Received: 21 October 2019 | Revised: 31 December 2019 | Accepted: 17 January 2020

DOI: 10.1002/mrm.28203

FULL PAPER

Magnetic Resonance in Medicine

Improved autoregressive model for correction of noise serial correlation in fast fMRI

Qingfei Luo[1] | Masaya Misaki[1] | Ben Mulyana[1] | Chung-Ki Wong[1] | Jerzy Bodurka[1,2]

[1]Laureate Institute for Brain Research, Tulsa, Oklahoma, USA

[2]Stephenson School for Biomedical Engineering, University of Oklahoma, Norman, Oklahoma, USA

Correspondence
Jerzy Bodurka, Laureate Institute for Brain Research, 6655 S Yale Ave, Tulsa, OK 74136-3326
Email: jbodurka@laureateinstitute.org

Funding information
National Institute of General Medical Sciences, Grant/Award Number: P20GM121312; William K. Warren Foundation; National Institute of General Medical Sciences; National Institutes of Health, Grant/Award Number: P20GM121312

Purpose: In rapidly acquired functional MRI (fast fMRI) data, the noise serial correlations (SC) can produce problematically overestimated T-statistics which lead to invalid statistical inferences. This study aims to evaluate and improve the accuracy of high-order autoregressive model (AR(p), where p is the model order) based prewhitening method in the SC correction.

Methods: Fast fMRI images were acquired at rest (null data) using a multiband simultaneous multi-slice echo planar imaging pulse sequence with repetition time (TR) = 300 and 500 ms. The SC effect in the fast fMRI data was corrected using the prewhitening method based on two AR(p) models: (1) the conventional model (fixed AR(p)) which preselects a constant p for all the image voxels; (2) an improved model (AR_{AICc}) that employs the corrected Akaike information criterion voxel-wise to automatically select the model orders for each voxel. To evaluate accuracy of SC correction, false positive characteristics were measured by assuming the presence of block and event-related tasks in the null data without image smoothing. The performance of prewhitening was also examined in smoothed images by adding pseudo task fMRI signals into the null data and comparing the detected to simulated activations (ground truth).

Results: The measured false positive characteristics agreed well with the theoretical curve when using the AR_{AICc}, and the activation maps in the smoothed data matched the ground truth. The AR_{AICc} showed improved performance than the fixed AR(p) method.

Conclusion: The AR_{AICc} can effectively remove noise SC, and accurate statistical analysis results can be obtained with the AR_{AICc} correction in fast fMRI.

KEYWORDS

Akaike information criterion, autocorrelation, autoregressive model, fast fMRI, noise serial correlation, prewhitening

1 | INTRODUCTION

With the development of simultaneous multi-slice (SMS) acceleration acquisition techniques,[1-3] echo planar imaging (EPI) images of the whole brain can be acquired with a short TR (<1 s) in functional MRI studies (fast fMRI). Compared to conventional whole brain fMRI experiments (TR ~ 2 s), fast fMRI can achieve higher signal-to-noise ratio efficiency and detection power.[4] However, higher noise serial correlations (SC) (ie, temporal autocorrelations) are present in fast fMRI time series.[5] Since the SC effect causes underestimation of noise variance,[6] the statistical analysis based on the general linear model (GLM)[7] can produce problematically large overestimates of *t*-scores. Consequently, the false positive rate (FPR) significantly exceeds the nominal value, and this can result in invalid statistical inferences and the difficulty of interpreting the findings in a study. Therefore, it is important to adequately suppress the noise SC in fast fMRI.[5]

One approach for correcting the SC effect is to use the ordinary least squares estimator to calculate the T-statistics and then adjust the degrees of freedom post hoc.[6,8] Alternatively, the SC in fMRI data can be suppressed with the prewhitening method, which applies a prewhitening filter to remove colored components from the noise.[9,10] The noise structure after prewhitening matches the assumption of the statistical test, that is, independent and identical distribution (i.i.d.). The advantage of prewhitening is that the degrees of freedom revert to their usual values (ie, the number of observations minus the number of parameters[11]) and the t and F values can be calculated with the classical equations,[11] so it has been widely used in fMRI data anlaysis. To ensure the efficacy of the prewhitening method, the autocorrelation function (ACF) of noise must be accurately measured.[12] In conventional fMRI experiments, one simple and widely used method for estimating the ACF is fitting the GLM residuals at image voxels to the first-order autoregressive process (AR(1)).[9] Nonetheless, the AR(1) is inadequate to model the noise in fast fMRI data and cannot effectively correct the overestimation of T-statistics.[5,13] More complicated noise models have been implemented in fMRI data analysis software to improve the accuracy of prewhitening. For instance, the AR(1) + white noise (a.k.a. ARMA(1,1)) model[14] has been used by AFNI[15] in its function "3dREMLfit" to correct the serial correlations. SPM (https://www.fil.ion.ucl.ac.uk/spm) suppresses the SC effect in fast fMRI data with the "FAST" model, which decomposes the noise autocovariance matrix into multiple basis covariance components.[16] Recently, the high-order AR model (AR(p), $P > 1$) has also been proposed for the SC correction in fMRI data acquired with short TRs.[5,13,17] Since the AR(p) is a direct extension to the AR(1) model, it is a straightforward and promising approach for removing the noise SC in fast fMRI.[5,13,17]

Previous studies[5,13,17] have shown that the AR(p)-based prewhitening method can reduce the t-score overestimation to a greater degree than the traditional AR(1) model. However, the accuracy of statistical analysis using the AR(p) correction has not yet been systematically investigated. It is still unknown whether the GLM analysis with the AR(p)-based prewhitening can produce accurate statistical inferences. In fMRI, an exact statistical test means that the FPR in the fMRI activation map is equal to the assumed significance level (α),[12,14] so the FPR-α relationship in fMRI images acquired at rest (null data) has been used as the quantitative criterion to evaluate the accuracy of SC correction.[10,14] Nevertheless, the false positive characteristics (FPC) (ie, FPR vs α curves) after the AR(p)-based prewhitening have not yet been measured in fast fMRI studies. In addition, note that the autocovariances of noise are not same as that of GLM residuals. A bias will creep into the measurement of noise ACF if this difference is not corrected in the AR(p) method.[18] Also, AR(p) model may fail if the selected model order is too low or too high relative to the true value.[19] In fMRI, the optimal order may vary with the voxel location even in the same tissue type[20] and also depends on imaging parameters such as TR.[5,13] Thus, to achieve accurate SC removal, it is necessary to reduce the bias in the estimation of noise ACF and to adaptively detect the optimal model orders at individual voxels. Unfortunately, these two factors have not been simultaneously considered in previous fast fMRI studies[5,13,17] using the AR(p)-based SC correction. Therefore, a robust AR(p)-based SC correction method is not available up to date, and it is difficult to ensure the accuracy of statistical analysis in fast fMRI studies. This can significantly limit the application of SMS acquisition techniques in fMRI.

This study aims to improve and evaluate the SC correction accuracy of AR(p)-based prewhitening method in fast fMRI data. We optimize the AR(p) framework by integrating a order selection method based on the corrected Akaike information criterion (AICc)[21] into the AR(p) model proposed by Worsley et al.[18] In Worsley's AR(p) model (fixed AR(p)), an efficient algorithm is implemented to reduce the bias in the estimation of noise autocovariances, but p is preselected by the user and is same at all the voxels. In our model (AR_{AICc}), the AICc identifies the optimal model order in voxel-wise according to the statistical property of noise,[21] so it is unnecessary to know *a priori* the relationships between optimal order, voxel location, and imaging parameters. Thus, this improved algorithm can automatically select the optimal model orders at individual voxels and also reduce the bias in the measurement of noise ACF.

In the present study, fast fMRI images were acquired at rest (null data) with TR = 300 and 500 ms to evaluate the performances of fixed AR(p) and AR_{AICc} models. The FPC of null data were measured in the images without

spatial smoothing and used as the criterion to evaluate the accuracy of SC removal.[10,14] Finally, we also examined the efficacy of AR(p)-based prewhitening in spatially smoothed images by injecting pseudo task fMRI signals into a region of interest (ROI) of the rest fMRI images. The activated clusters detected from the synthesized data were compared to the signal ROI (ie, ground truth) to assess the accuracy of statistical analysis. Through this study, we expect to develop a robust algorithm (AR_{AICc}) for correcting noise serial correlations in fast fMRI. This AR_{AICc} model would enable us to perform accurate statistical analyses on fast fMRI data. It would solve a fundamental question in fMRI experiments using short TRs.

2 | METHODS

A detailed description of the *MRI data acquisition* and *Image preprocessing* including citations to references[22-27] are included in the supplementary material.

2.1 | Data analysis

The SC effect was corrected using the prewhitening method based on the fixed AR(p)[18] and AR_{AICc} models. The theory regarding the prewhitening method based on these two models is described in Supporting Information. Briefly, the AR_{AICc} improves the fixed AR(p) model on two aspects: (1) In the fixed AR(p), the user is required to preselect a constant p which applies to all the voxels, while the AR_{AICc} algorithm automatically determines the optimal voxel-wise model orders with the AICc; (2) The cutoff lag (L) for the noise autocorrelation estimation = p in the fixed AR(p). In the AR_{AICc}, L depends on the TR used for image acquisition: L = int (T_{cutoff}/TR) where T_{cutoff} = 10 s and int() rounds a number to its nearest integer. Since the partial ACF (pACF)[28] and the Bayesian Information Criterion (BIC)[29] are also two common order selection critera used in previous studies on SC correction,[10,17] the AR_{AICc} algorithm also offers the options to use pACF or BIC for the order selection.

FPC in null fMRI data without spatial smoothing was used as the criterion to evaluate the accuracy of SC correction.[10,14] By assuming the presence of task induced signal in the resting data (ie, null data), t-maps were obtained from the GLM analysis with the assumed task regressors. At a given significance level (eg, $\alpha = 0.05$), the voxels above the p-threshold (ie, $P < \alpha$) were considered to be false positives, and the FPR was the ratio of the number of false postives to the total number of voxels in the brain. The FPRs were measured at different α values ($1 \times 10^{-4} \leq \alpha \leq 1$) and then FPC curves were obtained by plotting the FPR versus assumed α. If serial correlations are removed exactly, the noise of fMRI time series after the prewhitening will be in i.i.d. Thus, the FPR is equal to α in the theoretical FPC curve. An SC correction method has better performance if its FPC curve deviates less from the theoretical one.[10,14] In this study, one block and one event-related designed task paradigm was used in the FPC analysis. Task regressors were created by convolving the task paradigm with the canonical hemodynamic response function.[30] Since brain network activity, such as the default mode network,[31] may occur at the resting state, the neural activity can produce activated voxels (ie, true positives) in the null fMRI data if the brain network activity is correlated with the assumed task regressors. To minimize the impact of resting state networks on the FPC measurement, two tasks (A and B) are included in the assumed paradigms and the contrast between task A and B (A-B) was used in the GLM to test false positives.[32] The block design paradigm consisted of seven blocks plus 15-s resting epoch at the end. Each block was composed of one resting and two task epochs and each epoch lasted 15 s. In the event-related paradigm, the interstimulus intervals (ISIs) were randomized from a normal distribution with mean/SD of 6/2 s and all the ISIs were limited in the range of 2-10 s (interval = 1 s). The orders of tasks A and B were randomized in the pradigms and the duration of each paradigm was 330 s.

A continuous data segment was extracted from the fMRI scan to match the duration of the task paradigms and used as null data. To further reduce the bias of FPC measurement, the following null data selection strategy was adopted to avoid the true positives induced by the resting state brain network activity: from the beginning of the fMRI scan, the first 330-s image segment (ie, 0-330 s) was extracted and analyzed with the GLM without SC correction. The t-maps were thresholded at $P < .001$ uncorrected and visually inspected. If resting state fMRI networks were present in the t-maps, then the data segment starting from the next time point (ie, 1-331 s) was examined in the same way as the first one. This procedure continued until no resting state networks appeared in the t-maps and the current image segment was used as the null data.

GLM analyses with and without the prewhitening were performed on the null data with the assumed regressors. When using the fixed AR(p), p was set to different values (1-20) to investigate the impact of p on the SC removal. In the prewhitening, 0, 6, and 15 mm (recommended value in Worsley's study[18]) were used as the full-width half-maximum (FWHM) values in the autocorrelation regularization step (see Supporting Information). The FPC curves were calculated from the t-maps. To evaluate and compare the performances of different AR(p) models, the deviations of measured FPC from the theoretical curve (ie, FPR = α) were examined in the α range 5×10^{-4}-5×10^{-2}, which covers the significance levels usually used in fMRI data analysis. In addition to the AICc, the performances of BIC and pACF based order selection methods were also assessed using the FPC.

Similar to previous studies,[10,14] no spatial smoothing was applied to the fMRI images when measuring the FPC in the present study. Considering that image smoothing is commonly used in fMRI data analysis, the performances of AR(p) methods were also empirically assessed in spatially smoothed data through the following steps: first, task fMRI data were synthesized by adding pseudo blood oxygenation level dependent (BOLD) activations (with 2% signal change) to a ROI in the images acquired at rest. The task paradigms and hemodynamic response function used to create the pseudo BOLD signal were same as in the above FPC measurement. To simulate the data structure in task fMRI experiments, the paradigms were assumed to start with the MRI scan simultanouesly. The synthesized images were then smoothed with FWHM = 6 mm and analyzed with the AR(p) based SC correction. A p-threshold of .001 (uncorrected) was applied to the t-maps and the multiple comparison was corrected using the cluster size corresponding to FPR = 0.05.[33] Finally, the efficacy of SC correction was evaluated by inspecting the false positives in the BOLD activation map.[5,13] "Activated" clusters were considered to be false positives if their locations were distant from the true BOLD activation foci (ie, the ROI). A better SC correction method was expected to produce fewer false positives in the activation maps of individual subjects and to detect the true BOLD activation at the same time.[5,13]

3 | RESULTS

Figure 1A shows the t-maps with and without the SC correction at TR = 500 ms in one exemplar subject. In the t-maps without correction, a number of voxels beyond the p-threshold (0.001 uncorrected) (ie, false positives) were randomly distributed in the whole brain. When using the fixed AR(p) for SC correction, AR(5) or higher order models showed lower t-scores and fewer false positives than the traditional AR(1) model (Figure 1A). In the AR_{AICc} correction, the random activations also largely decreased and the t-maps were similar to the AR(5) case. In the power spectra of GLM residuals (Figure 1B), the data without correction and with AR(1) correction produced more power at the low frequencies (< 0.02 Hz) than at high frequencies. In contrast, the power distributions were approximately uniform in the data of AR(5) and AR_{AICc}, so the noise in these two cases can be considered as white noise. These findings suggest that noise SC can be effectively suppressed by using the high-order AR(p) or the AR_{AICc} algorithm.

The FPC curves with assumed block task design at TR = 500 ms are illustrated in Figure 2. In all the subjects, the FPC without the prewhitening largely deviated from the theoretical curve (Figure 2A). The FPC after the AR(1) correction reduced the deviation, but there were still relatively high overestimates of T-statistics (Figure 2B). When using the AR(5) (Figure 2C) and AR_{AICc} (Figure 2D), the FPC curves agreed well with the theoretical case. At $\alpha = 0.001$, the FPRs of the AR(5) and AR_{AICc} averaged over the subjects were about $(1.1 \pm 0.3) \times 10^{-3}$ and $(1.0 \pm 0.1) \times 10^{-3}$ (mean ± SD), respectively. Thus, they both provided enough accurate FPC at the typical α range (5×10^{-4}-5×10^{-2}) used in fMRI, but the AR_{AICc} showed smaller inter-subject variations on the correction accuracy. Thus, the AR_{AICc} had more reobust performance than the fixed AR(p) model. A similar result was also observed with the event-related task regressor (not shown).

The subject-averaged FPC curves (Figure 3) indicate that the performance of fixed AR(p) varied with the choice of AR order. The overestimates of t-scores were reduced

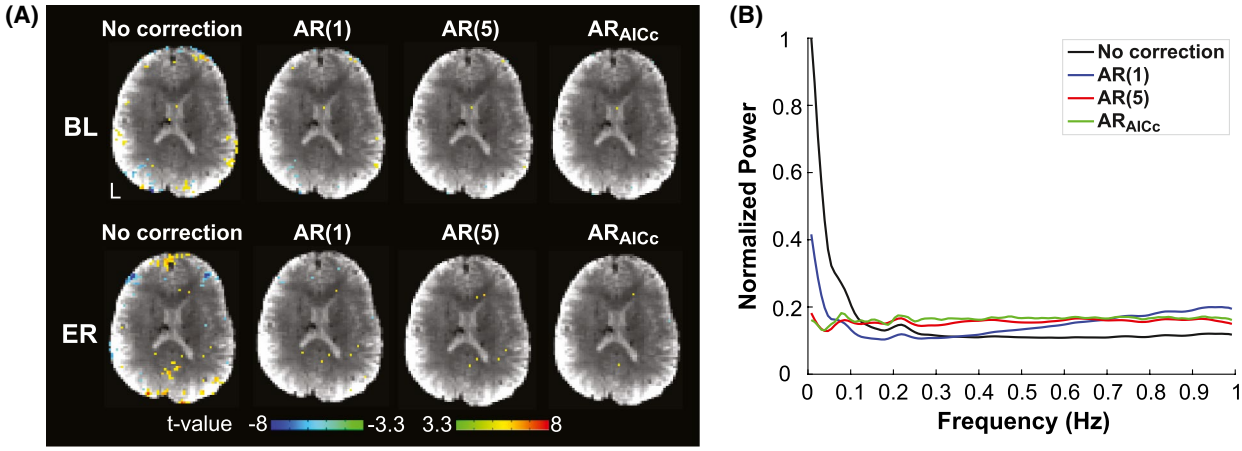

FIGURE 1 t-maps ($P < .001$ uncorrected) (A) and noise power spectra (B) from an exemplar subject with and without the AR(p)-based SC correction at TR = 500 ms. The corrected t-maps were separately obtained using the fixed AR(p) model with $p = 1$ (AR(1)) and 5 (AR(5)) and the improved AR(p) algorithm based on the AICc order selection (AR_{AICc}). The top and bottom rows in (A) show the result when assuming the block (BL) and event-related (ER) task designs, respectively. "L" indicates the left side of brain. The noise power spectra were normalized to the maximum peak of the power spectrum without the correction

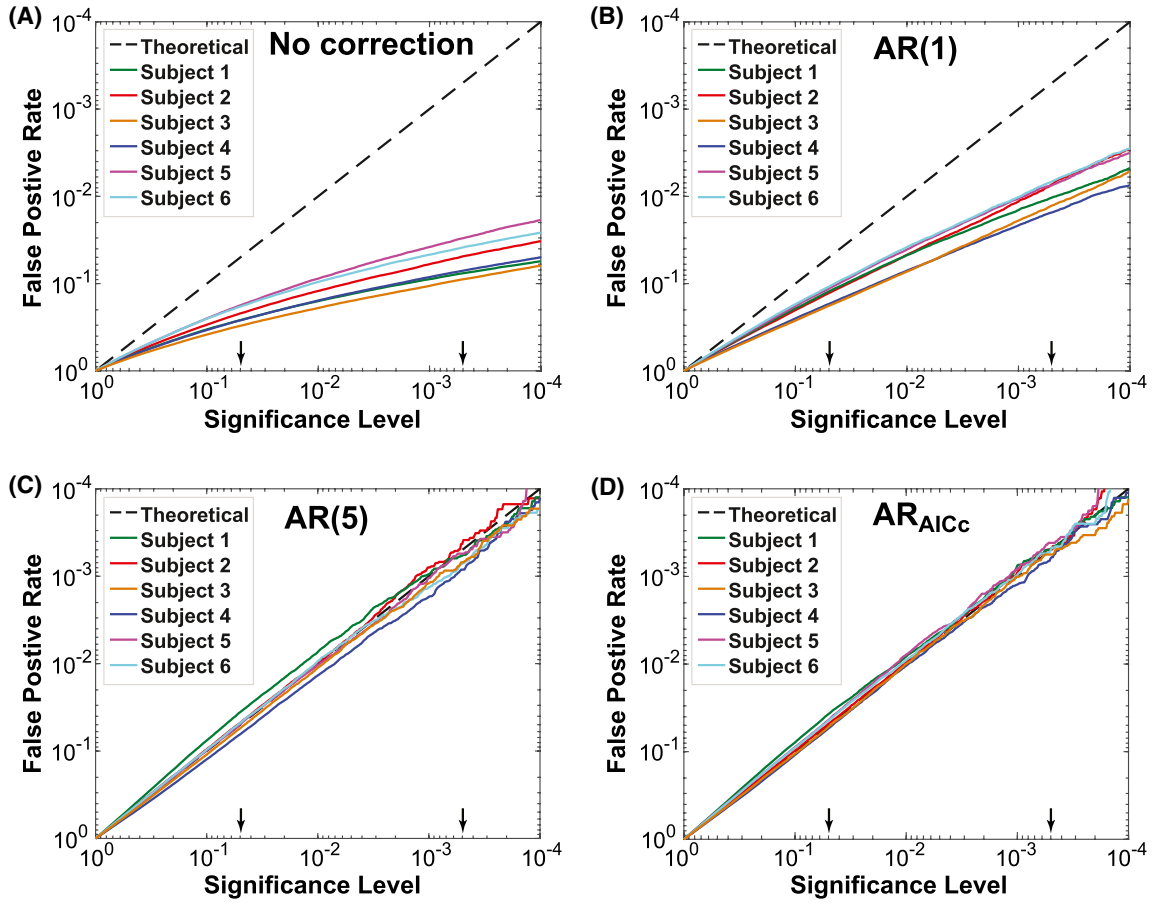

FIGURE 2 FPC curves of null data measured in the six subjects without correction (A) and corrected with the AR(1) (B), AR(5) (C), and AR_{AICc} (D) models. The arrows in (A) indicate the specificity range typically used in fMRI data analysis. The dashed line indicates the theoretical (true) FPC curve. If the measured data points are above/below the theoretical curve, it means that the *t*-values are underestimated/overestimated. The FPC curves were calculated by assuming the presence of the block task

with increasing p until the highest SC correction accuracy was achieved at the optimal order value. Further increase of p resulted in *t*-score overestimates again. In addition, the optimal AR order increased with a decrease in TR. The best performance was achieved approximately with $p = 5$ and 10 at TR = 500 and 300 ms, respectively. In contrast, the AR_{AICc} method was able to automatically and adaptively optimize the AR orders in individual voxels at the different TRs (Figure 4). It precisely removed the serial correlations in the data acquired with TR = 300 and 500 ms (Figure 3).

Figure 5 illustrates the impact of autocorrelation regularization on the performance of SC correction at TR = 500 ms. When smoothing the sample autocorrelations with FWHM = 6 and 15 mm, the FPC curves after the prewhitening deviated from the theoretical curve at all the orders. This phenomenon also appeared in the data acquired with TR = 300 ms. When the AR orders were selected with the BIC and pACF criteria, the BIC and pACF based AR(p) models showed similar SC correction accuracies and better performances than the traditional AR(1) model. However, the FPRs in the *t*-maps using the BIC and pACF were still higher than the theoretical values (Figure 6).

Figure 7 shows the performance of AR(p)-based prewhitening in the spatially smoothed fMRI data. When the GLM analysis was performed without the SC correction, the simulated BOLD activation was detected in the *t*-maps thresholded with the cluster extent, but various remote brain areas also showed significant clusters (ie, false positives). Although the AR(1)-based SC correction reduced the number of false positives, there were still some clusters outside the true activation region. In the *t*-maps obtained from the fixed AR(p) (with the optimal order) and the AR_{AICc} methods without autocorrelation regularization, no clusters above the threshold were found outside the ROI of BOLD signal. A few false positives were observed in the activation maps when smoothing the autocorrelations with FWHM = 6 and 8 mm (Figure 8). Thus, the AR(p)-based prewhitening without autocorrelation regularization was able to effectively remove SC in the smoothed fMRI images.

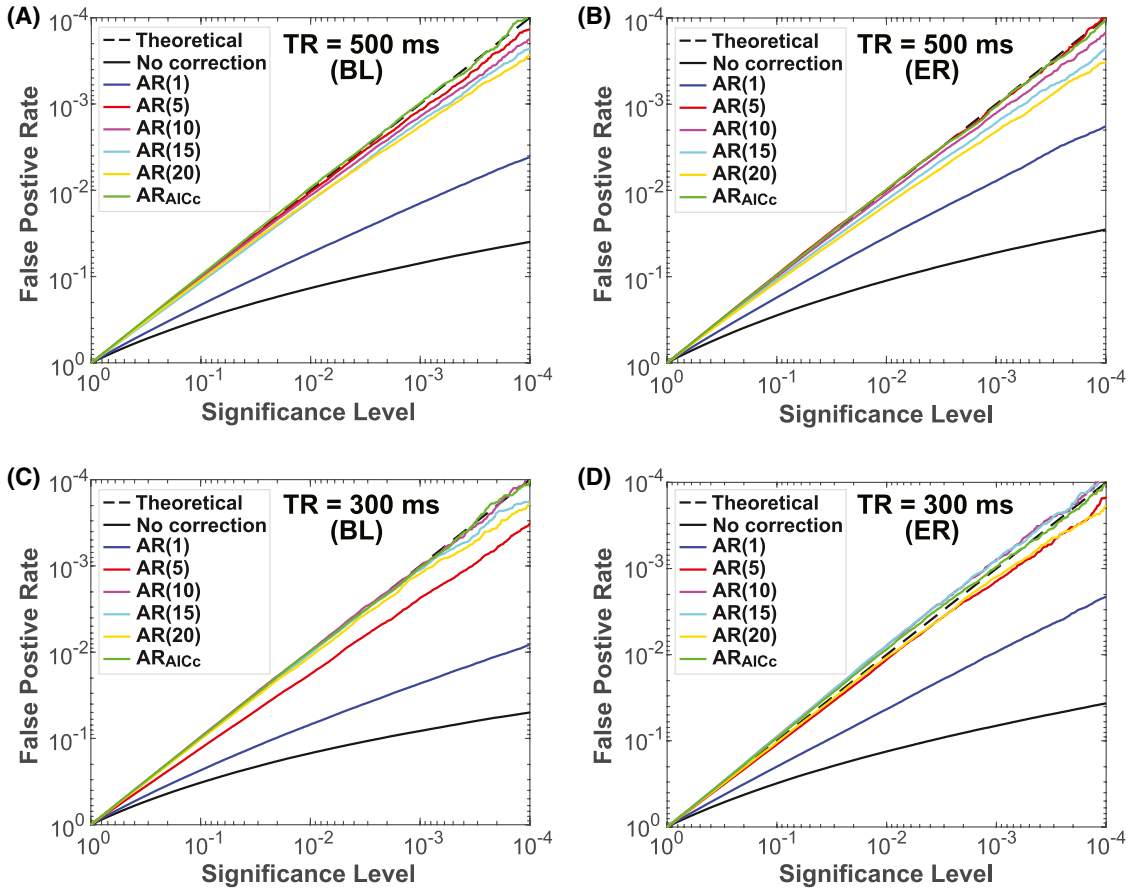

FIGURE 3 FPC curves averaged over the subjects when using the fixed AR(p) and the AR_{AICc} models for block (BL) and event-related designs (ER). Different model order values (p: 1, 5, 10, 15 and 20) were chosen in the fixed AR(p). (A) and (C) illustrate the results when assuming the block design at TR = 500 and 300 ms, and the cases of event-related design are shown in (B) and (D). No autocorrelation regularization was applied in the SC correction

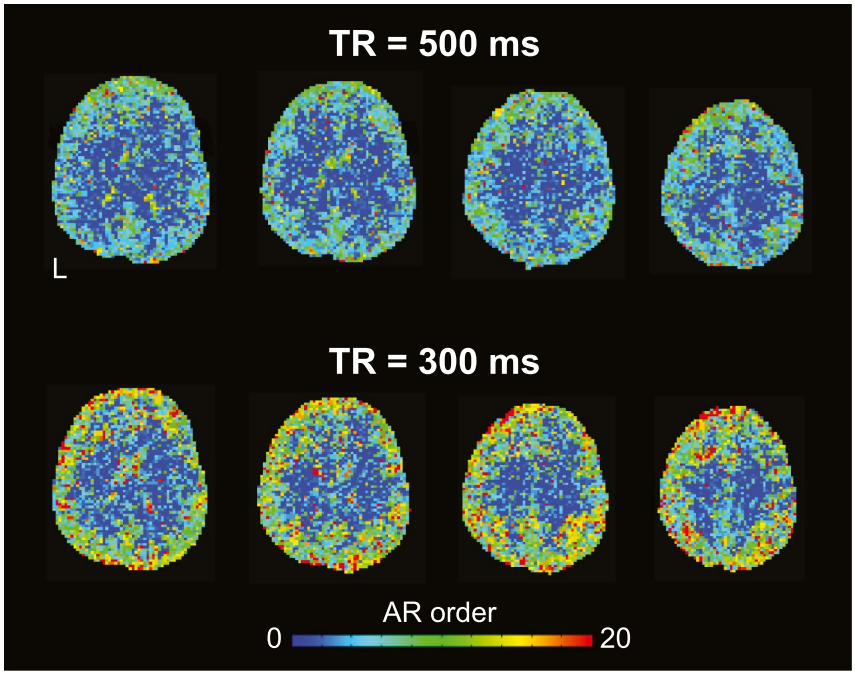

FIGURE 4 Maps of AR orders selected with the AICc at TR = 500 ms (top row) and 300 ms (bottom row) in the exemplar subject. "L" indicates the left side of brain

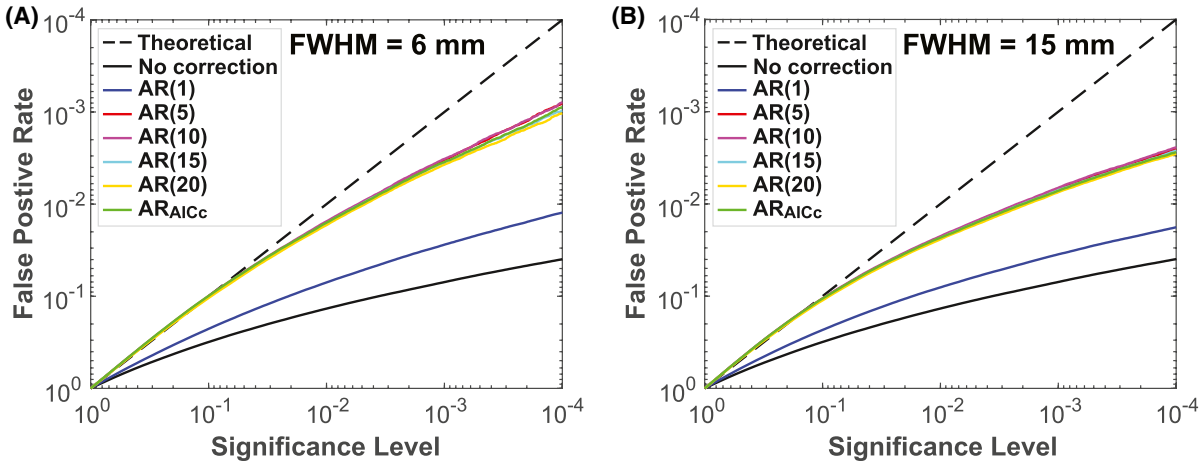

FIGURE 5 Subject averaged FPC curves at TR = 500 ms when regularizing the sample autocorrelations with spatial filters of FWHM = 6 mm (A) and 15 mm (B). The block task design was used to calculate the curves

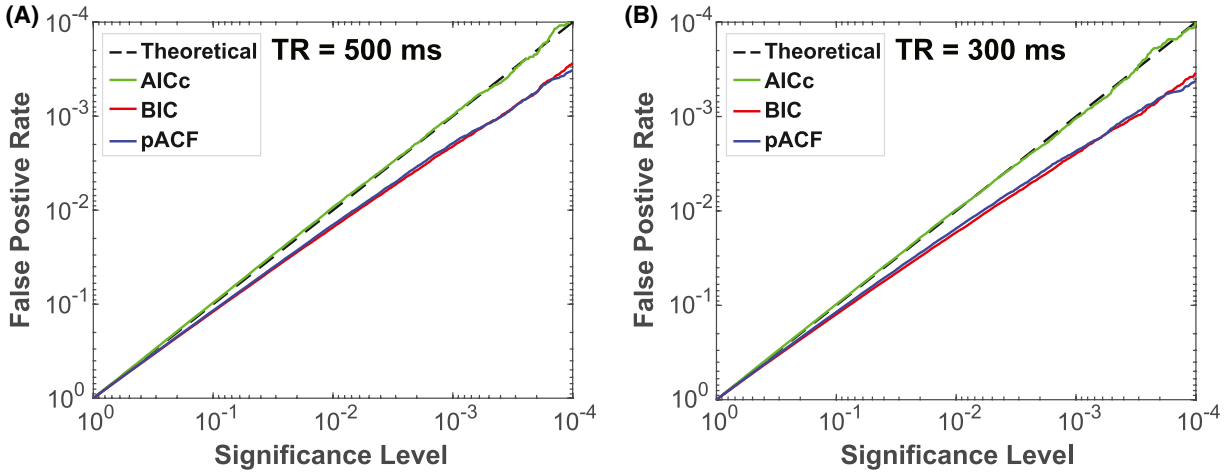

FIGURE 6 Subject averaged FPC curves at TR = 500 (A) and 300 ms (B) when using the AICc, BIC, and pACF order selection methods in the AR(p) correction. No autocorrelation regularization was applied in the correction and the block task design was assumed to be present in the null data.

4 | DISCUSSION

In this study, we improved the AR(p)-based prewhitening method and evaluated the performance of AR(p) on noise serial correlation correction in the fast fMRI data. The proposed AICc-based AR(p) algorithm (AR$_{AICc}$) allowed for automatic detection of optimal AR orders in voxel-wise at different TRs. By measuring the FPC in null data, we demonstrated that the GLM analysis with the AR$_{AICc}$ correction was able to suppress the SC effect from the fast fMRI images and produce accurate T-statistics.

Similar to a previous study,[5] we found that the performance of fixed AR(p) model[18] was affected by the choice of model order (Figure 3). At TR = 500/300 ms, the highest SC correction accuracy was achieved by setting the order to approximately 5/10. The selection of a lower (underfitting) or higher (overfitting) order than the optimal value resulted in a larger discrepancy between the measured and theoretical FPC curves. In addition, it is noted that the AR(p) model slightly underestimated the t-scores in a few of datasets, for example, the data at TR = 500 ms in Subject 1 after the AR(5) prewhitening (Figure 2C). Since motion or other artifacts could be present in the fMRI images, the assumption of AR(p) on the stationarity of data could not be completely satisfied. In such instances, the SC correction may produce smaller T-statistics than the true value. This phenomenon has also been observed in functional near-infrared spectroscopy studies.[34] Also, although t-value underestimation has no impact on the validity of statistical inference,[12] it reduces the detection sensitivity of the fMRI signal and the ability to identify weak neural activations. Hence, the optimal order must be choosen in the fixed AR(p) model.

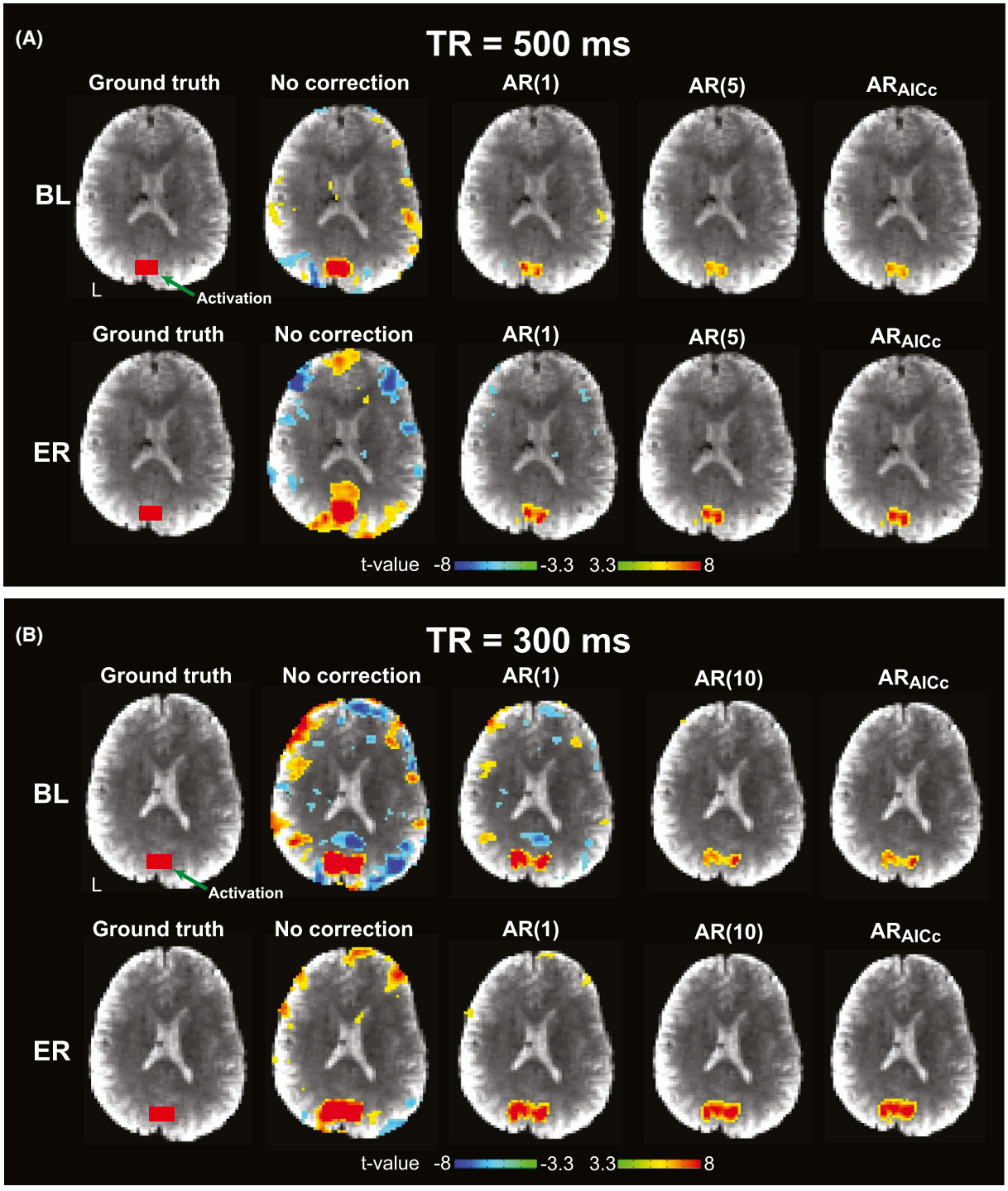

FIGURE 7 Activation maps of the examplar subject obtained from the synthesized fMRI data with the spatial image smoothing at TR = 500 ms (A) and TR = 300 ms (B). The red ROI in the ground truth images indicates the true fMRI activation foci, that is, the brain regions where the pseudo BOLD signal was injected. The SC effect was corrected with the AR(p)-based prewhitening and without the autocorrelation regularization. AR(5) and AR(10) were the optimal fixed AR(p) models at TR = 500 and 300 ms, respectively. The top/bottom row in (A) and (B) shows the results of block (BL)/event-related (ER)design. The t-maps were thresholded with $P < .001$ uncorrected and the multiple-comparison was corrected with the cluster size corresponding to FPR of 0.05

Our study showed that the optimal order in the fixed AR(p) model was higher at the shorter TR, and this finding agrees with previous reports.[5,13] In fMRI time series, the time interval between two adjacent data points is TR. The increase of AR order at shorter TRs suggests that a more complex AR model is required to describe the fMRI noise spectrum sampled with a higher frequency. When the data is acquired with a higher temporal resolution, more features

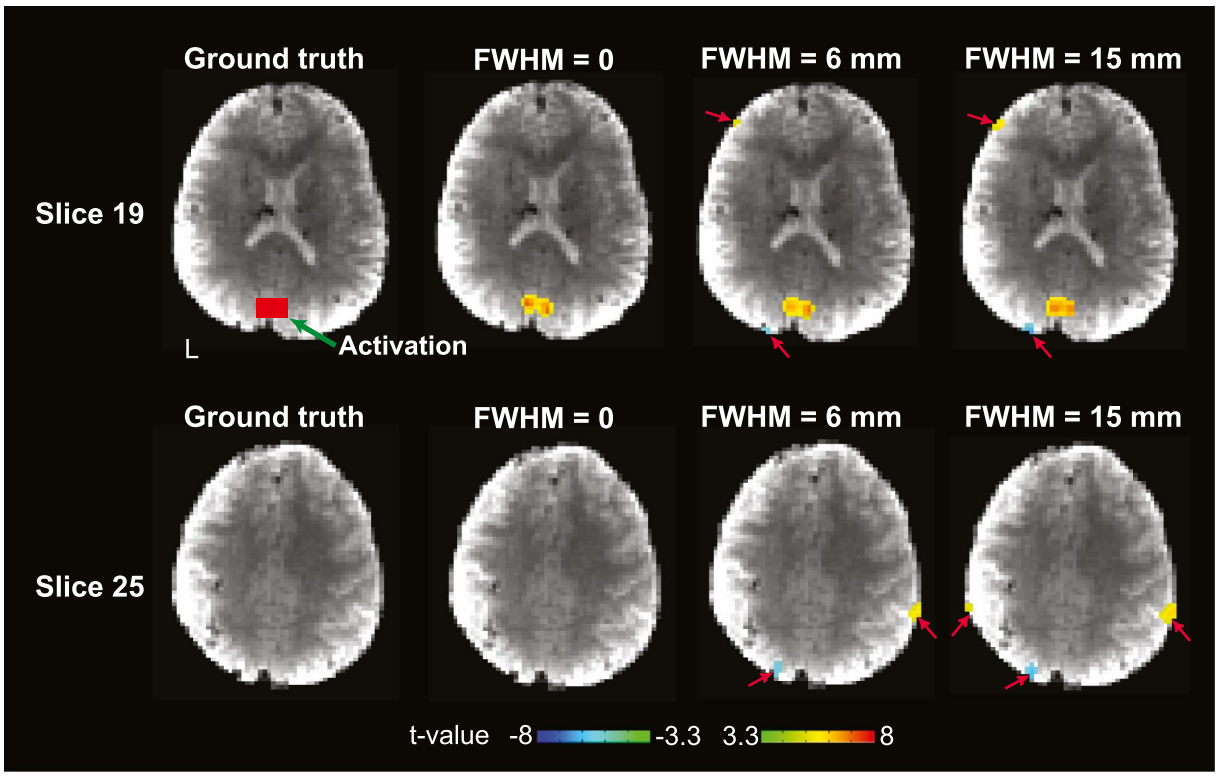

FIGURE 8 The activation maps of block design from the synthesized data with (FWHM = 6 and 15 mm) and without (FWHM = 0) the autocorrelation regularization. The red arrows indicate the false positives, which are remote to the true activation areas. The thresholded *t*-maps (cluster corrected FWE < 0.05) of two image slices (Slice 19 and Slice 25) are shown in the figure. The true BOLD activation area was located in Slice 19 but not in Slice 25. The results were obtained from the data corrected with the AR_{AICc} method

are resolved in its power spectrum. Thus, it is necessary to select a larger order value to allow higher model complexity and to better fit the spectral components.[35] In addition to the TR, other imaging parameters may also affect the optimal model order. In the prewhitening, the AR(p) is essentially used for fitting the temporal ACF of noise,[18] so optimal p is determined by the temporal/spectral property of noise. The SMS-EPI pulse sequence used in fast fMRI studies is the extension of parallel imaging techniques[36] in the slice direction.[37] Similar to parallel imaging, the noise property of SMS-EPI depends on the RF coil and the slice acceleration factor.[37] Also, fMRI noise includes three components: MR system noise, physiological noise, and thermal noise from the subject.[38] The ratio of physiological to thermal noise varies with the main magnetic field strength and imaging parameters such as the voxel size, echo time, and flip angle.[39-42] In addition, fMRI acquisitions with short TRs (TR < T_1 of cerebrospinal fluid) may cause incompletely spoiled spin coherence and temporal variations of MRI signal due to the disturbance of steady state.[43] Thus, if the MR hardware and pulse sequence parameters change significantly, the spectral property and ACF of noise may be altered accordingly, so the optimal fixed AR(p) model vary with experimental conditions.

The optimal model order must be known to achieve accurate SC corrections with the fixed AR(p) method. This requirement may limit its applications if prior knowledge about the optimal order is unavailable in some experimental conditions. The improved algorithm based on the AICc (AR_{AICc}) can overcome this disadvantage. Furthermore, as shown in Figure 2C,D, the subject-averaged correction accuracy of AR_{AICc} is comparable to that of the optimal fixed AR(p) model. However, the AR_{AICc} method yields smaller inter-subject variations in the FPC. For instance, the standard deviation of FPR across the subjects was about 0.0003 at $\alpha = 0.001$ in the AR(5), while the inter-subject variation decreased to 0.0001 in the AR_{AICc}. Therefore, the AR_{AICc} offers more convenience and more robust performance than the fixed AR(p) algorithm. On the other hand, the AR_{AICc} needs to additionally calculate the noise variance and Kullback-Leibler information at each voxel, which consumes more computation resources. In our study, it took approximatley one minute to process one dataset at TR = 500 ms with the AR(5), while the AR_{AICc} required about 2.1 min to finish the same data processing. Thus, the AR_{AICc} has lower computation efficiency.

If the autocorrelation matrix of noise (V) (Equation 2 in Supporting Information) is exactly known, the GLM with

prewhitening would provide accurate statistical test results (eg, T and F statistics),[18] and the measured FPC would perfectly match the theoretical curve (FPR = α).[10,14] However, in an fMRI experiment one can never know the true intrinsic correlation structure of noise,[12] so an accurate estimate of V is critical for the prewhitening approach.[10] Since the raw autocorrelations estimated with the cross-correlation method (Equations 8, 9 in Supporting Information Text S1) are noisy at large lags, the serial correlations cannot be effectively suppressed if the raw autocorrelations are directly used to construct V. The aim of modeling the noise with AR(p) is to more accurately estimate the noise ACF. Since the ACF calculated from the AR model varies with the model order and the corresponding coefficients,[12] it is necessary to select the optimal AR order to achieve the best accuracy in the estimation of ACF. AR models with lower (underfitting) or higher (overfitting) orders than the optimal value can produce larger deviations between the estimated V and the true V. Obviously, it is impossible to preselect the true model orders for individual image voxels because their intrinsic noise structures are unknown. Based on different assumptions about the statistical characteristics of noise, many different criteria have been proposed for the prediction of true AR order,[44,45] and the AICc is widely used one. According to the assumed noise properties, the AICc determines the optimal choice of model order by balancing the fitting error and overfitting. The present study shows that the AR(p) model selected the AICc provides high SC correction accuracy. This implies that the AICc method is able to accurately estimate the noise autocorrelation matrix V in fast fMRI data.

Due to the prediction errors, the selected AR models are not always the true noise models. Suboptimal order selections may occur in some cases when applying the criterion to a large number of time series, even in simulated realizations generated with known AR(p) models.[21] The human brain usually includes more than 100 000 image voxels in a fast fMRI scan, so the AICc or other criteria could select lower or higher orders than the true values at some voxels. However, it is important to note that the occurrence of suboptimal models is conceptually different from the multiple comparison (multiple testing) problem in fMRI data analysis. The aim of AR model selection is to enable noise after the prewhitening to approximately obey the assumed distribution (ie, i.i.d.). Although a suboptimal model results in imperfect prewhitening at a voxel, the estimated t-value will only slightly deviate from its true value if the discrepancy between the selected and true AR models is small. In such a case, suboptimal order selection will not produce a false activation at this voxel. The results from this study illustrate that the FPR after the AR_{AICc} correction agrees well with the theoretical value. This suggests that the selected AR models are enough close to the true fMRI noise structures and the occurrence of suboptimal order selection does not significantly affect the accuracy of SC correction.

In Worsley's study,[18] spatial smoothing is applied to regularize the sample autocorrelations at individual voxels. The aim of autocorrelation regularization is to reduce the variability of sample autocorrelation in conventional fMRI scans with ~100 time frames ($N \sim 100$).[18] However, the results from this study suggest that the regularization step can degrade the accuracy of SC correction. Note that a typical fast fMRI scan usually includes more than 600 image volumes ($N > 600$). The standard deviation of the sample autocorrelation from its true value is about $1/\sqrt{N}$,[18] so the variability of sample autocorrelation in fast fMRI is negligible compared to that in the conventional fMRI. In addition, although the regularization can decrease the variation of autocorrelation, it may lead to an increase in bias if the noise ACF varies substantially in the neighboring voxels, especially in the boundary areas between different tissue types or brain regions. Due to the inhomogeneous coil sensitivity, the SMS-EPI acquisition technique used in fast fMRI can further augment the nonuniformity in the spatial distribution of noise. As shown in Figure 9, large differnces may exist in the ACF of adjacent voxels, so the autocorrelation smoothing can decrease the accuracy of ACF estimation and SC correction. Our analysis results indicate that the AR(p) algorithm can produce enough accurate statistical maps without smoothing the sample autocorrelations. Hence, the autocorrelation regularization step should be avoided in fast fMRI.

In this study, the performance of AR(p) model was separately evaluated in the fMRI images with and without spatial smoothing. For the unsmoothed data, the FPC curves were calculated from the t-maps and compared to the theoretical FPC. The quality of SC correction in the smoothed data was assessed with an empirical method, which inspected whether false positives were present in

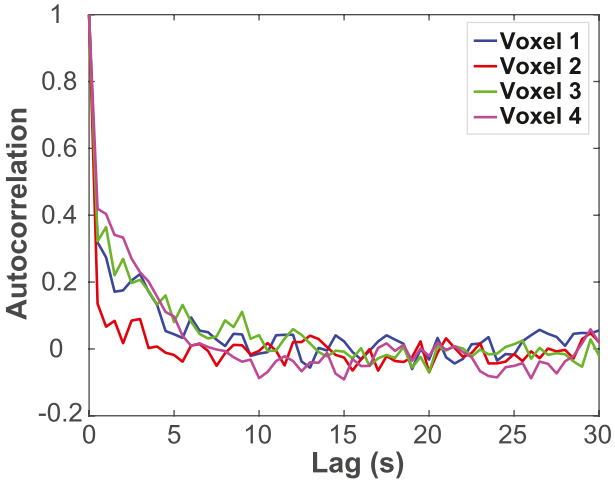

FIGURE 9 Autocorrelation curves in four neighboring voxels in the gray matter of one subject

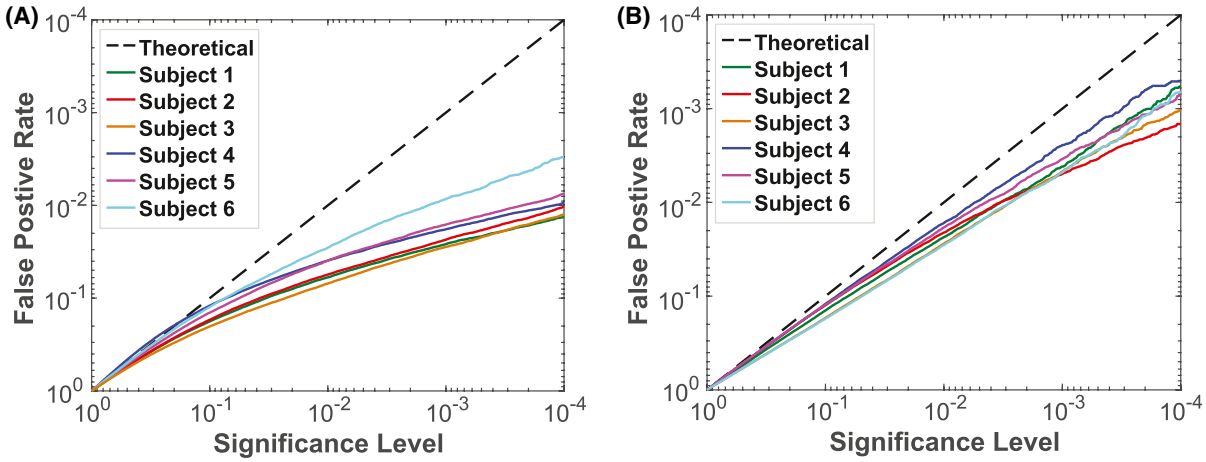

FIGURE 10 FPC curves of individual subjects when using the prewhitening algorithms in SPM based on the "FAST" model (A) and the ARMA(1, 1) model in AFNI (B). The default setting on the q value ($q = 6$) in SPM was used to construct the FAST model. The FPC curves were measured with the assumed block task design from the data of TR = 500 ms

the activation maps of individual subjects. Both the unsmoothed[10,14] and smoothed[5,13] data based criteria have been employed in the studies on SC correction. Compared to the visual inspection of false positives, the FPC can provide more quantitative information about the performance of an SC correction algorithm. However, the FPC criterion cannot be directly applied to smoothed fMRI data because the image smoothing leads to high correlations between the time courses of neighbor voxels. This violates the assumption of FPC criterion about the independence between voxels.[14] In addition, the familywise error (FWE) at a p-threshold corrected for the multiple-comparison has been used to evaluate the accuracy of statistical inference in smoothed data.[46] Theoretically the FWE method can also be used in this study to examine the performance of prewhitening in smoothed data. Nonetheless, the FWE is defined as the ratio of the number of datasets showing false positives to the total number of datasets, so a large number of fMRI datasets (~ 1000) is required to calculate the FWE.[46] Thus, the FWE method is inappropriate for our study due to the limited number of subjects and datasets (six datasets at each TR). In other words, we have chosen the the two best criteria available for evaluating the accuracy of prewhitening, and the AR_{AICc} algorithm showed high performance under the both criteria. Therefore, AR_{AICc} is a robust SC removal method for fast fMRI data analysis.

In addition to the AR(p) model, other advanced algorithms have also been implemented in some fMRI data analysis software to remove noise serial correlations. For example, SPM and AFNI provide prewhitening algorithms respectively based on the "FAST" and ARMA(1,1) models, and these SC correction methods have been applied in fast fMRI studies.[16,17,47] We conducted the statistical analysis on the same datasets using SPM and AFNI. In SPM's FAST model, a dictionary of covariance components of length $3q$ is constructed with q different exponential time constants.[16] We separately set $q = 2, 4, 6$ (default value), 8, 10 (maximum allowed value) in SPM to cover the range of selections of q. The FPC curves in individual subjects when using the FAST model with the different q values are shown in Figure 10A (default q value) and Supporting Information Figure S1 (the other q values), and Figure 10B illustrates the results from the ARMA(1,1) model. It can be seen that the deviations of FPC from the theoretical curve are larger in the FAST and ARMA(1,1) models than in the fixed AR(p) model (with the optimal order) (Figure 2C) and in the AR_{AICc} (Figure 2D). Thus, the AR(p)-based prewhitening algorithm provides higher performance.

In conclusion, the AR(p)-based prewhitening framework provides an accurate method for correcting serial correlations in fMRI images acquired with a TR < 1 s. The fixed AR(p) algorithm[18] with the optimal model order and the improved algorithm (AR_{AICc}) provide comparable SC correction accuracies on average, but the AR_{AICc} shows less inter-subject variations in the performance. Thus, we recommend using the AR_{AICc} algorithm in the data analysis of fast fMRI, considering its convenience and higher reliability. The fixed AR(p) method could be the favorable choice only if the optimal order is known and the computation efficiency is a crucial factor to be considered in a study.

ACKNOWLEDGMENTS

This work has been supported by Laureate Institute for Brain Research, the William K. Warren Foundation, and by National Institute of General Medical Sciences, National Institutes of Health Award P20GM121312. The funding sources has no influence in study design, the data collection, analysis, data interpretations, in the writing of the manuscript, and the decision to submit the article for publication.

DATA AVAILABILITY STATEMENT

The MATLAB code used for data analysis is available at https://github.com/librtulsa/ARp. As an example, the fMRI images of one subject in this study can be downloaded at this link: https://www.dropbox.com/s/4bm2wjvjvineh7n/example.zip?dl=0. The other subjects' data will be available upon direct request to the corresponding author.

ORCID

Jerzy Bodurka https://orcid.org/0000-0003-0053-9746

REFERENCES

1. Nunes R, Hajnal JV, Golay X, Larkman DJ. Simultaneous slice excitation and reconstruction for single shot EPI. The Annual Meeting of ISMRM; 2006; Seattle. p. 293.
2. Moeller S, Yacoub E, Olman CA, et al. Multiband multislice GE-EPI at 7 Tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magn Reson Med*. 2010;63:1144-1153.
3. Feinberg DA, Moeller S, Smith SM, et al. Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. *PLoS ONE*. 2010;5:e15710.
4. Chen L, Vu AT, Xu J, et al. Evaluation of highly accelerated simultaneous multi-slice EPI for fMRI. *NeuroImage*. 2015;104:452-459.
5. Sahib AK, Mathiak K, Erb M, et al. Effect of temporal resolution and serial autocorrelations in event-related functional MRI. *Magn Reson Med*. 2016;76:1805-1813.
6. Friston KJ, Jezzard PJ, Turner R. Analysis of functional MRI time-series. *Hum Brain Mapp*. 1994;1:153-171.
7. Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging: a general linear approach. *Hum Brain Mapp*. 1995;2:189-210.
8. Worsley KJ, Friston KJ. Analysis of fMRI time-series revisited—again. *NeuroImage*. 1995;2:173-181.
9. Bullmore E, Brammer M, Williams SC, et al. Statistical methods of estimation and inference for functional MR image analysis. *Magn Reson Med*. 1996;35:261-277.
10. Woolrich MW, Ripley BD, Brady M, Smith SM. Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage*. 2001;14:1370-1386.
11. Kiebel SJ, Holmes AP. Chapter 8 - the general linear model. In: Friston K, Ashburner J, Kiebel S, Nichols T, Penny W, eds. *Statistical parametric mapping: the analysis of funtional brain images*. Amsterdam; Boston: Elsevier/Academic Press; 2007.
12. Friston KJ, Josephs O, Zarahn E, Holmes AP, Rouquette S, Poline J. To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. *NeuroImage*. 2000;12:196-208.
13. Bollmann S, Puckett AM, Cunnington R, Barth M. Serial correlations in single-subject fMRI with sub-second TR. *NeuroImage*. 2018;166:152-166.
14. Purdon PL, Weisskoff R. Effect of temporal autocorrelations due to physiological noise stimulus paradigm on voxel-level false positive rates in fMRI. *Hum Brain Mapp*. 1998;6:239-249.
15. Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res*. 1996;29:162-173.
16. Corbin N, Todd N, Friston KJ, Callaghan MF. Accurate modeling of temporal correlations in rapidly sampled fMRI time series. *Hum Brain Mapp*. 2018;39:3884-3897.
17. Chen JE, Polimeni JR, Bollmann S, Glover GH. On the analysis of rapidly sampled fMRI data. *NeuroImage*. 2019;188:807-820.
18. Worsley KJ, Liao CH, Aston J, et al. A general statistical analysis for fMRI data. *NeuroImage*. 2002;15:1-15.
19. de Waele S, Broersen PMT. Order selection for vector autoregressive models. *IEEE Trans Signal Processing*. 2003;51:427-433.
20. Penny W, Kiebel S, Friston K. Variational Bayesian inference for fMRI time series. *NeuroImage*. 2003;19:727-741.
21. Hurvich CM, Tsai CL. Regression and time series model selection in small samples. *Biometrika*. 1989;76:297-307.
22. Setsompop K, Gagoski BA, Polimeni JR, Witzel T, Wedeen VJ, Wald LL. Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magn Reson Med*. 2012;67:1210-1224.
23. Glover G, Li TQ, Rees D. Image-based method for retrospective correction of physiological motions effects in fMRI: RETROICOR. *Magn Reson Med*. 2000;44:162-167.
24. Chang C, Glover GH. Effects of model-based physiological noise correction on default mode network anti-correlations and correlations. *NeuroImage*. 2009;47:1448-1459.
25. Chang C, Cunningham JP, Glover GH. Influence of heart rate on the BOLD signal: the cardiac response function. *NeuroImage*. 2009;44:857-869.
26. Birn RM, Smith MA, Jones TB, Bandettini PA. The respiration response function: the temporal dynamics of fMRI signal fluctuations related to changes inrespiration. *NeuroImage*. 2008;40:644-654.
27. Ohtsu K, Peng H, Kitagawa G. *Time series modeling for analysis and control: advanced autopilot and monitoring systems*. Tokyo: Springer; 2015.
28. Box GEP, Jenkins GM. *Time series analysis: forecasting and control*. San Francisco, CA: Holden-Day; 1976.
29. Schwartz G. Estimating the dimension of a model. *Ann Stat*. 1978;6:461-464.
30. Glover GH. Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*. 1999;9:416-429.
31. Fox MD, Raichle ME. Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat Rev Neurosci*. 2007;8:700-711.
32. Eklund A, Knutsson H, Nichols TE. Cluster failure revisited: impact of first level design and physiological noise on cluster false positive rates. *Hum Brain Mapp*. 2019;40:2017-2032.
33. Cox RW, Chen G, Glen DR, Reynolds RC, Taylor PA. FMRI clustering in AFNI: false-positive rates redux. *Brain Connect*. 2017;7:152-171.
34. Barker JW, Aarabi A, Huppert TJ. Autoregressive model based algorithm for correcting motion and serially correlated errors in fNIRS. *Biomed Opt Express*. 2013;4:1366-1379.
35. McFarland DJ, Wolpaw JR. Sensorimotor rhythm-based brain-computer interface (BCI): model order selection for autoregressive spectral analysis. *J Neural Eng*. 2008;5:155-162.
36. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. *Magn Reson Med*. 1999;42:952-962.
37. Larkman DJ, Hajnal JV, Herlihy AH, Coutts GA, Young IR, Ehnholm G. Use of multicoil arrays for separation of signal from

multiple slices simultaneously excited. *J Magn Reson Imaging*. 2001;13:313-317.
38. Kruger G, Glover GH. Physiological noise in oxygenation-sensitive magnetic resonance imaging. *Magn Reson Med*. 2001; 46:631-637.
39. Kruger G, Kastrup A, Glover GH. Neuroimaging at 1.5 T and 3.0 T: comparison of oxygenation-sensitive magnetic resonance imaging. *Magn Reson Med*. 2001;45:595-604.
40. Triantafyllou C, Hoge RD, Krueger G, et al. Comparison of physiological noise at 1.5 T, 3 T and 7 T and optimization of fMRI acquisition parameters. *NeuroImage*. 2005;26:243-250.
41. Bodurka J, Ye F, Petridou N, Murphy K, Bandettini PA. Mapping the MRI voxel volume in which thermal noise matches physiological noise-implications for fMRI. *NeuroImage*. 2007;34: 542-549.
42. Gonzalez-Castillo J, Roopchansingh V, Bandettini PA, Bodurka J. Physiological noise effects on the flip angle selection in BOLD fMRI. *NeuroImage*. 2011;54:2764-2778.
43. Zhao X, Bodurka J, Jesmanowicz A, Li SJ. B(0)-fluctuation-induced temporal variation in EPI image series due to the disturbance of steady-state free precession. *Magn Reson Med*. 2000;44:758-765.
44. Broersen PMT. The ABC of autoregressive order selection criteria. *IFAC Proc*. 1997;30:245-250.
45. Stoica P, Selen Y. Model-order selection: a review of information criterion rules. *IEEE Signal Process Mag*. 2004;21:36-47.
46. Eklund A, Nichols TE, Knutsson H. Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proc Natl Acad Sci USA*. 2016;113:7900-7905.
47. Olszowy W, Aston J, Rua C, Williams GB. Accurate autocorrelation modeling substantially improves fMRI reliability. *Nat Commun*. 2019;10:1220.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

FIGURE S1 FPC curves in individual subjects obtained with SPM's FAST model. (A), (B), (C), and (D) illustrate the results when respectively setting the q value = 2, 4, 8, and 10
TEXT S1 Supporting text and equations

How to cite this article: Luo Q, Misaki M, Mulyana B, Wong C-K, Bodurka J. Improved autoregressive model for correction of noise serial correlation in fast fMRI. *Magn Reson Med*. 2020;84:1293–1305. https://doi.org/10.1002/mrm.28203