Automatic, Fast and Robust Characterization of Noise Distributions for Diffusion MRI

Samuel St-Jean[(✉)] , Alberto De Luca , Max A. Viergever,
and Alexander Leemans

Center for Image Sciences, University Medical Center Utrecht,
Utrecht, The Netherlands
`samuel@isi.uu.nl`

Abstract. Knowledge of the noise distribution in magnitude diffusion MRI images is the centerpiece to quantify uncertainties arising from the acquisition process. The use of parallel imaging methods, the number of receiver coils and imaging filters applied by the scanner, amongst other factors, dictate the resulting signal distribution. Accurate estimation beyond textbook Rician or noncentral chi distributions often requires information about the acquisition process (e.g.coils sensitivity maps or reconstruction coefficients), which is not usually available. We introduce a new method where a change of variable naturally gives rise to a particular form of the gamma distribution for background signals. The first moments and maximum likelihood estimators of this gamma distribution explicitly depend on the number of coils, making it possible to estimate all unknown parameters using only the magnitude data. A rejection step is used to make the method automatic and robust to artifacts. Experiments on synthetic datasets show that the proposed method can reliably estimate both the degrees of freedom and the standard deviation. The worst case errors range from below 2% (spatially uniform noise) to approximately 10% (spatially variable noise). Repeated acquisitions of *in vivo* datasets show that the estimated parameters are stable and have lower variances than compared methods.

1 Introduction

Diffusion magnetic resonance imaging (dMRI) is a non invasive imaging technique which allows probing the microstructure of the brain. Recent advances in parallel imaging techniques and accelerated acquisitions have greatly reduced the inherently long scan time in dMRI. While it is known that the noise distribution found in magnitude dMRI data depends on the reconstruction algorithm used [1] and the number of channels in the receiver coils [2,3], noise correlation effects in adjacent channels change the noise distribution from its theoretical formulation. Assumption of the Rician or more general noncentral chi distributions with degrees of freedom equal to the number of receiver coils deviate due to these effects and other filtering applied by the scanner. The resulting

distribution usually exhibits a lower number of degrees of freedom N than the number of receiver coils and higher noise variance σ_g depending on the spatial location [2,4]. Correcting deviations from the theoretical noise distributions is challenging and oftentimes requires coils correlation maps or information about the complex signal combination process, which is not readily available on most scanners. While some recent algorithms for dMRI are developed to include information about the noise distribution [5,6], there is no method, to the best of our knowledge, providing a fully automatic way to characterize the noise distribution using information from the magnitude data itself only. Due to this gap between the physical acquisition process and noise estimation theory, noise distributions are either assumed as Rician or noncentral chi with N already known and concentrate in estimating the noise standard deviation σ_g [7–9]. We propose to estimate both σ_g and N from the magnitude data only by using a change of variable to a gamma distribution $\Gamma(N,1)$ [8], whose first moments directly depend on N. This makes the proposed method fast and easy to apply on existing data without additional information, while being robust to artifacts by only considering voxels adhering to the created gamma distribution.

2   Theory and Methods

Signal Distributions in Parallel MRI. To account for uncertainty in the acquisition process, the complex signal measured in k-space by the receiver coil can be modeled with a separate additive zero mean Gaussian noise for each channel, but assumed to have identical variance σ_g^2. When converted to the commonly used magnitude images, the resulting noise distribution follows a Rician or noncentral chi distribution, whose parameters depend on the employed reconstruction algorithm [2]. To account for signal correlations introduced by parallel imaging techniques, the case of the noncentral chi distribution is still valid with spatially varying parameters [4].

Parameter Estimation Using the Method of Moments and Maximum Likelihood. When the underlying signal intensity η is zero, the magnitude signal m reduces to a Rayleigh distribution or in the general case to a chi distribution. The pdf of magnitude noise over zero signal is given by $pdf(m|\eta = 0, \sigma_g, N) = (m^{2N-1})/(2^{N-1}\sigma_g^{2N}\Gamma(N)) \exp\left(-m^2/(2\sigma_g^2)\right)dm$ where $\Gamma(x)$ is the gamma function. With the change of variable $t = m^2/(2\sigma_g^2)$, the pdf can be rewritten as a gamma distribution $\Gamma(N,1)$ [8]. The pdf of the gamma distribution $\Gamma(\alpha,\beta)$ is defined as $pdf(t|\alpha,\beta) = 1/(\Gamma(\alpha)\beta^\alpha)t^{\alpha-1}\exp\left(-t/\beta\right)dt$ and has theoretical mean $\mu_{gamma} = \alpha\beta$ and variance $\sigma_{gamma}^2 = \alpha\beta^2$. For a gamma distribution $\Gamma(N,1)$, we obtain that the mean and the variance are equal with a value of N. Another useful identity is that the sum of gamma distributions is a gamma distribution such that if $t_i \sim \Gamma(\alpha_i,\beta)$, then $\sum_{i=1}^K t_i \sim \Gamma(\sum_{i=1}^K \alpha_i, \beta)$. We can therefore estimate the Gaussian noise standard deviation σ_g and the number of coils N from the moments of the magnitude image themselves where no signal from the imaged object is present. Any method suitable for computing σ_g can be used, or it can also be estimated from the moments once again with the relationship

$$\sigma_g = \frac{1}{\sqrt{2}} \sqrt{\frac{\sum_{k=1}^{K} m_k^4}{\sum_{k=1}^{K} m_k^2} - \frac{1}{K} \sum_{k=1}^{K} m_k^2} \tag{1}$$

where m_k is the magnitude signal for voxel k and K is the number of identified noise only voxels. Once σ_g is known, N can be estimated from the moments with

$$N = \frac{1}{K} \sum_{k=1}^{K} t_k = \frac{1}{2K\sigma_g^2} \sum_{k=1}^{K} m_k^2 \tag{2}$$

where $t_k = m_k^2/(2\sigma_g^2)$ is the change of variable for voxel k. Estimation based on the method of maximum likelihood yields two equations for estimating α and β. Rearranging the equations for a gamma distribution $\Gamma(N,1)$ [10] will give the same expression as Eq. (1) and a second implicit equation for N that is given by

$$\psi(N) = \frac{1}{K} \sum_{k=1}^{K} \log(m_k^2/2\sigma_g^2) \tag{3}$$

where $\psi(x)$ is the digamma function and can be numerically inverted using Newton's method to obtain N.

Estimating σ_g and N. For simplicity, we assume that each 2D slice with the same spatial location belongs to the same distribution throughout each 3D volume. This practical assumption allows selecting a large number of noise only voxels for computing statistics as well as discarding acquisition artifacts such as ghosting. Following a methodology similar to [8], it is possible to identify voxels belonging to the gamma distribution by checking if they fall inside a predefined probability threshold of the inverse cumulative distribution function (cdf). Taking the sum of all MRI volumes can therefore be used to separate the background signal belonging to the gamma distribution $\Gamma(KN,1)$ from the rest of the volume with a rejection step using the inverse cdf. In the particular case $\Gamma(\alpha,1)$ at a probability level p, the inverse cdf is $icdf(\alpha,p) = P^{-1}(\alpha,p)$ where P^{-1} is the inverse lower incomplete regularized gamma function. For the first iteration, initial bounds are set on the value of N and σ_g as they are unknown. We set a lower bound $N_{min} = 1$ and an upper bound $N_{max} = 12$ for the first iteration, noting that [2] reported values of N between 3 and 12 for a 32 channels receiver coil. Similar to [8], an upper bound of σ_g is given by $\sigma_{g_{max}} = median/\sqrt{2\,icdf(N_{max},1/2)}$ where $median$ is the median of the whole 4D dMRI dataset. From this upper bound $\sigma_{g_{max}}$, a search interval with a values is created, where we chose $a = 50$ as in [8]. Each point of the interval $\Phi = [1\sigma_{g_{max}}/a, 2\sigma_{g_{max}}/a, \ldots, a\sigma_{g_{max}}/a]$ is used as an initial value of σ_g in the change of variable $t = m^2/2\sigma_g^2$. With these initial values, an iterative search for σ_g and N is made as follow. The value of Φ which identifies the largest number of voxels between the lower bound given by $\lambda_- = icdf(KN_{min},p/2)$ and the upper bound given by $\lambda_+ = icdf(KN_{max},1-p/2)$ is accepted as σ_g. From those voxels, new values of σ_g are computed with Eq. (1) and N with Eq. (2) or Eq. (3). For

the next iteration, we set $\Phi = [0.95\sigma_g, 0.96\sigma_g, \ldots, 1.05\sigma_g]$ and recompute the *icdf* bounds λ_-, λ_+ with the new value of N. Voxels between λ_- and λ_+ belong to the distribution $\Gamma(KN, 1)$ and are recomputed until the values of σ_g and N reach convergence.

Synthetic Phantom Datasets. We generated synthetic datasets based on the ISBI 2013 HARDI challenge[1] with phantomas[2]. Two noiseless single shell phantoms with 64 gradient directions were generated at b = 1000 s/mm^2 and b = 3000 s/mm^2 with one b = 0 s/mm^2 each. The datasets were then corrupted with Rician ($N = 1$) and noncentral chi noise ($N = 4, 8$ and 12), both stationary and spatially varying, at a signal-to-noise ratio (SNR) of 30. The noisy data was generated according to $\hat{I} = \sqrt{\sum_{i=0,j=0}^{N} \left(\frac{I}{\sqrt{N}} + \tau\epsilon_i \right)^2 + \tau\epsilon_j^2}$, where $\hat{I}$ is the resulting noisy volume ϵ_i, ϵ_j are Gaussian distributed with mean 0 and variance $\sigma_g^2 = (mean(b0)/SNR)^2$. In the constant noise case, τ is set to 1 so that the noise is uniform. For the spatially varying noise case, τ is a sphere with a value of 1 in the center up to a value of 1.75 at the edges of the phantom, thus generating a stronger noise profile *outside* the phantom than for the stationary (constant) noise case. This noise profile mimics receiver coils disposed around the surface of the phantom, with an increase in the noise profile near each receiver. One important observation arising from choosing a single SNR level is that the noise standard deviation σ_g is the same for all datasets, while the magnitude standard deviation σ_{m_N} depends on the value of N and we have $\sigma_{m_N} < \sigma_g$.

In Vivo Datasets. We obtained four repetitions of a freely available dMRI dataset of a single subject[3] to assess the reproducibility of noise estimation without *a priori* knowledge. The acquisition was performed on a GE MR750 3T scanner at Stanford university, where a 3x slice acceleration with blipped-CAIPI shift of FOV/3 was used, partial Fourier 5/8 and a minimum TE of 81 ms. Two acquisitions were made in the anterior-posterior phase encoded direction and the two others in the posterior-anterior direction. The voxelsize was 1.7 mm isotropic with 7 b = 0 s/mm^2 images, 38 volumes at b = 1500 s/mm^2 and 38 volumes at b = 3000 s/mm^2.

Noise Estimation Algorithms for Comparison. To assess the performance of the proposed method, we used three other noise estimation algorithms [7–9] previously used in the context of diffusion MRI with their default parameters. The local adaptive noise estimation (LANE) algorithm [9] was designed to estimate the noise standard deviation over tissue for both Rician and noncentral chi noise while also taking into account the structure of the data for adaptive estimation. Since the method works on a 3D volume, we only used the b = 0 s/mm^2 image for all of the experiments as the signal does not vary spatially for the

---

[1] http://hardi.epfl.ch/static/events/2013_ISBI/.
[2] https://github.com/ecaruyer/phantomas.
[3] https://openfmri.org/dataset/ds000031.

same type of tissue in such image. We used the Marchenko-Pastur (MP) distribution fitting on the principal component analysis (PCA) decomposition of the diffusion data [7]. MPPCA estimates the magnitude noise standard deviation σ_{m_N} in small local windows by finding an optimal threshold in PCA space which separates the signal from the noise. This value of σ_{m_N} is slightly underestimated due to the discrete nature of the PCA decomposition. Finally, we compared our proposed method with the Probabilistic Identification and Estimation of Noise (PIESNO) [8], which originally proposed the change of variable to the gamma distribution that is at the core of our proposed method. PIESNO requires the value of N, which is used to iteratively estimate σ_g until convergence by removing voxels which do not belong to the distribution $\Gamma(N, 1)$ for a given slice. For our proposed algorithm, we set the probability level at $p = 0.05$ and initial values of $a = 50$, $N_{min} = 1$ and $N_{max} = 12$. To the best of our knowledge, ours is the first method which makes it possible to estimate both σ_g and N jointly without requiring any information about the reconstruction process of the MRI scanner. Finally, we quantitatively assessed the performance of each method on the synthetic datasets by measuring the percentage error inside the phantom against the known value of σ_g, where the error is computed as $percentage\,error = 100(\sigma_{g_{estimated}} - \sigma_{g_{true}})/\sigma_{g_{true}}$. We also show the estimated values of N using our method for each dataset.

3   Results

As MPPCA and LANE are designed to estimate σ_g over data, we report the estimation error computed only inside a mask excluding the background for all methods. Figure 1 shows the percentage error of each method on the synthetic datasets. The correct value of N was given to both LANE and PIESNO when σ_g was constant. All methods performed generally well, with our proposed method and PIESNO making less than 2% of errors for all cases. MPPCA and LANE commit larger errors (around 5% and 20% on average respectively) with increasing values of N, where LANE error is the largest when $N = 12$. For the case of spatially varying σ_g, we assumed N to be unknown and set $N = 1$ for LANE and PIESNO. Due to a misspecification of N, PIESNO errors are several orders of magnitude larger than the other methods except for the Rician noise case. MPPCA and LANE both underestimate σ_g (around 20% and between 10 to 15% respectively) while our proposed method resulted in the lowest error, which is around 10%. Figure 2 shows the estimated values of N by the proposed method for all cases of the synthetic datasets. Even when σ_g is underestimated, values of N are close to the real value. Estimating N using Eq. (2) or Eq. (3) gave similar results in both cases, so we used Eq. (2) in the present work. As limited information is available for the *in vivo* datasets, we assumed a Rician distribution for LANE and set $N = 1$ as suggested by [9]. For PIESNO, setting a Rician distribution with $N = 1$ returned less than 10 voxels identified per datasets. We instead assumed $N = 0.5$ since it corresponds to a half Gaussian distribution [2], which is the closest theoretical distribution estimated by our method.
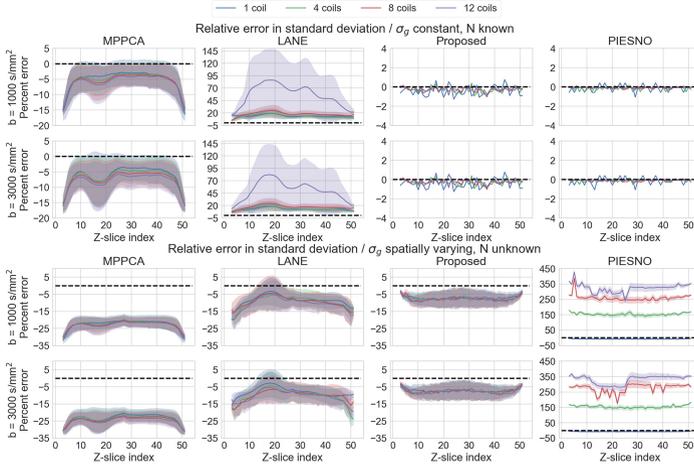
Fig. 1. Percentage of error in estimating the noise standard deviation for each slice along the Z axis with the mean (solid line) and standard deviation (shaded area). In the top image, $\sigma_g = 171$ is constant and N is known while in the bottom image σ_g varies spatially and N is unknown or assumed Rician distributed.
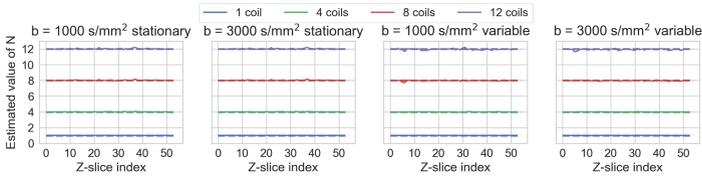

Fig. 2. Estimated value of N by the proposed method. Even for the spatially variable case where σ_g is slightly underestimated, the estimated values of N are stable and correspond to the real values used in the synthetic simulations in every case.

Figure 3 shows the mean (and standard deviation) value of σ_g on the *in vivo* datasets for each methods along axial slices. The value of N as computed by our proposed method is also reported and is stable across datasets. All methods recovered average stable values of σ_g on the four repetitions of the same subject. However, LANE recovered the highest values of σ_g amongst all methods with a large variance, which might indicate overestimation in some areas. Figure 4 shows an axial slice around the cerebellum corrupted by acquisition artifacts likely due to parallel imaging. Voxels containing artifacts were automatically discarded by our method. The values of N and σ_g computed from these voxels also offer a better qualitative fit than assuming a Rayleigh distribution or selecting non brain data. We also timed each method to estimate σ_g on one of the *in vivo* datasets using a standard desktop computer with a 3.5 GHz Intel Xeon processor. All methods were multi threaded while PIESNO was only single threaded. The runtime to estimate σ_g (and N) was around 10 secs for our proposed method, 11 secs for PIESNO, 3 min for MPPCA and 18 min for LANE.
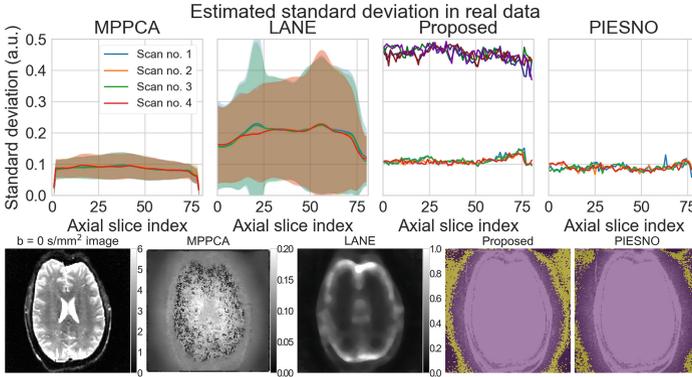
Fig. 3. At the top, estimated values of σ_g for the 4 *in vivo* datasets. For the proposed method, estimated values of N are shown in darker hues for each dataset. On the bottom, an axial slice of a b = 0 s/mm^2 image from one dataset and the estimated values of σ_g for MPPCA and LANE. For the proposed method and PIESNO, a mask of the identified background voxels (in yellow) overlaid on the data.
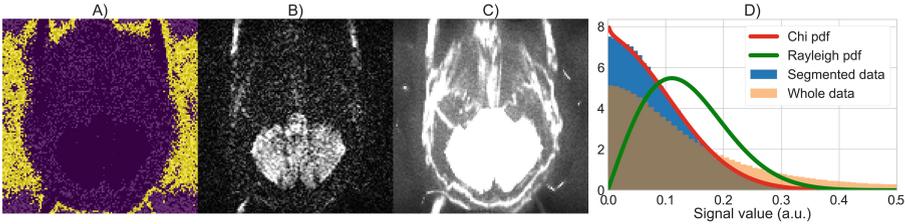

Fig. 4. An axial slice in the cerebellum from one of the *in vivo* datasets. Voxels identified in (A) as noise only (yellow) are free of artifacts in a single slice in (B) or along the sum of all volumes in (C). In (D), the normalized density histogram using the selected voxels from (A) (blue) fits well a chi distribution with $N = 0.47$ and $\sigma_g = 0.11$, while assuming a Rayleigh distribution (green) or using all non brain voxels (orange) leads to a worse visual fit.

4   Discussion and Conclusion

We have shown how a change of variable to a gamma distribution $\Gamma(N, 1)$ can be used to robustly and automatically identify background voxels. Once identified, the moments and maximum likelihood equations (Eqs. (1)–(3)) of the gamma distribution can be used iteratively to compute the number of degrees of freedom N and the Gaussian noise standard deviation σ_g relating to the original noise distribution. The presented equations are also fast to compute (around 10 s on *in vivo* data). Results on the synthetic datasets show that we can reliably estimate both parameters from the magnitude data itself. While the method we have presented assumes that each 2D slice contains a single noise distribution, N can be computed reliably on spatially varying noise and σ_g with an error

between 5 and 10%, which is less than the compared methods. On the *in vivo* datasets, our method is stable across the four repetitions and can automatically discard voxels corrupted by acquisition artifacts due to parallel acceleration. From the identified background voxels, without any specific assumption, the recovered distribution parameters fit well the histogram of the data. This distribution is close to a half Gaussian distribution ($N = 0.5$) while the Rician noise assumption would not be adequate in this case. Our method is also the first to identify any type of noise distribution from the magnitude data itself without requiring external information about the scanner or the reconstruction process. Interestingly, while we have shown results on dMRI datasets, the theory we presented applies to any other MRI weighting using large samples of magnitude data e.g.functional MRI. If measurements from the scanner without any object signals are acquired (i.e.noise maps), a local window estimation of our proposed method could be used to overcome the shortcoming of assuming stationary 2D noise distributions. Noise maps measurements could also be used for cases such as body or cardiac imaging where background voxels are usually not available in large quantities. Automatic identification of the noise distribution parameters could help multicenter studies which may not currently collect information about the acquisition and reconstruction process [4] or methods harmonizing data between different scanners and acquisition protocols [11]. Our method can also be used to provide prior knowledge beyond the textbook Rician distribution when computing local diffusion models [5,6].

References

1. Brown, R.W., et al.: Magnetic resonance imaging: physical principles and sequence design. Wiley, New York (2014)
2. Dietrich, O., et al.: Influence of multichannel combination, parallel imaging and other reconstruction techniques on MRI noise characteristics. Magn. Reson. Imaging **26**(6), 754–762 (2008)
3. Sotiropoulos, S.N., et al.: Effects of image reconstruction on fiber orientation mapping from multichannel diffusion MRI: reducing the noise floor using SENSE. Magn. Reson. Med. **70**(6), 1682–1689 (2013)
4. Aja-Fernández, S., Vegas-Sánchez-Ferrero, G., Tristán-Vega, A.: Noise estimation in parallel MRI: GRAPPA and SENSE. Magn. Reson. Imaging **32**(3), 281–290 (2014)
5. Collier, Q., et al.: Diffusion kurtosis imaging with free water elimination: a Bayesian estimation approach. Magn. Reson. Med. **80**(2), 802–813 (2018)
6. Sakaie, K., Lowe, M.: Retrospective correction of bias in diffusion tensor imaging arising from coil combination mode. Mag. Reson. Imaging **37**, 203–208 (2017)
7. Veraart, J., Fieremans, E., Novikov, D.S.: Diffusion MRI noise mapping using random matrix theory. Magn. Reson. Med. **76**(5), 1582–1593 (2016)
8. Koay, C.G., Özarslan, E., Pierpaoli, C.: Probabilistic identification and estimation of noise (PIESNO): a self-consistent approach and its applications in MRI. J. Magn. Reson. **199**(1), 94–103 (2009)
9. Tabelow, K., Voss, H.U., Polzehl, J.: Local estimation of the noise level in MRI using structural adaptation. Med. Image Anal. **20**(1), 76–86 (2015)

10. Thom, H.C.S.: A note on the gamma distribution. Mon. Weather Rev. **86**(4), 117–122 (1958)
11. Mirzaalian, H., et al.: Inter-site and inter-scanner diffusion MRI data harmonization. NeuroImage **135**, 311–323 (2016)