Non-White Noise in fMRI: Does modelling have an impact?

Torben E. Lund, [⋆,a] Kristoffer H. Madsen [a,b] Karam Sidaros [a]
Wen-Lin Luo [c] Thomas E. Nichols [d]

[a] *Danish Research Centre for Magnetic Resonance, Copenhagen University Hospital, Hvidovre, Denmark*

[b] *Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark*

[c] *Merck & Co., Inc., Whitehouse Station, New Jersey, USA*

[d] *Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA*

Abstract

The sources of non-white noise in Blood Oxygenation Level Dependent (BOLD) functional magnetic resonance imaging (fMRI) are many. Familiar sources include low-frequency drift due to hardware imperfections, oscillatory noise due to respiration and cardiac pulsation, and residual movement artefacts not accounted for by rigid body registration. These contributions give rise to temporal autocorrelation in the residuals of the fMRI signal and invalidate the statistical analysis as the errors are no longer independent. The low-frequency drift is often removed by high-pass filtering, and other effects are typically modelled as an autoregressive (AR) process. In this paper we propose an alternative approach: Nuisance Variable Regression (NVR). By inclusion of confounding effects in a general linear model (GLM), we first confirm that the spatial distribution of the various fMRI noise sources is similar to what has already been described in the literature. Subsequently we demonstrate, using diagnostic statistics, that removal of these contributions reduces first and higher order autocorrelation as well as non-normality in the residuals, thereby improving the validity of the drawn inferences. In addition we also compare the performance of the NVR method to the whitening approach implemented in SPM2.

Key words: Physiological noise, temporal autocorrelation, low-frequency drift.

Introduction

Non-white noise in fMRI has been a known issue ever since the earliest days of fMRI. Weisskoff et al. (1993) presented power-spectra from different regions of interest acquired using fast (7Hz) imaging of a single slice in visual cortex. The spectra showed white noise in white matter (WM), cardiac, respiratory and low-frequency noise in grey matter (GM) and cardiac noise in cerebrospinal fluid (CSF). Since a convenient assumption when making statistical inference is to assume that errors in the measurement are independent and identically normally distributed (i.i.d.) this observation is important and has had large impact on paradigm design and data analyses.

With non-white noise, the i.i.d. assumption is no longer fulfilled, and if this is ignored the estimated standard deviations will typically be negatively biased, resulting in invalid (liberal) statistical inferences. Another consequence is the difficulty in detecting signals when covered in noise. As we are normally interested in the GM signal it is problematic that this is the region where structured noise is most pronounced. With physiological noise increasing with field strength (Krüger & Glover, 2001; Krüger et al., 2001), the full benefit in signal to noise ratio (SNR) by using high field scanners cannot be achieved without proper handling of physiological noise.

While it has been argued that the low-frequency oscillations or drifts are physiological in nature they have also been observed in cadavers (Smith et al., 1999) and phantoms (Lund & Larsson, 1999). Noise contributions with these frequencies can be avoided by a combination of a high-pass filter and a stimulation paradigm fast enough to "escape" the region of the spectra where low-frequency noise is dominant. The system noise makes it difficult to study states of neural activity with BOLD fMRI, if the state remains unchanged for more than a minute. Being a differential, measurement arterial spinlabelling (ASL) experiments are less prone to these drift problems, and thus low paradigm frequency can be used in ASL (Wang et al., 2003).

In typical whole brain acquisitions the physiological noise components are

⋆ To whom correspondence should be addressed:
Torben E. Lund
Danish Research Centre for Magnetic Resonance, Copenhagen University Hospital, Hvidovre
Kettegaard Allé 30, 2650 Hvidovre, Denmark
Phone: +45 3632 3328
Fax: +45 3647 0302

Email address: torbenl@magnet.drcmr.dk.
URL: http://www.drcmr.dk.

heavily aliased and possibly non-stationary making it difficult to identify directly. This is perhaps the reason why no commonly accepted model of noise in fMRI exists. Recent versions of most fMRI analysis software try to estimate the covariance due to these effects (e.g. using an autoregressive model), and use the estimated process to pre-whiten the errors. If the estimated covariance structure is correct this procedure is unbiased and efficient. However, the modelling is time consuming and often global and/or low-order models need to be used in order to get a robust estimate. From the power spectra by Weisskoff et al. (1993) it is known that the noise in part consists of oscillations at respiratory and cardiac frequencies and their higher harmonics. It is also known (Dagli et al., 1999) that the non-white noise is structured in space. Nevertheless a first order process stationary in time and often in space as well, is often assumed when the coefficients of the model are to be estimated (e.g. SPM2 (Friston et al., 2002)). As it takes two AR coefficients to model a single oscillator the approach will, at best, only be valid when the noise is heavily aliased and smeared due to long repetition times (TR) and non-stationary heart and respiratory rates. Using a variational Bayesian approach for AR model-order estimation, Penny et al. (2003) have demonstrated that higher order processes are necessary to describe the auto-correlation in regions close to the vasculature e.g. medial cerebral artery (MCA), and circle of Willis (CW). However, with current high-end computing power, this estimation takes about 30 minutes per slice.

As an alternative to estimating the correlation in the errors, a number of techniques have been suggested which try to predict the effect of various noise sources on the measured signal. While these methods were primarily introduced for noise reduction purposes, they could potentially serve as whitening filters too and substitute autoregressive modelling which is both difficult and time consuming.

The first methods (Hu et al., 1995; Biswal et al., 1996) for removing cardiac and respiratory noise both used pulse-oximeter time-courses. The method by Biswal et al. (1996) uses the pulse-oximeter time-course to construct a band-stop filter, which is only appropriate when the noise is critically sampled or very stationary. The method by Hu et al. (1995) conversely uses retrospective gating of raw k-space data, and also works for long TR, and multi-shot sequences, as long as the noise is quasi-periodic. A drawback of this method is that it performs best for data points with high SNR which would typically mean mainly points in central k-space. Central k-space is where global effects like movement (e.g. due to respiration) manifest themselves. Effects of pulsating blood, on the other hand, are localised in real space to areas near vessels, and hence spread out over the entire k-space. This is perhaps the reason why Hu et al. (1995) find effects of respiration to be more dominant than effects of cardiac pulsation. On the other hand, the method by Hu et al. (1995) was used by Dagli et al. (1999) to show that cardiac noise is dominant near major

arteries e.g. MCA and CW. Josephs et al. (1997) and Glover et al. (2000) suggested performing retrospective correction in the reconstructed images in a way similar to that proposed by Hu et al. (1995). The two implementations differ slightly. Josephs et al. (1997) implemented the correction as a part of the design matrix in a GLM whereas Glover et al. (2000) performed the correction during the preprocessing. Both methods assume that physiological noise is fixed during acquisition of a single slice. While this is a good approximation for single-shot 2D echo planar imaging (EPI) it is not the case for multi-shot acquisitions in general [1] .

Using the property that respiration-induced noise is a global effect Frank et al. (2001) proposed estimating and correcting this effect from the k-space data itself. As global effects (centre of k-space) are actually sampled for each slice in an EPI volume, the sampling rate for global effects equals the number of slices per volume divided by TR instead of 1/TR, thereby making global noise-effects critically sampled. A similar approach is to use time-courses from regions of no interest to regress out behaviour similar to that of major vessels or CSF time-courses (Petersen et al., 1998; Lund & Hanson, 2001). This can be achieved by identifying voxels located in CSF and vessels by inspecting a variance image of the fMRI time-course (Lund & Hanson, 2001). Buonocore & Maddock (1997) used time-courses from "truly inactive" and "truly active" voxels to construct a Wiener-filter and demonstrated how this approach is superior to notch filtering. This filter is, however, still stationary and will not be optimal in the presence of variations in the heart and respiratory rates. Independent component analysis (ICA) has great potential for identifying patterns of structured noise in fMRI data. However, so far it has only been evaluated for TR's short enough to critically sample the cardiac noise (Thomas et al., 2002) in which case, a simple low-pass filter is adequate for removing physiological noise.

The purpose of this paper is to suggest a unified theory for physiological noise in fMRI. We will show how relevant nuisance regressors in a general linear model can be used to describe spatiotemporal behaviour of low-frequency, residual movement, respiratory and cardiac related effects. Using Statistical Parametric Mapping Diagnosis (SPMd) (Luo & Nichols, 2003) we hereafter demonstrate the impact of physiological noise correction on the i.i.d. assumption and compare the performance of the suggested method to the whitening approach implemented in SPM2. Finally the findings of this paper are discussed in relation to assumptions made during common fMRI preprocessing and analysis including slice-timing and functional connectivity analysis.

[1] Glover et al. (2000) stated that the method is effective up to 3-shot spiral acquisitions.

Materials and methods

General Linear Models for non-white noise in fMRI images

In this section we describe how the effect of different of non-white noise sources in fMRI can be described in the GLM framework where the significance of each noise source can be tested using an F-test.

Low-frequency drift due to hardware instabilities

A commonly used model for low-frequency drift is to include a basis-set of slowly varying functions in the design matrix. This will serve as a high-pass filter, removing oscillations which can be modelled as a linear combination of the basis-set. The set can be constructed using polynomials, as in `fmristat` (Worsley et al., 2002), or a discrete cosine-set as implemented in SPM2 (Worsley & Friston, 1995). In this paper we have chosen the discrete cosine-set and thus our model for the low-frequency noise due to hardware imperfections reads:

$$\begin{aligned} N_L(t_n) = \beta_{L,1}\cos(\phi_L(t_n)) + \beta_{L,2}\cos(2\phi_L(t_n)) + \dots \\ + \beta_{L,p}\cos(p\phi_L(t_n)), \end{aligned} \tag{1}$$

where $\phi_L(t_n) = \pi t_n/T$, and $p = \mathrm{div}(2T, T_L)$. Here, T is the duration of the entire time-series to be modelled T_L is the period of the fastest of the oscillations to be removed and t_n is the time when the first slice in volume n is sampled.

Residual movement effects

Even after rigid body realignment, residual movement artefacts can be present in the data. These have successfully been described by (Friston et al., 1996) using a first order Volterra expansion of the movement parameters from the realignment procedure. Let $m_i(t)$, $i = [1:6]$ be the 6 rigid-body movement parameters corresponding to the volume acquired at time t_n then the model for the residual movement effects is described by equation 2:

$$\begin{aligned} N_M(t_n) = \beta_{M,1}m_1(t_n) + ... + \beta_{M,6}m_6(t_n) \\ + \beta_{M,7}m_1(t_{n-1}) + ... + \beta_{M,12}m_6(t_{n-1}) \\ + \beta_{M,13}m_1^2(t_n) + ... + \beta_{M,18}m_6^2(t_n) \\ + \beta_{M,19}m_1^2(t_{n-1}) + ... + \beta_{M,24}m_6^2(t_{n-1}) \end{aligned} \tag{2}$$

Modelling aliased physiological noise

From basic signal processing it is known that the modulus of the Fourier transform of a real signal, S, sampled at the frequency, f_s, is symmetric around 0 and repeated with a period of f_s. A signal containing frequencies larger than the Nyquist frequency $f_n = f_s/2$, is not critically sampled and will, therefore, be aliased i.e. folded into the frequency range $[-f_n, f_n[$. Contrary to what is sometimes stated e.g. (Frackowiak et al., 2004, p795-796) or (Thomas et al., 2002; Li et al., 2000) there is in general no single frequency band into which the aliasing happens. If $f = kf_s/2 + \nu$ is the frequency of a signal, and k is the largest integer to make $0 \leq \nu < f_s/2$ the aliased frequency, f_a, will be given by equation 3 (see e.g.(Franckowiak et al., 1997, p. 475 Fig.2)):

$$f_a = \begin{cases} \nu & \text{for } k \text{ even} \\ f_s/2 - \nu & \text{for } k \text{ odd} \end{cases} \tag{3}$$

In the case of multislice fMRI a typical sampling frequency $f_s = 1/\mathrm{TR}$ is around 0.5Hz, and with fundamental frequencies of the cardiac noise f_c in the interval $[0.6Hz, 1.5Hz]$ this gives $2 \leq k \leq 6$ i.e. the cardiac noise is from 2 to 6 times under-sampled. Correspondingly this interval of cardiac frequencies is four times as large as the sampled bandwidth!

By using external measures of respiration and cardiac pulsation, e.g. respiratory belt and pulse-oximeter, it is possible to assign a cardiac and respiratory phase to each volume and thereby identify physiological noise even when it is aliased. One way of performing this retrospective image correction is the RETROICOR method described by Glover et al. (2000). The RETROICOR method models the physiological noise as a basis set of sines and cosines representing the aliased frequencies of the cardiac and respiratory oscillations and their higher harmonics. The method accounts for non-stationarity in the cardiac and respiratory time-series by letting the phase of the reference time-course be the phase of a quasiperiodic oscillation.

The respiratory noise has at least three different components, and two fundamentally different phases are of potential interest in relation to modelling its effect. Part of the respiration-induced noise will manifest itself as rigid body motion of the head, which, in part, is likely to also be reflected in the movement parameters. Ideally, by modelling respiratory effects, within volume movements could also be accounted for. For TR values short enough to sample the respiration critically alignment is expected to account for this. Other respiratory effects include susceptibility changes in the brain due to movement of organs in the abdomen (Raj et al., 2000, 2001) and respiration-dependent vasodilation or oxygenation difference (Windischberger et al., 2002). With regard to the first two of the described phenomena the phase does not increase linearly with time but changes dramatically around the time of inspiration and

expiration, and stays fairly constant in between. Conversely the phase of the effects of oxygenation differences and vasodilation increases linearly with the time since the last inspiration. Glover et al. (2000) the non-linearly increasing phase is used, but as we have used fairly short TR and comprehensive modelling of residual movement effects we have chosen primarily to model the oxygenation dependent part of the respiration induced noise, and thus only the phase increasing linearly with time is used.

Each volume in an fMRI time-course is assigned a cardiac and respiratory phase, ϕ_c and ϕ_r. The phases are the temporal distance from the first slice in the volume to the last peak in the cardiac or respiratory reference time-course, divided with the peak to peak interval at that timepoint, for each reference time-course, at the time of the acquisition the first slice. Here it is assumed that the pulse rate is constant within the acquisition of a volume. Based on power-density spectra from low TR data, where the physiological noise is critically sampled, we have chosen a model with 5 harmonics for the cardiac and 3 harmonics for the respiratory signal, leading to a total of 16 physiological noise regressors as described in Equation 4:

$$\begin{aligned} N_p(t_n) = \beta_{p,1}\sin(\phi_c(t_n)) + \beta_{p,2}\cos(\phi_c(t_n)) + \ldots \\ + \beta_{p,15}\sin(3\phi_r(t_n)) + \beta_{p,16}\cos(3\phi_r(t_n)), \end{aligned} \tag{4}$$

where $\beta_{p,1-10}$ are coefficients of sines and cosines describing the five harmonics of the cardiac noise and $\beta_{p,11-16}$ are the coefficients of sines and cosines describing the three harmonics of the respiratory noise. By calculating an F-test on the set, or a subset, of the β_p-values it is possible to map the regions in the brain showing a significant effect of one or more of the physiological noise regressors e.g. the first harmonic of the cardiac noise or all harmonics of the respiratory noise. In the following we will use Nuisance Variable Regression (NVR) as a generic term for models including different combinations of movement and RETROICOR regressors.

Datasets

To investigate the properties and performance of our nuisance models simulated, phantom and *in vivo* data were used.

Simulated data

The performance of the RETROICOR method as a whitening approach is expected to depend critically on at least three factors: 1) the precision with

which the phase of the externally recorded physiological signal can be assigned, 2) the white noise level and 3) The frequency of the underlying physiological noise source in question. In order to get a rough estimate of the required precision we generated an artificial dataset containing white noise and aliased harmonic oscillations. A total of 13 datasets each containing 381 volumes of four slices were generated. The mean of each voxel timecourse was set to 100 and added normal distributed white-noise with a dataset specific white noise level. The standard deviation of the added white noise was $W_i = [0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]$. Three oscillatory time-courses with frequencies $f_1 = 1$Hz, $f_2 = 2$Hz, $f_3 = 3$Hz, were created and sampled at TR=2.37s giving rise to aliased frequencies of $f_1^a = 0.1561$Hz, $f_2^a = 0.1097$Hz, $f_3^a = 0.0464$Hz. These oscillations were now added to the time-courses of selected voxels as illustrated in figure 1. In slice 1 voxels in a pattern marking the digit "1" were added a single sine with frequency f_1^a, and similar for slice 2 and 3. In slice 4 the pattern of oscillatory behaviour was constructed by adding the patterns of the slices 1 to 3. Where two or more patterns from these slices intersect the corresponding oscillations were added to the time-course of the particular voxel. These datasets were then analysed with a simple model: (mean+white noise) and several NVR models containing RETROICOR nuisance regressors corresponding to Equation 4 but with different amounts of jitter introduced to the phase of the reference time-courses. The added phase jitter was normal distributed white noise with standard deviations of: [0 1 2 3 ... 10 15 20 ... 40 50 60 ... 100 120 140 ... 200] ms. The parameters of these models were estimated using Ordinary Least Squares.

Phantom data

In order to investigate the effects of system noise in the absence of physiological noise a phantom experiment was performed. A phantom consisting of three gel-filled bottles were positioned in the headcoil and scanned using the same echo planar imaging (EPI) sequence as in the *in vivo* data described below (but with slightly different parameters 500 volumes of 44 slices with a TR=2.61s). Gel-phantoms were selected to minimise convection effects which we have previously observed using other phantoms. The dataset was analysed both with a simple model: (mean+white noise) and a model extended to include a high-pass filter as described in equation 1 with $T_L = 60$ s.

In vivo data

Sixteen datasets (8 subjects and 2 paradigms) each consisting of 381 volumes of forty slices (matrix size 64x64) were acquired on a 3T scanner (Magnetom Trio, Siemens, Erlangen Germany) using a gradient echo EPI sequence: TE=30 ms, TR=2.37 s, in-plane resolution 3 mm$\times$ 3 mm, number of slices 40, flipangle:

90°, slice thickness 3mm, interleaved acquisition without gaps. During the scans the subjects were stimulated visually using 8Hz reversing checkerboard (expanding ring (R) and rotating wedge (W)). Each rotation/expansion lasted 30 seconds. Following rigid-body realignment using SPM2, each dataset was subsequently analysed with eight different general linear models described in Table 1. All models were estimated using SPM2. For all models except "SPM2-AR(1)" errors were assumed to be i.i.d. In the "SPM2-AR(1)" model the SPM2 "AR(1) + white noise" noise-model was used.

Testing assumptions in the general linear model

After the analyses, Statistical Parametric Mapping diagnosis (SPMd) was used to test the whiteness and normality of the residuals. One test for dependent noise (Dep) is based on the cumulative power spectrum of the BLUS residuals (Best Linear Unbiased residuals with a Scalar (diagonal) covariance matrix). The cumulative spectrum is tested for linearity, as it should be a straight line under whiteness. Instead of usual residals $e = Y - \hat{Y}$, BLUS residuals are used because $\mathrm{Var}(e) = (I - (X'X)^{-1}X')\sigma^2 \neq \sigma^2$ even under whiteness. The BLUS residuals are a sequence of $n - p$ residuals that are, under i.i.d. assumptions, white and have smallest expected distance to the unobservable errors ϵ. Another test of dependence is the Durbin-Watson statistic, specifically designed to detect AR(1) autocorrelation (Corr). The Durbin-Watson is the uniformly most powerful test of Gaussian AR(1) noise versus an alternative of white Gaussian noise. The Shapiro-Wilk test of normality (Norm) is essentially the square of the correlation between the ordered standardized residuals and their expected value. (See Luo & Nichols (2003) for complete references).

Results

Simulated data

In Figure 2 we show as a function of SNR, for different values of jitter in the phase of the reference time-course, the result of the SPMd analysis as applied to the residuals after model fitting. With regard to the whiteness of the errors (Dep and Corr) it is seen that for a range of white noise levels the nominal number of rejections is achieved with a phase precision around 30ms whereas 70ms will lead to 10 times more voxels being rejected. With regard to the normality of errors (Norm) it is noted that the test seems to lose power when the noise level increases. In Figure 3 we show the corresponding output from SPMd of the dataset with a standard deviation of the white noise of 0.5.

Three things are observed from the figure. The simple model does not give white errors (Dep and Corr) in the voxels where oscillations were added. At jitter-values lower than 30ms the NVR method seems capable of removing the aliased noise of even the 3Hz oscillation, whilst with increased jitter values only the slower oscillation can be removed. At this noise level the normality test (Norm) show problems in identifying the aliased 1 Hz oscillation as non-normal.

Phantom study

The results of the SPMd on the two analyses (with and without high-pass filter) of the phantom dataset are shown in Figure 4 together with periodograms, of standardised residuals, averaged over all voxels in the phantom. It is seen from the figure that the use of the 60s high-pass filter adequately removes non-white and non-normal noise.

In vivo data

The F-test maps of the different nuisance effects are shown in Figure 5 to Figure 8. The maps are all from the same session for the expanding ring paradigm in subject 3 (R3), however maps from other sessions had similar structures. Both low-frequency noise (Figure 5) and residual movement artefacts (Figure 6) are seen to be pronounced near the edges of the brain (F-test of respective nuisance regressors thresholded at p=0.05 corrected using Gaussian Random Fields (GRF)). The map of the respiratory noise (Figure 7) shows significant effect in ventricles and veins as expected from previous studies, but the effect was less pronounced and for display purposes the F-test of respiration regressors was thresholded at p=0.05 corrected using False Discovery Rate (FDR). The cardiac induced noise is dominant near the major arteries of the brain e.g. CW and MCA, in fact the map shown in Figure 8 has similarities with a coarse arterial angiogram. The corresponding activation map (F-test of the 6 sinosoidals describing the paradigm thresholded at p=0.05 GRF corrected) is shown in Figure 9, where activation in the visual cortex and the laterale geniculate nuclei is seen.

The results of the SPMd of the eight different analyses of the 16 sessions are summarized in Figure 10. The figure shows for the different analyses across the different sessions the factor by which rejections of the null-hypothesis exceeded the expected number when the SPMd maps are thresholded at p-value of p=0.001. In Figure 11 the SPMd maps for selected representative sessions (R3, W3 and R6) are shown.

The first thing to notice in the maps in Figure 11 is that the lowpass filter is not sufficient to provide white normal noise *in vivo*. This is in agreement with the histograms in Figure 12 showing that in our study almost all cardiac induced noise is aliased down to frequencies escaping the high-pass filter. Next thing to note is that the "SPM2-AR(1)" model fails to whiten the noise in the regions where we have shown especially cardiac noise to be dominant e.g. MCA and CW. In fact while the "SPM2-AR(1)" method seems to reduce first order temporal correlation significantly it also seems to increase higher order correlations.

The "NVR-ECG" and "NVR-ox" models have the best overall performance and in general seem to be able to remove the cardiac-induced noise, but there is virtually no difference between the performance of these two "NVR" models. "NVR-Phys" and "NVR-Motion" seem to have similar performance. Next to the "Simple" model the "NVR-rp" model has the poorest overall performance, only the normality test seems to loose power when extra random regressors are included. Yet even when this drop in power of the normality test is taken into consideration no model performs as well as the "NVR-ECG" and "NVR-ox" models.

Discussion

Simulated data

Our simualations show that aliased physiological noise can give rise to non-white non-normal noise, suggesting that these can be modelled satisfactorily using RETROICOR or similar methods. The results show, as expected, that the necessary precision of the phase measurement provided by external recording (ECG, pulse-oximeter, respiratory belt) increases with the frequency of the oscillation in question. Correspondingly a longer TR will also require higher precision. The results indicate that 1Hz oscillations sampled at 1/(2.37s) should be almost completely removable with a phase precision of 100ms. This can be matched by the quality of the time-courses provided by scanner vendors (e.g. on newer Siemens systems ECG is sampled at 400Hz and pulse-oximetry 50Hz). For higher frequencies, or longer TR's, the results are less optimistic. Given the flat nature of the peaks observed with pulse oximetry a better sampling rate is probably not going to help. The ECG is sampled at 400Hz, but is typically contaminated by the oscillating gradients of the EPI sequence. If filtering techniques known from simultaneous EEG-fMRI (Allen et al., 2000) were applied it should be possible to increase the precision of the phase estimated from the ECG time-courses.

Phantom study

From Figure 4 it is seen that the high-pass filter modelled as a discrete cosine set adequately removes non-normal and correlated noise. This allows us to suggest that if high-pass filtering is not enough for ensuring white normal distributed noise in the *in vivo* data, physiological sources are needed to model the remaining correlation.

In vivo data

Not only do the "NVR-ECG" and "NVR-ox" models seem to be capable of removing correlations and non-normality in the residual errors, they also have the best overall performance in the SPMd tests. The "SPM2-AR(1)" model is better at removing AR(1) type correlations but is not able to robustly reduce the correlation in errors due to cardiac-induced noise. The observation that the "NVR-ox" model gives almost similar results to the "NVR-ECG" model indicates that the phase of the cardiac cycle is not better determined in the ECG than in the pulse-oximeter time-course. Two explanations exist for this phenomenon. The first explanation is that the poor ECG quality does not allow for precise phase determination. If this is the case we could in future expect the "NVR-ECG" to perform better if more precise phase determination can be made. Alternatively the phase of the pulse-wave is poorly defined because of dispersion of the pulse-wave through the vasculature, and in this case there is little hope for improved performance. The "NVR-Phys" and "NVR-Motion" models perform similar, but both worse than the full model indicating that all effects need to be modelled if the technique is to be successful. The poor performance of the "NVR-rp" demonstrates that the performance obtained by the "NVR-ECG" and "NVR-ox" models is not just an effect of the SPMd tests losing power when a large number of regressors is included in the design matrix.

Impact on fMRI techniques and data analysis

The findings in this paper are of special concern if areas such as hippocampus, amygdala, insula and thalamus are among the regions of interest. However since cardiac induced noise is found in 27.5% of all grey matter voxels (Dagli et al., 1999) the problem of non-white noise is of general concern when valid statistics are required.

In addition to residual whitening the findings in this paper are relevant to other topics in fMRI including slice-timing and connectivity measurements in

fMRI.

The main assumption behind slice-timing is that the measured signal and noise are critically sampled thereby making it possible to use sinc-interpolation to predict the signal between samples. While the signal itself is probably sampled at a high enough rate, the cardiac noise etc. is typically not. Therefore slice-timing is problematic in regions where we have shown cardiac and potentially respiratory noise to contribute to the measured signal. In our study it was possible to use a GLM which can account for differences in acquisition time without slice-timing. Based on the findings in this and other studies, this is the recommended procedure, if possible. Alternatively methods should be developed which account for physiological noise as a part of the slice-timing procedure.

Functional connectivity mapping is an interesting technique for studying spontaneous neural activity by correlating the time-course of a seed voxel to all other voxels in the brain. In its original version (Biswal et al., 1995) short TR was used to assure critical sampling of cardiac and respiratory noise, making it possible to eliminate these noise sources using simple high pass filtering. For long TR, however, connectivity mapping is complicated by aliasing of physiological noise (Lund, 2001) which could lead to vascular components (Lund et al., 2002) in what is interpreted as a map of functional connectivity. The F-test of the effect of the nuisance regressors described in this paper illustrates clearly that several non-neural effects can give rise to structured "activity" patterns in fMRI time-courses. On the other hand, the suggested nuisance model (or similar models) could be used to study functional connectivity even in the case where cardiac noise is not critically sampled. This approach is along the same path as that used by Peltier & Noll (2002) who have used the method proposed by Hu et al. (1995) to correct for physiological noise sources.

Conclusion

In the current paper we have shown that the NVR model composed of a comprehensive set of nuisance regressors is sufficient to whiten otherwise correlated residuals. The NVR model is based on a number of effects which are known to contribute to the non-white noise in fMRI (hardware drift, residual movement artefacts, respiration and cardiac pulsation). In fact the proposed NVR model is only new in the sense that we for the first time have used a combination of several already published models in the same analysis.

It was furthermore found that our approach, in general, was superior to the covariance estimation currently implemented in SPM2. In particular, we found the global AR(1) model of SPM to be inadequate near larger arteries which

is not surprising given the inability of a first-order AR model to acount for oscillatory noise.

Acknowledgements

Mark Griffin, Montreal Neurological Institute, Canada and Gunnar Krüger, Siemens Medical, Germany, are acknowledged for their help with accessing the physiological recordings from the scanner. The Simon Spies Foundation is acknowledged for donation of the Siemens Trio scanner.

References

Allen, P., Josephs, O., & Turner, R. (2000). A method for removing imaging artifact from continuous EEG recorded during functional MRI. *NeuroImage*, **12**(2), 230–9.

Biswal, B., Yetkin, F. Z., Haughton, V. M., & S., H. J. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn Reson Med*, **34**(4), 537–41.

Biswal, B., Edgar, A., Yoe, D., & Hyde, J. S. (1996). Reduction of Physiological Fluctuations in fMRI Using Digital filters. *Magn Reson Med*, **35**, 107–113.

Buonocore, M. & Maddock, R. (1997). Noise suppression digital filter for functional magnetic resonance imaging based on image reference data. *Magn Reson Med*, **38**(3), 456–69.

Dagli, M. S., Ingeholm, J. E., & Haxby, J. V. (1999). Localization of cardiac-induced signal change in fMRI. *NeuroImage*, **9**(4), 407–15.

Frackowiak, R., Friston, K., C.D., F., Dolan, R., Price, C., Zeki, S., Ashburner, J., & Penny, W. (2004). *Human Brain Function*. Elsevier Academic Press, London, second edition.

Franckowiak, R. S. J., Friston, K. J., Frith, C. D., Dolan, R. J., & Mazziotta, J. C. (1997). *Human Brain Function*. Academic Press, London.

Frank, L. R., Buxton, R. B., & Wong, E. C. (2001). Estimation of respiration-induced noise fluctuations from undersampled multislice fMRI data. *Magn Reson Med*, **45**, 635–44.

Friston, K. J., Williams, S., Howard, R., Franckowiak, R. S. J., & Turner, R. (1996). Movement-Relatet effects in fMRI time series. *Magn Reson Med*, **35**, 346–355.

Friston, K. J., Glaser, D. E., Henson, R. N., Kiebel, S., Phillips, C., & Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: applications. *NeuroImage*, **16**, 484–512.

Glover, G. H., Li, T. Q., & Ress, D. (2000). Image-based method for retro-

spective correction of physiological motion effects in fMRI: RETROICOR. *Magn Reson Med*, **44**, 162–7.

Hu, X., Le, T. H., Parrish, T., & Erhard, P. (1995). Retrospective Estimation and Correction of Physiological Fluctuation in Functional MRI. *Magn Reson Med*, **34**, 201–212.

Josephs, O., Howseman, A., Friston, K., & Turner, R. (1997). Physiological noise modelling for multi-slice EPI fMRI using SPM. *In: Proceedings of the 5th Annual Meeting of ISMRM*, page 1682.

Krüger, G. & Glover, G. H. (2001). Physiological noise in oxygenation-sensitive magnetic resonance imaging. *Magn Reson Med*, **46**, 631–7.

Krüger, G., Kastrup, A., & Glover, G. H. (2001). Neuroimaging at 1.5 T and 3.0 T: comparison of oxygenation-sensitive magnetic resonance imaging. *Magn Reson Med*, **45**, 595–604.

Li, S., Biswal, B., Li, Z., Risinger, R., Rainey, C., Cho, J., Salmeron, B., & Stein, E. (2000). Cocaine administration decreases functional connectivity in human primary visual and motor cortex as detected by functional MRI. *Magn Reson Med*, **43**(1), 45–51.

Lund, T. E. (2001). fcMRI -Mapping Functional Connectivity or Correlating Cardiac Induced Noise. *Magn Reson Med*, **46**, 628.

Lund, T. E. & Hanson, L. G. (2001). Physiological noise correction in fMRI using vessel time-series as covariates in a general linear model. *In: Proceedings of the 9th Annual Meeting of ISMRM*, page 22.

Lund, T. E. & Larsson, H. B. W. (1999). Spatial distribution of low-frequency noise in fMRI. *In: Proceedings of the 7th Annual Meeting of ISMRM*, page 1705.

Lund, T. E., Pagsberg, A. K., & Barré, W. (2002). Spatial distribution of cardiac induced noise - implications for fMRI and functional connectivity mapping. *In: Proceedings of the 10th Annual Meeting of ISMRM*, page 304.

Luo, W. L. & Nichols, T. E. (2003). Diagnosis and exploration of massively univariate neuroimaging models. *NeuroImage*, **19**, 1014–32.

Peltier, S. J. & Noll, D. C. (2002). T2* dependence of low frequency functional connectivity. *NeuroImage*, **16**, 985–92.

Penny, W., Kiebel, S., & Friston, K. (2003). Variational Bayesian inference for fMRI time series. *NeuroImage*, **19**, 727–41.

Petersen, N. V., Jensen, J. L., Burchardt, J., & Stødkilde-Jørgensen, H. (1998). State space models for physiological noise in fMRI time series. *NeuroImage*, **7**(4), s592.

Raj, D., Paley, D. P., Anderson, A. W., Kennan, R. P., & Gore, J. C. (2000). A model for susceptibility artefacts from respiration in functional echo-planar magnetic resonance imaging. *Phys Med Biol*, **45**, 3809–20.

Raj, D., Anderson, A. W., & Gore, J. C. (2001). Respiratory effects in human functional magnetic resonance imaging due to bulk susceptibility changes. *Phys Med Biol*, **46**, 3331–40.

Smith, A. M., Lewis, B. K., Ruttimann, U. E., Ye, F. Q., Sinnwell, T. M., Yang, Y., Duyn, J. H., & Frank, J. A. (1999). Investigation of low frequency drift

in fMRI signal. *NeuroImage*, **9**, 526–33.

Thomas, C. G., Harshman, R. A., & Menon, R. S. (2002). Noise reduction in BOLD-based fMRI using component analysis. *NeuroImage*, **17**, 1521–37.

Wang, J., Aguirre, G. K., Kimberg, D. Y., Roc, A. C., Li, L., & Detre, J. A. (2003). Arterial spin labeling perfusion fMRI with very low task frequency. *Magn Reson Med*, **49**, 796–802.

Weisskoff, R. M., Baker, J., Belliveau, J., Davis, T. L., Kwong, K. K., Cohen, M., & Rosen, B. R. (1993). Power spectrum analysis of functionally weigthed MR data: What's in the noise. In *Proc., SMRM*, number 12, page 7, New York.

Windischberger, C., Langenberger, H., Sycha, T., Tschernko, E. M., Fuchsjager-Mayerl, G., Schmetterer, L., & Moser, E. (2002). On the origin of respiratory artifacts in bold-epi of the human brain. *Magn Reson Imaging*, **20**, 575–82.

Worsley, K. J. & Friston, K. J. (1995). Analysis of fMRI time-series revisited–again. *NeuroImage*, **2**, 173–81.

Worsley, K. J., Liao, C. H., Aston, J., Petre, V., Duncan, G. H., Morales, F., & Evans, A. C. (2002). A general statistical analysis for fMRI data. *NeuroImage*, **15**, 1–15.

Table 1
The different analysis performed on the *in-vivo* data.

Model	Description
Simple	A model including baseline plus sine and cosine of the first three harmonics of the (1/30s) oscillation. This model is capable of modelling all possible phases of the haemodynamic response, and thus eliminate the need of performing slice-timing. Moreover the two higher harmonics of the paradigm frequency should eliminate correlation in the residuals due to shape differences between a canonical haemodynamic response function and the observed response.
60s-HP	The "Simple" model extended to include a high-pass filter modelled as a discrete cosine-set (30 regressors) with a minimum period of 60s.
SPM2-AR(1)	The "60sec-HP" model with whitened residuals using a global AR(1) model estimated in a mask defined by the voxels where a significant effect of the paradigm was observed (Friston et al., 2002). (This is the recommended SPM2 procedure).
NVR-ECG	The "60sec-HP" model extended to include several extra nuisance regressors for modelling the autocorrelation. A Volterra expansion of the movement parameters was used to model residual movement effects including spin-history effects (Friston et al., 1996) (24 regressors). Respiration and cardiac noise was modelled using 16 RETROICOR (Glover et al., 2000) regressors (5 cardiac harmonics and 3 respiratory harmonics). The cardiac frequency and phase was determined using the scanner ECG system. The respiratory phase and frequency was measured using the scanner respiratory belt.
NVR-ox	A model similar to "NVR-ECG" but using the scanner pulse-oximeter, instead of ECG, to measure the cardiac phase and frequency.
NVR-rp	A model similar to "NVR-ECG" but here the nuisance regressors are permuted to provide a set of random nuisance regressors of similar size as in the "NVR-ECG" model.
NVR-Phys	A model similar to "NVR-ECG" but without the movement regressors.
NVR-Motion	A model similar to "NVR-ECG" but without the physiological regressors.

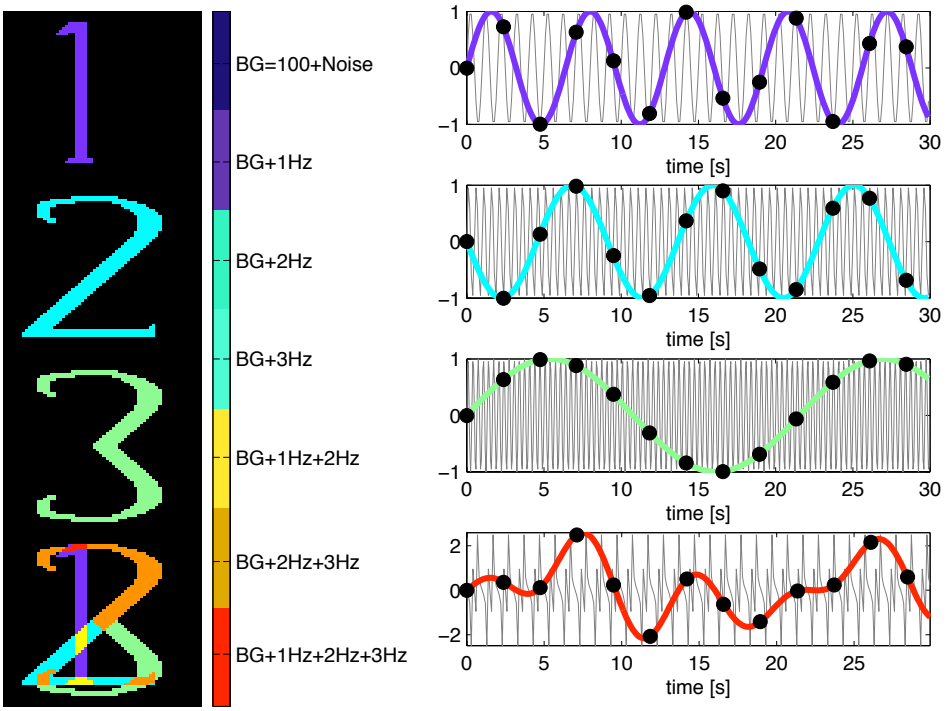

Fig. 1. **Generation of Simulated data.** The figure shows how the data used in the simulation were generated. Slice 1 contains an aliased 1Hz oscillation in all voxels located in a pattern showing the digit "1", slice 2 contains an aliased 2Hz oscillation in all voxels located in a pattern showing the digit "2", slice 3 contains an aliased 3Hz oscillation in all voxels located in a pattern showing the digit "3". In slice 4 the pattern of oscillatory behaviour was constructed by adding the patterns of the slices 1 to 3. Where two or more patterns from these slices intersect the corresponding oscillations were added to the time-course of the particular voxel, as illustrated for the voxels where all three digits intersect (red curve). In addition to the harmonic oscillations the time-course of all voxels were added a constant value of 100 and normal-distributed white noise with standard deviation specific to each dataset.

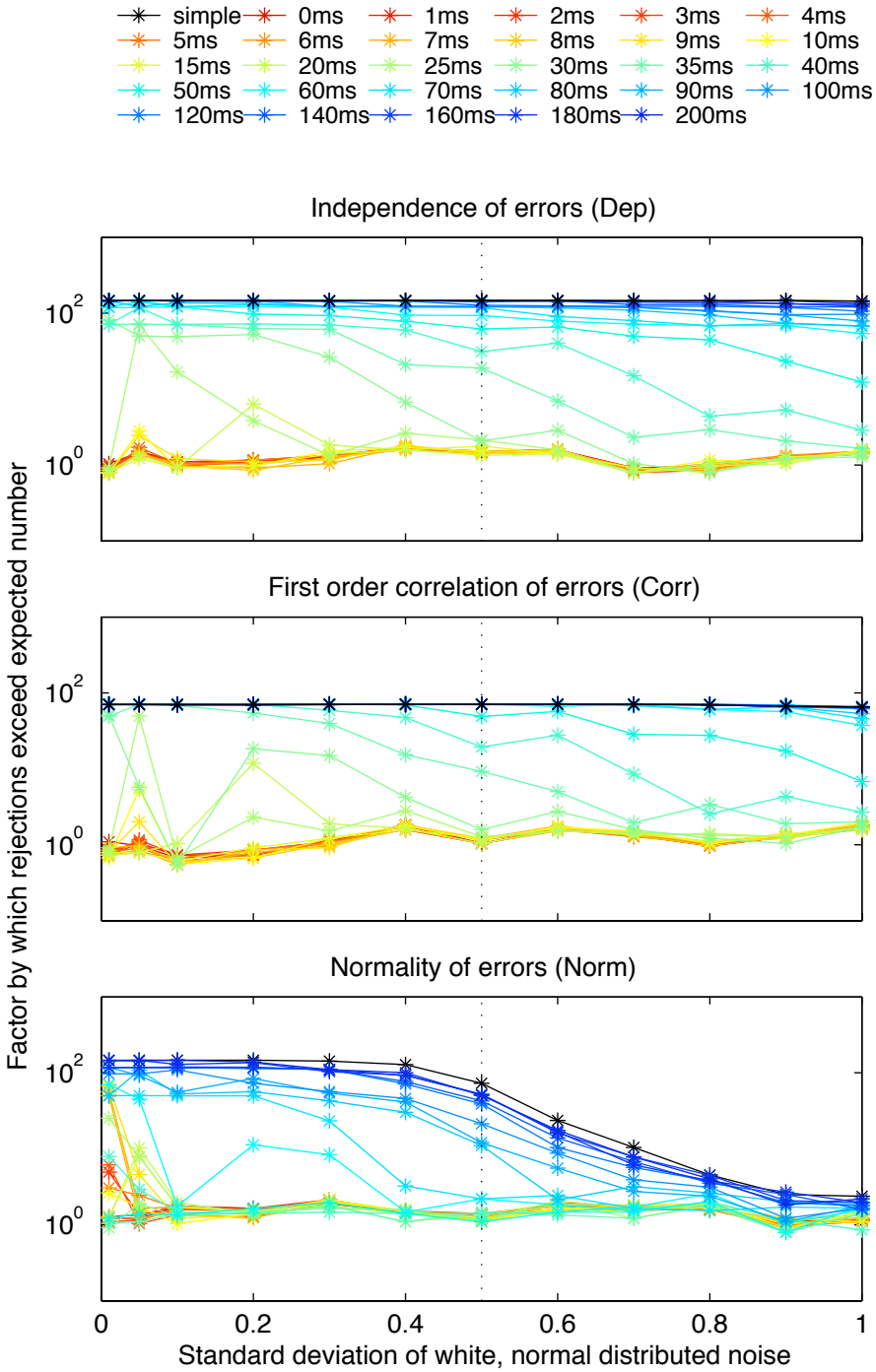

Fig. 2. **Non-white noise in simulated data.** The figure shows for the three different tests, as a function of increasing standard deviation of the white noise, the factor by which rejections exceed the expected number. The different curves correspond to a simple analysis and several analyses using an external reference time-course to whiten the periodic behaviour. Different colours indicate the standard deviation of the jitter introduced to the phase of the reference time-course. The vertical dotted black line at 0.5 indicates the noise-level of the dataset used in Figure 3.

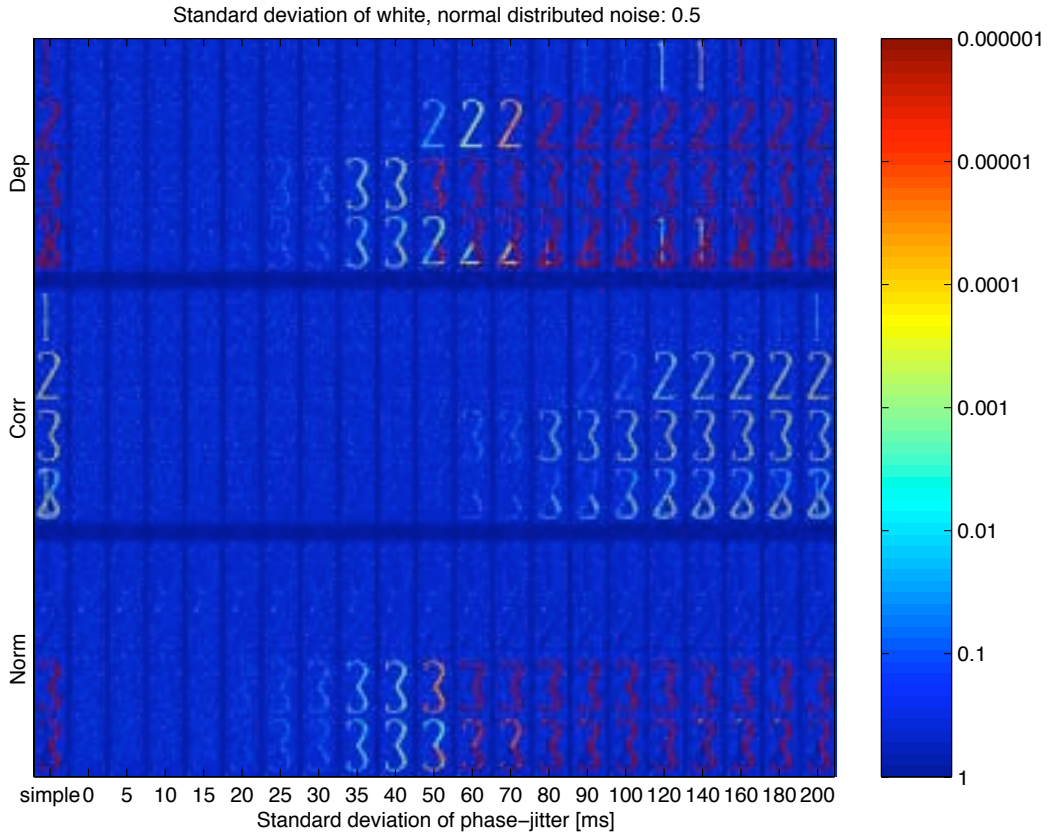

Fig. 3. **Non-white noise in simulated data.** The figure shows the output of the SPM-diagnosis from the simple analysis and from several analyses using an external reference time-course to whiten the periodic behaviour. In the horizontal direction increasing jitter has been introduced to the phase of the reference time-course. The colours correspond to the p-value at which the null hypothesis of the corresponding SPMd test (Dep, Corr or Norm) was rejected.

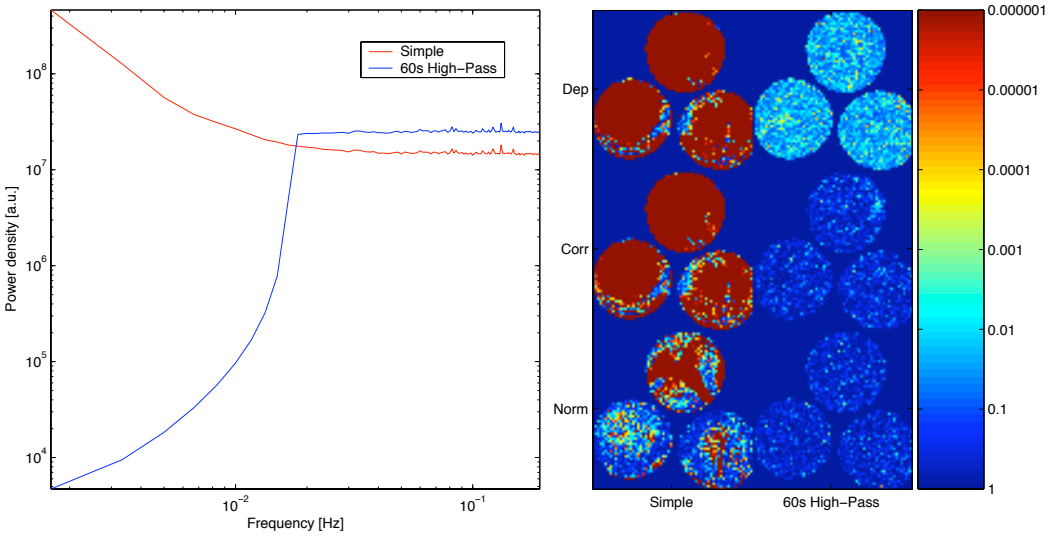

Fig. 4. **Non-white noise in Phantom data.** The figure shows **(left)** averaged periodograms of the standardised residuals from all voxel inside the phantom and **(right)** SPM-diagnosis maps from analysis with and without high-pass filter included in the design matrix. The colours correspond to the p-value at which the null hypothesis of the corresponding SPMd test (Dep, Corr or Norm) was rejected. It is seen from the spectrum and the image that non-white noise in phantom data can be modelled adequately using a discrete cosine-set (with shortest period of 60s) as high-pass filter.

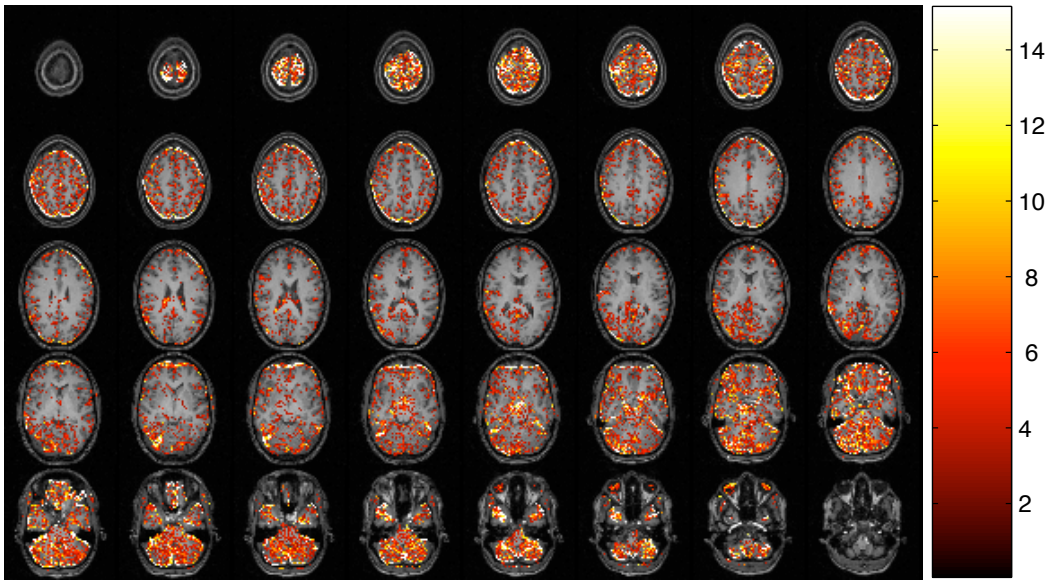

Fig. 5. **Low-frequency noise** *in vivo*. The figure shows (in colour) F-test values of the voxels showing significant (p=0.05 corrected using GRF) effect of a linear combination of the discrete cosine-set constituting the high-pass filter (F-test). It is seen that the low-frequency noise is dominant near the edges of the brain.

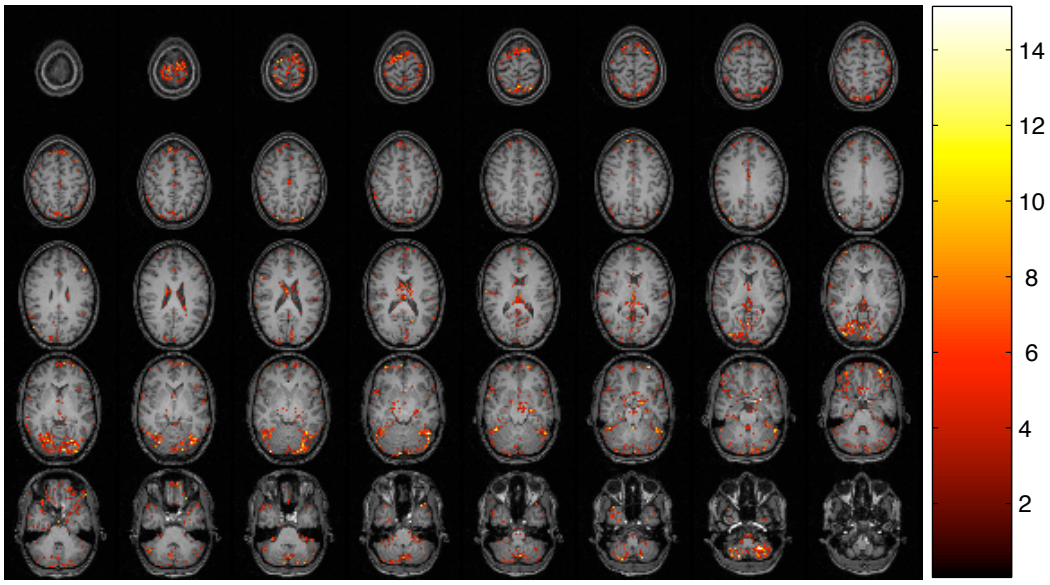

Fig. 6. **Residual movement effects**. The figure shows (in colour) F-test values of the voxels showing significant (p=0.05 corrected using GRF) effect of a linear combination of the regressors describing residual movement effects (F-test). It is seen that the residual movement effects are dominant near the edges of the brain.

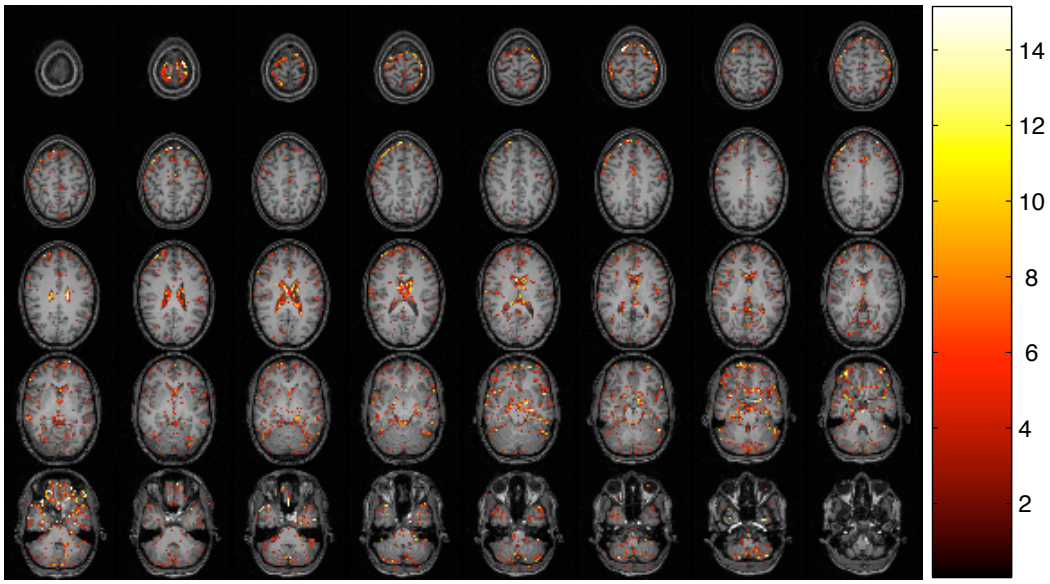

Fig. 7. **Respiratory induced noise.**The figure shows (in colour) F-test values of the voxels showing significant (p=0.05 corrected using FDR) effect of a linear combination of the regressors describing the aliased respiratory oscillation (F-test). It is seen that the respiratory induced noise is dominant near the edges of the brain as well as near in the larger veins and in the ventricles.

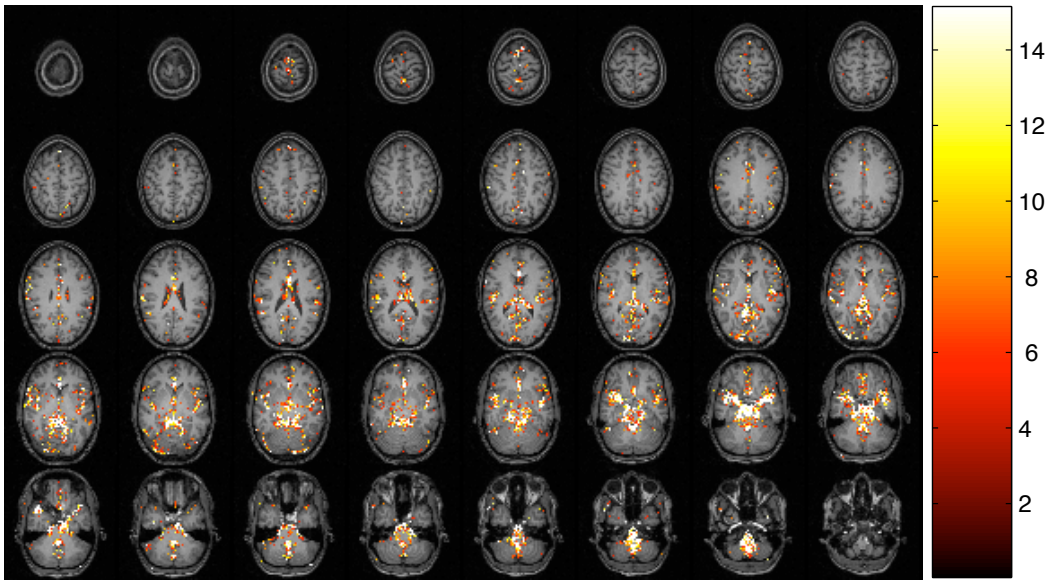

Fig. 8. **Cardiac-induced noise.** The figure shows (in colour) F-test values of the voxels showing significant (p=0.05 corrected using GRF) effect of a linear combination of the regressors describing the aliased cardiac oscillation (F-test). It is seen that the cardiac-induced noise is dominant near larger vessels (e.g. medial cerebral artery and Circle of Willis).

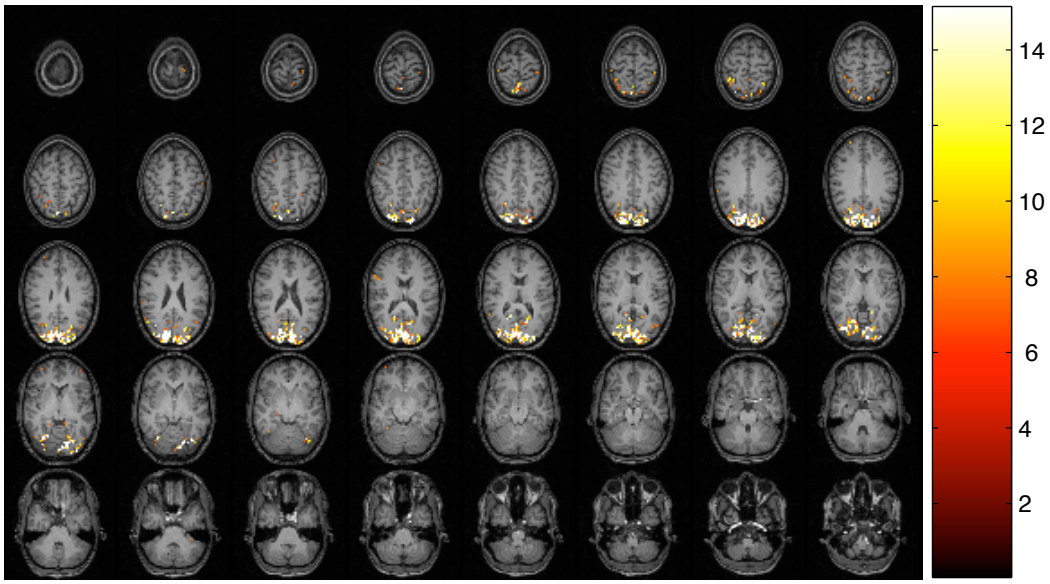

Fig. 9. **Visual activation.** The figure shows (in colour) F-test values the voxels showing significant (p=0.05 corrected using GRF) effect of a linear combination of the regressors describing the stimulation with the expanding ring paradigm (F-test). Primary and higher order visual cortices are seen as well as Lateral Geniculate Nuclei.

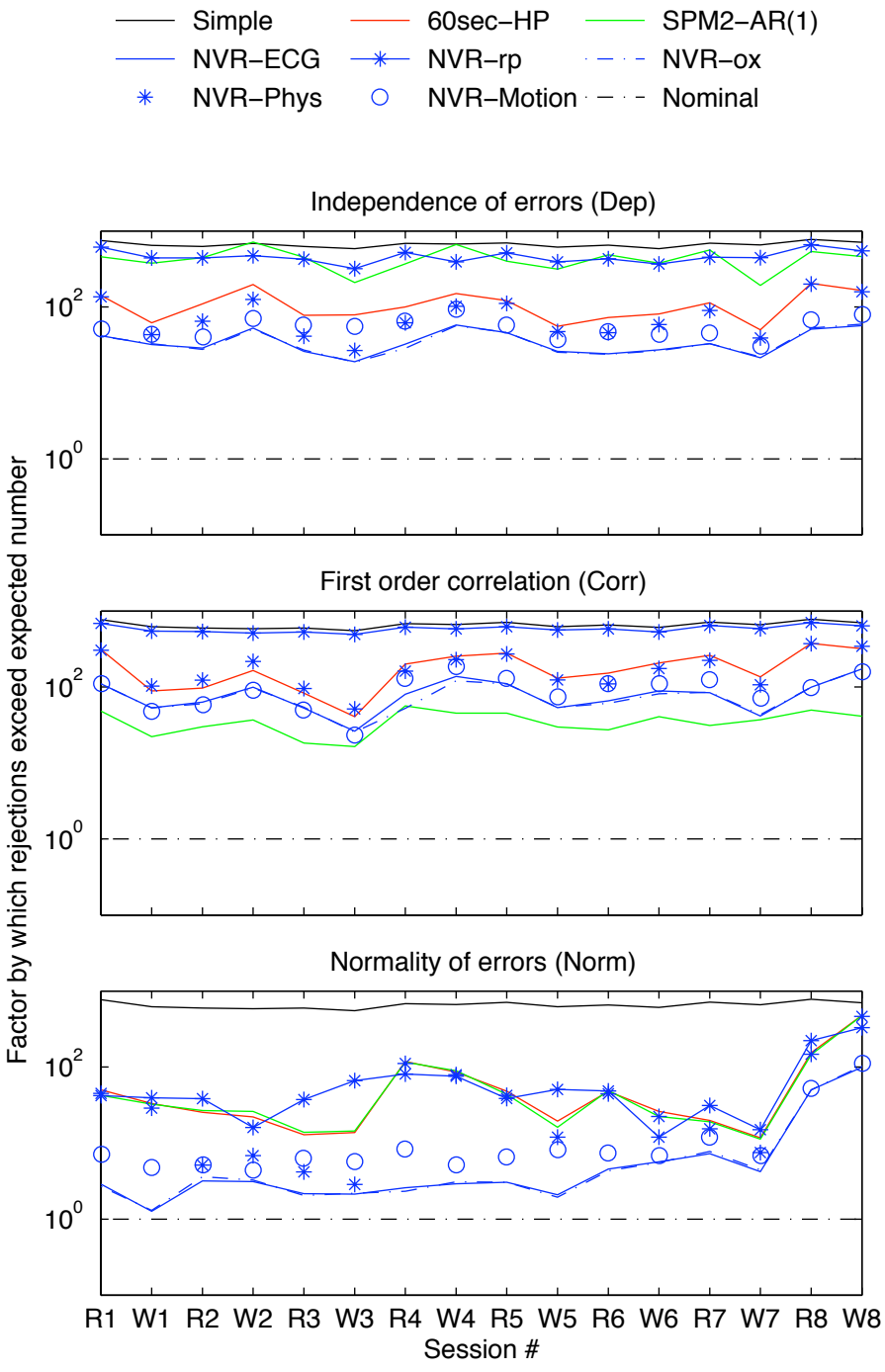

Fig. 10. **Non-white** *in vivo***.** The figure shows for the three different tests, for each of the 16 sessions, the factor by which rejections exceed the expected number. The different curves correspond to various models. The maps corresponding to session R3, W3 and R8 are shown in Figure 11.

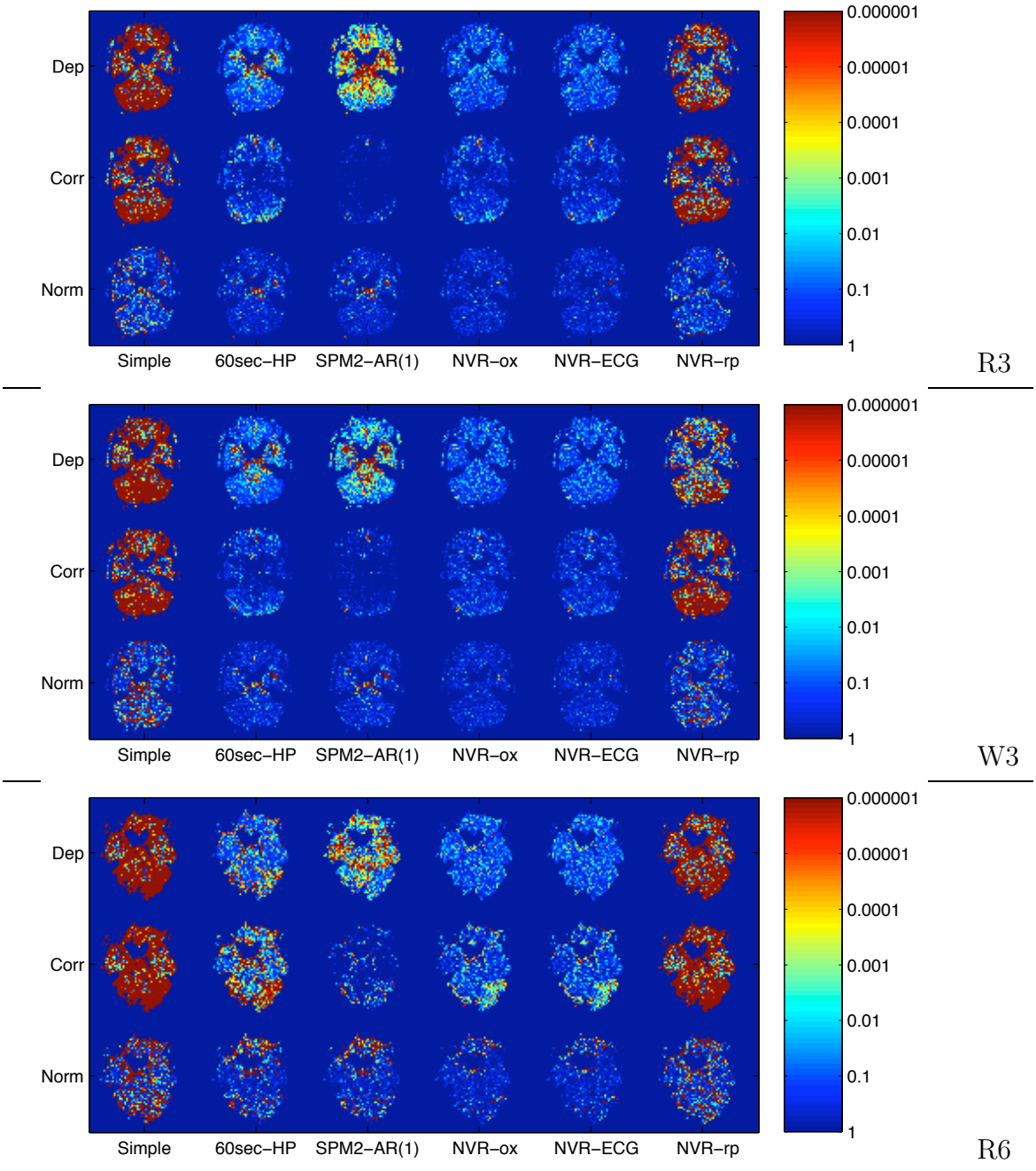

Fig. 11. **Non-white noise** *in vivo***.** The figure shows SPM-diagnosis maps from several different analyses of the data from 3 different sessions. The colours correspond to the p-value at which the null hypothesis of the corresponding SPMd test (Dep, Corr or Norm) was rejected. It is seen that white noise can be obtained by modelling effects of cardiac, respiratory, motion and low-frequency drift (NVR-ECG or NVR-ox), but not using the standard SPM2 whitening approach (SPM2-AR(1)).

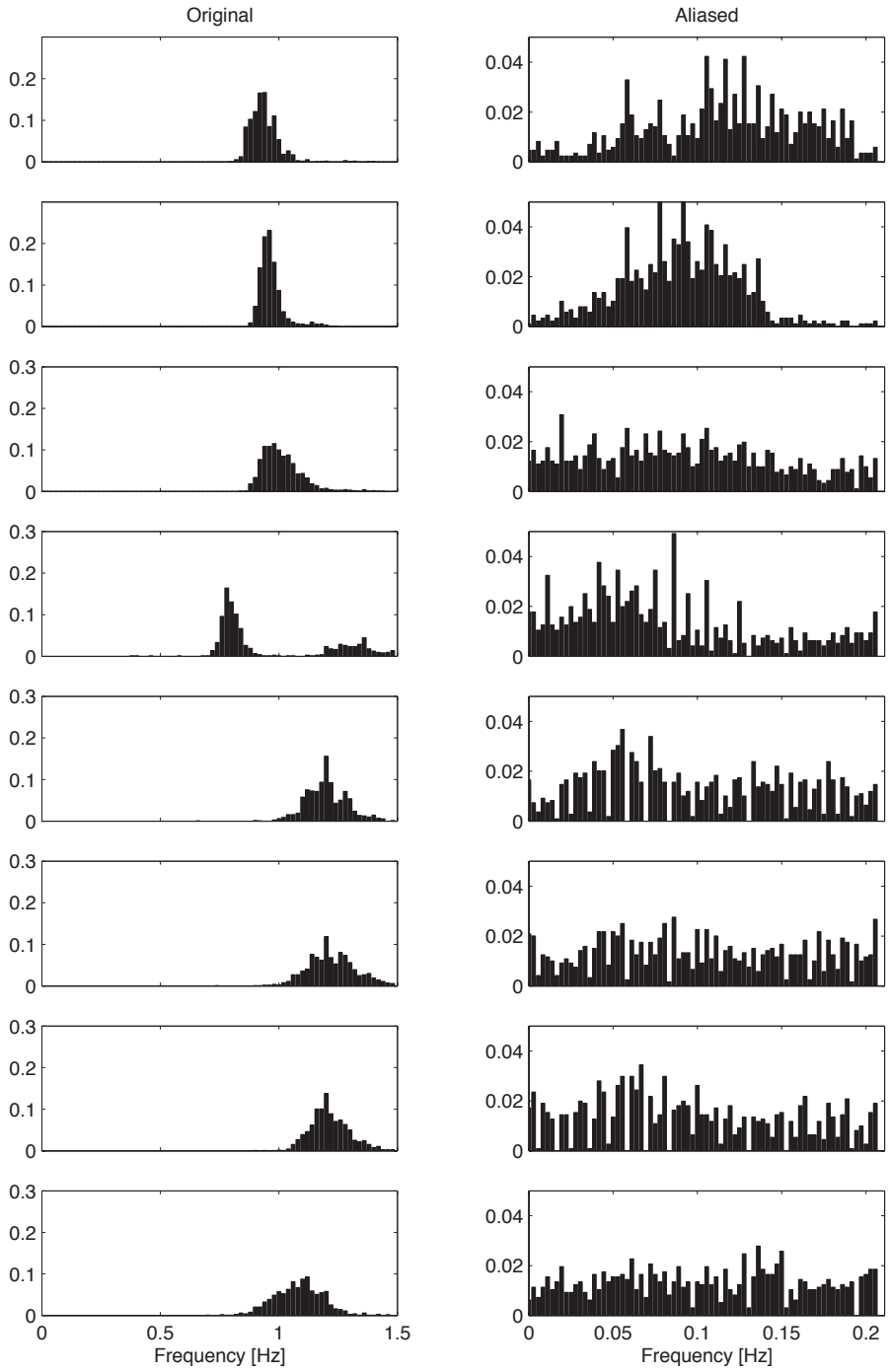

Fig. 12. **Aliasing of cardiac pulsation.** The figure shows normalised histograms of beat to beat frequencies, for the eight different subjects, before and after aliasing (aliased frequencies are calculated using Eaquation 3). The frequencies are calculated as one divided by the beat to beat interval. It is seen from the histogram that the aliased beat to beat frequencies are distributed across the entire sampled bandwidth.