Dynamic cerebral autoregulation reproducibility is affected by physiological variability

AUTHORS

Marit L. Sanders[1], Jan Willem J. Elting[2], Ronney B. Panerai[3], Marcel Aries[4], Edson Bor-Seng-Shu[5], Alexander Caicedo[6], Max Chacon[7], Erik D. Gommer[8], Sabine Van Huffel[9], José L. Jara[7], Kyriaki Kostoglou[10], Adam Mahdi[11], Vasilis Z. Marmarelis[12], Georgios D. Mitsis[13], Martin Müller[14], Dragana Nikolic[15], Ricardo C. Nogueira[5], Stephen J. Payne[11], Corina Puppo[16], Dae C. Shin[12], David M. Simpson[15], Takashi Tarumi[17], Bernardo Yelicich[16], Rong Zhang[17], Jurgen A.H.R. Claassen[1*]

[1] Department of Geriatric Medicine, Radboudumc Alzheimer Centre and Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, The Netherlands

[2] Department of Neurology, University Medical Center Groningen, The Netherlands

[3] Department of Cardiovascular Sciences and Leicester NIHR Biomedical Research Centre in Cardiovascular Sciences, Glenfield Hospital, Leicester, UK

[4] Department of Intensive Care, University of Maastricht, Maastricht University Medical Center, Maastricht, The Netherlands

[5] Department of Neurology, Hospital das Clinicas University of Sao Paulo, Brazil

[6] Mathematics and Computer Science, Faculty of Natural Sciences and Mathematics, Universidad del Rosario, Bogotá, Colombia

[7] Institute of Biomedical Engineering, department of Engineering Informatics, University of Santiago, Chile

[8] Department of Clinical Neurophysiology, University Hospital Maastricht, Netherlands

[9] Department of Electronic Engineering (ESAT), STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Belgium; imec, Leuven, Belgium

[10] Department of Electrical, Computer and Software Engineering, McGill University, Montreal, Canada

[11] Department of Engineering Science , University of Oxford, UK

[12] Department of Biomedical Engineering, University of Southern California, Los Angeles, USA

[13] Department of Bioengineering, McGill University, Montreal, Canada

[14] Department of Neurology, Luzerner Kantonsspital, Switzerland

[15] Institute of Sound and Vibration Research, University of Southampton, UK

[16] Departamento de Emergencia, Hospital de Clínicas, Universidad de la República, Montevideo, Uruguay

[17] IEEM, Presbyterian Hospital Dallas, University of Texas Southwestern Medical Center, USA

CORRESPONDING AUTHOR

Jurgen AHR Claassen, MD, PhD

Department of Geriatric Medicine, Radboud Alzheimer Centre and Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center

Reinier Postlaan 4, Postbox 9101, 6500 HB Nijmegen, The Netherlands

Tel. : +31 24-3616772, Email: Jurgen.Claassen@radboudumc.nl

Number of words: 3.327

Number of figures: 3

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ABSTRACT

Parameters describing dynamic cerebral autoregulation (DCA) have limited reproducibility. In an international, multi-centre study, we evaluated the influence of multiple analytical methods on the reproducibility of DCA. Fourteen participating centers analyzed repeated measurements from 75 healthy subjects, consisting of five minutes of spontaneous fluctuations in blood pressure (BP) and cerebral blood flow velocity (CBFv) signals, based on their usual methods of analysis. DCA- methods were grouped into three broad categories, depending on output types: 1. Transfer function analysis (TFA); 2. Autoregulation index (ARI); and 3. correlation coefficient. Only TFA gain in the low frequency (LF) band showed good reproducibility in approximately half of the estimates of gain, defined as an intraclass correlation coefficient (ICC) of > 0.6. None of the other DCA metrics had good reproducibility. For TFA-like and ARI-like methods, ICCs were lower than values obtained with surrogate data ($p<0.05$). For TFA-like methods, ICCs were lower for the very low frequency (VLF) band (gain 0.38 ± 0.057, phase 0.17 ± 0.13) than for LF band (gain 0.59 ± 0.078, phase 0.39 ± 0.11, $p\leq0.001$ for both gain and phase). For ARI-like methods, the mean ICC was 0.30 ± 0.12 and for the correlation methods 0.24 ± 0.23. Based on comparisons with ICC estimates obtained from surrogate data, we conclude that physiological variability or non-stationarity is likely to be the main reason for the poor reproducibility of DCA parameters.

KEYWORDS

ARI index

Cerebral blood flow

Cerebral heamodynamics

Transcranial Doppler

Transfer function analysis

INTRODUCTION

The importance of cerebral autoregulation (CA) has been clearly established, as a cerebro-protective mechanism to alterations in blood pressure (BP) by keeping cerebral blood flow (CBF) relatively constant (van Beek, Claassen et al. 2008). Dynamic cerebral autoregulation (DCA) is the transient cerebrovascular response to rapid changes in BP (Aaslid, Lindegaard et al. 1989). Compared to the more classical modality of 'static' autoregulation, that often requires the use of pharmacological agents to induce steady-state changes in BP (Tiecks, Lam et al. 1995), DCA has benefitted from recent developments in non-invasive techniques to record CBF and BP, and it is now the preferred approach for assessment of CA in physiological and clinical studies.

Despite its many advantages, protocols to reliably assess DCA remain the object of considerable debate (Simpson and Claassen 2018, Simpson and Claassen 2018, Tzeng and Panerai 2018, Tzeng and Panerai 2018). On one hand, maneuvers that induce relatively large and rapid changes in BP, such as the sudden release of compressed thigh cuffs (Aaslid, Lindegaard et al. 1989), lead to recordings with better signal-to-noise ratio and the possibility of visualizing and quantifying the DCA response with measurements as short as 30 seconds. On the other hand, using the spontaneous fluctuations in BP and CBF, that can be observed in most individuals, allows estimation of DCA parameters at rest, without the need for a physiological disturbance or challenge. This can lead to better acceptance and feasibility in most clinical conditions.

Which road to take? The answer to this fundamental question is not straightforward as it is unlikely that a single protocol will be suitable for all different scenarios of patient care and physiological intervention (Simpson and Claassen 2018, Simpson and Claassen 2018, Tzeng and Panerai 2018, Tzeng and Panerai 2018).

A definition of an optimal protocol could be one which, combined with robust modeling techniques (Panerai 2008), leads to the best sensitivity and specificity performance for detection of CA disturbances, as well as predictive accuracy for patient prognosis.

Before reaching this stage though, it is essential that measurement reproducibility is demonstrated as a key property of any method of assessment. This target is at the forefront of the collaborative initiatives promulgated by the International Cerebral Autoregulation Network (CARNet) as part of the effort to identify potential sources of methodological disparity (Meel-van den Abeelen, Simpson et al. 2014) and encourage technical standardization (Claassen, Meel-van den Abeelen et al. 2016). The most recent stage of this pathway is described in this article and involves an international, multi-centre assessment of the reproducibility of the main parameters that are currently available to assess DCA based on spontaneous fluctuations of BP and CBF.

Examining the reproducibility of DCA parameters, obtained from spontaneous fluctuations at rest, is important due to the widespread use of this approach for both physiological and clinical studies. Early assessments of the reproducibility of the spontaneous fluctuations approach were not encouraging (Brodie, Atkins et al. 2009, Gommer, Shijaku et al. 2010, Smirl, Hoffman et al. 2015), but were not regarded as the definitive answer, only as indicative of a single method, handled by a single centre. This limitation was addressed in the current multi-center study. An initial report (Sanders, Claassen et al. 2018), described the influence of different methods of analysis on the reproducibility of synthetic data, where surrogate time-series of CBF velocity (CBFv) were generated based on real measurements of BP, coupled with a realistic signal-to-noise ratio. These generated CBFv data were based on a linear model. Thus, compared to real CBFv data, these generated data are free of any physiological influences on the BP-CBFv relationship. Such physiological influences could include non-stationary behavior of autoregulatory function (i.e. variations in function over time), and

factors known to influence CBFv (e.g. $PaCO_2$, cognitive activity, autonomic nervous activity, temperature, breathing pattern).

The present communication therefore had as aim to provide a much broader description of the reproducibility of 'real' estimates of DCA from fourteen leading international centers, using a diversity of analytical methods. In particular, this study addressed two main objectives 1) to compare the reproducibility of DCA parameters from these real physiological measurements to that of surrogate data, and 2) to establish the influence of different analytical methods used by a variety of research centers worldwide on the reproducibility of DCA metrics.

MATERIALS AND METHODS

Subjects

A database was created from available datasets of cerebral hemodynamic measurements from participating centers (Table S1). Included were healthy adults > 18 years of age. Exclusion criteria were uncontrolled hypertension, smoking, cardiovascular disease, diabetes, irregular heart rhythm, TIA/stroke or significant pulmonary disease. The study has been carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Written informed consent was obtained from all subjects.

Description of datasets

Six of a total of 14 centers (Table S1) provided datasets that consisted of two measurements from 10-15 healthy volunteers in each centre, resulting in a total of 75 healthy subjects. Time between the two measurements varied between centers, from minutes to a maximum of three months. Data sets consisted of five minutes of beat-to-beat artifact free mean CBFv (transcranial Doppler ultrasound, TCD), mean BP (digital artery volume clamping) and end-tidal CO_2 ($EtCO_2$, capnography) measurements at rest. Beat-to-beat parameters were re-sampled at 10 Hz. In 22 subjects, the TCD data were unilateral. The dataset was as follows: N=55 left side signals, N=71 right side signals.

DCA Analysis

Data analyses were performed by 14 participating centers. The following DCA analysis methods were used: TFA (Panerai, Rennie et al. 1998, Zhang, Zuckerman et al. 1998, Mitsis, Zhang et al. 2002, Muller, Bianchi et al. 2003, Reinhard, Muller et al. 2003, Liu, Simpson et al. 2005, Gommer, Shijaku et al. 2010, van Beek, Olde Rikkert et al. 2010, Meel-van den Abeelen, Simpson et al. 2014, Muller and Osterreich 2014, Panerai 2014), Laguerre expansion of 1^{st}-order Volterra kernels or finite impulse response models (Marmarelis 2004, Mitsis, Poulin et al. 2004, Mitsis, Zhang et al. 2009, Marmarelis, Shin et al. 2013, Marmarelis, Shin et al. 2014, Marmarelis, Shin et al. 2014), wavelet analysis (Torrence and Webster 1999, Grinsted, Moore et al. 2004, Peng, Rowley et al. 2010), parametric finite-impulse response filter based methods (Panerai, Simpson et al. 2000, Simpson, Panerai et al. 2001), ARI analysis (Panerai, White et al. 1998), autoregressive moving average (ARMA) based ARI methods and variant ARI methods (Panerai, Eames et al. 2003), autoregressive with exogenous input (ARX) methods (Liu and Allen 2002, Liu, Birch et al. 2003, Panerai, Eames et al. 2003) and correlation coefficient-like indices (Heskamp, Meel-van den Abeelen et al. 2013, Caicedo, Varon et al. 2016). A summary of the methods and corresponding references are given in Table 2.

Reproducibility of DCA metrics

For the reproducibility and variability analysis of the DCA parameters, DCA- methods were grouped into three broad categories: 1. TFA-like output; 2. ARI-like output and 3. correlation coefficient-like outputs. These categories were created from the perspective of similar output parameters, not because of similarity on mathematical grounds. In general, all centers were free to use their own settings to cover the standard frequency range between 0-0.5 Hz. In the majority of cases though, for the TFA-like output methods, the settings for TFA were similar to what was later proposed in the CARNet White Paper (Claassen, Meel-van den Abeelen et al. 2016). In summary, this involved spectral estimates using the Welch method with multiple segments of data of at least 100 s, 50% superposition and cosine windowing to reduce spectral leakage. Individual method settings are listed in Table S4. Estimates of gain and phase were averaged for different frequency bands, very-low frequency (VLF) and low frequency (LF) bands (Table S4) (Claassen, Meel-van den Abeelen et al. 2016).
The ARI-like output methods consisted of time domain estimates of the impulse or step response, using the inverse Fourier transform of gain and phase, or ARMA models (Panerai, White et al. 1998, Liu and Allen 2002, Liu, Birch et al. 2003, Panerai, Eames et al. 2003).
Finally, the correlation coefficient-like outputs consisted of a single parameter, obtained by linear regression or similar methods (Heskamp, Meel-van den Abeelen et al. 2013, Caicedo, Varon et al. 2016).

Statistical analysis

We assessed reproducibility as follows: To quantify the level of agreement between first and second measurement, we applied the Bland-Altman method to obtain mean difference (or bias) and to determine limits of agreement (LOA). This was done for the methods in the TFA-like, ARI-like and correlation-like category. A non-parametric Wilcoxon signed rank test was used to check if there were significant differences between left and right side results. Left and right output results were averaged for further analyses. To correct for abnormal data distributions, Box-Cox transformations were performed, which is a power transformation with different power levels (Box and Cox 1964). Within one analysis method, the same transformation was applied to both the first and second measurement, but different transformations may be used for different methods and different variables.

Further quantification of agreement between the repeated measurements for all DCA analysis methods was determined by one way intraclass correlation coefficient analysis (ICC). ICC results of TFA-like methods combined for the parameters gain and phase were compared for VLF and LF. Furthermore, the differences between the ICC results of previously obtained surrogate data (Sanders, Claassen et al. 2018) and physiological data were analyzed for the methods combined in parameters gain VLF, gain LF, phase VLF, phase LF, ARI and correlation. These differences between ICC parameter values were tested with the paired Wilcoxon signed rank test, considering that most parameters, such as TFA estimates, are not normally distributed. SPSS 22 was used for all analyses, a value of $p<0.05$ was adopted to indicate statistical significance.

Interpretation of the absolute and maximal values of ICC were based on often quoted guidelines : Poor (ICC<0.40); Fair (0.40 to 0.59; Good (0.60 to 0.74); Excellent (0.75 to 1.00).(Cicchetti 1994)

RESULTS

Subject characteristics are listed in Table 1. No significant differences were found for MAP, CBFv and $EtCO_2$ for the two measurements (T1 and T2).

The scatterplots of Figure 1(a) show examples of TFA- like metrics of the estimated LF gain and Figure 1(b) of the ARI-like results of the repeated measurements for both physiological and surrogate data. The figures show a difference in distribution of the data between Figure 1(a) and 1(b), with a higher correlation between the repeated measurements for lower gain values only in the TFA like results. Despite the lower number of cases in the surrogate results, it is clearly shown that there is less variability in the surrogate data (bottom) compared to physiological data (top) for all TFA-like methods (Figure 1(a)) and the ARI an IR-filter methods (Figure 1(b)).

Comparing different autoregulation metrics with Bland Altman analysis, we see a difference between gain variables and all the other variables (Figure 2). Both gain VLF and LF show a strong increase in the difference between two measurements on the y-axis for higher values of mean gain on the x-axis. For the smallest values of gain, where the DCA is considered most effective, the agreement is the strongest. Results for T1, T2, bias (T1-T2) and the LOA of the different method categories per method group are listed in Table 3. Each method group corresponds to results of several methods combined (Table 2, Table S3(a-c)).

Left and right ICC results were not different. ICC analysis of physiological data is shown in Figure 3. Despite minor differences in ICC values between methods, 12 methods qualified as having good reproducibility (ICC >0.6). TFA-like and ARI-like methods scored significantly higher ICC for surrogate data compared to physiological data, combined for centres using the same methods, for gain VLF ($p<0.001$), gain LF ($p<0.001$), phase VLF ($p<0.001$), phase LF ($p<0.001$) and ARI ($p=0.018$) (Sanders, Claassen et al. 2018). Only the correlation like methods did not score higher ICC values for surrogate data compared to physiological data ($p=0.18$). ICC results of the surrogate data are presented in Table S5.

For the TFA-like methods, ICC gain VLF (mean (SD)) was lower than ICC gain LF, respectively 0.38 (0.057) and 0.59 (0.078), $p<0.001$. Also for phase, the corresponding ICC values were lower for VLF than for LF, 0.17 (0.13) and 0.39 (0.11) respectively, $p=0.001$. For ARI-like methods the mean (SD) ICC results were 0.30 (0.12) and for the correlation-like 0.24 (0.21).

DISCUSSION

With this multi-center, multi-method study we aimed to provide an internationally representative and broader evaluation of the reproducibility of many DCA assessment methods. By comparing real physiological measurements with those where physiological variability was reduced by use of surrogate data, we have been able to assess the contribution of physiological non-stationary to the reproducibility of DCA parameters. For surrogate data, with realistic CBFv signals generated from measured BP data, we had demonstrated good to excellent reproducibility for most DCA methods. We now hypothesized that in real recordings of BP and CBF, non-stationarity in the BP-CBF relationship would reduce reproducibility for these DCA methods.

We asked researchers from various centers with expertise in DCA to apply their DCA method(s) to a common dataset with repeated physiological measurements of BP and CBFv. Participating centers, and respective analytical methods, are representative of the literature on

DCA assessment (Panerai, Rennie et al. 1998, Panerai, White et al. 1998, Zhang, Zuckerman et al. 1998, Torrence and Webster 1999, Panerai, Simpson et al. 2000, Simpson, Panerai et al. 2001, Liu and Allen 2002, Mitsis, Zhang et al. 2002, Liu, Birch et al. 2003, Muller, Bianchi et al. 2003, Panerai, Eames et al. 2003, Reinhard, Muller et al. 2003, Grinsted, Moore et al. 2004, Marmarelis 2004, Mitsis, Poulin et al. 2004, Liu, Simpson et al. 2005, Mitsis, Zhang et al. 2009, Gommer, Shijaku et al. 2010, Peng, Rowley et al. 2010, van Beek, Olde Rikkert et al. 2010, Heskamp, Meel-van den Abeelen et al. 2013, Marmarelis, Shin et al. 2013, Marmarelis, Shin et al. 2014, Marmarelis, Shin et al. 2014, Meel-van den Abeelen, Simpson et al. 2014, Muller and Osterreich 2014, Panerai 2014, Caicedo, Varon et al. 2016).

Main findings

Two main outstanding findings came out of the study: i) the reproducibility of most DCA metrics, independently of the analytical approach adopted, should be regarded as 'poor', given the prevailing values of ICC<0.4,(Cicchetti 1994); and ii) physiological variability is likely to be the main reason for the degradation in reproducibility, when compared to results obtained from surrogate data (Sanders, Claassen et al. 2018).

Strictly speaking, these results indicate that, at this moment, most DCA metrics do not meet criteria for individual and clinical use for diagnostic and/or monitoring purposes. Despite the high variability across DCA parameters, only TFA and ARX scored ICC results that could be categorized as 'good' (ICC>0.6, Figure 3) for approximately half of the gain metrics in the LF band (Cicchetti 1994). As discussed in more detail below though, these findings need to be placed into perspective, taking into account methodological issues and current knowledge of the wider application of DCA assessment metrics.

Methodological considerations

Although indicative of the deterioration of DCA metrics, from what was obtained with surrogate data, to the case of 'real' physiological measurements, the ICC can be misleading when estimated using only healthy subjects. Differently from the intra-subject standard error, the ICC takes into account both intra- and inter-subject variability. Given that healthy subjects would be expected to cluster around values indicative of a good working DCA, this would reduce inter-subject variability, in comparison with intra-subject variance, thus putting a bias towards reduced values of ICC. However, as can be observed in Figure 1, there was wide inter-subject variability, indicating that this alone cannot explain the low ICC results. Nonetheless, despite the indication that most DCA metrics have limited reproducibility, it would be premature to use our findings to put a halt on their use in physiological and clinical studies, before further research is conducted, ideally assessing the ICC for much larger cohorts of both patients and healthy individuals.

The analysis of physiological data presents large within and between subject variability, similar to what has been reported before in patient data (Gommer, Shijaku et al. 2010, van Beek, Olde Rikkert et al. 2010, Elting, Aries et al. 2014, Smirl, Hoffman et al. 2015). Non-Gaussian distributions were corrected by the Box-Cox transformations (Box and Cox 1964). The ICC values were much lower than what was found when these same methods were applied to analyze surrogate data (Sanders, Claassen et al. 2018). In that study, physiological variability was reduced to only the BP signal, because the CBF signal was software-generated using the repeated BP signals as input. Even though realistic levels of noise were added to the generated CBF signal, all DCA methods demonstrated good to excellent reproducibility (ICC 0.6-1.00) on those surrogate data, whereas the majority of these same methods had poor

reproducibility (ICC <0.4) for the current dataset where both BP and CBF signals represented physiological data. One interpretation of these results is that the poor reproducibility of DCA is not solely explained because the methods provide poor accuracy or poor precision. With surrogate data, all methods showed accuracy and precision, leading to good reproducibility.

Comparable with results of Smirl et al.(Smirl, Hoffman et al. 2015), the highest ICC results were obtained with gain LF parameters, although Figure 2 shows that reproducibility differs for different gain values, with highest reproducibility for lower gain values. This is a proportional increase in variability, recognizable by the arrowhead shape in Figure 2. ICC for gain and phase parameters are decreased in VLF compared to LF, and may be explained by the lower coherence between BP and CBFv in VLF oscillations, resulting in wider confidence limits for VLF and lower ICC values. Comparing gain ICC results with phase, one can see decreased reproducibility in the phase results over both frequency bands. This does not immediately favor gain parameters as more suitable DCA metrics, since a lower ICC value for phase can be expected purely based on the definition and dependence between the two parameters (Bendat and Piersol 1986). This explains that confidence limits will automatically be wider for phase compared to gain. We recommend to routinely plot confidence limits when creating TFA results.

To improve reproducibility, it may be beneficial to use measurement conditions where the DCA regulatory system is maximally activated, for example in sit-to-stand measurements (Simpson and Claassen 2018) or squat-stand measurements (Smirl, Hoffman et al. 2015). This may result in minimal gain values in the LF band and improve reproducibility. However, it remains an ongoing debate whether TFA gain is the most suitable parameter to reflect state of DCA, or if phase may be more physiologically relevant.

Clinical implications

Given the limited reproducibility shown by most indices of DCA, to what extent should we trust their use in clinical studies? This is a crucial question given the stage of research on DCA, with many centers advocating the use of DCA metrics in clinical decision-making and patient management. In this context, the results of this study might be a watershed. Until recently, the prevailing view has been that, amongst a plethora of DCA metrics, there could be one that could become a 'gold standard' based on its reproducibility, as well as its sensitivity and specificity, to detect changes in DCA, either due to disease or physiological status. What this study is showing though, is that none of the methods in use could fulfill this role, at least not as reproducibility is concerned. Furthermore, the comparison between physiological and surrogate data, also suggests that it is unlikely that other current or future methods will have an outstanding reproducibility either. The reason for this somber perspective lies with the growing awareness that regulation of CBF, not only in response to BP changes, but also due to changes in CO_2 or neural stimulation, is a highly non-stationary phenomenon, thus requiring an entirely different conceptual paradigm to ascertain their clinical usefulness (Panerai 2014). On the other hand, it is not all gloom and doom. Looking back into a vast literature, too extensive to be enumerated here, reporting on clinical applications of most of the DCA metrics included in this study, there is plenty of evidence to suggest their sensitivity to detect worsening DCA in a range of cerebrovascular and, increasingly, also systemic conditions. To study reproducibility in the presence of disease is a major challenge though, as patient conditions are either worsening or improving on a daily basis. Nevertheless, several follow up studies have been able to use diverse indices of DCA to describe the natural history of conditions like severe head injury (Czosnyka, Smielewski et al. 1997), ischemic stroke (Salinet, Panerai et al. 2014) or intracerebral hemorrhage (Ma, Guo et

al. 2016) which is also reassuring. Certainly much more research is needed, mainly to understand the nature of DCA non-stationarity and how this is affected by, and manifested in, clinical conditions, to improve the reliability and usefulness of DCA assessment for patient care.

Limitations and future directions

Only methods that could be applied to short data segments (5 min) were evaluated, therefore the correlation-like methods were underrepresented, The correlation-like methods clearly showed reduced reproducibility compared to the other categories (Figure 3) under these conditions.

It is difficult to select a suitable method to assess reproducibility of DCA analysis parameters. We selected ICC, although this method being sensitive to outliers. This has probably affected phase VLF results the strongest in a negative way, since high variability and outliers were most present in phase VLF.

The time interval differences between repeated measurements were not considered in the analysis. A dataset consisting of rest measurements was used, with limited BP fluctuations, resulting in a low power of BP and CBFv oscillations. At rest, cerebral perfusion is usually well maintained and DCA may not be activated, while during a physical challenge, when sufficient DCA functioning is crucial, will give more meaningful results (Simpson and Claassen 2018, Simpson and Claassen 2018, Tzeng and Panerai 2018, Tzeng and Panerai 2018). Moreover, it will be relevant to add clinical data to the healthy controls to have a greater spread of inter-subject variability.

It could not yet be answered what the precise reason is for low reproducibility of DCA assessment in physiological data. It is necessary to study physiological variation in DCA function within individuals in repeated measurements. From a theoretical perspective, the variability in DCA results can be reduced in two ways: Increase the coherence or increase the number of averages (Bendat and Piersol 1986, Halliday, Rosenberg et al. 1995). To increase the coherence, oscillations could be induced and included in the measurement protocol. Increased coherence could also be achieved by selection of the data used for DCA analysis based on the power of BP oscillations. This line of investigation will be pursued as part of this wider project. To increase the number of averages, more or longer measurement protocols should be used, although duration of recordings is usually limited in most clinical settings.

Selecting the most promising DCA parameter is complex, since the most reproducible parameter is not necessarily the best parameter to reflect DCA status. Although there was not a single method that outperformed others both linear and non-linear, there are inter-method differences that are worth investigating. In particular, future studies could look to the influence of measurement length or increased oscillations in the measurement protocol or data selection (Simpson and Claassen 2018).

Furthermore, the question to answer is to what extent does reproducibility depend on autoregulation status. Are DCA parameters less reproducible in case of worse DCA status and functioning? One interesting and relatively easy next step could be to perform repeated measurements in hypercapnic data (Katsogridakis, Bush et al. 2013), as a model for impaired DCA, and compare these with repeated measurements in normocapnia to assess differences in reproducibility.

CONCLUSION

The physiological nature of these measurements strongly reduced reproducibility of DCA when assessed in short data recordings in healthy subjects. This conclusion is not affected by the choice of analytical method used to derive different DCA metrics, or by local procedures in multiple international centres which participated in this study. Further investigation is needed to improve our understanding of how physiological variability affects DCA reproducibility in health and disease.

AUTHOR CONTRIBUTIONS

MS, JWE, RBP, JC developed the idea for the study and drafted the manuscript . All authors performed data analyses,. participated in revising the manuscript and have approved the final version of this paper prior to submission.

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ACKNOWLEDGEMENTS

We would like to thank all subjects who contributed with data for this study.

DECLARATION OF CONFLICTING INTERESTS

The authors declared no potential conflict of interest.

REFERENCES

Aaslid, R., K. F. Lindegaard, W. Sorteberg and H. Nornes (1989). "Cerebral autoregulation dynamics in humans." Stroke **20**(1): 45-52.

Bendat, J. S. and A. G. Piersol (1986). Random Data: Analysis and Measurement Procedures. New York, Wiley.

Box, G. E. P. and D. R. Cox (1964). "An Analysis of Transformations." Journal of the Royal Statistical Society Series B-Statistical Methodology **26**(2): 211-252.

Brodie, F. G., E. R. Atkins, T. G. Robinson and R. B. Panerai (2009). "Reliability of dynamic cerebral autoregulation measurement using spontaneous fluctuations in blood pressure." Clin Sci (Lond) **116**(6): 513-520.

Caicedo, A., C. Varon, B. Hunyadi, M. Papademetriou, I. Tachtsidis and S. Van Huffel (2016). "Decomposition of Near-Infrared Spectroscopy Signals Using Oblique Subspace Projections: Applications in Brain Hemodynamic Monitoring." Front Physiol **7**: 515.

Cicchetti, D. (1994). Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instrument in Psychology.

Claassen, J. A., A. S. Meel-van den Abeelen, D. M. Simpson, R. B. Panerai and N. international Cerebral Autoregulation Research (2016). "Transfer function analysis of dynamic cerebral autoregulation: A white paper from the International Cerebral Autoregulation Research Network." J Cereb Blood Flow Metab **36**(4): 665-680.

Czosnyka, M., P. Smielewski, P. Kirkpatrick, R. J. Laing, D. Menon and J. D. Pickard (1997). "Continuous assessment of the cerebral vasomotor reactivity in head injury." Neurosurgery **41**(1): 11-17.

Elting, J. W., M. J. Aries, J. H. van der Hoeven, P. C. Vroomen and N. M. Maurits (2014). "Reproducibility and variability of dynamic cerebral autoregulation during passive cyclic leg raising." Med Eng Phys **36**(5): 585-591.

Gommer, E. D., E. Shijaku, W. H. Mess and J. P. Reulen (2010). "Dynamic cerebral autoregulation: different signal processing methods without influence on results and reproducibility." Med Biol Eng Comput **48**(12): 1243-1250.

Grinsted, A., J. C. Moore and S. Jevrejeva (2004). "Application of the cross wavelet transform and wavelet coherence to geophysical time series." Nonlinear Processes in Geophysics **11**(5-6): 561-566.

Halliday, D. M., J. R. Rosenberg, A. M. Amjad, P. Breeze, B. A. Conway and S. F. Farmer (1995). "A framework for the analysis of mixed time series/point process data - Theory and application to the study of physiological tremor, single motor unit discharges and electromyograms." Progress in Biophysics & Molecular Biology **64**(2-3): 237-278.

Heskamp, L., A. Meel-van den Abeelen, E. Katsogridakis, R. Panerai, D. Simpson, J. Lagro and J. Claassen (2013). "Convergent cross mapping: a promising technique for future cerebral autoregulation estimation." Cerebrovascular Diseases **35**: 15-16.

Katsogridakis, E., G. Bush, L. Fan, A. A. Birch, D. M. Simpson, R. Allen, J. F. Potter and R. B. Panerai (2013). "Detection of impaired cerebral autoregulation improves by increasing arterial blood pressure variability." J Cereb Blood Flow Metab **33**(4): 519-523.

Kostoglou, K., C. T. Debert, M. J. Poulin and G. D. Mitsis (2014). "Nonstationary multivariate modeling of cerebral autoregulation during hypercapnia." Med Eng Phys **36**(5): 592-600.

Liu, J., D. M. Simpson and R. Allen (2005). "High spontaneous fluctuation in arterial blood pressure improves the assessment of cerebral autoregulation." Physiol Meas **26**(5): 725-741.

Liu, Y. and R. Allen (2002). "Analysis of dynamic cerebral autoregulation using an ARX model based on arterial blood pressure and middle cerebral artery velocity simulation." Med Biol Eng Comput **40**(5): 600-605.

Liu, Y., A. A. Birch and R. Allen (2003). "Dynamic cerebral autoregulation assessment using an ARX model: comparative study using step response and phase shift analysis." Med Eng Phys **25**(8): 647-653.

Ma, H., Z. N. Guo, J. Liu, Y. Xing, R. Zhao and Y. Yang (2016). "Temporal Course of Dynamic Cerebral Autoregulation in Patients With Intracerebral Hemorrhage." Stroke **47**(3): 674-681.

Marmarelis, V. Z. (2004). Nonlinear Dynamic Modeling of Physiological Systems. New Jersey, Wiley-Interscience.

Marmarelis, V. Z., D. C. Shin, M. Orme and Z. Rong (2014). "Time-varying modeling of cerebral hemodynamics." IEEE Trans Biomed Eng **61**(3): 694-704.
Marmarelis, V. Z., D. C. Shin, M. E. Orme and R. Zhang (2013). "Model-based quantification of cerebral hemodynamics as a physiomarker for Alzheimer's disease?" Ann Biomed Eng **41**(11): 2296-2317.
Marmarelis, V. Z., D. C. Shin, M. E. Orme and R. Zhang (2014). "Model-based physiomarkers of cerebral hemodynamics in patients with mild cognitive impairment." Med Eng Phys **36**(5): 628-637.
Meel-van den Abeelen, A. S., D. M. Simpson, L. J. Wang, C. H. Slump, R. Zhang, T. Tarumi, C. A. Rickards, S. Payne, G. D. Mitsis, K. Kostoglou, V. Marmarelis, D. Shin, Y. C. Tzeng, P. N. Ainslie, E. Gommer, M. Muller, A. C. Dorado, P. Smielewski, B. Yelicich, C. Puppo, X. Liu, M. Czosnyka, C. Y. Wang, V. Novak, R. B. Panerai and J. A. Claassen (2014). "Between-centre variability in transfer function analysis, a widely used method for linear quantification of the dynamic pressure-flow relation: the CARNet study." Med Eng Phys **36**(5): 620-627.
Mitsis, G. D., M. J. Poulin, P. A. Robbins and V. Z. Marmarelis (2004). "Nonlinear modeling of the dynamic effects of arterial pressure and CO2 variations on cerebral blood flow in healthy humans." IEEE Trans Biomed Eng **51**(11): 1932-1943.
Mitsis, G. D., R. Zhang, B. D. Levine and V. Z. Marmarelis (2002). "Modeling of nonlinear physiological systems with fast and slow dynamics. II. Application to cerebral autoregulation." Ann Biomed Eng **30**(4): 555-565.
Mitsis, G. D., R. Zhang, B. D. Levine, E. Tzanalaridou, D. G. Katritsis and V. Z. Marmarelis (2009). "Autonomic neural control of cerebral hemodynamics." IEEE Eng Med Biol Mag **28**(6): 54-62.
Muller, M., O. Bianchi, S. Erulku, C. Stock, K. Schwerdtfeger and G. Homburg Traumatic Brain Injury (2003). "Changes in linear dynamics of cerebrovascular system after severe traumatic brain injury." Stroke **34**(5): 1197-1202.
Muller, M. W. and M. Osterreich (2014). "A comparison of dynamic cerebral autoregulation across changes in cerebral blood flow velocity for 200 s." Front Physiol **5**: 327.
Panerai, R. B. (2008). "Cerebral autoregulation: from models to clinical applications." Cardiovasc Eng **8**(1): 42-59.
Panerai, R. B. (2014). "Nonstationarity of dynamic cerebral autoregulation." Med Eng Phys **36**(5): 576-584.
Panerai, R. B., P. J. Eames and J. F. Potter (2003). "Variability of time-domain indices of dynamic cerebral autoregulation." Physiol Meas **24**(2): 367-381.
Panerai, R. B., J. M. Rennie, A. W. Kelsall and D. H. Evans (1998). "Frequency-domain analysis of cerebral autoregulation from spontaneous fluctuations in arterial blood pressure." Med Biol Eng Comput **36**(3): 315-322.
Panerai, R. B., D. M. Simpson, S. T. Deverson, P. Mahony, P. Hayes and D. H. Evans (2000). "Multivariate dynamic analysis of cerebral blood flow regulation in humans." IEEE Trans Biomed Eng **47**(3): 419-423.
Panerai, R. B., R. P. White, H. S. Markus and D. H. Evans (1998). "Grading of cerebral dynamic autoregulation from spontaneous fluctuations in arterial blood pressure." Stroke **29**(11): 2341-2346.
Peng, T., A. B. Rowley, P. N. Ainslie, M. J. Poulin and S. J. Payne (2010). "Wavelet phase synchronization analysis of cerebral blood flow autoregulation." IEEE Trans Biomed Eng **57**(4): 960-968.
Reinhard, M., T. Muller, B. Guschlbauer, J. Timmer and A. Hetzel (2003). "Transfer function analysis for clinical evaluation of dynamic cerebral autoregulation--a comparison between spontaneous and respiratory-induced oscillations." Physiol Meas **24**(1): 27-43.
Salinet, A. S., R. B. Panerai and T. G. Robinson (2014). "The longitudinal evolution of cerebral blood flow regulation after acute ischaemic stroke." Cerebrovasc Dis Extra **4**(2): 186-197.
Sanders, M. L., J. Claassen, M. Aries, E. Bor-Seng-Shu, A. Caicedo, M. Chacon, E. D. Gommer, S. Van Huffel, J. L. Jara, K. Kostoglou, A. Mahdi, V. Z. Marmarelis, G. D. Mitsis, M. Muller, D. Nikolic, R. C. Nogueira, S. J. Payne, C. Puppo, D. C. Shin, D. M. Simpson, T. Tarumi, B. Yelicich, R. Zhang, R. B. Panerai and J. W. J. Elting (2018). "Reproducibility of dynamic cerebral autoregulation parameters: a multi-centre, multi-method study." Physiol Meas **39**(12): 125002.

Simpson, D. and J. Claassen (2018). "CrossTalk opposing view: dynamic cerebral autoregulation should be quantified using induced (rather than spontaneous) blood pressure fluctuations." J Physiol **596**(1): 7-9.
Simpson, D. and J. Claassen (2018). "Rebuttal from David Simpson and Jurgen Claassen." J Physiol **596**(1): 13.
Simpson, D. M., R. B. Panerai, D. H. Evans and A. R. Naylor (2001). "A parametric approach to measuring cerebral blood flow autoregulation from spontaneous variations in blood pressure." Ann Biomed Eng **29**(1): 18-25.
Smirl, J. D., K. Hoffman, Y. C. Tzeng, A. Hansen and P. N. Ainslie (2015). "Methodological comparison of active- and passive-driven oscillations in blood pressure; implications for the assessment of cerebral pressure-flow relationships." J Appl Physiol (1985) **119**(5): 487-501.
Tiecks, F. P., A. M. Lam, R. Aaslid and D. W. Newell (1995). "Comparison of static and dynamic cerebral autoregulation measurements." Stroke **26**(6): 1014-1019.
Torrence, C. and P. J. Webster (1999). "Interdecadal changes in the ENSO-monsoon system." Journal of Climate **12**(8): 2679-2690.
Tzeng, Y. C. and R. B. Panerai (2018). "CrossTalk proposal: dynamic cerebral autoregulation should be quantified using spontaneous blood pressure fluctuations." J Physiol **596**(1): 3-5.
Tzeng, Y. C. and R. B. Panerai (2018). "Rebuttal from Y. C. Tzeng and R. B. Panerai." J Physiol **596**(1): 11-12.
van Beek, A. H., J. A. Claassen, M. G. Rikkert and R. W. Jansen (2008). "Cerebral autoregulation: an overview of current concepts and methodology with special focus on the elderly." J Cereb Blood Flow Metab **28**(6): 1071-1085.
van Beek, A. H., J. Lagro, M. G. Olde-Rikkert, R. Zhang and J. A. Claassen (2012). "Oscillations in cerebral blood flow and cortical oxygenation in Alzheimer's disease." Neurobiol Aging **33**(2): 428 e421-431.
van Beek, A. H., M. G. Olde Rikkert, J. W. Pasman, M. T. Hopman and J. A. Claassen (2010). "Dynamic cerebral autoregulation in the old using a repeated sit-stand maneuver." Ultrasound Med Biol **36**(2): 192-201.
Zhang, R., J. H. Zuckerman, C. A. Giller and B. D. Levine (1998). "Transfer function analysis of dynamic cerebral autoregulation in humans." Am J Physiol **274**(1 Pt 2): H233-241.

Figure legends

Figure 1(a). Gain LF results of TFA-like methods for repeated measurements. Top row: physiological data, bottom row: surrogate data. For each method group (TFA, Laguerre, Wavelet, IR-filter, ARX) the results of similar methods are combined (Table 2). TFA: black dots are 10 methods (cm/s/mmHg), grey dots are 3 methods (%/% or %/mmHg); Laguerre: 4 methods (cm/s/mmHg); Wavelet: 1 method (cm/s/mmHg); IR-filter: 2 methods (%/%); ARX: 2 methods (cm/s/mmHg). See Figure S1-S3 for Phase VLF/LF and Gain VLF.

Figure 1(b). ARI-like results of different methods for repeated measurements. Top row: physiological data, bottom row: surrogate data. For each method group (ARI/ARMA, ARX, IR-filter, correlation) the results of similar methods are combined (Table 2). ARI: black dots are 3 methods (ARI 0-9 arbitrary units); grey dots are 2 methods (ARMA-ARI 0-9 arbitrary units); ARX: 1 method (ARX coefficient); IR-filter: 1 method (arbitrary units); Correlation: 2 methods.

Figure 2. Bland-Altman plot of TFA like parameters: gain VLF (top left), gain LF (top right), phase VLF (middle left) and phase LF (middle right); ARI-like parameters (bottom left); correlation-like parameters (bottom right). Units are similar to Figure 1A and B.

Figure 3. ICC values for methods using TFA or similar approaches with gain VLF and LF (top), phase VLF or LF (middle) and ARI or correlation-like methods (bottom). Results are shown per method (Table 2). ICC values less than 0.40: Poor, between 0.40 and 0.59: Fair, between 0.60 and 0.74: Good, between 0.75 and 1.00: Excellent (Cicchetti 1994).

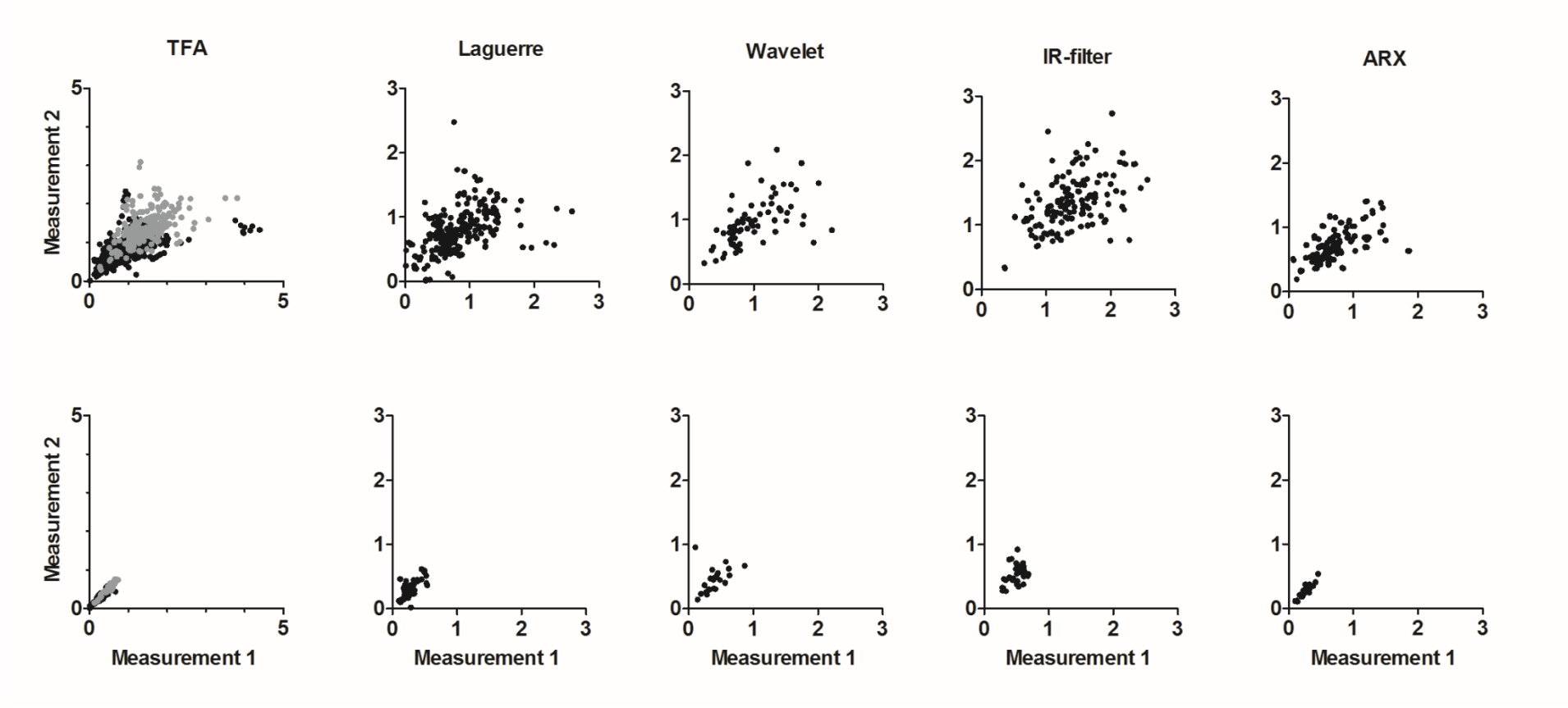
TFA
Laguerre
Wavelet
IR-filter
ARX
Measurement 2
Measurement 1
0
1
2
3
5

Figure 1(a)

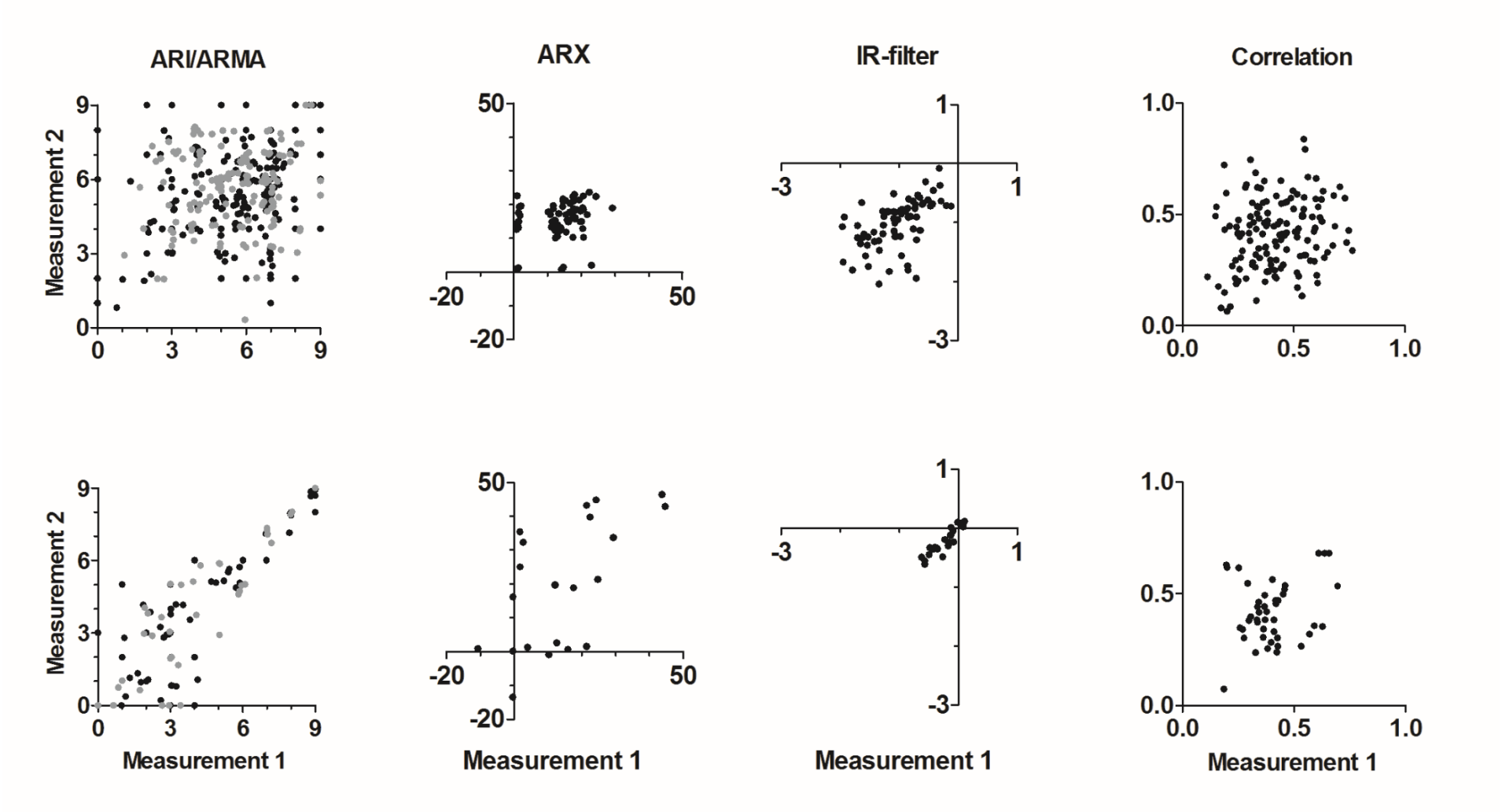
ARI/ARMA
ARX
IR-filter
Correlation
Measurement 2
Measurement 1
9
6
3
0
50
-20
1
-3
1.0
0.5
0.0

Figure 1 (b)

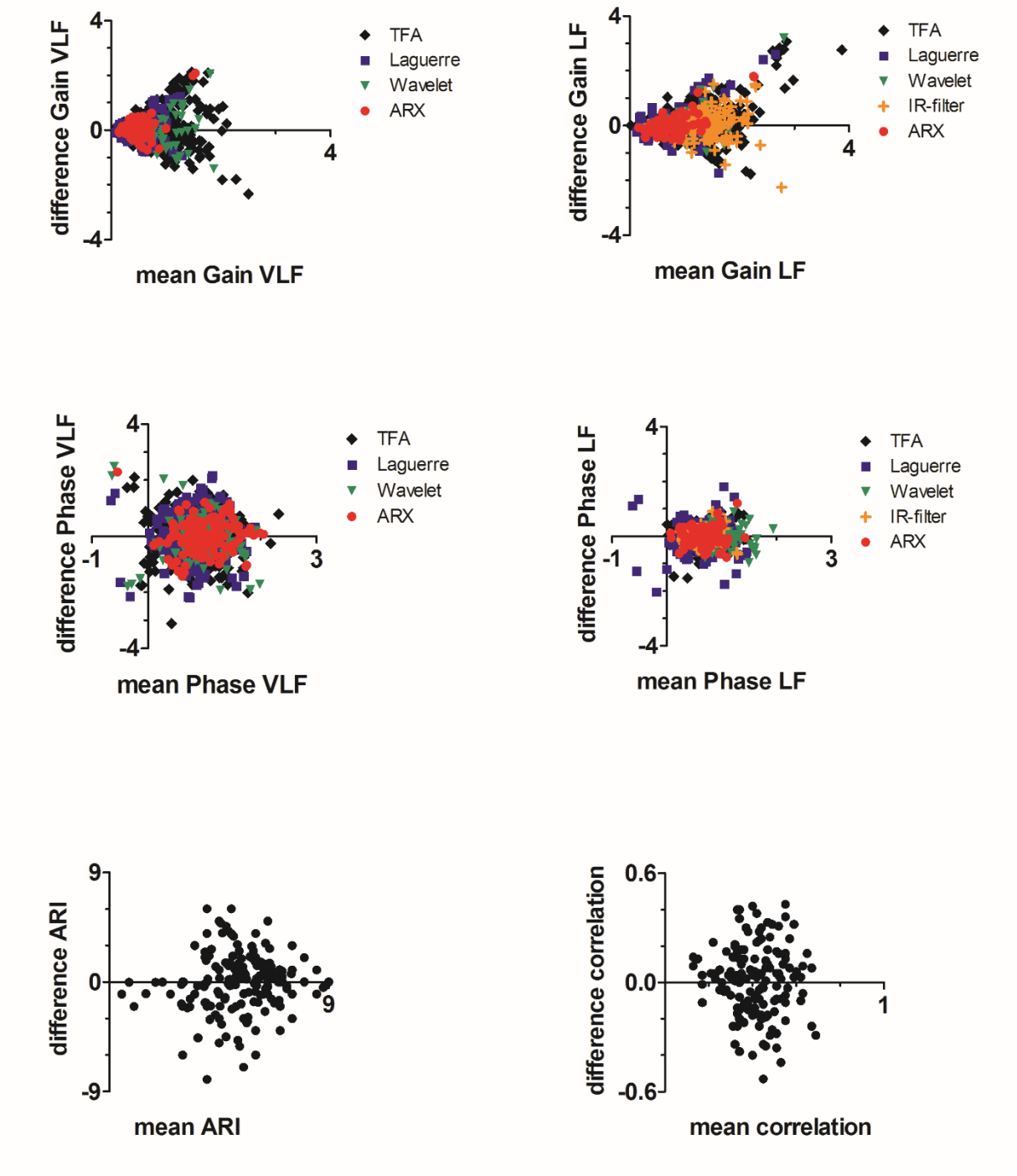
difference Gain VLF
mean Gain VLF
TFA
Laguerre
Wavelet
ARX
difference Gain LF
mean Gain LF
TFA
Laguerre
Wavelet
IR-filter
ARX
difference Phase VLF
mean Phase VLF
TFA
Laguerre
Wavelet
ARX
difference Phase LF
mean Phase LF
TFA
Laguerre
Wavelet
IR-filter
ARX
difference ARI
mean ARI
difference correlation
mean correlation

Figure 2

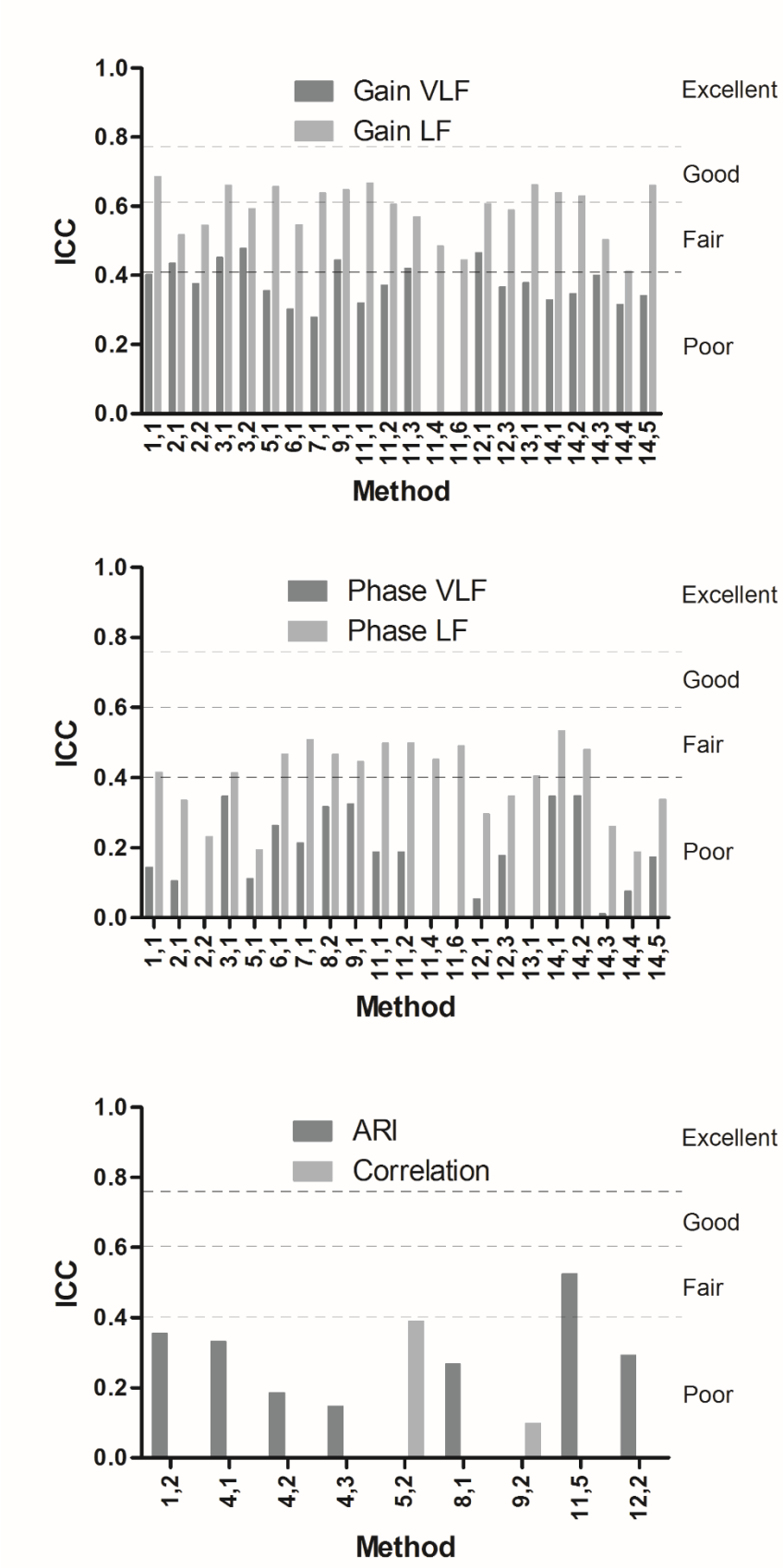
Gain VLF
Gain LF
Excellent
Good
Fair
Poor
ICC
1.0
0.8
0.6
0.4
0.2
0.0
1,1 2,1 2,2 3,1 3,2 5,1 6,1 7,1 9,1 11,1 11,2 11,3 11,4 11,6 12,1 12,3 13,1 14,1 14,2 14,3 14,4 14,5
Method
Phase VLF
Phase LF
Excellent
Good
Fair
Poor
ICC
1.0
0.8
0.6
0.4
0.2
0.0
1,1 2,1 2,2 3,1 5,1 6,1 7,1 8,2 9,1 11,1 11,2 11,4 11,6 12,1 12,3 13,1 14,1 14,2 14,3 14,4 14,5
Method
ARI
Correlation
Excellent
Good
Fair
Poor
ICC
1.0
0.8
0.6
0.4
0.2
0.0
1,2 4,1 4,2 4,3 5,2 8,1 9,2 11,5 12,2
Method

Figure 3

Tables

Table 1. Subject characteristics and hemodynamic parameters

n	75	
Age, years	47.8 ± 18.6	
Female, n (%)	33 (44)	
Use of AHD, n (%)	5 (6.7)	
Use of NSAID, n (%)	4 (5.3)	
MCI, n (%)	4 (5.3)	
	T1	**T2**
MAP, mmHg	90.1 ± 14.9	87.6 ± 14.8
MCBFv, cm/s	56.3 ± 13.4	56.2 ± 12.5
$EtCO_2$, kPA	5.0 ± 0.5	5.0 ± 0.5

Values are presented as mean ± SD or n (%). AHD, antihypertensive drugs; NSAID, nonsteroidal anti-inflammatory drug; MCI, mild cognitive impairment; MAP, mean arterial pressure; MCBFv, mean blood flow velocity; $EtCO_2$, end-tidal CO_2 of measurement 1 (T1) and measurement 2 (T2).

Table 2. Methods with corresponding output variables per centre

Center number	Method	Output Variables	Category	Method group	References
1	1.1 Transfer Function Analysis 1.2 Autoregulation index	Coherence, Gain (cm/s/mmHg) and Phase (rad) in VLF, LF ARI	1 2	1 6	(Zhang, Zuckerman et al. 1998) (Panerai, White et al. 1998)
2	2.1 Laguerre expansion of 1st-order Volterra kernels, single input (BP) 2.2 Laguerre expansion of 1st-order Volterra kernels, dual input (BP, CO_2)	Gain (cm/s/mmHg) and Phase (rad) in VLF, LF Gain (cm/s/mmHg) and Phase (rad) in VLF, LF	1 1	2 2	(Marmarelis 2004, Marmarelis, Shin et al. 2013, Marmarelis, Shin et al. 2014, Marmarelis, Shin et al. 2014)
3	3.1 Transfer Function Analysis 3.2 Transfer Function Analysis	Coherence, Gain (cm/s/mmHg), Phase (rad) in VLF, LF Coherence, Gain (%/%) in VLF, LF	1 1	1 1	(Zhang, Zuckerman et al. 1998)
4	4.1 Autoregulation index (FFT) 4.2 Autoregulation index (Moving Average 1) 4.3 Autoregulation index (Moving Average 2)	ARI ARI ARI	2 2 2	6 7 7	(Panerai, White et al. 1998, Panerai, Eames et al. 2003)
5	5.1 Transfer Function Analysis 5.2 Oblique and Orthogonal Subspace Projections	Coherence, Gain (cm/s/mmHg), Phase (rad) in VLF, LF Subspace Ratio's	1 3	1 10	(Zhang, Zuckerman et al. 1998) (Caicedo, Varon et al. 2016)
6	6.1 Transfer Function Analysis	Coherence, Gain (cm/s/mmHg), Phase (rad) in VLF, LF	1	1	(Muller, Bianchi et al. 2003, Muller and Osterreich 2014)
7	7.2 Transfer Function Analysis	Coherence, Gain (cm/s/mmHg), Phase (rad) in VLF, LF	1	1	(Gommer, Shijaku et al. 2010)
8	8.1 ARX 8.2 Wavelet Analysis	ARX Coefficient (3rd) Synchronization index, Phase (rad) in VLF,LF	2 1	7 3	(Liu and Allen 2002, Liu, Birch et al. 2003, Panerai, Eames et al. 2003) (Peng, Rowley et al. 2010)
9	9.1 Transfer Function Analysis 9.2 Convergent cross mapping	Coherence, Gain (cm/s/mmHg), Phase (rad) in VLF, LF CCM correlation coefficient	1 3	1 10	(van Beek, Olde Rikkert et

					al. 2010, van Beek, Lagro et al. 2012) (Heskamp, Meel-van den Abeelen et al. 2013)
11	11.1 Transfer Function Analysis, 11.2 Transfer Function Analysis 11.3 Transfer Function Analysis 11.4 Univariate Transfer Function Analysis (parametric method) 11.5 Univariate Impulse Response (parametric method) 11.6 Multivariate Transfer Function Analysis (parametric method)	Coherence, Gain (cm/s/mmHg), Phase (rad) in VLF, LF Coherence, Gain (%/mmHg), Phase (rad) in VLF, LF Coherence, Gain (%/%) in VLF, LF Coherence, Gain (%/%), Phase (rad) in LF The second filter coefficient (h_1) of the estimated FIR Gain (%/%) and Phase (rad) for LF band	1 1 1 1 2 1	1 1 1 4 9 4	(Panerai, Simpson et al. 2000, Simpson, Panerai et al. 2001)
12	12.1 Transfer Function Analysis 12.2 Autoregulation index 12.3 Wavelet Coherence Analysis	Coherence, Gain (cm/s/mmHg), Phase (rad) in VLF, LF ARI Gain (cm/s/mmHg) and Phase (rad) in VLF, LF	1 2 1	1 6 3	(Zhang, Zuckerman et al. 1998) (Panerai, White et al. 1998) (Torrence and Webster 1999, Grinsted, Moore et al. 2004)
13	13.1 Transfer Function Analysis	Coherence, Gain (cm/s/mmHg), Phase (rad) in VLF, LF	1	1	(Panerai, Rennie et al. 1998)
14	14.1 ARX models: 1 input 14.2 ARX models: 2 inputs 14.3 Laguerre expansion FIR models, single input (BP) 14.4 Laguerre expansion FIR models, dual input (BP, CO_2) 14.5 Transfer function analysis	Gain (cm/s/mmHg), Phase (rad) in VLF, LF Gain (cm/s/mmHg), Phase (rad) in VLF, LF Gain (cm/s/mmHg), Phase (rad) in VLF, LF Gain (cm/s/mmHg), Phase (rad) in VLF, LF Coherence, Gain (cm/s/mmHg), Phase (rad) in VLF, LF	1 1 1 1 1	5 5 2 2 1	(Mitsis, Zhang et al. 2002, Mitsis, Zhang et al. 2009) (Mitsis, Poulin et al. 2004, Kostoglou, Debert et al. 2014) (Meel-van den Abeelen, Simpson et al. 2014)

Category: 1= TFA-like methods, 2= ARI-like methods, 3= correlation-like methods, Method group: 1=TFA, 2=Laguerre expansions, 3=Wavelets, 4=IR-filter, 5=ARX, 6=ARI, 7=ARMA-ARI/ARX, 9=IR-filter, 10=correlation coefficient; VLF: very low frequency; LF: low frequency; BP: blood pressure; FFT: fast Fourier transform; ARI: autoregulation index; ARX: autoregressive model with exogenous input; Center names are listed in Table S1, individual method settings are listed in Table S4.

Table 3. Bland Altman results for each method sub category and variable.

		Left						Right					
Method groups	Variable	T1	T2	bias	INT	LLOA	ULOA	T1	T2	bias	INT	LLOA	ULOA
TFA-like													
TFA	Gain VLF	0.68 ± 0.43	0.59 ± 0.30	0.09 ± 0.40	0.78	-0.69	0.87	0.68 ± 0.46	0.60 ± 0.31	0.07 ± 0.42	0.82	-0.75	0.88
	Gain LF	1.02 ± 0.58	0.94 ± 0.46	0.08 ± 0.45	0.89	-0.81	0.97	1.02 ± 0.66	0.92 ± 0.46	0.10 ± 0.50	0.98	-0.88	1.08
	Phase VLF	0.87 ± 0.44	0.86 ± 0.50	0.01 ± 0.58	1.14	-1.13	1.15	0.87 ± 0.46	0.89 ± 0.52	-0.02 ± 0.65	1.27	-1.29	1.25
	Phase LF	0.68 ± 0.25	0.69 ± 0.23	-0.01 ± 0.24	0.46	-0.47	0.45	0.69 ± 0.27	0.69 ± 0.24	0.01 ± 0.26	0.52	-0.51	0.52
Laguerre	Gain VLF	0.50 ± 0.29	0.43 ± 0.18	0.07 ± 0.29	0.57	-0.50	0.65	0.49 ± 0.29	0.43 ± 0.19	0.06 ± 0.27	0.54	-0.48	0.60
	Gain LF	0.86 ± 0.44	0.77 ± 0.31	0.09 ± 0.40	0.78	-0.69	0.88	0.86 ± 0.51	0.77 ± 0.30	0.10 ± 0.42	0.83	-0.74	0.93
	Phase VLF	0.81 ± 0.51	0.89 ± 0.51	-0.08 ± 0.70	1.37	-1.45	1.29	0.81 ± 0.52	0.94 ± 0.52	-0.13 ± 0.72	1.42	-1.55	1.29
	Phase LF	0.65 ± 0.30	0.65 ± 0.31	0.00 ± 0.39	0.77	-0.77	0.77	0.64 ± 0.32	0.69 ± 0.34	-0.05 ± 0.41	0.79	-0.84	0.75
Wavelet	Gain VLF	0.91 ± 0.47	0.79 ± 0.36	0.11 ± 0.54	1.05	-0.94	1.16	0.89 ± 0.49	0.83 ± 0.36	0.05 ± 0.53	1.04	-1.00	1.09
	Gain LF	1.04 ± 0.51	0.97 ± 0.37	0.08 ± 0.47	0.93	-0.85	1.00	1.06 ± 0.63	0.95 ± 0.37	0.11 ± 0.55	1.08	-0.97	1.19
	Phase VLF	0.89 ± 0.62	1.05 ± 0.55	-0.12 ± 0.70	1.38	-1.49	1.26	0.96 ± 0.49	1.06 ± 0.66	-0.10 ± 0.70	1.36	-1.46	1.27
	Phase LF	0.91 ± 0.32	0.95 ± 0.30	-0.05 ± 0.30	0.58	-0.63	0.54	0.94 ± 0.31	0.97 ± 0.32	-0.03 ± 0.30	0.58	-0.61	0.55
IR-filter	Gain LF	1.46 ± 0.55	1.28 ± 0.40	0.18 ± 0.18	0.35	-0.17	0.52	1.40 ± 0.55	1.27 ± 0.40	0.12 ± 0.46	0.90	-0.77	1.02
	Phase LF	0.59 ± 0.20	0.63 ± 0.18	-0.04 ± -0.04	-0.08	0.04	-0.12	0.61 ± 0.21	0.63 ± 0.23	-0.02 ± 0.25	0.50	-0.52	0.47
ARX	Gain VLF	0.48 ± 0.29	0.42 ± 0.17	0.06 ± 0.29	0.58	-0.52	0.64	0.48 ± 0.36	0.42 ± 0.18	0.06 ± 0.34	0.67	-0.61	0.74
	Gain LF	0.81 ± 0.39	0.74 ± 0.27	0.07 ± 0.30	0.58	-0.51	0.66	0.81 ± 0.50	0.73 ± 0.27	0.08 ± 0.38	0.74	-0.65	0.82
	Phase VLF	1.05 ± 0.47	1.07 ± 0.43	-0.02 ± 0.50	0.99	-1.01	0.97	1.07 ± 0.47	1.05 ± 0.50	0.01 ± 0.58	1.14	-1.13	1.16
	Phase LF	0.73 ± 0.30	0.74 ± 0.25	-0.01 ± 0.32	0.62	-0.63	0.61	0.73 ± 0.32	0.73 ± 0.26	0.00 ± 0.32	0.62	-0.62	0.62
ARI-like													
ARI		5.48 ± 1.92	5.74 ± 1.62	-0.26 ± 2.12	4.15	-4.40	3.89	5.72 ± 1.89	5.74 ± 1.58	-0.03 ± 2.36	4.63	-4.66	4.60
ARMA-ARI/ARX		8.38 ± 5.32	8.38 ± 5.30	0.00 ± 4.74	9.30	-9.30	9.30	8.27 ± 5.91	9.15 ± 5.92	-0.88 ± 4.49	8.80	-9.68	7.92
IR-filter		-1.07 ± 0.56	-1.02 ± 0.47	-0.06 ± 0.56	1.10	-1.15	1.04	-1.06 ± 0.53	-0.99 ± 0.42	-0.07 ± 0.50	0.98	-1.05	0.92
Correlation-like													
Correlation		0.45 ± 0.14	0.42 ± 0.15	0.03 ± 0.19	0.37	-0.33	0.40	0.44 ± 0.14	0.42 ± 0.15	0.02 ± 0.19	0.38	-0.35	0.40

For each method group the results of similar methods are combined. Methods and units are listed in Table 2. T1: measurement 1; T2: measurement 2; bias: T1-T2; INT: interval (=1.96*SD_{bias}); LLOA: upper limit of agreement (=$mean_{bias}$-interval); ULOA: lower limit of agreement (=$mean_{bias}$+interval); TFA: transfer function analysis; IR-filter: impulse response filter; ARX: autoregressive model with exogenous input; ARI: autoregulation index; VLF: very low frequency; LF: low frequency;