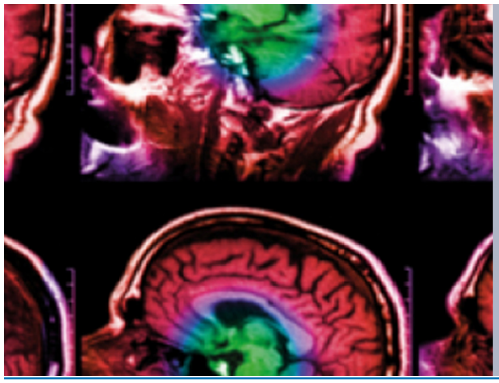
PAPER • **OPEN ACCESS**

Estimating confidence intervals for cerebral autoregulation: a parametric bootstrap approach

To cite this article: Jack E D Bryant *et al* 2021 *Physiol. Meas.* **42** 104004



IOP Publishing Physiol. Meas. 42 (2021) 104004 https://doi.org/10.1088/1361-6579/ac27b8

Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

OPEN ACCESS

RECEIVED
12 May 2021

REVISED
10 September 2021

ACCEPTED FOR PUBLICATION
17 September 2021

PUBLISHED
29 October 2021

PAPER

Estimating confidence intervals for cerebral autoregulation: a parametric bootstrap approach

Jack E D Bryant[1], Anthony A Birch[2], Ronney B Panerai[3], Dragana Nikolic[1], Diederik Bulters[4] and David M Simpson[1]

[1] Faculty of Engineering, University of Southampton, Highfield, Southampton, United Kingdom
[2] Department of Medical Physics, University Hospital Southampton NHS Foundation Trust, Southampton, Hampshire, United Kingdom
[3] Department of Cardiovascular Sciences, University of Leicester, United Kingdom
[4] Wessex Neurological Centre, University Hospital Southampton NHS Foundation Trust, Southampton, Hampshire, United Kingdom

E-mail: jedb1g14@soton.ac.uk

Keywords: cerebral autoregulation, blood pressure, cerebral blood flow, phase estimates, confidence intervals, bootstrap, physiological model estimation

Abstract

Cerebral autoregulation (CA) refers to the ability of the brain vasculature to control blood flow in the face of changing blood pressure. One of the methods commonly used to assess cerebral autoregulation, especially in participants at rest, is the analysis of phase derived from transfer function analysis (TFA), relating arterial blood pressure (ABP) to cerebral blood flow (CBF). This and other indexes of CA can provide consistent results when comparing groups of subjects (e.g. patients and healthy controls or normocapnia and hypercapnia) but can be quite variable within and between individuals. The objective of this paper is to present a novel parametric bootstrap method, used to estimate the sampling distribution and hence confidence intervals (CIs) of the mean phase estimate in the low-frequency band, in order to optimise estimation of measures of CA function and allow more robust inferences on the status of CA from individual recordings. A set of simulations was used to verify the proposed method under controlled conditions. In 20 healthy adult volunteers (age 25.53.5 years), ABP and CBF velocity (CBFV) were measured at rest, using a Finometer device and Transcranial Doppler (applied to the middle cerebral artery), respectively. For each volunteer, five individual recordings were taken on different days, each approximately 18 min long. Phase was estimated using TFA. Analysis of recorded data showed widely changing CIs over the duration of recordings, which could be reduced when noisy data and frequencies with low coherence were excluded from the analysis (Wilcoxon signed rank test $p = 0.0065$). The TFA window-lengths of 50s gave smaller CIs than lengths of 100s ($p < 0.001$) or 20s ($p < 0.001$), challenging the usual recommendation of 100s. The method adds a much needed flexible statistical tool for CA analysis in individual recordings.

1. Introduction

Cerebral autoregulation (CA) is a control mechanism that ensures cerebral blood flow (CBF) remains relatively constant when arterial blood pressure (ABP) varies (Greisen 2005, Caldas *et al* 2017). Impairment of CA has been linked with several serious medical conditions, including stroke, sub-arachnoid haemorrhage and traumatic brain injury (Newell *et al* 1996, Dawson *et al* 2000, Vavilala *et al* 2002, Budohoski *et al* 2013) and a large body of research has been developed in this field (Aaslid *et al* 1989, Panerai *et al* 1998, 2008, Liu *et al* 2010) over many years. Clinical interest is driven by the desire to identify and optimise the treatment of vulnerable patients with impaired CA. The measurement of CA has however remained a challenge. Initial studies assessed the blood flow responses in the brain to sustained changes in ABP, which is now known as static CA. The advent of transcranial Doppler (TCD) with its high temporal resolution, allowed investigating the responses to transient changes in

ABP, known as dynamic cerebral autoregulation (dCA). Such transient changes can be provoked by a range of experimental techniques including the inflation and deflation of pressure cuffs on a subject's thighs, having a subject alternate between standing and either a squatting or sitting positions or tilting a subject's body up and down. Even the spontaneous fluctuations in blood pressure can be exploited to assess dCA. Unlike methods requiring a sustained blood pressure change, dynamic measurements are generally less intrusive and thus more acceptable for clinical use and such studies have come to dominate the field in recent years.

The analysis of dCA requires continuous simultaneous measurement of arterial blood pressure (ABP—typically measured using non-invasive methods applied to a finger) and CBF velocity (typically TCD ultrasound is used to measure cerebral blood flow velocity—CBFV). Many methods with a wide range of complexity have been proposed to analyse these signals and extract parameters that reflect CA status (Tiecks *et al* 1995, Panerai *et al* 1998), but no gold standard has yet emerged that is generally accepted in the research or clinical communities. One popular method, known as transfer function analysis (TFA), performs the analysis in the frequency domain, and involves the estimation of either the phase or gain of the frequency response when relating changes in pressure (as 'input') to those in flow (as 'output'). Evidence that peaks (or troughs) in flow precede those in pressure (positive phase angles in the frequency response in the frequency range say from about 0.07 to 0.2 Hz), or that the gain in this frequency range is relatively low, are deemed to reflect active physiological processes attenuating fluctuations in CBF, and thus indicate active autoregulation (Claassen *et al* 2015). However, estimates of these indexes of autoregulation show considerable variability over time in the same subject (including in measurements made in quick succession), between repeated experiments and across cohorts of healthy subjects and patients (Birch *et al* 2001, Liu *et al* 2005, Budohoski *et al* 2013, Panerai 2014, Sanders *et al* 2019, Elting *et al* 2020). This makes it difficult to assess individual's autoregulatory impairment, and test for any significant changes over time. Fluctuations in CA may be due to physiological changes (even on a time-scale of only a few minutes (Panerai *et al* 2003, Rowley *et al* 2007)), but 'noise' in the measurements is also expected to lead to estimation errors and broad confidence intervals (CIs) for any indices of CA estimated. This 'noise' may include noise in the recorded signals due to imperfect measurement techniques and equipment, the effects of physiological variables not taken into account in the model (e.g. fluctuations in arterial CO_2 level or intracranial pressure) and inaccurate assumptions about the relationship between ABP and CBFV (e.g. linearity) (Panerai *et al* 1998, Mahdi *et al* 2017). It is thus highly desirable to obtain measures of the robustness of CA indices estimated in a given recording, expressed for example by estimates of confidence limits for phase estimates. This would permit any inferences to be made taking the likely precision of estimates into account and subsequently to assess the statistical significance of any changes observed over time, or between measurement conditions. The assessment of CA can thus become more nuanced, with a measure of the confidence in the results from individual recordings. Recordings may then also be identified where the confidence range is so broad that reliable inference of dCA function cannot be made.

The current paper presents a method for the estimation of CIs for estimates of average phase obtained from TFA. This method is based on a parametric bootstrap approach. Performance is then evaluated using simulated signals and data recorded from healthy human volunteers. Some examples of applications will be presented that demonstrate the power of the method in identifying recordings (or sections of recordings) that do not permit robust inferences of dCA, tracking changes in dCA over time and assessing the effect different TFA parameters have on the results obtained. This paper thus addresses a long-standing need to obtain indicators of the quality of dCA assessments in individual recordings, where the variability of results has long been a challenge. While the current paper is focused on the mean phase of TFA, the approach could readily be expanded to other indices of dCA or even the assessment of other physiological control systems such as baroreceptor sensitivity where system identification is also commonly used.

2. Methods

2.1. Data acquisition and processing

The collection of data for this study was performed at the University of Southampton Hospital Trust and was approved by the NHS research ethics committee for Northern Ireland *ref:14/NI/1146*. Recordings of 18.5 ± 1.1 (mean $\pm$ standard deviation) minutes were collected from 20 healthy young adult volunteers (25.5 ± 3.5 years, height 168.3 ± 12.3 cm, weight 64.5 ± 16.4 kg, body mass index 22.6 ± 4.5 kg m^{-2}, systolic blood pressure 119.6 ± 15.4 mmHg, diastolic blood pressure 70.5 ± 8.3 mmHg, 10 female) during rest. All volunteers had no history of cardiovascular or neurological disorders.

ABP (see figure 1 for an example of signals acquired) was recorded non-invasively using a finger plethysmography device (Finometer MIDI, Finapres Medical Systems, Amsterdam, The Netherlands). CBFV in both the left and right middle cerebral arteries (MCA) was recorded non-invasively using TCD ultrasound with a 2 MHz transducer (Dopplerbox, DWL, Compumedics Germany GmbH). A three-lead electrocardiogram

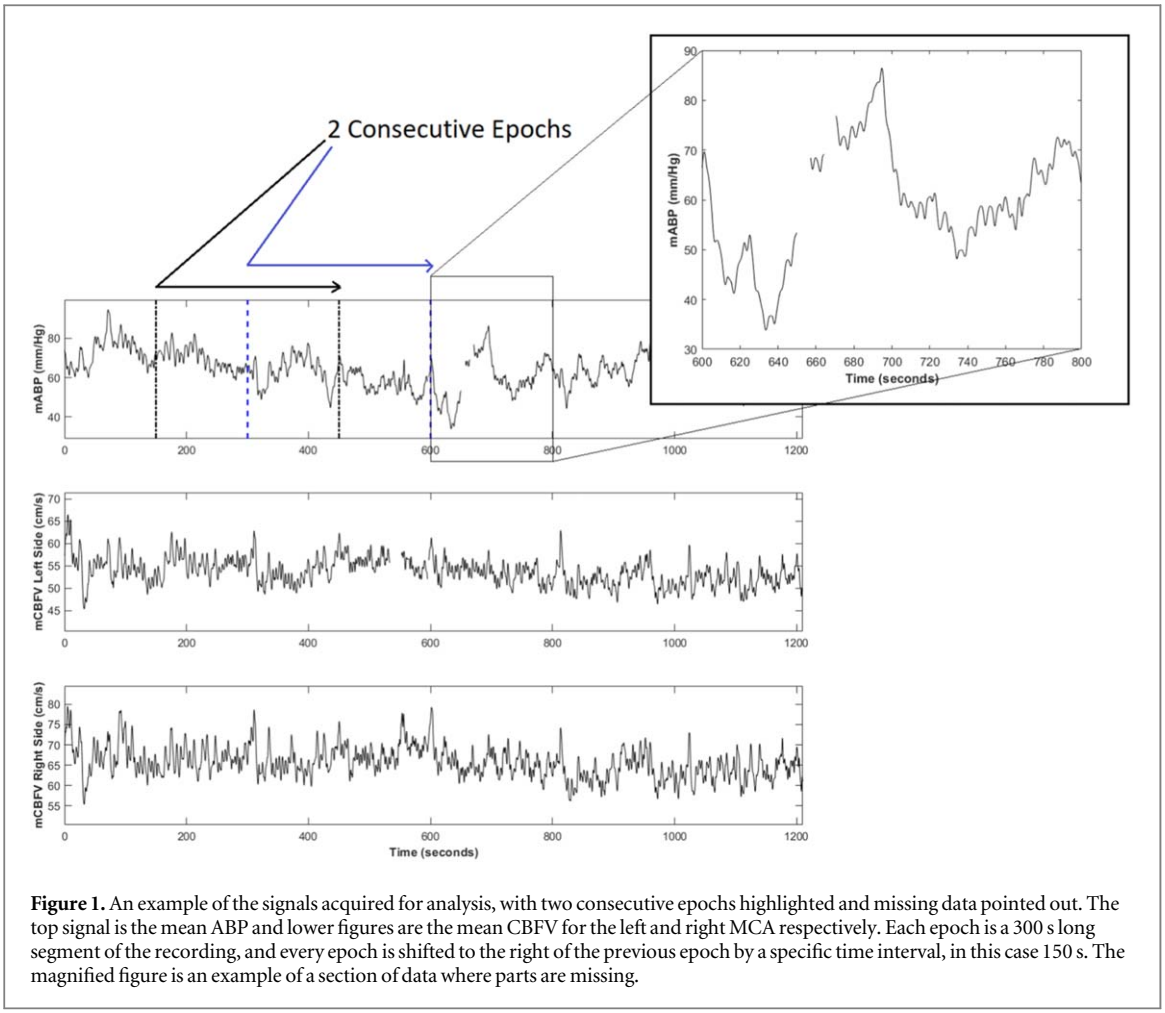


Figure 1. An example of the signals acquired for analysis, with two consecutive epochs highlighted and missing data pointed out. The top signal is the mean ABP and lower figures are the mean CBFV for the left and right MCA respectively. Each epoch is a 300 s long segment of the recording, and every epoch is shifted to the right of the previous epoch by a specific time interval, in this case 150 s. The magnified figure is an example of a section of data where parts are missing.

(ECG) was also acquired for the duration of the recording. Recordings were taken on 5 separate occasions approximately one week apart (8.9 $\pm$ 6.4 d). All signals were sampled at 125 Hz, except the ECG, which was sampled at 250 Hz, in order to increase temporal accuracy in identifying heart-beats, and stored for later offline analysis. Custom software built in Matlab® was used to pre-process and edit the signals before analysis was performed. Two recordings out of the 100 were omitted from the analysis due to their quality being too poor.

Processing of the signals began with a 9th order median filter applied to the CBFV signals to remove any isolated spikes that are a common occurrence in these signals. The CBFV signals were then visually inspected and any remaining spikes were replaced by linear interpolation. This was followed by a 5th order Butterworth low-pass filter with a cut-off frequency of 20 Hz applied both forwards and in reverse to negate the effect of any phase shifts introduced by the filtering. R-peaks from the ECG signal were detected based on Pan-Tompkins algorithm, and were used to compute the beat-to-beat average values of the ABP and CBFV signals, denoted henceforth as mABP, mCBFV-L and mCBFV-R (for left and right middle cerebral arteries respectively). These were then interpolated using a 3rd order polynomial and resampled at 10 Hz.

Because artefacts due to factors such as participant movements are unavoidable when collecting data, and have been shown to negatively impact results (Meel-van den Abeelen *et al* 2016), every signal was visually inspected and any sections of data with artefacts were marked as bad data. Any such sections shorter than 3 s were replaced by linear interpolation, whilst the rest were treated as gaps in the data by replacing them with NaNs (not-a-number) in Matlab®; those gaps were excluded in further analysis, with the effect of doing so on CIs also being considered in this paper. Finally, the signals were normalised and expressed as relative change (in percent) with respect to the mean values within the recording.

2.2. Data analysis

In this section we will first outline the TFA method and how phase is estimated from it. Then the new parametric bootstrap method for estimating the sampling distribution and CIs will be explained. The sampling distribution refers to a set of individual phase estimates generated from multiple simulations (the parametric bootstrap) of the CBFV signals in the current recording. The approach taken to test and evaluate the methods will then be presented, followed by methods used in exploring applications of the approach.

A TFA based approach is used to estimate the phase. Phase estimates of the transfer function (H) between pressure and flow signals provide an indication of the lag between the two signals, with a positive phase indicating a functional autoregulatory system. Phase has been chosen for this paper due to its previous validation as a metric with a strong relationship to dCA (Birch *et al* 1995, Diehl *et al* 1995) and its wide acceptance in current dCA research (Claassen *et al* 2015). H is based on the relation between mABP, denoted now as p, and mCBFV, denoted as v:

$$H(f) = \frac{S_{pv}(f)}{S_{pp}(f)}, \tag{1}$$

where f denotes frequency (in Hz or rads s^{-1}) and S_{pp} the power spectrum of p and S_{pv} the cross-spectrum between p and v. The average phase over a specific frequency band is commonly used as a metric of dCA. For this study the low-frequency (LF) band from 0.07 to 0.2 Hz is adopted (Claassen *et al* 2015), though other frequency bands are also commonly chosen (Meel-van den Abeelen *et al* 2014). In order to avoid the need for phase unwrapping (which can give aberrant results especially in noisy data) the complex values of H in the chosen frequency range are first all given a constant magnitude and then averaged, with the phase of the resultant complex number providing the average phase value. The cross- (S_{pv}) and autospectra (S_{pp}) are usually evaluated using the Welch method (Barbé *et al* 2010), where the signals are divided into overlapping 'windows', tapered at both ends by a window function, and then Fourier transformed. The length of these windows and the function used to taper them, as well as the frequency band chosen, are some of numerous parameters that can affect the results of TFA. Another choice is the use of a coherence cut-off (C-C-O) such that any TFA estimates with a coherence below this threshold are deemed to provide unreliable results for phase estimates and are therefore excluded from further analysis. Coherence is also based on the relationship between the spectra, and is given by the following equation:

$$\gamma^2 = \frac{|S_{pv}(f)|^2}{S_{pp}(f)S_{vv}(f)}. \tag{2}$$

A more detailed description of the parameters used in estimating TFA and their recommended values can be found in Claassen *et al* (2015).

Due to the nature of the Fourier transform, the gaps in the recordings (missing data identified by NaNs) have to be excluded from analysis. In order to do this, the overlap between windows are adjusted so that no windows include the gaps and windows start and end immediately before or after the gaps, thus maximising the data used. For example in figure 1, where some data is missing between approximately 650 and 670 s. Window overlap would thus be adjusted so that a window stops at the last good sample around 650 s, and the next one starts at approximately 670 s where good data is again available; the short data segment at around 660s is too short to fit a window (in the current work 20, 50 or 100 s long) and is lost to the analysis. Auto- and cross-spectral analysis is then performed using these windows, as indicated in (1), using these modified window locations.

To track the phase throughout the entire recording, the recordings were divided into overlapping blocks (known as 'epochs', see figure 1), each one shifted by a specific time interval compared to the previous one. For each epoch, the recommended TFA process as described above (see also (Claassen *et al* 2015)) was performed to acquire an estimate of the phase and subsequently its confidence interval.

2.3. Parametric bootstrap for phase estimates

To estimate the CIs, a parametric bootstrap based approach is used, with figure 2 illustrating the process. Through TFA (block $\boldsymbol{H}$ in the figure), the best fit relationship between ABP ($\boldsymbol{p}$) and CBFV ($\boldsymbol{v}$) is first estimated. The phase (averaged over the selected frequency band) provides the estimate of phase (point estimate) for this recording. The output of $\boldsymbol{H}$ provides a 'clean' estimate of $\boldsymbol{v}$, known as $\hat{\boldsymbol{v}}$, based on the contribution that can be explained by $\boldsymbol{p}$. The 'noise' $\boldsymbol{e}(\boldsymbol{t})$ (also known as the residual) is estimated as the difference between the 'clean' and the measured CBFV. If the linear model fitted the data perfectly and there were no noise present in the recordings, the residual would be zero. Simulated signals of CBFV $\tilde{\boldsymbol{v}}$ are then obtained by adding simulated noise $\tilde{\boldsymbol{e}}(\boldsymbol{t})$ to the 'clean' CBFV signal. This simulated noise has the same power spectrum as the residual, this being achieved by filtering random white noise with a filter $\boldsymbol{G}$, whose frequency response is determined by the spectrum of $\boldsymbol{e}(\boldsymbol{t})$; an autoregressive model and the Burg algorithm (Rodríguez-Liñares and Simpson 2019) are used for this purpose. To determine the order of model used, the Bayesian information criterion (BIC) was used to select the order which resulted in the lowest BIC in each individual recording. Visual inspection of the power spectrum confirmed that the simulated noise signals provided a close approximation to the power spectrum of the residual signal. Using this method, $\boldsymbol{M} = 200$ simulated signals $\tilde{\boldsymbol{v}}(\boldsymbol{t})$ are generated, each with independent random noise. TFA is then applied to each of these signals, using the original $\boldsymbol{p}(\boldsymbol{t})$ as input. The average phase value over the LF frequency range (Zhang *et al* 1998) is then found in each of these signal pairs in the same way as

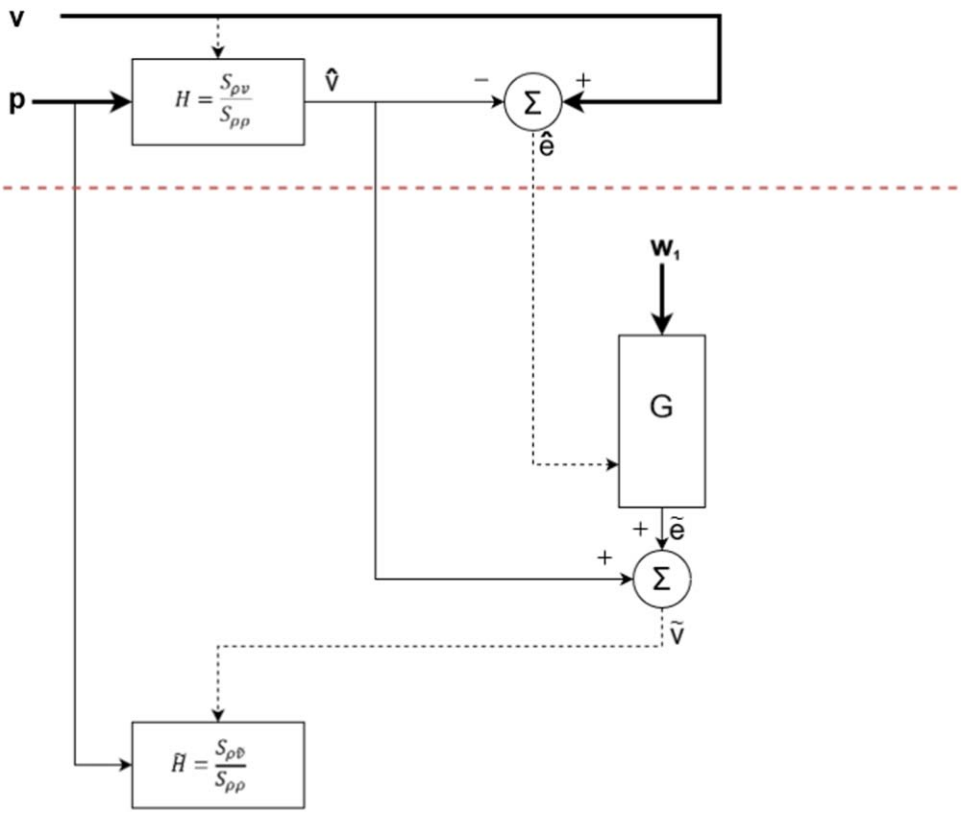


Figure 2. Parametric bootstrap process: the frequency response (transfer function) $\boldsymbol{H}$ is estimated and used to generate a prediction of flow $\boldsymbol{v}$ ($\hat{\boldsymbol{v}}$). The error $\boldsymbol{e}$ (also known as residual) between this prediction and the original flow is calculated, and its spectrum is used to determine a filter that, when applied to white noise $\boldsymbol{w}$, provides a simulated error signal. This is then added to $\hat{\boldsymbol{v}}$ to generate a simulation of $\boldsymbol{v}$ ($\tilde{\boldsymbol{v}}$). This, together with $\boldsymbol{p}$ is then used in as input to TFA to obtain one bootstrap estimate of $\boldsymbol{H}$ ($\tilde{\boldsymbol{H}}$). By repeating the operation below the bold (red) dotted line multiple times (each time with different random noise), the bootstrap distribution of $\tilde{\boldsymbol{H}}$ is found. The dotted lines signify control lines, used in estimating $\boldsymbol{H}$, $\tilde{\boldsymbol{H}}$ and $\boldsymbol{G}$, respectively.

is done for the original recordings. This provides the estimated sampling distribution for the phase estimates, with the 2.5th and 97.5th percentiles being taken as the confidence limits.

2.4. Verification of the method

2.4.1. Verifying confidence interval estimates

In order to verify that confidence interval estimates are reliable, a further simulation is performed, with the process illustrated in figure 3. For this, the 'true' confidence interval should be known. To this end, first one recording from one participant is randomly chosen, and the confidence interval is obtained, using the method described above. This is deemed the 'correct' confidence interval for this set of simulated signals. If the proposed parametric bootstrap method is reliable, then our method should provide good estimates of this confidence interval when the method is also applied to any one of these simulated signals. As shown in figure 7, we therefore randomly select 50 of the simulated signals $\tilde{\boldsymbol{v}}_{\boldsymbol{i}}(\boldsymbol{t})$, $i = 1..50$) and apply the parametric bootstrapping method to each of these (i.e. a second level of simulation) with $\boldsymbol{p}(\boldsymbol{t})$ as the input signal, to find CIs. The width of these 50 CIs is then compared to the 'correct' confidence interval from the original simulations, and the percentage error in confidence interval recorded. This analysis is repeated on 5 different recordings.

2.4.2. Validating assumptions

The method outlined previously relies on the assumption that noise is present in the output of the signal and not in the input i.e. it is present in the mCBFV recordings but not in the mABP recordings. However, blood pressure measurements are also likely to be contaminated by some noise. In order to assess the potential impact of this, the method outlined in 2.3 is applied once again, but now simulating noisy input signals $\tilde{\boldsymbol{p}}(t)$ rather than noisy output signals $\tilde{\boldsymbol{v}}(\boldsymbol{t})$ using an approach equivalent to that discussed above.

2.5. Estimates of CIs for phase

Having verified the performance of the bootstrap method in simulated data, it is then applied to the recorded data from healthy individuals in order to assess CIs of phase estimates and compare them within recordings (i.e.

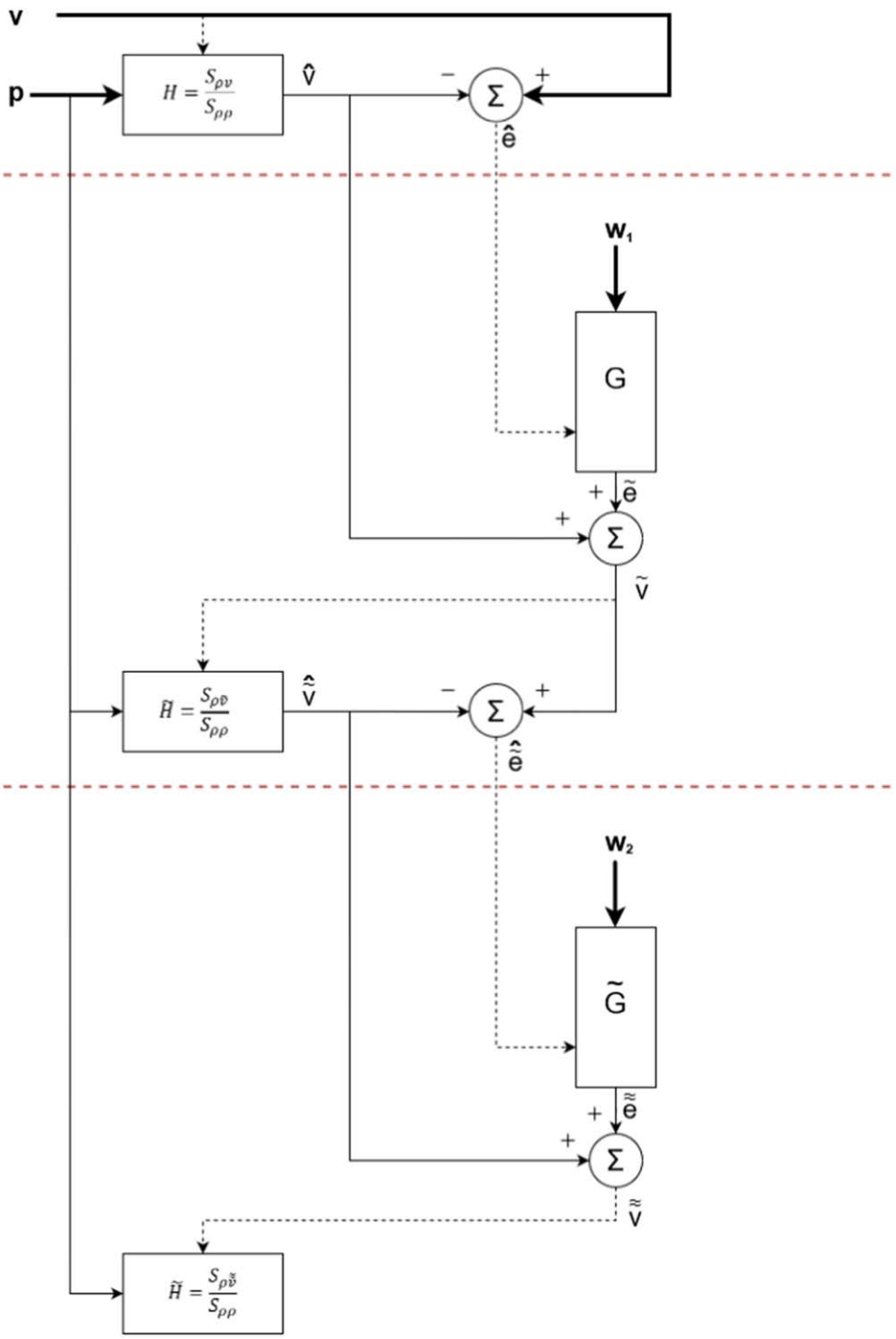


Figure 3. A double bootstrap process. The output of the original bootstrap ($\tilde{v}$) is treated as the input signal and used in a second bootstrap which follows the same procedure as described in figure 2. Note that the processing below the top dotted line is repeated 50 times to generate 50 $\tilde{v}$ and the loop below that line is repeated 200 times to generate 200 bootstrap estimates for each of these 50 simulations.

between overlapping epochs) and between recordings (within and between individual variations). The standard deviation of bootstrapped phase estimates from different epochs within one recording are averaged, and compared to the dispersion of phase estimates (point estimates) from the different epochs in that recording. One might expect these to be similar, but any time-varying behaviour of dCA may lead to greater dispersion between epochs than predicted from the bootstrap, which only assumes additive noise in the data.

2.6. Assessment of the impact of data quality and tfa analysis parameters

Following on from the descriptive analysis of CI for phase estimates in a cohort of healthy volunteers at rest, the potential of the bootstrap method in improving CA analysis is tested. In recordings where some data had

previously been visually identified as containing bad segments (originally marked as gaps with NaNs) it is expected that these segments will provide poor estimates of $\boldsymbol{H}$ and large CIs. The ability to thus 'automatically identify' bad data informs on the potential of the bootstrap method to identify poor data segments without the need for extensive visual analysis. A minimum length of 45 seconds of continuous missing data in mCBFV (either left or right) is chosen for this analysis. This length is chosen as a trade-off between ensuring that a large part of data in a window is of poor quality, and maximising the amount of data files available to perform the investigation on. Any signals that fit this criteria are used in the test, with the remaining files being excluded, resulting in 17 recordings of the left velocity and 9 recordings of the right velocity being used. For each of these the average confidence interval size across all epochs is calculated under three different conditions: (1) after removing bad data ('normal' analysis as described previously in Methods 2.2), (2) including all data in the analysis and using the recommended coherence cut-off threshold (Claassen *et al* 2015), and finally (3) including all data but performing analysis without applying the recommended coherence cut-off thresholding described in Claassen *et al* (2015).

Wilcoxon signrank tests are carried out on the set of 26 recordings being analysed to determine the significance of any difference between CIs from these three analyses.

It is desirable that TFA estimates provide precise estimates of phase, i.e. narrow CIs. In order to test the effect of different parameter choices when using the TFA on the CI, a range of different window sizes and epoch lengths are tested, as outlined in section 3.3; epoch length refers to the length of data over which each TFA analysis is performed, and window-length refers to the length of (Hanning) windows used when applying the Welch method for the auto- and cross-spectral density estimation. For each combination of window size and epoch length, analysis is performed on all 98 recordings, from which the average CI across all epochs in a recording was calculated and used to compare methods. This resulted in 98 CIs, from which the best-performing choices are identified.

3. Results

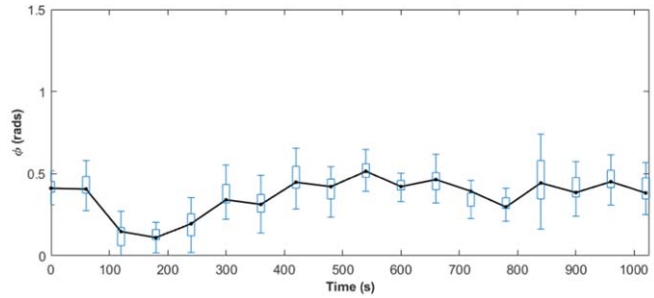
3.1. Analysis of recorded signals

Illustrative examples of bootstrap distributions of LF phase estimates over the duration of the recording are shown in figure 4. The dots connected by straight lines indicate the phase estimates (point estimates) in each epoch using only the recorded data, and the box and whiskers show the median, quartiles and 95% CIs (2.5th and 97.5th percentiles) of the bootstrapped distributions. Estimates were obtained from overlapping 5 min epochs in three 18 min recordings. It is evident that the phase varies considerably over time, and that the CIs (given by the length of the whiskers in each epoch) can also fluctuate strongly in many of the recordings. However, due to the often large size of the CIs only large changes in phase would be statistically significant, with this being explored further in a follow-on paper. Figure 5(a) shows the range of phase estimates (point estimates, i.e. without using the bootstrap method) obtained from the standard deviation of the overlapping epochs within each recording, and compares them to the average of the standard deviations obtained from the bootstrap method. This figure shows that the values from the two methods are similar, but differ considerably between recordings, with generally lower values obtained from the bootstrap method. This is confirmed by the scatter plot in figure 5(b), where most points lie below the line of identity. The discrepancy between the estimates is not unexpected, and can probably be explained by the bootstrap method only taking random variability ('noise') over 5 min epochs into account. The estimates from the dispersion from the point estimates of phase would however also be affected by any physiological change in autoregulatory status that may occur over the duration of the recordings. Figure 6 shows the estimated phase dispersion (average of the standard deviation of the bootstrap estimates across the epochs within each recording) when noise is either added to the output (mCBFV) (as described in figures 2 and 3) or the input (mABP) signal (see methods 2.4.2). It is evident that both lead to similar results in this data. This alleviates the concern that the assumption of noise only in the output might distort results.

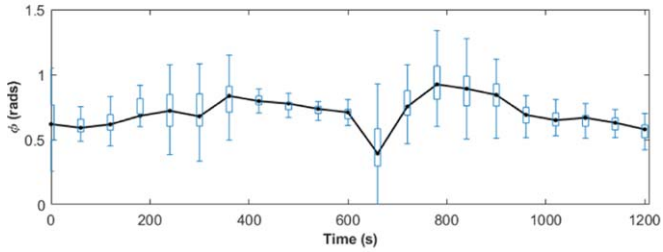
Having explored the proposed method on some recordings, it will now be validated on simulated signals after which the impact that different parameter choices in TFA analysis have on estimation errors for the phase will be considered.

3.2. Verification of the bootstrap method

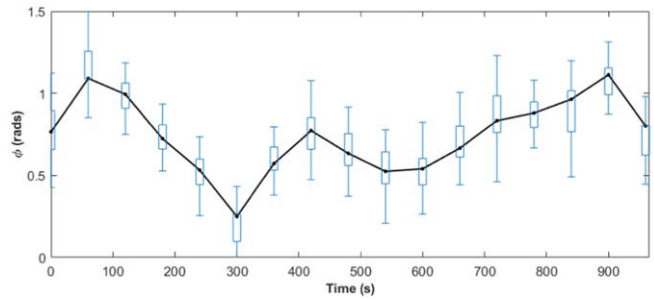
Figure 7(a) shows a histogram of $\boldsymbol{M} = 200$ bootstrapped phase estimates for a single epoch in one recording, with the CIs (95% range) derived for the epoch given by the dotted lines. Each phase has been derived after adding simulated noise. For the purpose of verification, fifty of these simulated data epochs have been processed as if they were the original data to produce $\boldsymbol{K} = 50$ sub-distributions. The central sub-figure 7(b) shows the size of the CIs for each of the $\boldsymbol{K} = 50$ distributions as a vertical line with the mean phase estimate marked with a *. Given



(a)

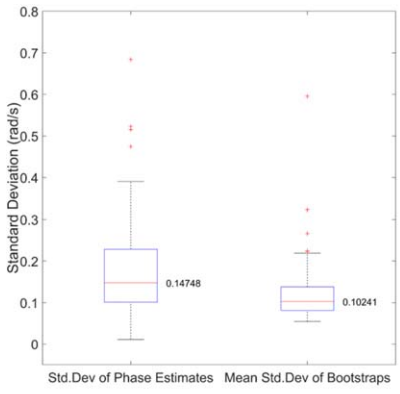


(b)

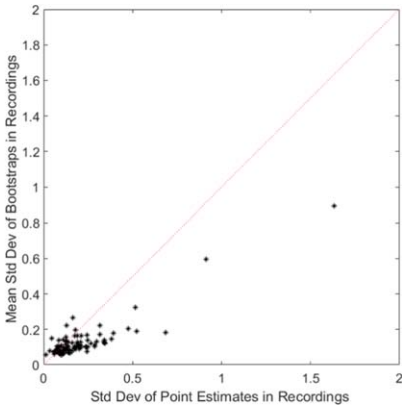


(c)

Figure 4. Examples of phase estimates from three recordings together with confidence intervals, from overlapping epochs (3 min long epochs with a time-shift of 60 s) along recordings of approximately 18 min duration. The box shows the upper and lower quartiles, and the whiskers show the 95% confidence intervals for the estimates. The top figure 4(a) is an example of a recording with more consistent point estimates and smaller confidence intervals, whereas the central figure 4(b) has more variation, with larger confidence intervals highlighting areas of dramatic variation in the estimates. The final figure 4(c) provides an example of very large variation in point estimates, with consistently large confidence intervals.



(a) Spread of the standard deviation of both the point estimates across recordings and the mean standard deviation of bootstraps across recordings.



(b) Standard deviation of phase estimates across the entire signal against range of standard deviations of phase estimates from individual 5 minute epochs

Figure 5. Comparison of standard deviations of point estimates and bootstraps.

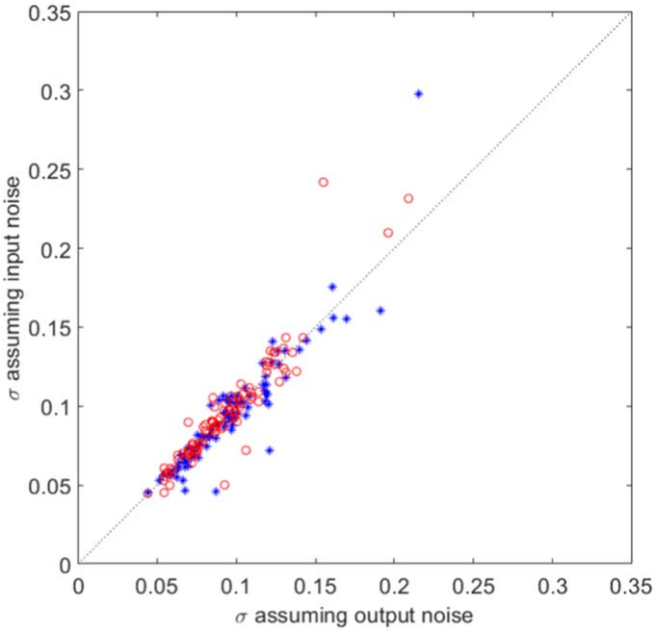
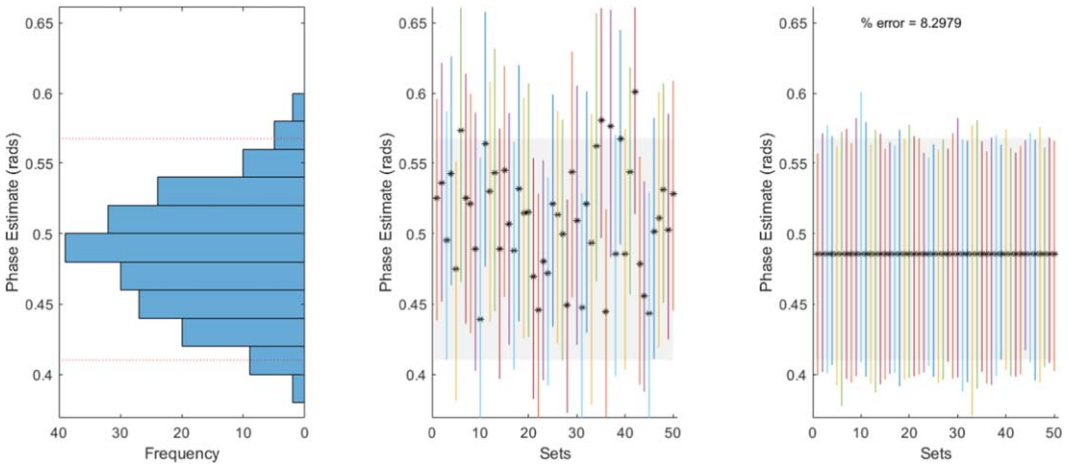


Figure 6. Standard deviation of confidence intervals estimated from the same recordings used in 2.6, for both the left and right MCA (red circles indicate left and blue crosses indicate right). The standard deviation of each recording when assuming noisy input signals and when assuming noisy output signals are plotted against each other. The identifiable approximately linear relationship indicates errors on the input or output have little effect on the overall size of the confidence interval when compared to each other.



(a) Histogram of phase estimates **(b)** Confidence intervals of 50 simulations **(c)** Aligned confidence intervals of 50 simulations

Figure 7. Verification of the parametric bootstrap method. Figure 7(a): a histogram of the distribution of phase estimates acquired from one parametric bootstrap simulation. The two dotted lines indicate the lower and upper confidence limits (95%). Figure 7(b): Confidence intervals for $\boldsymbol{K} = 50$ second level parametric bootstrap simulations, performed on $\boldsymbol{K}$ random phase estimates taken from the distribution in the left figure. The grey band indicates the confidence limits from the original parametric bootstrap, and the * indicate the central (mode) phase estimates from each new distribution. Figure 7(c): the same $\boldsymbol{K}$ confidence intervals as in the previous figure (7)b)), with their central phase estimates aligned, to facilitate visual assessment. It is noted that the $\boldsymbol{K} = 50$ estimates of the CI all agree quite closely with the expected width (grey band).

that each of the bootstraps is based on only one of the 50 simulated recordings, which each give different point estimates, it is not surprising that their bootstrap distributions have distinct mean values. The estimates of CIs from each of the $\boldsymbol{K}$ recordings are expected to provide an approximation to the CIs shown in figure 7(a). In order to make this clearer, in the right sub-figure 7(c), the CIs have been realigned, so that the mean phase estimates agree. It can now be noted that the confidence interval in 7(a) (shown as a grey band) is of a similar size to the intervals estimates from each of the 50 simulations. Similar results were obtained in the other four sets of similar simulations carried out on other randomly selected recordings. The average absolute error between initial CIs

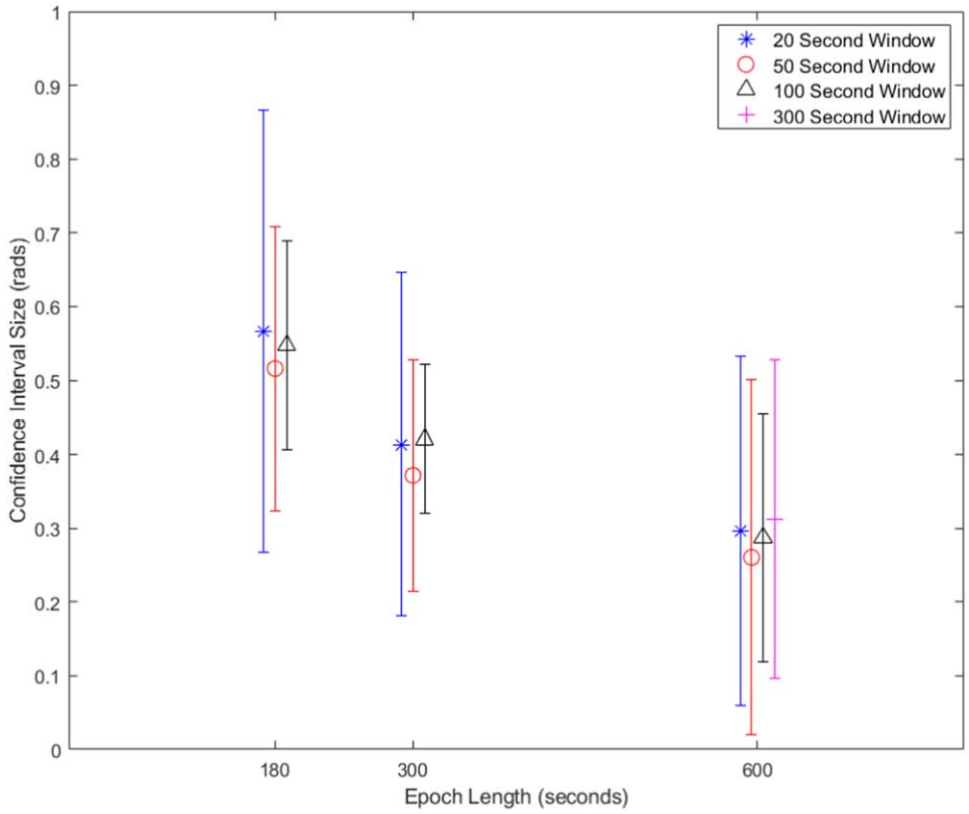


Figure 8. The average size of the confidence intervals across all recordings as a function of the length of window used and the length of epochs. The dotted–dashed lines show the average of the estimates for each epoch length. Window lengths of 20, 50 and 100s were applied to all epochs, and for the 10 min epoch, 300s was also investigated.

(grey band in figure 7) and the sub estimates (plots in figure 7(c)) was 7.6%. These results are very encouraging for the use of the technique in the recorded data sets.

3.3. Effect of tfa parameters on CIs of phase estimates

Figure 8 shows the effect of window size and epoch length on the bootstrap-estimated CIs. As might be expected, longer epochs give more robust estimates (smaller confidence intervals). Less expected was that window lengths of 50s consistently give smaller CIs than those of either 20 or 100s. For the commonly used five minute epochs, a window length of 50 s resulted in significantly smaller CIs compared to 20 s (Wilcoxon signed rank $p < 0.001$) and 100 second windows (Wilcoxon signed rank $p < 0.001$).

3.4. The effect of data quality on CIs

Given the variable data quality that can be achieved in acquiring ABP and CBFV signals, it is recommended (Claassen *et al* 2015) that data is selected so as to only process artefact-free segments. This, however, has to be balanced against the impact of having less data in each epoch, which will tend to increase CIs of phase estimates (see 3.1). Removing frequencies in which ABP and CBFV have low coherence is another means of controlling the results of the analysis (Claassen *et al* 2015), but here again there may be a loss of information, that may degrade or add bias to phase estimates. We therefore used the bootstrap confidence interval estimates to compare the different options in applying TFA. Three alternatives were compared, firstly we excluded poor quality data segments and applied a minimal threshold for coherence estimates, which corresponds to the standard approach recommended in Claassen *et al* (2015) and is outlined in the methods section. It may be noted that with our approach of fitting windows in between gaps, fewer windows or greater overlap between them will result from removing data from some (fixed length) epochs. Secondly, we did not remove bad data, but did apply the minimal coherence criterion. Thirdly, we neither removed bad data from analysis, nor applied the coherence criterion. The results on mean phase estimates and mean CIs for this set of 98 recordings is shown in table 1.

The smallest CIs are obtained in this data set with the first method, with a non-significant increase in CIs when bad data was included but the coherence criterion was still applied. Without the latter, CIs increased significantly (Wilcoxon signed rank test $p < 0.02$) against both alternative approaches. It may also be noted that the mean of the phase estimates did not change significantly between the first two cases, but did between cases

Table 1. Comparison of mean phase estimates and mean confidence interval estimates when excluding bad data in time and frequency domains.

Excluding poor quality data	
$\bar{\phi}$ *(rads)*	0.62
Confidence interval size (rads)	0.59
Including poor quality data, with coherence Cut-Off	
$\bar{\phi}$ *(rads)*	0.70
Confidence interval size (rads)	0.66
Including poor quality data, with no coherence Cut-Off	
$\bar{\phi}$ *(rads)*	0.58
Confidence interval size (rads)	1.11[a]

[a] The confidence intervals when excluding data and not using a cut-off coherence were significantly different to both other test cases.

two and three (Wilcoxon signed rank test $p = 0.044$), with the mean phases being significantly lower when coherence criterion was not used. These results confirm the importance of using the coherence criterion to improve the robustness of estimates, as previously recommended (Claassen *et al* 2015), especially in the presence of poor quality data.

4. Discussion

Confidence intervals of autoregulation estimates are routinely used in investigations into CA, when presenting results (Eames *et al* 2002, Parthasarathy *et al* 2018). However, these are almost always based on estimates made across a cohort, rather than referring to each recording individually. The former captures the between-subject variability, as well as any additional estimation errors in TFA analysis in individual recordings. The current work focuses on estimating these latter errors. We thus achieve a long-standing goal of obtaining an indication of how robust estimates of phase from any given recording may be. By using a parametric bootstrap, a single recording can effectively be used to generate an approximation of the sampling distribution of phase estimates obtained from that recording, under the assumption of additive noise. Bootstrap based methods of estimating CIs have been used extensively in many applications, including signal analysis (Efron 1985, Zoubir and Iskander 2004), however their usage in regards to autoregulation remains limited. A bootstrap method was proposed in Simpson *et al* (2004) to determine the spread of ARI based estimates from individual recordings, but the approach employed there used resampling of recorded data rather than generating new surrogate signals with additive noise (a parametric bootstrap approach), as employed here. In Simpson *et al* (2004), the results were used to objectively select specific signals that provide more robust results. A parametric bootstrap approach more akin to that used here was proposed in Beda *et al* (2017), but in the context of extracting frequency domain features from heart-rate variability, rather than for TFA.

An alternative to estimating the CIs through the method shown is to use the theoretical variance of phase estimates, based on the coherence and number of windows used in the TFA (Piersol and Bendat 2010). However, this method relies on the assumption that recordings are of sufficient length to negate any resolution bias error that may have arisen in estimating the cross-spectra. A detailed evaluation of this potential alternative method is beyond the scope of the current paper.

The results presented here confirm the importance of considering the robustness of phase estimates from individual recordings. It is evident from figures 4 and 5—showing time-varying changes in point estimates and CIs as well as dispersion between recordings—that some recordings provide much more reliable results than others, with considerable variability evident along time even in the same recording (figure 5(a)) and between recordings (figure 5(b)). That some recordings provide much more robust phase estimates than others is also evident from the range of phase estimates obtained from multiple epochs within the same recording, as evident from the range of standard deviation values shown along the horizontal axis in figure 5(b).

One may well envisage that decisions as to whether or not a patient is deemed to have impaired autoregulation should be based not simply on whether or not the point estimate of phase in a given recording is below a set threshold, but also whether the width of the confidence interval allows a clear decision to be made.

4.1. Determination of data quality

Whilst it has been widely understood since early assessments of CA that poor quality data, be that from noise or artefacts, has a negative impact on the results gained and can lead to unreliability (Meel-van den Abeelen *et al* 2014, Claassen *et al* 2015), identification of such data seems to primarily revolve around visually inspecting the

recorded signals. Whilst this is an effective method at improving confidence, it is not a quantitative method. The need for detail visual inspection of data prior to TFA also makes it difficult to translate this approach to CA assessment into clinical practice. One possible alternative to this visual identification is the use of the parametric based CIs that can be estimated periodically throughout a recording. As demonstrated in 3.4, these CIs are able to track changes in the performance of dCA over time, and give an indication of the reliability of the results obtained. Further work is required to investigate the performance in specific challenging settings, such as when there is data drop-out (flat data) or spike-like activity. It may be necessary that additional parameters, in addition to the width of the CIs, are used in deciding whether specific estimates of phase should be deemed reliable.

This method highlights the impact of using a cut-off value of coherence, below which estimates can be said to be not reliable, in reducing confidence interval size and thus improving the robustness of phase estimates. Results showed that including poor quality data did not significantly increase the size of CIs. A possible explanation for this is that removing data segments of poor quality lead to loss of data and thus fewer windows, which also degrades results. It may be that having no data at all from these segments is almost as bad as having poor quality data from these segments, especially when these segments may also contain some data that is less degraded. The increase in CIs with lower numbers of windows was shown in 3.3, when different length epochs were also analysed. Small amounts of artefacts were found to be acceptable in the work of Claassen (Meel-van den Abeelen *et al* 2016) and also Deegan (Deegan *et al* 2011).

4.2. Parameters of TFA

Claassen *et al* (2015) recommend a 100 s window and at least five minute long recordings to achieve reliable estimates of autoregulation using TFA. Mahdi *et al* (2017) also investigated the required data lengths for stable estimates of autoregulation, and recommended that five minutes was the minimum required. Our results found that when using five minute long epochs, the use of a 50 s window gave lower CIs than 100 and 20 s windows, which counters the recommendation of 100 s windows given in Claassen *et al* (2015). Although ten minute epochs were found to result in smaller CIs than five minute epochs, the difference between the interval size when using three minute epochs in comparison to five minute epochs is considerably larger than the difference between five and ten minute epochs, suggesting that the stabilising effect of using longer recordings diminishes as the recording becomes longer. Five minute recordings appear to be an adequate length to achieve reliable results, provided sensible choice of window length is made. It may also be noted that the benefits of shorter window sizes have been pointed out in other recent work (Panerai *et al* 2020). These observations are just a small demonstration of the capabilities of this method in optimising parameters for autoregulation assessment. Any number of parameter combinations could be tested, including parameters for other methods such as ARI. This could extend work such as that by Chacón *et al* (2008) assessing other parameter combinations for the ARI model to find increased stability.

4.3. Limitations of the study

One of the main limitations of this study, in common with much work on CA, was the potential for errors in the measurement methods. CBFV was used as an approximation for CBF, however, this is only a reasonable approximation if the diameter of the blood vessel—in this case the MCA—is constant throughout the recording. Whilst numerous studies have suggested this is a reasonable assumption under specific conditions (Schreiber *et al* 2000, Serrador *et al* 2000), several authors have shown that MCA diameter changes do compromise the use of CBFV as a surrogate for CBF (Altman 2005). Additionally, the Finometer device used to measure blood pressure continuously can lead to errors due to drifting and smooth muscle activity in the finger during recordings (Birch and Morris 2003).

The cohort chosen for the study was made up of only healthy adult volunteers without including patients or others likely to have impaired autoregulation. However, the aim of the current work was to present and assess the method to estimate CIs, rather than comparing results between different cohorts or experimental conditions. That will be considered in a follow-on paper. The results presented illustrate the wide CIs that are sometimes achieved, and how the method can be exploited in controlling estimation errors. It will be of great interest to see if different experimental conditions could reduce the CIs and if some patient groups (e.g. with stroke or TBI) provide more or less robust results.

The bootstrap approach is based on the hypothesis that errors in phase estimates are due to additive random noise in the output signal. The CIs thus only reflect this source of error, and not others, such as time-varying physiological behaviours. It would be very challenging to find robust models that could include this aspect in estimates of CIs. When using the bootstrap method to find the sampling distribution or CIs, it should always be highlighted that these are estimates based on the assumption or additive random noise rather than other possible sources of error.

5. Conclusion

A new method for assessing the reliability of phase estimates has been introduced and results validated in a series of tests. Some descriptive statistics of CIs in recordings from healthy adult volunteers have been provided and some examples of how the methods can be used have been shown. The use of these as a tool for identifying poor quality sections of recordings that do not allow for robust analysis has been discussed. Finally, some of the recommendations of Claassen *et al* (2015) have been tested with a view to optimising the TFA method (optimising parameter combinations). In conclusion, the study provides very encouraging results and promise that this method will provide a useful additional new tool in the goal to be able to robustly identify CA and changes in CA from individual recordings.

ORCID iDs

Jack E D Bryant https://orcid.org/0000-0003-4864-7543
Anthony A Birch https://orcid.org/0000-0002-2328-702X
Ronney B Panerai https://orcid.org/0000-0001-6983-8707
Dragana Nikolic https://orcid.org/0000-0002-9925-4814
Diederik Bulters https://orcid.org/0000-0001-9884-9050
David M Simpson https://orcid.org/0000-0001-9072-5088

References

Aaslid R, Lindegaard K F, Sorteberg W and Nornes H 1989 Cerebral autoregulation dynamics in humans *Stroke* **20** 45–52
Altman D G 2005 Why we need confidence intervals *World J. Surg.* **29** 554–6
Barbé K, Pintelon R and Schoukens J 2010 Welch method revisited: nonparametric power spectrum estimation via circular overlap *IEEE Trans. Signal Process.* **58** 553–65
Beda A, Simpson D M and Faes L 2017 Estimation of confidence limits for descriptive indexes derived from autoregressive analysis of time series: methods and application to heart rate variability *PLoS One* **12** e0183230
Birch A A, Dirnhuber M J, Hartley-Davies R, Iannotti F and Neil-Dwyer G 1995 Assessment of autoregulation by means of periodic changes in blood pressure *Stroke* **26** 834–7
Birch A A and Morris S L 2003 Do the finapres and colin® radial artery tonometer measure the same blood pressure changes following deflation of thigh cuffs? *Physiol. Meas.* **24** 653-60
Birch A A, Neil-Dwyer G and Murrills A J 2001 The repeatability of cerebral autoregulation assessment using sinusoidal lower body negative pressure *Physiol. Meas.* **23** 73–83
Budohoski K P, Czosnyka M, Kirkpatrick P J, Smielewski P, Steiner L A and Pickard J D 2013 Clinical relevance of cerebral autoregulation following subarachnoid haemorrhage *Nat. Rev. Neurol.* **9** 152–63
Caldas J R *et al* 2017 Cerebral blood flow autoregulation in ischemic heart failure *Am. J. Physiol.—Regulatory Integr. Comparative Physiol.* **312** R108–13
Chacón M, Nũez N, Henríquez C and Panerai R B 2008 Unconstrained parameter estimation for assessment of dynamic cerebral autoregulation *Physiol. Meas.* **29** 1179–93
Claassen J A *et al* 2015 Transfer function analysis of dynamic cerebral autoregulation: a white paper from the international cerebral autoregulation research network *J. Cerebral Blood Flow Metab.* **36** 665–80
Dawson S L, Blake M J, Panerai R B and Potter J F 2000 Autoregulation is impaired in acute ischaemic stroke *Cerebrovascular Dis.* **10** 126–32
Deegan B M, Serrador J M, Nakagawa K, Jones E, Sorond F A and ÓLaighin G 2011 The effect of blood pressure calibrations and transcranial doppler signal loss on transfer function estimates of cerebral autoregulation *Med. Eng. Phys.* **33** 553–62
Diehl R R, Linden D, Lücke D and Berlit P 1995 Phase relationship between cerebral blood flow velocity and blood pressure: a clinical test of autoregulation *Stroke* **26** 1801–4
Eames P J, Blake M J, Dawson S L *et al* 2002 Dynamic cerebral autoregulation and beat to beat blood pressure control are impaired in acute ischaemic stroke *Journal of Neurology, Neurosurgery & Psychiatry* **72** 467–72
Efron B 1985 Bootstrap confidence intervals for a class of parametric problems *Biometrika* **72** 45–58
Elting J W *et al* 2020 Assessment of dynamic cerebral autoregulation in humans: is reproducibility dependent on blood pressure variability? *PLoS One* **15** 1-17 e0227651
Greisen G 2005 Autoregulation of cerebral blood flow in newborn babies *Early Human Dev.* **81** 423–8
Liu J, Simpson D M and Allen R 2005 High spontaneous fluctuation in arterial blood pressure improves the assessment of cerebral autoregulation *Physiol. Meas.* **26** 725–41
Liu J, Simpson M D, Yan J and Allen R 2010 Tracking time-varying cerebral autoregulation in response to changes in respiratory PaCO2 *Physiol. Meas.* **31** 1291–307
Mahdi A, Nikolic D, Birch A A and Payne S J 2017 At what data length do cerebral autoregulation measures stabilise? *Physiol. Meas.* **38** 1396–404
Meel-van den Abeelen A S S, de Jong D L K, Lagro J, Panerai R B and Claassen J A H R 2016 How measurement artifacts affect cerebral autoregulation outcomes: a technical note on transfer function analysis *Med. Eng. Phys.* **38** 490–7
Meel-van den Abeelen A S S, van Beek A H E A, Slump C H, Panerai R B and Claassen J A H R 2014 Transfer function analysis for the assessment of cerebral autoregulation using spontaneous oscillations in blood pressure and cerebral blood flow *Med. Eng. Phys.* **36** 563–75
Newell D W, Weber J P, Watson R, Aaslid R and Winn H R 1996 Effect of transient moderate hyperventilation on dynamic cerebral autoregulation after severe head injury *Neurosurgery* **39** 35–44
Panerai R B 2014 Nonstationarity of dynamic cerebral autoregulation *Med. Eng. Phys.* **36** 576–84

Panerai R B, Eames P J and Potter J F 2003 Variability of time-domain indices of dynamic cerebral autoregulation *Physiol. Meas.* **24** 367–81
Panerai R B, Intharakham K, Minhas J S, Llwyd O, Salinet A S M, Katsogridakis E, Maggio P and Robinson T G 2020 Cohmax: an algorithm to maximise coherence in estimates of dynamic cerebral autoregulation *Physiol. Meas.* **41** 085003
Panerai R B, Rennie J M, Kelsall A W R and Evans D H 1998 Frequency-domain analysis of cerebral autoregulation from spontaneous fluctuations in arterial blood pressure *Med. Biol. Eng. Comput* **36** 315–22
Panerai R B, Sammons E L, Smith S M, Rathbone W E, Bentley S, Potter J F and Samani N J 2008 Continuous estimates of dynamic cerebral autoregulation: influence of non-invasive arterial blood pressure measurements *Physiol. Meas.* **29** 497–513
Parthasarathy A B, Gannon K P, Baker W B, Favilla C G, Balu R, Kasner S E, Yodh A G, Detre J A and Mullen M T 2018 Dynamic autoregulation of cerebral blood flow measured non-invasively with fast diffuse correlation spectroscopy *J. Cerebral Blood Flow Metab.* **38** 230–40
Piersol A G and Bendat J S 2010 *9.1 Cross-Spectral Density Function Estimates* 4th edn (New York, NY: Wiley) pp 289–98
Rodríguez-Liñares L and Simpson D M 2019 Spectral estimation of HRV in signals with gaps *Biomed. Signal Process. Control* **52** 187–97
Rowley A B, Payne S J, Tachtsidis I, Ebden M J, Whiteley J P, Gavaghan D J, Tarassenko L, Smith M, Elwell C E and Delpy D T 2007 Synchronization between arterial blood pressure and cerebral oxyhaemoglobin concentration investigated by wavelet cross-correlation *Physiol. Meas.* **28** 161–73
Sanders M L *et al* 2019 Dynamic cerebral autoregulation reproducibility is affected by physiological variability *Front. Physiol.* **10** 865-76
Schreiber S J, Gottschalk S, Weih M, Villringer A and Valdueza J M 2000 Assessment of blood flow velocity and diameter of the middle cerebral artery during the acetazolamide provocation test by use of transcranial doppler sonography and MR imaging *Am. J. Neuroradiol.* **21** 1207–11
Serrador J M, Picot P A, Rutt B K, Shoemaker J K and Bondar R L 2000 MRI measures of middle cerebral artery diameter in conscious humans during simulated orthostasis *Stroke* **31** 1672–8
Simpson D M, Panerai R B, Ramos E G, Lopes J M A, Marinatto M N V, Nadal J and Evans D H 2004 Assessing blood flow control through a bootstrap method *IEEE Trans. Biomed. Eng.* **51** 1284–6
Tiecks F P, Lam A M, Aaslid R and Newell D W 1995 Comparison of static and dynamic cerebral autoregulation measurements *Stroke* **26** 1014–9
Vavilala M S, Lee L A and Lam A M 2002 Cerebral blood flow and vascular physiology *Anesthesiol. Clin. North Am.* **20** 247–64
Zhang R, Zuckerman J H, Giller C A and Levine B D 1998 Transfer function analysis of dynamic cerebral autoregulation in humans *Am. J. Physiol.—Heart Circ. Physiol.* **274** H233–41
Zoubir A M and Iskander D R 2004 *Bootstrap Techniques for Signal Processing* (Cambridge: Cambridge University Press) (https://doi.org/10.1017/CBO9780511536717)