.6471	0.9949	0.7788
	CGC	-	0.8555	0.9697	0.4706	0.9949	0.6337
D	GraphEM	0.121	0.9247	0.9601	0.7547	0.9862	0.8421
	MLEM	0.239	0.2656	0.2656	1	0	0.4198
	PGC	-	0.8906	0.9	0.6618	0.9734	0.7627
	CGC	-	0.8477	0.9394	0.4559	0.9894	0.6139

Table 2. Results for GraphEM, MLEM, PGC and CGC.

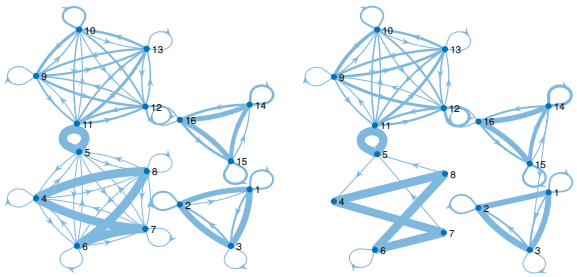

Fig. 1. True graph (left) and GraphEM estimate (right) for dataset C.

de-seasonalized the time series and applied Varimax-rotated PCA to monthly averages. Then used the obtained weights to generate daily component time series, averaging to a 5-day resolution. We randomly picked N component time series, standardised them, and constructed the ground truth by fitting a VAR model as in [26]. Ground truth was obtained by thresholding coefficients < 0.22 and keeping random draws of N components with $L = N$ links. We then constructed experiments scenarios modelling climate [26]. We use linear models and two sample sizes ($T \approx 100 - 250$ and $N = 5, 40$) and accounted for non-stationarity and observational noise, resulting in 12 experiments with 200 realisations each. We implemented GraphEM in our CAUSEME web platform and compared the performance to standard algorithms for time series data: VAR [32], GC [14], and PCMCI [25]. Results in Table 3 show that the GraphEM methods outperform VAR and GC in all performance metrics (accuracy, precision, recall, specificity and F1) and PCMCI in recall and overall F1 score. Figure 2 shows the obtained graphs for a particular example and evidences the good detection of links $2, 4 \longrightarrow 5$ unlike with PCMCI, as well as the much sparser (and less convoluted) solution attained compared to VAR and GC.

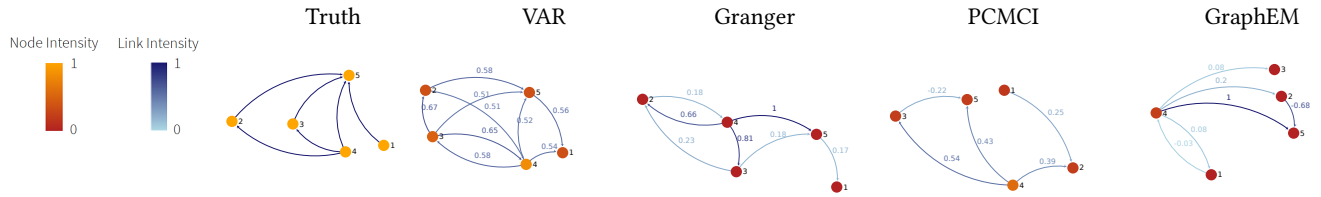

Fig. 2. Causal graphs extracted by different methods in an illustrative climate problem.

method	best hyperparameters	accur.	prec.	recall	spec.	F1
GraphEM [12]	$\sigma_{\mathbf{R}} = 0.1, \sigma_{\mathbf{P}} = 10^{-4}, \gamma_1 = 50$	**0.72**	0.75	**0.55**	0.86	**0.63**
VAR [32]	$\ell = 8$	0.56	0.50	0.46	0.64	0.48
Granger [14]	$\ell = 8$	0.6	0.57	0.36	0.79	0.44
PCMCI [25]	$\tau_{\max} = 8, \alpha_{\mathrm{PC}} = 0.05$, ParCorr	0.72	**0.83**	0.45	**0.93**	0.59

Table 3. Results for GraphEM [12] (Lasso), VAR [32], GC [14] and PCMCI [25] for a set of climate problems. Best results are highlighted in bold.

5 CONCLUSION

GraphEM is an EM method for estimating the linear operator in linear-Gaussian state-space models. GraphEM incorporates a sparsity constraint using a Lasso penalty term, which makes it well suited for modelling the state entries interactions as a compact and interpretable graph. The inner minimisation problem is solved using a proximal splitting algorithm. Numerical results show that GraphEM outperforms several other techniques for graphical modelling. Results in more challenging real problems in the Earth, Social and Climate sciences will be illustrated at the time of the conference.

6 ACKNOWLEDGEMENTS

The work of V. E. is supported by the *Agence Nationale de la Recherche* of France under PISCES (ANR-17-CE40-0031-01), the Leverhulme Research Fellowship (RF-2021-593), and by ARL/ARO under grants W911NF-20-1-0126 and W911NF-22-1-0235. É.C. acknowledges support from the European Research Council (ERC) Starting Grant MAJORIS ERC-2019-STG-850925, and G.C-V. acknowledges support from the ERC Synergy Grant USMILE (grant agreement 855187), the Fundación BBVA with the project 'Causal inference in the human-biosphere coupled system (SCALE)', and the Microsoft Climate Research Initiative through the Causal4Africa project.

REFERENCES

[1] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski. 2012. Optimization with Sparsity-Inducing Penalties. *Foundations Trends Machine Learning* 4, 1 (Jan. 2012), 1–106.
[2] F. R. Bach and M. I. Jordan. 2004. Learning graphical models for stationary time series. *IEEE Transactions on Signal Processing* 52, 8 (Aug. 2004), 2189–2199.
[3] D. Barber and A. T. Cemgil. 2010. Graphical Models for Time-Series. *IEEE Signal Processing Magazine* 27, 6 (Nov 2010), 18–28.
[4] L Barnett and A.K. Seth. 2015. Granger causality for state-space models. *Physical Review E* 91, 4 (2015), 040101.
[5] H. H. Bauschke and P. L. Combettes. 2017. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York. 2nd edition.
[6] A. Benfenati, E. Chouzenoux, and J.-C. Pesquet. 2018. *A Proximal Approach for a Class of Matrix Optimization Problems*. Technical Report.
[7] S. L. Bressler and A. K. Seth. 2011. Wiener–Granger causality: a well established methodology. *NeuroImage* 58, 2 (2011), 323–329.
[8] E. Chouzenoux and V. Elvira. 2020. GraphEM: EM Algorithm for Blind Kalman Filtering under Graphical Sparsity Constraints. In *ICASSP*. 5840–5844.
[9] P.L. Combettes and J.C. Pesquet. 2007. A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE Journal of Selected Topics in Signal Processing* 1, 4 (Dec. 2007), 564–574.
[10] P.L. Combettes and J.C. Pesquet. 2010. *Proximal Splitting Methods in Signal Processing*. Springer-Verlag, New York.
[11] M. Eichler. 2012. Graphical modelling of multivariate time series. *Probability Theory and Related Fields* 153, 1 (01 Jun. 2012), 233–268.
[12] V. Elvira and E. Chouzenoux. 2022. Graphical Inference in Linear-Gaussian State-Space Models. *IEEE Transactions on Signal Processing* 70 (2022), 4757–4771.
[13] V. Eyring, S. Bony, G.A. Meehl, C.A. Senior, Bjorn Stevens, R.J. Stouffer, and K.E. Taylor. 2016. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development* 9, 5 (2016), 1937–1958.
[14] C. W. J. Granger. 1969. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* 37, 3 (1969), 424–438.
[15] J. Hicks et al. 1980. *Causality in economics*. Australian National University Press.
[16] A. Hyvärinen, K. Zhang, S. Shimizu, and P.O. Hoyer. 2010. Estimation of a structural vector autoregression model using non-gaussianity. *Journal of Machine Learning Research* 11, May (2010), 1709–1731.
[17] M.W. Jacobson and J.A. Fessler. 2007. An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms. *IEEE Transactions on Image Processing* 16, 10 (2007), 2411–2422.
[18] R. E. Kalman. 1960. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering* 82 (1960), 35–45.
[19] X. San Liang. 2016. Information flow and causality as rigorous notions ab initio. *Phys. Rev. E* 94 (Nov 2016), 052201. Issue 5.
[20] D. Luengo, G. Rios-Mu noz, V. Elvira, C. Sánchez, and A. Artés-Rodríguez. 2019. Hierarchical Algorithms for Causality Retrieval in Atrial Fibrillation Intracavitary Electrograms. *IEEE Jour. Biomed. Hea. Inf.* 12, 1 (Jan 2019), 143–155.
[21] M.M. Marini and B. Singer. 1988. Causality in the social sciences. *Sociological methodology* 18 (1988), 347–409.
[22] J. Massey. 1990. Causality, feedback and directed information. In *Proc. Int. Symp. Inf. Theory Applic.(ISITA-90)*. Citeseer, 303–305.
[23] J. Peters, D. Janzing, and B. Schölkopf. 2013. Causal inference on time series using restricted structural equation models. In *Advances in Neural Information Processing Systems*. 154–162.
[24] A.T. Reid, D.B. Headley, R.D. Mill, R. Sanchez-Romero, L.Q. Uddin, D. Marinazzo, D. J Lurie, P.A. Valdés-Sosa, S.J. Hanson, and B.B. Biswal. 2019. Advancing functional connectivity research from association to causation. *Nature Neuroscience* 1, 10 (2019).
[25] J. Runge, S. Bathiany, E. Bollt, G. Camps-Valls, D. Coumou, E. Deyle, C. Clymour, M. Kretschmer, M. Mahecha, J. Muñoz-Marí, E. van Nes, J. Peters, R. Quax, M. Reichstein, M. Scheffer, B. Schölkopf, P. Spirtes, G. Sugihara, J. Sun, K. Zhang, and J. Zscheischler. 2019. Inferring causation from time series with perspectives in Earth system sciences. *Nature Communications* 10, 2553 (2019).
[26] J. Runge, X.-A. Tibau, M. Bruhns, J. Muñoz Marí, and G. Camps-Valls. 2020. The Causality for Climate Competition. In *NeurIPS 2019 Competition and Demonstration Track*, Hugo Jair Escalante and Raia Hadsell (Eds.), Vol. 123. PMLR, 110–120.
[27] S. Sarkka. 2013. *Bayesian Filtering and Smoothing* (3 ed.).
[28] M. Schmidt, G. Fung, and R. Rosales. 2007. Fast Optimization Methods for L1 Regularization: A Comparative Study and Two New Approaches. In *Machine Learning: ECML 2007*, J. N. Kok, J. Koronacki, R.L. Mantaras, S. Matwin, D. Mladenič, and A. Skowron (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 286–297.
[29] Thomas Schreiber. 2000. Measuring Information Transfer. *Phys. Rev. Lett.* 85 (Jul 2000), 461–464. Issue 2.
[30] G. Sugihara, R. May, H. Ye, C.-h. Hsieh, E. Deyle, M. Fogarty, and S. Munch. 2012. Detecting Causality in Complex Ecosystems. *Science* 338, 6106 (2012), 496–500.
[31] R. Tibshirani. 1996. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 58, 1 (1996), 267–288.
[32] H. Toda. 1991. *Vector autoregression and causality*. Yale University.
[33] H. White, K. Chalak, X. Lu, and others. 2011. Linking Granger Causality and the Pearl Causal Model with Settable Systems.. In *NIPS Mini-Symposium on Causality in Time Series*. 1–29.

Received X; revised Y; accepted Z