Brain Multiphysics 3 (2022) 100044

Contents lists available at ScienceDirect

Brain Multiphysics

journal homepage: www.sciencedirect.com/journal/brain-multiphysics

"Technical considerations on the use of Granger causality in neuromonitoring"

Michał M. Placek [a,*], Erta Beqiri [a], Marek Czosnyka [a,b], Peter Smielewski [a]

[a] Brain Physics Laboratory, Division of Neurosurgery, Department of Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom
[b] Institute of Electronic Systems, Warsaw University of Technology, Poland

ARTICLE INFO

Keywords:
Traumatic brain injury
Granger causality
Neurocritical care
Physiological signal analysis
Brain monitoring
Neuromonitoring
Multivariate time series analysis

ABSTRACT

Neuromonitoring-derived indices play an important role in implementing personalised medicine for traumatic brain injury patients. A well-established example is the pressure reactivity index (PRx), calculated from spontaneous fluctuations of arterial blood pressure (ABP) and intracranial pressure (ICP). PRx assumes causal relationship between ABP and ICP but lacks the check for this assumption. Granger causality (GC) — a method of assessing causal interactions between time series data — is gaining popularity in neurosciences. In our work, we used ABP and ICP data recorded at the frequency of 100 Hz or higher from 235 traumatic brain injury patients. We focused on time domain GC. Analysis was first performed directly on high-resolution data, which included pulse waves. We showed that due to the measurement delay in high-resolution ABP data, GC analysis may erroneously indicate strong ICP→ABP causal relation. Subsequently, the data were downsampled to 0.1 Hz, effectively removing pulse and respiratory waves. We aimed to investigate how different ways of calculating GC influence results and which way should be recommended for ABP-ICP recordings. We considered aspects like selecting autoregressive model order and dealing with data non-stationarity. In addition, we generated simulated signals to investigate the influence of gaps and different procedures of missing data imputation on GC estimation. We showed that unlike methods which interpolate missing data, replacing missing data by white Gaussian noise did not increase the rate of false GC detection. Python source code used in this study is available at: https://github.com/m-m-placek/python-icmplus-granger-causality.

Statement of significance: Assessing causality between time series data is of particular interest when neuromonitoring indices are derived from those time series and causal interaction between them is assumed. Causality assessment can improve reliability of such indices and open pathways for their safe clinical implementation. Granger Causality (GC) has recently been investigated in data collected from traumatic brain injury patients. However, there are two main issues related to applications suggested in these studies. Firstly, they considered GC for entire multi-day data recordings or for 24-h long episodes. There is interest in considering causal relationships in finer granularity, also in terms of their potential real-time applications at the bedside. Secondly, GC calculation requires selecting some parameters and there is no unique nor standardised way of doing that. Many papers often provide very brief description of data pre-processing and GC calculation. For this reason, it can be harder to reproduce and compare results derived from GC application. Different ways of obtaining GC may potentially lead to inconsistent results. Here, we attempted to explore possibility of time-varying GC of finer granularity and to provide general guidelines for application of GC to neurocritical care time series affected by periods of missing values.

Abbreviations

AIC	– Akaike information criterion
ABP	– arterial blood pressure
BIC	– Bayesian information criterion
GC	– Granger causality
ICP	– intracranial pressure
PRx	– pressure reactivity index
Q1–Q3	– the first and the third quartile
SD	– standard deviation

* Corresponding author.
E-mail address: mp963@cam.ac.uk (M.M. Placek).

https://doi.org/10.1016/j.brain.2022.100044
Received 8 October 2021; Received in revised form 22 December 2021; Accepted 21 January 2022
Available online 23 January 2022

TBI – traumatic brain injury
VAR – vector autoregressive

1. Introduction

Traumatic brain injury (TBI) is the most common cause of death in young adults and a major cause of disability worldwide [1]. There is growing interest in treating TBI patients according to personalised medicine. Data-driven approaches and especially neuromonitoring-derived indices have been investigated within this scope [2]. In particular, monitoring cerebral autoregulation using the pressure reactivity index (PRx) has gained momentum in the recent years [3–6].

Cerebral autoregulation is a physiological protective mechanism that maintains an adequate cerebral blood flow despite changes in cerebral perfusion pressure [7,8], which can be defined as the difference between arterial blood pressure (ABP) and intracranial pressure (ICP). In the acute critical phase of TBI pathophysiology, cerebral autoregulation is often disturbed, leading to high risk of both ischaemic or hyperaemic secondary insults. The impairment of cerebral autoregulation was shown to be strongly associated with poor outcome for severe TBI patients admitted in critical care unit [4,9].

Cerebral autoregulation relies on the brain vessels capacity to constrict or dilate in response to blood pressure fluctuations. Variations in vessels diameter (which happen at low frequencies [10]) causes changes in cerebral blood volume, which ultimately induces changes in ICP, according to the brain pressure-volume curve. The pressure reactivity index (PRx) is defined as the moving correlation coefficient between spontaneous, vasogenic waves in ABP and ICP. Thus, PRx provides means for continuous assessment of cerebral autoregulation in TBI patients who require ICP monitoring.

The clinical importance of PRx is multifactorial and well established. PRx has an independent impact on outcome when added to well-defined TBI prognostic models [9] and was validated against more direct measures of cerebral blood flow regulation in humans [11]. Individualised, dynamic cerebral perfusion pressure targets can be derived from PRx in order to optimize cerebral autoregulation [4–6]. However, there are pitfalls in the current methodology of PRx [12]. For instance, PRx relies on the assumption that the variability in ICP slow waves is entirely driven by the variability in extracranial ABP.

The causal relationship between spontaneous oscillations of ABP and ICP varies across the frequencies taken into consideration. The causal ABP→ICP relationship is expected to be maintained when looking at phenomena happening at higher frequencies corresponding to the heart rate. In fact, pulsations generated by the beating heart are transmitted via the circulatory system to other organs, including the brain [13]. For lower frequencies, however, ICP waves can be driven not only by ABP but also by physiological mechanisms such as neurovascular coupling and metabolic adjustments [3,8,14]. The causal ABP→ICP relation can manifest itself when there are changes in ICP resulting from controlled manipulations in ABP, e.g. pharmacologically [15]. On the other hand, when only spontaneous slow fluctuations of ABP and ICP are analysed, signal-to-noise ratio becomes low, and the nature of the relationship might be questionable and more subject to confounding influences [5]. This would potentially seriously diminish robustness of indices that assume such causal relationship, like PRx. This, in turn, would ultimately lead to risk of incorrect decision making in managing patients according to individualised, PRx-based cerebral perfusion pressure targets [4–6]. Therefore, the scientific community has investigated ways of improving reliability and robustness of the PRx index. One example was the recent attempt to filter out unreliable estimates of PRx using auxiliary measures like coherence [16]. Coherence, however, as a non-directional measure of coupling between variables, cannot judge about the direction of information flow. Hence, there is a need for implementation of the assessment of causality between ABP and ICP signals. This would naturally translate in improvement of reliability of metrics that are based on the relationship between the above-mentioned time series.

Granger causality (GC) is a method of assessing causal interactions between time series data. Originally developed in the context of econometrics [17], GC achieved broad application in other fields, including neurosciences [18,19]. Studying causal links using GC does not have to be limited to a single organ, e.g. signals recorded from different regions of the brain [20]. Recently, GC has also been applied in data collected from TBI patients to study causal relations between the cardiovascular, cerebrovascular, and autonomic systems described by ABP, ICP [21], and heart rate time series [22,23]. In those papers, however, GC was applied to the entire patient's recording [21] or limited to one day [22,23], which would not be appropriate for guiding a sliding-window-based analysis of ABP and ICP signals.

One of the essential issues that needs to be considered in analysing neurocritical care monitoring data is presence of artefacts and gaps. The longer gaps in neurocritical care monitoring data arise when the patient is taken for a brain scan or a surgery. Shorter intervals of invalid data may originate from sensor disconnection, misplacement, recalibration routines, arterial line flushing or movement artefacts (nursing activities). In addition, when ICP is measured using an external ventricular drain, the pressure waveform is frequently distorted, particularly during the drainage of cerebrospinal fluid. A robust method applied in real time using a sliding window approach must take effects of these data flow interruptions into account.

In this paper, we have set out to explore properties of GC when applied to shorter data segments, and to propose a set of recommendations for its use. In particular, we investigated properties, like the model order selection, data stationarity, the method of parameters estimation, and data gaps treatment. Detailed mathematical background for these properties can be found in Appendix.

The main objectives were: 1) to explore viability of time-varying GC of finer granularity; 2) to recommend a standardised way of GC estimation in ABP-ICP data and investigate how GC inference is impacted when different calculation methods are used; 3) to provide general guidelines for application of GC in neurocritical care data affected by gaps and artefacts.

2. Materials & methods

Detailed mathematical background for Granger causality (GC) can be found in Appendix. Here, we only mention basic terms. GC quantifies the ability of unique past information in one time series to improve prediction of future values of another time series [17–19,24–26]. The higher the magnitude of GC is, the stronger that ability is. Significance of GC is assessed in terms of hypotheses testing, where the null hypothesis is 'zero GC.' GC formalism is based on the vector autoregressive (VAR) modelling and hence, it requires selecting the model order m. Fitted VAR model should be stable, meaning that it constitutes a covariance-stationary process. This assumption can be verified by checking if the spectral radius $\rho(A)$ is lower than 1 [26,27]. When data does not fulfil stationarity assumption, differencing can be applied to make data stationary and then perform GC analysis [28].

2.1. Dataset

As a representative real neurocritical care dataset, we used data recorded from 235 moderate to severe TBI patients. The cohort consisted of 77.9% males and 22.1% females. Median (Q1–Q3) age was 49 (30–62) years and Glasgow Coma Scale at admission was 6 (3–9). The patients were admitted to neuro-intensive care unit in Addenbrooke's Hospital (Cambridge, UK). Multi-day, high-resolution time series, including ABP and ICP, were recorded from patients' bedside monitors as a part of standard patient care. The use of the anonymised data was approved by the institutional Research Ethics Committee (29 REC 97/291). Separate informed consent for this analysis was not required. ABP was monitored invasively through either the radial or femoral artery

using a standard pressure monitoring kit (Baxter Healthcare, Cardio-Vascular Group, Irvine, CA, USA). ICP was monitored using an intra-parenchymal probe (Codman ICP MicroSensor, Codman & Shurtleff, Raynham, MA, USA) inserted into the frontal cortex. ABP and ICP were sampled at the frequency no lower than 100 Hz and recorded using ICM+ software (Cambridge Enterprise, Cambridge, UK). Artefacts were marked in the first stage automatically and then manually. The automatic detection involved simple heuristics including expected ranges of physiological variables and their peak-to-peak values, the presence of pulse waveform. Artefacts segments were removed from the data.

2.2. Illustrative Granger causality on high-resolution ABP-ICP data

Illustrative GC analysis was first performed directly on one recording containing high-resolution data resampled to the frequency of 50 Hz, which preserved the shape of ABP and ICP pulse waveforms well enough. This case study was done to check for the influence of measurement delays on the reliability of GC metric. Complementary ABP-ICP GC analyses on the high-resolution data were done when ICP signal was additionally delayed by 100 and 200 milliseconds to attempt to compensate for the measurement delay in ABP. Data were divided into 10-seconds long non-overlapping slices. For each slice, the optimal model order was selected individually using the Bayesian information criterion (BIC) from the range 1–50, and then ABP→ICP and ICP→ABP GC magnitudes were calculated.

2.3. Downsampling and data processing

Apart from one illustrative analysis performed on high-resolution data, all other analyses on real data were conducted upon downsampling ABP and ICP signals to the frequency of 0.1 Hz by averaging them in the 10-seconds long non-overlapping slices, effectively removing pulse and respiratory waves [5,12,22,23,29]. Downsampled time series were divided into segments with 75% overlap to mimic the way the indices like PRx are calculated using the sliding window approach. Unless otherwise specified, the majority of analyses were performed on three different segment lengths: 20 min (120 samples), 1 h (360 samples), and 4 h (1440 samples). In each individual data segment, ABP and ICP were standardised to zero mean and unit standard deviation (SD) prior to calculation of ABP→ICP and ICP→ABP GC estimates. When evaluating significance of GC, the threshold of $p = 0.05$ was assumed.

2.4. The issue of missing data

The purpose of the data gaps analysis in the ABP-ICP recordings was to obtain descriptive statistics about gaps occurring in practice. Missing data for a pair of two signals were identified in such time instants where at least one signal had missing data point. Two measures describing missing data in a data segment were introduced: 1) the longest gap (expressed in number of samples) in the data segment and 2) the percentage of missing data points in data segment. Data segments that contained no valid data points at all were omitted and did not count towards the total number (100%) of all data segments. We examined how the number of data segments that have to be excluded depends on assumed threshold for missing data limits.

The influence of gap lengths and different missing data handling procedures on the rejection rates of the null-hypothesis of zero GC were investigated using Monte Carlo simulations approach. We defined the rejection rate as the percentage of cases when the null hypothesis was rejected to the number of all Monte Carlo repetitions. Process Y was assumed to be white Gaussian noise with zero mean and unit standard deviation (SD). Process X was created by adding two components: Y lagged by 3 samples and the innovation process. For the latter, another white Gaussian noise was used, with the SD two times higher than SD of Y. X was subsequently standardised to unit SD. Such a scenario corresponds to $GC_{Y\to X}$ of a relatively high magnitude. In the second scenario, both Y and X were assumed to be uncorrelated white Gaussian noise processes. $GC_{Y\to X}$ was tested in the two scenarios at a significance level of 0.05. In the first scenario, we can expect very high null-hypothesis rejection rates, close to 100%. In the second scenario, on the other hand, we expect the rejection rate close to 5% because of the defined significance level of 0.05. However, missing data and the way of handling them may inflate the rates of false positives (erroneous causality) and false negatives (erroneous non-causality). To investigate this effect, gaps, i.e. continuous blocks of missing data points, were created in the generated data. Missing data were then replaced using four strategies: 1) previous non-missing value, 2) nearest neighbour, 3) linear interpolation of neighbouring non-missing values, or 4) imputation by white Gaussian noise with an SD equal to the SD of the remaining valid data in the segment. After dealing with missing data, GC was tested at a significance level of 0.05, assuming model order $m = 3$. The rejection rates of the null hypothesis of zero GC were estimated based on 10,000 repetitions.

2.5. Determination of the optimal model order in real data

Optimal model orders according to the Akaike (AIC) [30] and the Bayesian information criterion (BIC) [31] were calculated for each ABP-ICP segment having no more than 10% of missing data. Then, the median value of the optimal order was computed across all data segments for each recording. Finally, the median (Q1–Q3) of the optimal order was calculated in the analysed cohort.

Prior to GC calculation, the optimal model order can be selected individually for each single ABP-ICP data segment, i.e. a varying-order approach. In order to reduce the computation time, we have pursued another approach where one fixed model order would be pre-selected based on a training set and applied equally to all ABP-ICP data segments of a particular length. To make sure that the fixed order approach can be used instead of the varying order approach, we checked if there were substantial differences in the median magnitudes of $GC_{ABP\to ICP}$ as well as in the proportion of time with significant $GC_{ABP\to ICP}$ between these two approaches. The latter was defined as the percentage of data subsegments where there was a statistically significant GC in respect to all valid (for GC analysis) data segments. This auxiliary analysis was performed only for 1-hour long data segments and using the BIC as the optimal order selection criterion.

2.6. Stationarity of data subsegments

Assuming VAR model orders of real ABP-ICP data based on the results of the previous section analysis, stationarity was investigated by checking if the spectral radius $\rho(A)$ was lower than 1 for each data segment. The percentage of data segments fulfilling this criterion as well as median spectral radius were calculated for each recording. Median and interquartile range (Q1–Q3) values of these metrics were reported. Finally, we also counted the number of recordings containing at least one ABP-ICP data segment that could not be made covariance stationary after one application of differencing.

To investigate potential influence of differencing on GC estimates in ABP-ICP data, we compared within-subject median GC magnitudes obtained for data without and after one application of differencing, regardless of stationarity checking, using Wilcoxon signed-rank test. Within-subject correlation coefficients were calculated between data without and after differencing. Between-subject median (Q1–Q3) values of these coefficients were reported. This comparison was performed for 20-minutes long data segments.

2.7. Comparison of single vs double regression method of GC estimation

ABP→ICP and ICP→ABP GC estimates for 20-minutes long data segments obtained using the two different algorithms, i.e. the double

and single regression approach, were compared using Wilcoxon signed-rank test. Apart from GC magnitudes, also a measure of volatility of GC significance was calculated for each recording. This measure was defined as the number of transitions in statistical significance of GC divided by the number of all data segments that were valid for GC analysis. Obviously, the higher its value, the higher the number of transitions within the recording.

Since the number of iterations in the single regression algorithm grows to infinity when the spectral radius $\rho(A)$ approaches 1 (see Appendix for more detail), the threshold for $\rho(A)$ was diminished from 1 to 0.99. To make a fair comparison between the two methods of GC estimation, the same diminished value of the threshold for $\rho(A)$ was also applied in the double regression method in this particular sub-analysis.

2.8. Statistical comparisons

The study involved comparing GC-based indices obtained from the same data sample using different algorithms (the fixed vs the varying model order approach, with vs without differencing, and the double vs the single regression method). Not all variables met assumption of normal distribution. Subsequently, nonparametric Wilcoxon signed-rank test was chosen. Apart from p-values, effect size r was reported

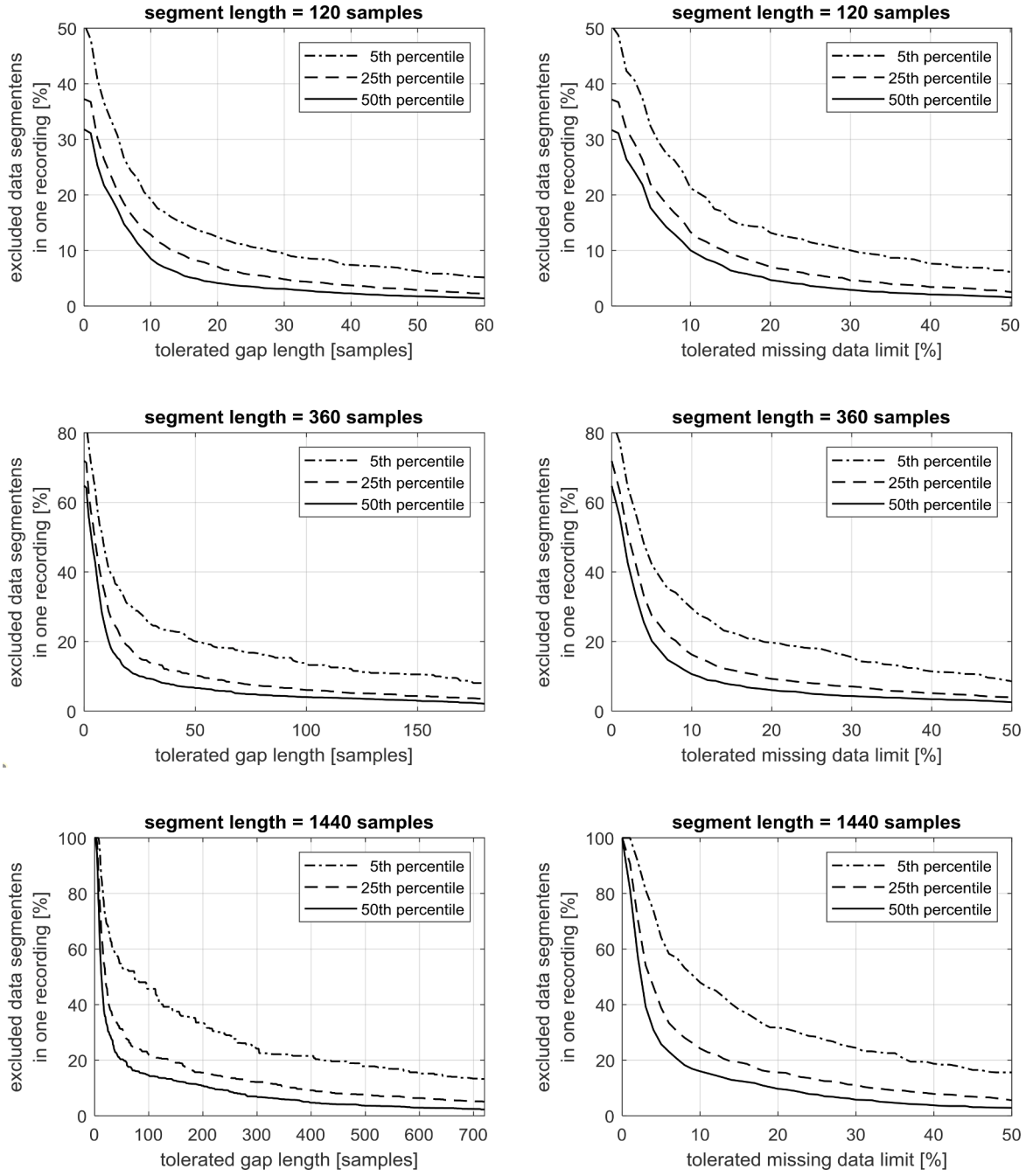

Fig. 1. Percentage of excluded data segments depending on the assumed threshold for missing data limits, i.e. the longest tolerated gap (the left column) and the percentage of missing data points in data segment (the right column). Lines represent different percentiles of recordings that will have more than a vertical-axis-specified percentage of their data segments excluded when a threshold given in the horizontal axis is assumed. For instance, if the maximal tolerated gap length is 10 samples, then a half of the recordings (50^{th} percentile) will have more than 8.5% of their 20-minutes long data segments excluded, and 5% of recordings (5^{th} percentile) will have more than 19.2% of their data segments rejected (the upper left-hand corner).

($r = Z/\sqrt{n}$, where Z stands for the Z-score and n is the sample size) [32, 33]. Similarly to Pearson's correlation coefficient, r is bounded from –1 to +1. Its sign indicates the direction, and the higher its magnitude, the larger the effect size. Statistical comparisons were performed in MATLAB® R2020a (MathWorks®, Natick, MA, USA).

2.9. Granger causality implementation in Python and ICM+

Algorithms for GC estimation together with auxiliary functions were implemented in the form of the extension for ICM+ environment, through the recently developed Python–ICM+ plugin interface [34]. The double regression approach was implemented using the Python open-source function (*statsmodels.tsa.stattools.grangercausalitytests*, v0.11.1) with small modifications allowing *addConst=False* option, which led to obtaining practically the same results as the GCCA MATLAB toolbox [25]. The single regression approach was developed in Python for bivariate models only, basing on the MVGC v1.0 MATLAB toolbox [26]. Python source code used in this study is available at: https://github.com/m-m-placek/python-icmplus-granger-causality.

3. Results

3.1. Granger causality in high-resolution ABP-ICP data

When Granger causality (GC) analysis was performed directly on high-resolution data, which included pulse waves, persistent, strong and significant $GC_{ICP\rightarrow ABP}$ was detected. Moreover, in this case, the magnitude of $GC_{ICP\rightarrow ABP}$ was many times higher than the one of $GC_{ABP\rightarrow ICP}$ (for one illustrative recording, it was 1.71 (1.20–2.18) vs 0.11 (0.07–0.39)). The character of causality, however, was easily reversed by introducing delay to the recorded ICP signal. When ICP was delayed by 100 milliseconds, the magnitudes of $GC_{ICP\rightarrow ABP}$ and $GC_{ABP\rightarrow ICP}$ became close to each other, and when ICP was delayed by 200 milliseconds, the magnitude of $GC_{ABP\rightarrow ICP}$ turned out to be many times higher than the one of $GC_{ICP\rightarrow ABP}$.

3.2. Missing data and its influence on GC estimation

Fig. 1 illustrates how many ABP-ICP data segments have to be excluded when given threshold for missing data limit is assumed. The greater the maximal length of tolerated gap or percentage of missing data points is, the smaller part of data needs to be excluded.

Fig. 2 presents how the rejection rates of the null hypothesis of zero Granger causality (GC) on simulated data depend on gap lengths and a way of imputing missing data. Fig. 2a shows the rejection rates when non-zero GC is simulated, whereas Fig. 2b refers to the rejection rates when both time series are unrelated and hence no causal relation is expected.

3.3. Optimal model order

Table 1 shows the optimal vector autoregressive (VAR) model order analysis for ABP-ICP data and three different segment lengths.

Table 2 presents a comparison between the values of ABP→ICP GC-related parameters when the fixed and the varying model order strategy

Table 1
Optimal orders for ABP-ICP vector autoregressive model.

Segment length	Median optimal model order according to information criterion	
	AIC	BIC
20 min	4 (4–5)	2 (2–2)
1 h	8 (7–10)	3 (3–4)
4 h	15 (13–18)	5 (5–6)

Values are: between-subject median (Q1–Q3). AIC – Akaike information criterion; BIC – Bayesian information criterion.

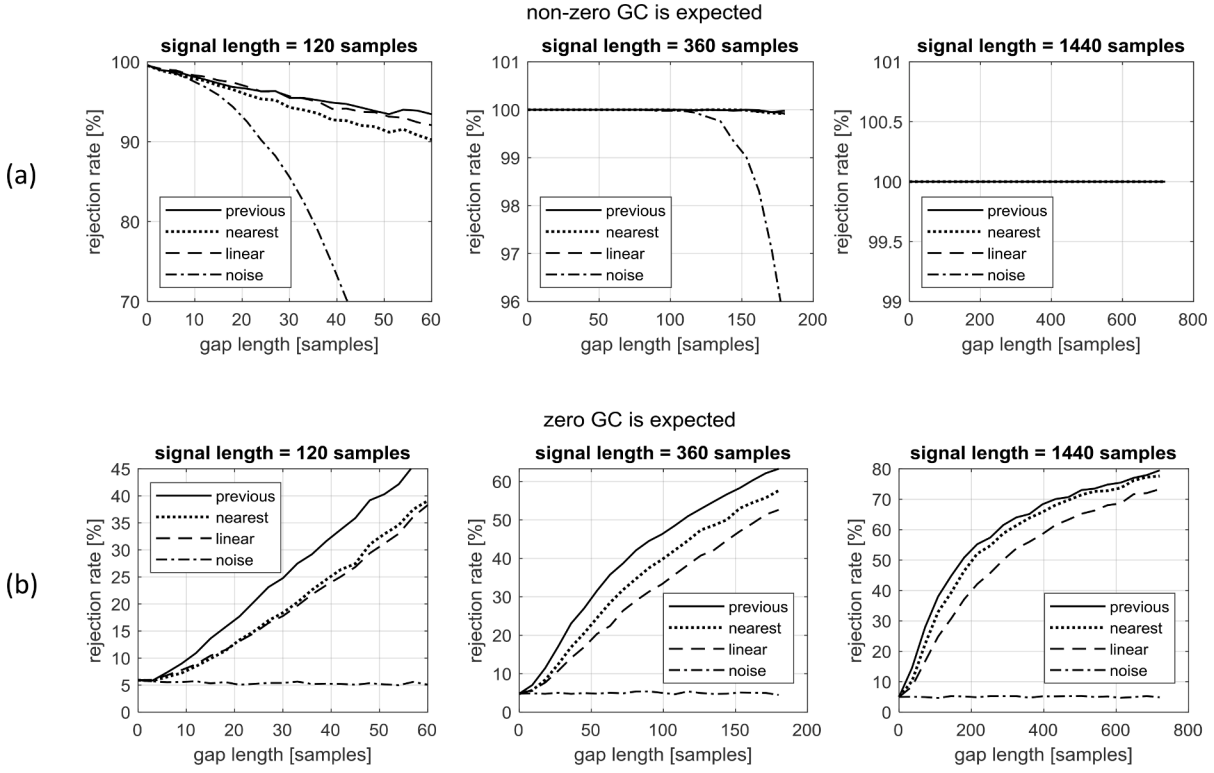

Fig. 2. The rejection rates of the null hypothesis of zero Granger causality (GC) (a) when non-zero GC is simulated and (b) when both time series are unrelated and hence no causal relation is expected. Three different data segment lengths were considered. Gaps of the lengths specified on the vertical axes were replaced using four strategies: previous non-missing value, nearest neighbour, linear interpolation, or imputing noise.

Table 2
Results obtained using the two different strategies for selecting the model order GC estimation in the real ABP-ICP data.

Parameter	Fixed model order ($m = 3$)	Varying model order (according to the BIC)	p-value	Effect size r
Magnitude of ABP→ICP GC	0.095 (0.058–0.144)	0.100 (0.064–0.154)	$\ll 0.0001$	–0.659
Time percentage of significant ABP→ICP GC	86.83% (76.20–93.65)	86.50% (76.66–93.54)	0.11	–0.104

Values are: between-subject median (Q1–Q3). Time percentage of significant ABP→ICP GC was defined as the percentage of data subsegments where there is a statistically significant GC in respect to all data segments that were valid for GC analysis. BIC – Bayesian information criterion.

is applied. This part is limited to the BIC and data segment equal to 1 h. The fixed model order $m = 3$ was taken as the median optimal model order according to the BIC (see Table 1). Even though the magnitudes of ABP→ICP GC obtained with the varying model order (according to BIC) were significantly higher than those obtained with the fixed order $m = 3$, there were no statistically significant differences in the time percentage of detecting significant ABP→ICP GC between the two approaches.

3.4. Stationarity of real-data subsegments and the influence of differencing on GC estimates

Table 3 shows the median spectral radius and percentage of data segments fulfilling the stationarity criterion for three different segment lengths and their model orders, which were established in the previous analysis. Even though the ratio of data segments exhibiting non-stationarity was lower for 1-h and 4-h long ABP-ICP data segments than for 20-minutes long ones, spectral radiuses were generally closer to 1 for those longer data segments. Situations when one application of differencing could not make ABP-ICP data covariance stationary were rare in practice. Out of 235 recordings, this happened in 24, 4, and 1 case for 20-minutes, 1-h, and 4-h long data segments, respectively, and it always applied to less than 1% of data segments in the recording.

Application of differencing significantly decreased GC magnitudes of ABP-ICP data. For both ABP→ICP and ICP→ABP, within-subject median GC magnitudes were statistically significantly lower for data after one application of differencing than without it (see Table 4). Within-subject correlation coefficients between GC magnitudes of non-differenced and differenced data were 0.783 (0.650–0.849) for $GC_{ABP\rightarrow ICP}$ and 0.590 (0.446–0.700) for $GC_{ICP\rightarrow ABP}$.

3.5. Comparison of the two different methods of GC estimation

Fig. 3 shows illustrative time courses of GC magnitudes. Both algorithms produced very similar results, but there were some particular data segments where the single regression approach yielded lower GC magnitudes than the double regression approach.

A comparison between the values of GC-related parameters obtained in the real ABP-ICP data using the two different algorithms is presented in Table 5. Even though the single regression approach gave significantly lower ABP→ICP and ICP→ABP GC magnitudes than the double regression approach, there was no significant difference in the volatility of ABP→ICP GC. Only in ICP→ABP did the single regression approach exhibit lower GC volatility than the double regression approach.

4. Discussion

In this study, we examined several technical aspects of Granger causality (GC). We considered GC on three ABP-ICP data segment lengths: 20 min, 1 h, and 4 h. We showed that performing GC analysis on high-resolution data may lead to spurious conclusions. We looked at the issue of missing data imputation, model order selection, and potential data non-stationarity. We compared two different methods of GC estimation. The following sections discuss all these issues in detail. When evaluating significance of GC, the threshold of $p = 0.05$ was assumed with no correction for multiple comparisons, since studying clinical associations was not a goal of this work.

4.1. Granger causality on high-resolution ABP-ICP data

An example of spurious conclusions drawn about causality is the apparent strong $GC_{ICP\rightarrow ABP}$ when analysing high-resolution data waveforms (including pulse waves). This is likely not to be a genuine ICP→ABP causal relation but rather the effect of ABP being delayed in respect to ICP (Fig. 4). ICP is changed nearly immediately with cerebral blood volume, whereas ABP measured from peripheral circulation is delayed in comparison to blood inflow into main cerebral arteries [26]. While this delay is too small to affect GC analysis of 10-seconds averaged ABP and ICP time series, it becomes a serious issue when dealing with high resolution data. The fact that the strong $GC_{ICP\rightarrow ABP}$ in high-resolution data disappears when ICP is manually delayed by 200 ms seems to confirm deceptive nature of that causal relation. Way of dealing with this issue is to downsample the data to the rate that makes the measurement delay insignificant.

4.2. Dealing with missing data

Gap analysis showed that this issue cannot be simply ignored. Gaps

Table 3
Stationarity of real ABP-ICP data.

Segment length	Model order	Percentage stationarity	Spectral radius
20 min	2	98.96% (98.39–99.40)	0.868 (0.828–0.896)
1 h	3	99.16% (98.54–99.65)	0.960 (0.947–0.968)
4 h	5	100.00% (99.69–100.00)	0.985 (0.980–0.989)

Values are: between-subject median (Q1–Q3). Stationarity was assessed by checking if the spectral radius < 1 (see section A.1 for more detail).

Table 4
GC magnitudes obtained for real ABP-ICP data without and after one application of differencing.

Parameter	No differencing	One application of differencing	p-value	Effect size r
Magnitude of ABP→ICP GC	0.111 (0.072–0.168)	0.085 (0.052–0.137)	$\ll 0.0001$	0.816
Magnitude of ICP→ABP GC	0.048 (0.039–0.064)	0.037 (0.030–0.050)	$\ll 0.0001$	0.721

Values are: between-subject median (Q1–Q3).

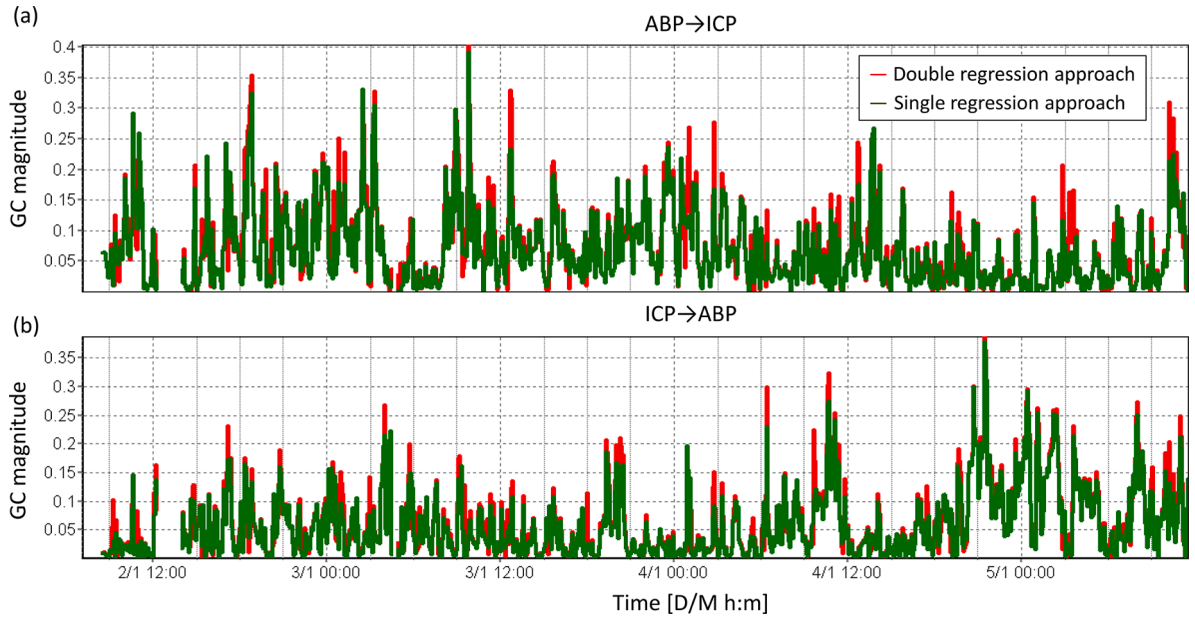

Fig. 3. Illustrative time courses of Granger causality (GC) magnitudes calculated using the two different algorithms, i.e. the double and single regression approach. (a) ABP→lCP GC (b) ICP→ABP GC.

Table 5
Results obtained using the two different algorithms for GC estimation in the real ABP-ICP data.

Parameter	Direction	Double regression algorithm	Single regression algorithm	*p*-value	Effect size *r*
Magnitude of GC	ABP→ICP	0.108 (0.071–0.165)	0.010 (0.065–0.154)	≪ 0.0001	0.867
	ICP→ABP	0.047 (0.038–0.063)	0.042 (0.034–0.056)	≪ 0.0001	0.867
Volatility of GC	ABP→ICP	0.202 (0.152–0.254)	0.203 (0.156–0.249)	0.77	0.019
	ICP→ABP	0.261 (0.241–0.276)	0.252 (0.233–0.269)	≪ 0.0001	0.518

Values are: between-subject median (Q1–Q3). Volatility of GC was defined as the number of transitions in statistical significance of GC divided by the number of all data segments that were valid for GC analysis.

in the data are not necessarily due to discontinuities in recording, but also due to removal of artefacts, like the arterial line flushes or possible drainage when ICP is measured using an external ventricular drain. Therefore application of any automated analysis, like the proposed GC, needs to take such gaps into account. The zero-tolerance policy, i.e. excluding all data segments that contain any missing data, would lead to unacceptably high data exclusion rate, particularly for longer data segments (Fig. 1). On the other hand, the simulation study indicated that too inclusive policy would distort GC inference (Fig. 2). Therefore, a compromise for missing data tolerance level is needed.

In our simulation experiments, the focus was put on continuous blocks of missing data. However, the chances are that multiple gaps fall into one segment, particularly when longer data segments are considered. Thus, the percentage of missing data points and the longest gap in the data segment complement each other in the quantitative description of the structure of missing data.

The most common approaches of imputing missing data in time series involve linear or non-linear interpolation, sample & hold filters, or

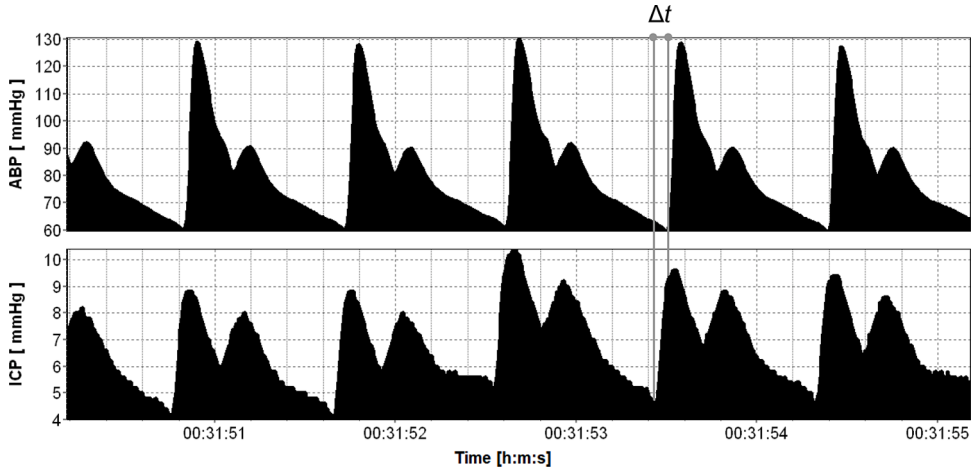

Fig. 4. Illustrative high-resolution data waveforms (including pulse waves). ABP appears to be delayed in respect to ICP due to the peripheral circulation.

the nearest neighbour technique. The strategy of replacing point in a gap by the previous non-missing value is worse than nearest neighbour and linear interpolation in terms of the rejection rates. Linear interpolation might be seen superior to nearest neighbour. The latter, however, seems to be more universal, as linear interpolation must reduce to nearest neighbour when there are trailing or leading missing points in data segment. As an alternative, we proposed replacing gaps by white Gaussian noise with a standard deviation (SD) equal to the SD of the remaining valid data in the segment. The most important observation from the Monte Carlo simulation is that replacing missing data by white Gaussian noise does not inflate false positives at all, unlike the other methods which interpolate missing data (Fig. 2b). On the other hand, it does inflate false negatives stronger than interpolation methods (Fig. 2a). Therefore, the practical consequence is that the missing data limit can be relaxed if one chooses replacing missing data by noise and can accept higher false negatives rate. The other observation from the simulation study is that the longer the data segment is, the longer gap can be tolerated to achieve similar increase in the rate of false positives or negatives. A minor difference between the two types of data imputation is that the interpolation methods are fully deterministic, whereas replacing missing data by random noise may yield results dependant on properties of a random numbers generator. To sum up, we recommend replacing gaps by white Gaussian noise when one wants to avoid false positives and applying linear interpolation or nearest neighbour when one wants to limit false negatives or have ideally deterministic results.

Since replacing gaps by white Gaussian noise does not inflate false positives, it seems to be a suitable data imputation method for the potential clinical application of GC to assess reliability of the pressure reactivity index (PRx). Avoiding false positives is preferred in this context because false indication of ABP→ICP causal relation would incorrectly increase reliability of PRx. In conservative approach, no evidence of ABP→ICP causal relation (regardless if this is true negative or false negative) should be interpreted that PRx value may be questionable and hence unsuitable for, e.g. calculating PRx-based cerebral perfusion pressure targets. If such conservative approach is assumed, then replacing missing data by noise should not introduce additional error.

4.3. Optimal model order

The optimal model order according to both AIC and BIC increased with the ABP-ICP segment length (see Table 1). Since Akaike tends to yield relatively high model order for data segments consisting of large number of data points [25,35,36], the BIC criterion seems to be preferable to the ABP-ICP data.

Our analysis encourages assuming fixed model order for a given data segment length, instead of selecting optimal model order individually for every single ABP-ICP data segment. This is supported by the fact that in the analysed cohort, interquartile range of the optimal model order according to the BIC was not greater than one (see Table 1). Another point in favour of the simpler and more computationally-efficient fixed order approach is that the time proportion of significant $GC_{ABP\rightarrow ICP}$ did not differ significantly between the two approaches (see Table 2).

4.4. Stationarity and differencing

When data do not fulfil stationarity assumption, it is advisable to apply differencing before GC analysis [28]. However, one should keep in mind that differencing changes interpretation of causal interactions, as they then reflect relationship not between original time series but rather between the rates of their changes [25]. Repeated applications of differencing required to obtain stationary data is fortunately not expected in case of physiological data like ABP and ICP, as that would make the interpretation even more difficult. Our results seem to support this. On the contrary, when an individual data segment cannot be made stationary despite single differencing, this could be taken as an indication of presence of major artefacts or distortions. Another aspect is that the differencing of ABP-ICP data significantly decreases GC magnitudes (see Table 4). Therefore, we advise against mixing GC estimates of data which were partly differenced and partly not, prior to calculating average or median values. One should either perform single differencing for each data segment, or keep original data but exclude from average or median GC values segments which violate stationarity assumption. The influence of differencing on GC magnitudes was investigated for 20-minutes long data segments, since they did not fulfil the stationarity criterion more often than longer segments.

If data segment contains any missing data, there is a need to replace them by some substituted values to calculate the spectral radius and assess stationarity. This assessment, however, may indicate non-stationarity and the need to perform differencing. In such cases, we decided not to apply differencing on imputed data, but rather perform differencing on the original data and then fill the gaps. This is particularly important when filling by white noise is chosen, as differencing intensifies high frequencies and increases standard deviation of noise.

4.5. Comparison of the two different methods of GC estimation

The single regression approach is less prone to false positives, i.e. mistaken rejection of the zero-GC null hypothesis [26]. Since $GC_{ICP\rightarrow ABP}$ is weaker than $GC_{ABP\rightarrow ICP}$, the advantage of the single regression approach over the double regression approach becomes more noticeable in this case. Lower volatility of $GC_{ICP\rightarrow ABP}$ significance seems to mean less false positives, what is in favour of the single regression approach. Computation time of the single regression algorithm depends heavily on the spectral radius of fitted vector autoregressive (VAR) models [26]. The number of iterations of this algorithm grows to infinity when the spectral radius approaches 1. The fact that spectral radiuses for 20-minutes long ABP-ICP data segments were lower than those for longer data segments, encouraged attempting to apply the single regression approach in this case. In practice, however, this algorithm turned out to be over a dozen times slower than the double regression approach. Even mandatory data differencing, which usually decreases spectral radiuses, did not improve the single regression approach satisfactory enough in terms of computation time. Since from our point of view, this performance penalty outweighs the benefits of the single regression algorithm, we conclude that the double regression approach is more suitable for ABP-ICP data, at least when simple bivariate (i.e. involving two variables only) GC is considered.

4.6. Limitations

This paper considered only time-domain GC. Spectral decomposition, however, allows for considering GC as a function of frequency [19,24]. Nevertheless, time-domain GC is prevalent in the literature, also amongst existing studies regarding GC in neurocritical care data [21–23]. We used the most obvious way of implementing time-varying GC, which is a sliding window approach. In our approach, GC in each data segment was estimated independently of previous data. However, there are time-varying GC methods where coefficients of VAR model are updated adaptively based on least mean squares estimation or Kalman filtering [19,37]. Another limitation of our work is that we focused only on two physiological variables: ABP and ICP. Therefore, simple, bivariate GC was considered. There are methods that consider causal relationships between more variables. Conditional GC can be used to study e.g. causality between ABP and ICP after controlling the influence of heart rate [22]. There are further multivariate GC extensions where a group of variables can have joint predictive power on another group of variables [26]. The recommendations we provided in the paper may not necessarily apply to multivariate cases. Lastly, methods which allow for detecting non-linear causal interactions [38] were beyond the scope of this paper. However, they may be of interest in the context ABP-ICP relationship, as it exhibits non-linearities [3].

5. Conclusions

The paper offers a set of recommendations regarding Granger causality (GC) analysis in ABP-ICP data. By decreasing the duration of analysed ABP-ICP data segments from days to hours and below, we proposed a concept of time-varying GC. High-resolution data (including pulse waves) need to be downsampled prior to GC estimation. Otherwise, GC analysis may erroneously indicate strong, artificial ICP→ABP causal relation. We recommend replacing gaps by white Gaussian noise to avoid false positives (erroneous causality). On the other hand, if one wants to limit false negatives (erroneous non-causality), linear interpolation or nearest neighbour should be applied. We advise assuming fixed model order for a given data segment length, instead of selecting optimal model order individually for every single ABP-ICP data segment. We do not recommend mixing GC estimates of data which were partly differenced and partly not, prior to calculating average or median values. One should either perform single differencing for each data segment, or keep original data but exclude from average or median GC values segments which violate stationarity assumption. We justify that the double regression method of GC estimation is more suitable for ABP-ICP data than the single regression approach. The methods presented in the paper should allow for reliable investigation of datasets obtained from intensive care units.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: P. Smielewski and M. Czosnyka receive part of licensing fees for the ICM+ software (https://icmplus.neurosurg.cam.ac.uk; Cambridge Enterprise, Cambridge, UK).

Acknowledgement

M. M. Placek was supported by the European Union seventh Framework Program (grant 602150) for Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTRE-TBI) until March 2021 and by Action Medical Research (grant GN2609) for Studying Trends of Auto-Regulation in Severe Head Injury in Paediatrics (STARSHIP). E. Beqiri is supported by the Medical Research Council (grant MR N013433–1) and by the Gates Cambridge Scholarship. Python source code used in this study is available at: https://github.com/m-m-placek/python-icmplus-granger-causality.

>Appendix. Granger causality – theory & definitions

A time series Y is said to *Granger-cause* (or *G-cause*) a time series X if information contained in the past of Y allows for more accurate prediction of the future of X than when only information contained in the past of X itself is taken into account [17]. Formal definition of Granger causality (GC) involves autoregressive models. Let us assume that the behaviour of time series $X[t]$ and $Y[t]$ can be modelled by the following vector autoregressive (VAR) model:

$$X[t] = \sum_{j=1}^{m} a_{xx,j} X[t-j] + \sum_{j=1}^{m} a_{xy,j} Y[t-j] + \xi_x[t], \tag{1a}$$

$$Y[t] = \sum_{j=1}^{m} a_{yx,j} X[t-j] + \sum_{j=1}^{m} a_{yy,j} Y[t-j] + \xi_y[t], \tag{1b}$$

where m is the model order, i.e. the number of past observations included in the model, $A = (a_{\cdot,\cdot,\cdot})$ constitutes the coefficients of the model, and ξ_x, ξ_y stand for the prediction errors (residuals) for respective time series.

The omission of the potential influence of past Y on X, by assuming $a_{xy,j} = 0$, $j \in \{1, \ldots, m\}$, leads to the so-called reduced model:

$$X[t] = \sum_{j=1}^{m} a'_{xx,j} X[t-j] + \xi'_x[t] \tag{2}$$

(note that, in general, $a'_{xx,j} \neq a_{xx,j}$ and $\xi_x \neq \xi'_x$). The idea to examine GC is to fit the full model (a.k.a. the unrestricted model, Eq. (1a)) and the reduced model (a.k.a. the restricted model, Eq. (2)), and then compare their prediction errors. Since the mean value of the prediction error is zero, its variance is equivalent to the mean square error. The reduced model is nested in the full model, hence $\mathrm{var}(\xi_x) \leq \mathrm{var}(\xi'_x)$, where $\mathrm{var}(\cdot)$ stands for variance.

The magnitude of GC, i.e. the measure of linear feedback from Y to X ($Y \to X$), is defined as the natural logarithm of the ratio of the prediction error variances for the reduced and full models [24–26]:

$$\mathscr{F}_{Y \to X} = \ln \frac{\mathrm{var}(\xi'_x)}{\mathrm{var}(\xi_x)}. \tag{3}$$

In terms of statistical significance and hypotheses testing, the null hypothesis is 'zero GC,' i.e. $\mathrm{var}(\xi_x) = \mathrm{var}(\xi_x')$, which is equivalent to $a_{xy,j} = 0$ for each $j \in \{1, \ldots, m\}$ jointly [19], and it can be approached using an F-test [25]. If more tests for GC are performed on the same dataset, e.g. bidirectional GC testing ($Y \to X$ and $X \to Y$) or more pairs of variables, it is advised to apply a correction for multiple comparisons [25].

The most obvious method of estimating GC requires fitting both the full and the reduced model [25]. This can be called a double regression approach. There is also another way to estimate GC that does not need explicit identification of the reduced model [26]. That, in turn, can be named a single regression approach.

A.1. Single regression algorithm for GC estimation

The conventional method of estimating GC requires fitting both a full and a reduced model (see previous section) [25]. There is, however, another way to estimate GC, without explicit identification of the reduced model [26]. That method exploits the equivalence of different representations of a

VAR model: regression parameters (i.e. the regression coefficients A and the residuals covariance matrix) and the autocovariance function (ACF). These two representations of a VAR model are related to each other via the Yule–Walker equations [39]. There are numerical routines for solving these equations for the ACF when the regression parameters are known and vice versa [26].

The idea to estimate GC without explicit identification of the reduced model is as follows. One needs to fit the full VAR model and then, knowing its regression parameters, solve the Yule–Walker equations for the ACF. Having the ACF of the full VAR model, it is trivial to extract the part of the ACF corresponding to the reduced model. Subsequently, the regression parameters of the reduced model ($a'_{xx,j}$ for $j \in \{1,\ldots,m\}$ and $\mathrm{var}(\xi'_x)$, see Eq. (2)) can be found by solving the Yule–Walker equations with known ACF of the reduced model. Finally, having $\mathrm{var}(\xi'_x)$, the magnitude of GC is calculated using Eq. (3), like it is done in the double regression approach. Simulations performed in [26] showed that by avoiding explicit estimation of the reduced model, the single regression approach eliminates a source of estimation error and improves statistical power.

A non-obvious aspect of that method is selecting the number of lags q to consider in the ACF (note that this is a different parameter than the VAR model order m). Assuming covariance stationarity, the ACF 1) is dependant only on lag but not on time, 2) is symmetrical, 3) decays exponentially with the increasing lag. The factor of that exponential decay is just the spectral radius $\rho(A)$ of the full VAR model. For a large enough lag, the ACF becomes numerically insignificant and hence can be truncated. The number of lags q in the ACF sufficient to satisfy a numerical tolerance ε can be calculated according to the formula:

$$q = \left\lceil \frac{\ln\varepsilon}{\ln\rho(A)} \right\rceil, \tag{4}$$

where $\lceil\cdot\rceil$ stands for the ceiling function. It was proposed to set ε at the fixed level of 10^{-8} [26]. In practice, however, when the spectral radius $\rho(A)$ is close to 1, i.e. the VAR model is nearly unstable, the number lags q in the ACF goes to infinity. In order to prevent a situation that calculations become unacceptably long, we decided to diminish the threshold for $\rho(A)$ from 1 to 0.99 in the single regression algorithm. This limited the maximum number of lags q to $\lceil \ln(10^{-8}) / \ln(0.99) \rceil = 1833$, assuming the numerical tolerance $\varepsilon = 10^{-8}$.

A.2. Stationarity and spectral radius

Valid GC estimation requires that the VAR model defined by coefficients A is stable [27], meaning that it constitutes a covariance-stationary (a.k.a. wide-sense stationary) process. VAR process is stationary if and only if the roots reciprocals of its characteristic polynomial lie inside the unit circle in the complex plane [27]. Defining the spectral radius $\rho(A)$ of the VAR process as the maximal absolute value of the roots reciprocals of its characteristic polynomial, the criterion for stability is: $\rho(A) < 1$ [26]. When data does not fulfil stationarity assumption, differencing, i.e. filtering of the form $\dot{U}[t] = U[t] - U[t-1]$, can be applied, potentially multiple times, to make data stationary and then perform GC analysis [28].

A.3. Model order selection

The fitting of VAR processes requires assumption of the model order. As for all multivariate regression problems, there is a balance between model overfitting (too high order, that is too many model coefficients) and underfitting (too low order, not enough parameters) that must be found using some sort of regularisation terms. The two most popular criteria for selecting the optimal model order are: the Akaike information criterion (AIC) [30] and the Bayesian information criterion (BIC) [31].

References

[1] A.I.R. Maas, et al., Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research, Lancet Neurol. 16 (2017) 987–1048, https://doi.org/10.1016/S1474-4422(17)30371-X.

[2] N. Stocchetti, M. Carbonara, G. Citerio, A. Ercole, M.B. Skrifvars, P. Smielewski, T. Zoerle, D.K. Menon, Severe traumatic brain injury: targeted management in the intensive care unit, Lancet Neurol. 16 (2017) 452–464, https://doi.org/10.1016/S1474-4422(17)30118-7.

[3] M. Czosnyka, P. Smielewski, P. Kirkpatrick, R.J. Laing, D. Menon, J.D. Pickard, Continuous Assessment of the Cerebral Vasomotor Reactivity in Head Injury, Neurosurgery 41 (1997) 11–19, https://doi.org/10.1097/00006123-199707000-00005.

[4] L.A. Steiner, M. Czosnyka, S.K. Piechnik, P. Smielewski, D. Chatfield, D.K. Menon, J.D. Pickard, Continuous monitoring of cerebrovascular pressure reactivity allows determination of optimal cerebral perfusion pressure in patients with traumatic brain injury, Crit. Care Med. 30 (2002) 733–738, https://doi.org/10.1097/00003246-200204000-00002.

[5] M. Czosnyka, K. Brady, M. Reinhard, P. Smielewski, L.A. Steiner, Monitoring of cerebrovascular autoregulation: facts, myths, and missing links, Neurocritical Care 10 (2009) 373–386, https://doi.org/10.1007/s12028-008-9175-7.

[6] J. Tas, E. Beqiri, R.C. van Kaam, M. Czosnyka, J. Donnelly, R.H. Haeren, I.C.C. van der Horst, P.J. Hutchinson, S.M.J. van Kuijk, A.L. Liberti, D.K. Menon, C.W. E. Hoedemaekers, B. Depreitere, P. Smielewski, G. Meyfroidt, A. Ercole, M.J. H. Aries, Targeting Autoregulation-Guided Cerebral Perfusion Pressure after Traumatic Brain Injury (COGiTATE): a Feasibility Randomized Controlled Clinical Trial, J. Neurotrauma 11 (2021) 1–11, https://doi.org/10.1089/neu.2021.0197.

[7] N.A. Lassen, Autoregulation of cerebral blood flow, Circ. Res. 15 (1964) 201–204. http://www.ncbi.nlm.nih.gov/pubmed/14206303.

[8] P. Brassard, L. Labrecque, J.D. Smirl, M.M. Tymko, H.G. Caldwell, R.L. Hoiland, S. J.E. Lucas, A.Y. Denault, E.J. Couture, P.N. Ainslie, Losing the dogmatic view of cerebral autoregulation, Physiol. Rep. 9 (2021), https://doi.org/10.14814/PHY2.14982.

[9] F.A. Zeiler, A. Ercole, E. Beqiri, M. Cabeleira, E.P. Thelin, N. Stocchetti, E. W. Steyerberg, A.I.R. Maas, D.K. Menon, M. Czosnyka, P. Smielewski, Association between Cerebrovascular Reactivity Monitoring and Mortality Is Preserved When Adjusting for Baseline Admission Characteristics in Adult Traumatic Brain Injury: a CENTER-TBI Study, J. Neurotrauma 37 (2020) 1233–1241, https://doi.org/10.1089/neu.2019.6808.

[10] M. Czosnyka, J.D. Picard, Monitoring and interpretation of intracranial pressure, J. Neurol. Neurosurg. Psychiatry 75 (2004) 813–821, https://doi.org/10.1136/jnnp.2003.033126.

[11] L.A. Steiner, J.P. Coles, M. Czosnyka, P.S. Minhas, T.D. Fryer, J.C. Clark, P. Smielewski, D.A. Chatfield, T. Donovan, J.D. Pickard, Cerebrovascular pressure reactivity is related to global cerebral oxygen metabolism after head injury, J. Neurol. Neurosurg. Psychiatry 74 (2003) 765–770, https://doi.org/10.1136/jnnp.74.6.765.

[12] M. Czosnyka, Z. Czosnyka, P. Smielewski, Pressure reactivity index: journey through the past 20 years, Acta Neurochir 159 (2017) 2063–2065, https://doi.org/10.1007/s00701-017-3310-1.

[13] C. Vlachopoulos, M. O'Rourke, W.W. Nichols, McDonald's Blood Flow in Arteries : Theoretical, Experimental and Clinical Principles, CRC Press, 2011, https://doi.org/10.1201/B13568.

[14] A. Spiegelberg, M. Preuß, V. Kurtcuoglu, B-waves revisited, Interdisciplinary Neurosurgery 6 (2016) 13–17, https://doi.org/10.1016/j.inat.2016.03.004.

[15] J.P. Muizelaar, J.D. Ward, A. Marmarou, P.G. Newlon, A. Wachi, Cerebral blood flow and metabolism in severely head-injured children: part 2: autoregulation, J. Neurosurg. 71 (1989) 72–76, https://doi.org/10.3171/JNS.1989.71.1.0072.

[16] X. Liu, J. Donnelly, M. Czosnyka, M.J.H. Aries, K. Brady, D. Cardim, C. Robba, M. Cabeleira, D.J. Kim, C. Haubrich, P.J. Hutchinson, P. Smielewski, Cerebrovascular pressure reactivity monitoring using wavelet analysis in traumatic brain injury patients: a retrospective study, PLoS Med. 14 (2017) 1–19, https://doi.org/10.1371/journal.pmed.1002348.

[17] C.J.W. Granger, Investigating Causal Relations by Econometric Models and Cross-spectral Methods Authors, Econometrica 37 (1969) 424–438. http://www.jstor.org/stable/1912791.

[18] A.K. Seth, A.B. Barrett, L. Barnett, Granger causality analysis in neuroscience and neuroimaging, J. Neurosci. 35 (2015) 3293–3297, https://doi.org/10.1523/JNEUROSCI.4399-14.2015.
[19] S. Cekic, D. Grandjean, O. Renaud, Time, frequency, and time-varying Granger-causality measures in neuroscience, Stat. Med. 37 (2018) 1910–1931, https://doi.org/10.1002/sim.7621.
[20] K. Sameshima, L.A. Baccala, Methods in Brain Connectivity Inference Through Multivariate Time Series Analysis, CRC Press, 2014.
[21] E.P. Thelin, R. Raj, B.M. Bellander, D. Nelson, A. Piippo-Karjalainen, J. Siironen, P. Tanskanen, G. Hawryluk, M. Hasen, B. Unger, F.A. Zeiler, Comparison of high versus low frequency cerebral physiology for cerebrovascular reactivity assessment in traumatic brain injury: a multi-center pilot study, J. Clin. Monit. Comput. 34 (2020) 971–994, https://doi.org/10.1007/s10877-019-00392-y.
[22] L. Gao, P. Smielewski, M. Czosnyka, A. Ercole, Early Asymmetric Cardio-Cerebral Causality and Outcome after Severe Traumatic Brain Injury, J. Neurotrauma 34 (2017) 2743–2752, https://doi.org/10.1089/neu.2016.4787.
[23] L. Gao, P. Smielewski, P. Li, M. Czosnyka, A. Ercole, Signal Information Prediction of Mortality Identifies Unique Patient Subsets after Severe Traumatic Brain Injury: a Decision-Tree Analysis Approach, J. Neurotrauma 37 (2020) 1011–1019, https://doi.org/10.1089/neu.2019.6631.
[24] J. Geweke, Measurement of linear dependence and feedback between multiple time series, J. Am. Statist. Assoc. 77 (1982) 304–313, https://doi.org/10.1080/01621459.1982.10477803.
[25] A.K. Seth, A MATLAB toolbox for Granger causal connectivity analysis, J. Neurosci. Methods 186 (2010) 262–273, https://doi.org/10.1016/j.jneumeth.2009.11.020.
[26] L. Barnett, A.K. Seth, The MVGC multivariate Granger causality toolbox: a new approach to Granger-causal inference, J. Neurosci. Methods 223 (2014) 50–68, https://doi.org/10.1016/j.jneumeth.2013.10.018.
[27] Helmut. Lütkepohl, New Introduction to Multiple Time Series Analysis, Springer Science & Business Media, 2005.
[28] B.S. Cheng, T.W. Lai, An investigation of co-integration and causality between energy consumption and economic activity in Taiwan, Energy Economics 19 (1997) 435–444, https://doi.org/10.1016/S0140-9883(97)01023-2.
[29] C. Sortica da Costa, M.M. Placek, M. Czosnyka, B. Cabella, M. Kasprowicz, T. Austin, P. Smielewski, Complexity of brain signals is associated with outcome in preterm infants, J. Cereb. Blood Flow Metab. 37 (2017), https://doi.org/10.1177/0271678X16687314.
[30] H. Akaike, A New Look at the Statistical Model Identification, IEEE Trans. Autom. Control 19 (1974) 716–723, https://doi.org/10.1109/TAC.1974.1100705.
[31] G. Schwarz, Estimating the Dimension of a Model, Ann. Statistics 6 (1978) 461–464. https://www.jstor.org/stable/2958889.
[32] C.O. Fritz, P.E. Morris, J.J. Richler, Effect size estimates: current use, calculations, and interpretation, J. Experimental Psychol.: General 141 (2012) 2–18, https://doi.org/10.1037/a0024338.
[33] M. Tomczak, E. Tomczak, The need to report effect size estimates revisited. An overview of some recommended measures of effect size, Trends in Sport Sciences 21 (2014) 19–25. http://tss.awf.poznan.pl/files/3_Trends_Vol21_2014_no1_20.pdf.
[34] M.M. Placek, A. Khellaf, B.L. Thiemann, M. Cabeleira, P. Smielewski, Python-Embedded Plugin Implementation in ICM+: novel Tools for Neuromonitoring Time Series Analysis with Examples Using CENTER-TBI Datasets, in: B. Depreitere, G. Meyfroidt, F. Güiza (Eds.), Intracranial Pressure and Neuromonitoring XVII. Acta Neurochirurgica, Supplementum, Springer, Cham, 2021, pp. 255–260, https://doi.org/10.1007/978-3-030-59436-7_48.
[35] S. Marple, Digital Spectral analysis: With Applications, Prentice-Hall, Englewood Cliffs, N.J., 1987.
[36] S. Kay, Modern Spectral estimation: Theory and Application, Prentice-Hall, Englewood Cliffs N.J., 1988.
[37] W. Hesse, E. Möller, M. Arnold, B. Schack, The use of time-variant EEG Granger causality for inspecting directed interdependencies of neural assemblies, J. Neurosci. Methods 124 (2003) 27–44, https://doi.org/10.1016/S0165-0270(02)00366-7.
[38] A. Müller, J.F. Kraemer, T. Penzel, H. Bonnemeier, J. Kurths, N. Wessel, Causality in physiological signals, Physiol. Meas. 37 (2016) R46–R72, https://doi.org/10.1088/0967-3334/37/5/R46.
[39] T.W. Anderson, The Statistical Analysis of Time Series, Wiley, 1971, https://doi.org/10.1002/9781118186428.