

Article

## Source Localization by Entropic Inference and Backward Renormalization Group Priors

Nestor Caticha

Departamento de Física Geral, Instituto de Física, Universidade de São Paulo, CP 66318, 05315-970, São Paulo-SP, Brazil; E-Mail: nestor@if.usp.br

Academic Editor: Raúl Alcaraz Martínez

Received: 3 February 2015 / Accepted: 16 April 2015 / Published: 23 April 2015

---

**Abstract:** A systematic method of transferring information from coarser to finer resolution based on renormalization group (RG) transformations is introduced. It permits building informative priors in finer scales from posteriors in coarser scales since, under some conditions, RG transformations in the space of hyperparameters can be inverted. These priors are updated using renormalized data into posteriors by Maximum Entropy. The resulting inference method, backward RG (BRG) priors, is tested by doing simulations of a functional magnetic resonance imaging (fMRI) experiment. Its results are compared with a Bayesian approach working in the finest available resolution. Using BRG priors sources can be partially identified even when signal to noise ratio levels are up to  $\sim -25$ dB improving vastly on the single step Bayesian approach. For low levels of noise the BRG prior is not an improvement over the single scale Bayesian method. Analysis of the histograms of hyperparameters can show how to distinguish if the method is failing, due to very high levels of noise, or if the identification of the sources is, at least partially possible.

**Keywords:** inverse problems; maximum entropy; prior distribution; renormalization group; fMRI; EEG

---

### 1. Introduction

This paper presents a method of constructing priors for spatially distributed systems which draws from ideas of the renormalization group (RG) [1] and can be used to improve the performance of maximum entropy or Bayesian inference based methods. Both RG and multigrid methods [2] separate

the computational effort into separate simpler problems at different scales. However, they are quite different in the implementation. The multigrid methods deal with distributions of probability of variables at one scale, conditional on variables in another scale. The RG deals with the marginal distribution of the variables in one scale, after the joint distributions is integrated over the degrees of freedom in a finer resolution scale.

While the method is general and can be adapted to several source inference scenarios, a particular application was chosen to exemplify its power. A simulation of functional Magnetic Resonance imaging is presented, showing that for high noise levels it is significantly better than a pure Bayes approach working in the finest available resolution. Earlier versions of these ideas, without the Entropic inference and RG justifications, were used to design what were then dubbed multigrid priors used for fMRI [3]. Applications to identify active cortical sources using magneto or electroencephalography (M/EEG) and the information connectome using EEG will be published elsewhere.

Our general method of inference proceeds by iteration over a set of lattices of increasing spatial resolution. The method will at several stages assign probability distribution to groups of variables representing the sources. To start consider the same prior uniform distribution at every region of the coarsest lattice. That is, uniform distribution within a region, and the same distribution for all regions. Two steps, Entropic Inference and Backward Renormalization Group, are iterated. Using the data, Maximum Relative Entropy updates to a posterior. A backward renormalization group transformation generates a prior on the next level of spatial resolution. These steps are repeated until the last lattice. Since the small resolution stages have only a few degrees of freedom, the inverse problem inference is easy. As the dimension of the problem increases the solution is still fast because the now very informative prior permits a very good starting point.

This paper is organized as follows. In Section 2 the theory of Backward Renormalization Group priors is presented. In Section 3 the renormalization group transformation and the Entropic inference steps are calculated for a simple model of fMRI in a task-driven paradigm. The results of simulations under different conditions of noise corruption are presented in Section 4. Averaging solutions obtained by shifting the lattice's origin leads to an algorithm that can identify activity isles at signal to noise levels of less than  $-20$  dB. The mechanisms that lead to the clear separation of signal from noise in the construction of the inferred images can be identified permitting the introduction of a criterion to decide when the noise is so large that no reliable identification of active regions is possible.

## 2. Methods

### 2.1. Overview of the Method

In a spatially extended system a source  $x(\mathbf{r})$  gives rise to a measurable quantity  $y(\mathbf{r}')$ . We restrict the description to a discretized spatial lattice and consider that measurements subject to noise  $\xi$  are modeled by a linear matrix model  $y = Gx + \xi$ . For  $M$  sensors,  $T$  the length of the time series of measurements,  $N$  the number of sites of the lattice where  $x$  lives, the dimensions of the matrices  $y$ ,  $x$ ,  $G$  and  $\xi$  are respectively  $M \times T$ ,  $N \times T$ ,  $M \times N$  and  $M \times T$ . In the case of fMRI  $x$  may be taken as a variable that takes one of two values, active or inactive and  $G$  is the Bold hemodynamic response. For

EEG,  $x$  represents electric dipoles and  $G$  the Green function of the brain and the  $y$ 's are scalp voltage measurements. Call the determination of  $y$  given  $x$  the forward problem. The inverse problem is to determine the sources from the data. The representation of the space where the  $x$  variables live can be done at different levels of resolution, and a set of renormalized lattices  $\{\Lambda_d\}_{d=0,\dots,D}$  is considered. Each lattice is composed by sites  $r^d(i) \in \Lambda_d$  with  $i = 1, \dots, |\Lambda_d|$ . The particular form of how a coarser or renormalized lattice  $\Lambda_{d-1}$  is obtained from the previous  $\Lambda_d$  depends on the particular type of problem. At each scale  $d$  the set of source variables is collectively denoted  $\mathbf{x}_d = \{x_{d,i}\} = \{x_d(r^d(i))\}$  and the integration measure by  $d\mathbf{x}_d$  which includes the possibility of sums over discrete variables.

The method consists of three steps, called 1.a, 1.b and 2 respectively, to be iterated from the coarser to the finer scales, which are schematically described as follows. Choose a convenient parametric family of distributions, denoted  $Q(\mathbf{x}_d|\boldsymbol{\theta}_d)$ , where  $\boldsymbol{\theta}_d$  is the set of parameters for the entire lattice. For  $d = 0$  start with a prior  $Q(\mathbf{x}_0|\boldsymbol{\theta}_0^0)$ . It can be shown in simulations that this choice is not critical. Suppose now that the method has been applied and a distribution  $Q(\mathbf{x}_{d-1}|\boldsymbol{\theta}_{d-1}^0)$  is available. To update to the next level  $Q(\mathbf{x}_d|\boldsymbol{\theta}_d^0)$ , the first step 1.a uses the data to determine a posterior distribution  $P(\mathbf{x}_{d-1})$ . This requires the use of Maximum Relative Entropy methods and it might lead to a distribution not in the same parametric family as the original  $Q(\mathbf{x}_{d-1}|\boldsymbol{\theta}_{d-1}^0)$ . The step 1.b projects it back to the parametric space resulting in  $Q(\mathbf{x}_{d-1}|\boldsymbol{\theta}_{d-1}^f)$ , to be considered a posterior still in the same lattice, where parameters  $\boldsymbol{\theta}_{d-1}^f$  incorporate information from the data. This is also a Maximum Relative Entropy step, but with the reverse Kullback-Leibler divergence, since it leads back to the parametric space. The effect of these two steps is an update on the parameters at the same resolution:

$$(\boldsymbol{\theta}_{d-1}^0) \xrightarrow{\text{Learning}} (\boldsymbol{\theta}_{d-1}^f) \tag{1}$$

Step 2, called the Backward Renormalization Group, permits enlarging the space and leads to the required  $Q(\mathbf{x}_d|\boldsymbol{\theta}_d^0)$ :

$$(\boldsymbol{\theta}_{d-1}^f) \xrightarrow{\text{Backward RG}} (\boldsymbol{\theta}_d^0). \tag{2}$$

This completes the transition from prior at level  $d - 1$  to prior at level  $d$ . In the next paragraph these steps will be described in detail.

### 2.2. Details of the Method

Instead of looking at the distribution of probability at a single level of resolution at each point in the iterative process the state of knowledge will be represented by a distribution  $P(\mathbf{x}_d, \mathbf{x}_{d-1}, y)$ , to be obtained from a prior distribution  $Q(\mathbf{x}_d, \mathbf{x}_{d-1}, y)$ . The method of choice is the Maximum Relative Entropy (ME) because of two reasons: (i) when constraints are imposed and the prior already satisfies the constraints ME will result in a posterior equal to the prior and (ii) when the results of a measurement are considered as constraints, ME gives the same results as updating with Bayes rule [4]. The Bayesian usual update can be thought of as a special case of Maximum Entropy update.

To show (i) we perform a simple Maximum Entropy exercise: Consider a variable  $\zeta$  that takes values  $z$  and that  $Q(z)$  represents our prior state of knowledge. New information is obtained, e.g., that the

expected value  $\langle f(z) \rangle$  is known to have a particular value:  $\langle f(z) \rangle = E$ . The update from  $Q(z)$  to  $P(z)$  is done by maximizing

$$S[P||Q] = - \int P(z) \log \frac{P(z)}{Q(z)} d\mu(z) + \lambda \left( \int P(z) f(z) d\mu(z) - E \right) + \lambda_0 \left( \int P(z) d\mu(z) - 1 \right) \quad (3)$$

The result, after satisfying normalization is the Boltzmann-Gibbs probability density  $P(z) = Q(z) \frac{\exp \lambda f(z)}{Z(\lambda)}$ , with  $\lambda$  to be chosen to satisfy  $\langle f(z) \rangle_P = E$ . What is the value of  $\lambda$  if  $\langle f(z) \rangle_Q = E$ , if the prior already satisfied the constraint? It is simple to see that  $\lambda = 0$  and  $Z(\lambda) = 1$ , so the reuse of old data in the form of constraints, again and again doesn't change a density  $Q(z)$  that already satisfies the constraint. While this sounds trivial, it will be useful since data in Bayes updates can be written as constraints. But the reuse of data using Bayes rules should be avoided.

To prove condition (ii) we follow [4]. Consider the problem where a distribution  $P(x)$  has to be obtained, first, from the knowledge that a measurement of  $Y$  has yielded a datum  $y'$ ; second, that prior to the inclusion of such information our knowledge of  $x$  is codified by a distribution  $Q(x)$  and third, that the relation between  $Y$  and  $X$  is codified by a likelihood  $Q(y|x)$ . Thus we have to determine  $P(x, y)$  subject to constraints  $\int dx P(x, y) = P(y) = \delta(y - y')$ , by maximizing

$$S[P||Q] = - \int P(x, y) \log \frac{P(x, y)}{Q(x, y)} dy dx + \int \lambda(y) \left( \int P(x, y) dx - \delta(y - y') \right) dy + \lambda_0 \left( \int P(y) dy - 1 \right), \quad (4)$$

where the Lagrange multiplier is a function  $\lambda(y)$  since there is a constraint for each possible value of  $y$ .

From the resulting joint density  $P(x, y)$  follows the desired marginal  $P(x) = \int dy P(x, y)$ . The result is

$$P(x) = Q(x|y'), \quad (5)$$

where  $Q(x|y') = \frac{Q(x)Q(y'|x)}{Q(y')}$  is the Bayes posterior  $Q(x|y')$  given by Bayes theorem, which just follows from the rules of probability. This proves that Maximum Relative Entropy as an inference engine justifies the use of Bayes theorem as the update procedure when the constraint is a datum such as knowing that a measurement of  $Y$  turned out to give a value  $y'$ . Maximum Relative Entropy justifies that Bayes theorem should be used in the inference process.

Going back to the spatially extended problem, consider that the problem's solution has advanced up to a coarse level of resolution  $d - 1$  where the degrees of freedom are  $\mathbf{x}_{d-1}$ . Now the data  $y_d$  will be used to update the information about  $\mathbf{x}_d$ . Hence, and this is probably the most important assumption, the relevant space for inference is formed by the degrees of freedom at the two scales and the data at the level of description  $d$ :  $(\mathbf{x}_d, \mathbf{x}_{d-1}, y_d)$ .

Thus, we seek the maximization of

$$S[P||Q] = - \int P(\mathbf{x}_d, \mathbf{x}_{d-1}, y_d) \log \frac{P(\mathbf{x}_d, \mathbf{x}_{d-1}, y_d)}{Q(\mathbf{x}_d, \mathbf{x}_{d-1}, y_d)} d\mathbf{x}_d d\mathbf{x}_{d-1} dy_d \quad (6)$$

to obtain  $P(\mathbf{x}_d, \mathbf{x}_{d-1}, y_d)$ , which in addition to normalization, is subject to

(A) The marginal of  $y_d$  is known:  $P(y) = \delta(y_d - y'_d)$ , since the measured data is  $y'_d$ .

(B) The marginal prior  $Q(\mathbf{x}_{d-1})$  belongs to a parametric family and hence shall be written as  $f_{d-1}(\mathbf{x}_{d-1}|\boldsymbol{\theta}_{d-1}^f)$

(C)  $Q(\mathbf{x}_d|\mathbf{x}_{d-1})$  codes knowledge about the backward renormalization, the relation between the  $\mathbf{x}$  variables at different renormalization stages (see Section 3.1.)

(D) Given  $\mathbf{x}_d$ , knowledge of  $\mathbf{x}_{d-1}$  is irrelevant for  $y_d$ :  $Q(y_d|\mathbf{x}_d\mathbf{x}_{d-1}) = Q(y_d|\mathbf{x}_d)$ , since coarser information is unnecessary in the presence of finer information.

Marginalization and the product rule of probability give, for the prior

$$Q(\mathbf{x}_d) = \int f_{d-1}(\mathbf{x}_{d-1}|\boldsymbol{\theta}_{d-1}^f)Q(\mathbf{x}_d|\mathbf{x}_{d-1})d\mathbf{x}_{d-1}. \quad (7)$$

which can be calculated from (B) and (C). To return to the chosen parametric family, determine the closest member  $f_d(\mathbf{x}_d|\boldsymbol{\theta}_d^0)$  by obtaining the set of  $\boldsymbol{\theta}_d^0$  which maximizes the entropy  $S[Q(\mathbf{x}_d)||f_d(\mathbf{x}_d|\boldsymbol{\theta}_d^0)]$

$$f_d(\mathbf{x}_d|\boldsymbol{\theta}_d^0) = \text{maxarg } S[Q(\mathbf{x}_d)||f_d(\mathbf{x}_d|\boldsymbol{\theta}_d)], \quad (8)$$

which is solved by imposing the partial derivatives of the entropy with respect to the parameters are zero and leads to the new parameters being given in terms of moments of  $Q(\mathbf{x}_d)$ . This is essentially what is called a Variational Bayes step in [5] and is a mean field approximation. Equations (7) and (8) implement the mapping in Equation (2) and how to do this within the realm of the Renormalization Group is addressed in Section 3.1.

We have all the ingredients to go back to the problem of solving the maximization problem (Equation (6)). Taking the marginal of the maximum entropy distribution we obtain

$$P(\mathbf{x}_d) = \frac{f_d(\mathbf{x}_d|\boldsymbol{\theta}_d^0)Q(y'_d|\mathbf{x}_d)}{Q(y'_d)}, \quad (9)$$

which is given by what we would have expected: Bayes theorem is used to incorporate the information in the data  $y'_d$  through the use of the likelihood of the data  $Q(y'_d|\mathbf{x}_d)$ . But there is an extra and important ingredient brought in by Equations (7) and (8), that the prior at this new scale is obtained by whatever information we had on  $\mathbf{x}_{d-1}$ , *i.e.*  $f_{d-1}(\mathbf{x}_{d-1}|\boldsymbol{\theta}_{d-1}^f)$  and by the backward renormalization procedure  $Q(\mathbf{x}_d|\mathbf{x}_{d-1})$  that links the two scales.

Now the less dramatic point (1.b). Yet another maximum entropy projection leads to the final result of this scale resolution

$$f_d(\mathbf{x}_d|\boldsymbol{\theta}_d^f) = \text{maxarg } S[P(\mathbf{x}_d)||f_d(\mathbf{x}_d|\boldsymbol{\theta}_d)] \quad (10)$$

which implements the mapping of equation that permits starting all over again. This step might not be necessary if the renormalization and learning steps do not lead outside the  $f(\mathbf{x}|\boldsymbol{\theta})$  parametric family. Equations (9) and (10) implement the step in display (1).

Note that condition (ii) ensures that while we are using the measurements of  $y_d$  at each scale we are not reusing the data in a forbidden manner.

### 3. Results: an Application to fMRI

The high impact that the techniques of neuroimaging like fMRI (e.g., [6] for and overview and [7] for statistical analysis) have on current research justifies that this method should be applied in that context.

In the case of fMRI it is reasonable to consider an Ising spin variable  $x = s_{d,i}$ , which takes two values (1 for active,  $-1$  for inactive). It describes the state of activity of lattice site  $i$  when the system is represented in the lattice  $\Lambda_d$ . The probability of the activities is  $f_d(s_d|\theta_d)$ . The renormalized activity variables in a coarser grain lattice  $\Lambda_{d-1}$  are given by the renormalization group transformation  $s_{d-1,i} = R(\{s_{d,i}\}_{i \in I})$ . Different renormalization group transformation choices are possible, for example a majority rule for spins inside a block of sites  $i$  that gives rise to the coarser scale variable  $s_{d-1,I}$  or a Migdal Kadanoff dization. The renormalized parameters can be calculated from Equations (7) and (8). This is in general a difficult procedure but several approximate schemes are available.

### 3.1. Backward Migdal Kadanoff Renormalization

The direct RG transformation is

$$f_{d-1}(\{s_{d-1,I}\}|\theta_{d-1}) = \sum_{\{s_{d,i}\}_{i \in \Lambda_d}} f_d(\{s_{d,i}\}_{i \in I}|\theta_d) Q(s_{d-1,I}|\{s_{d,i}\}_{i \in I}), \tag{11}$$

$$f_{d-1}(\{s_{d-1,I}\}|\theta_{d-1}) = \sum_{\{s_{d,i}\}_{i \in \Lambda_d}} f_d(\{s_{d,i}\}_{i \in I}|\theta_d) \delta(s_{d-1,I} - R(\{s_{d,i}\}_{i \in I})), \tag{12}$$

a real space renormalization group mapping that leads from the finer scale parameters to the renormalized parameters. In general there are parameters  $\theta_d$  coupling spins in all sorts of manners. In order to proceed a simple parametric family has to be considered. We consider simply pair and single site interactions within  $2 \times 2$  plaquettes and the couplings are  $\theta = (K, h)$ . The real space renormalization group version of of Migdal [8] and Kadanoff [9] is the simplest possible but sufficient for purposes at hand. It leads to the evolution of the parameters [10]

$$K_{d-1}^f = \frac{1}{2} \ln \cosh(8K_d^0), \tag{13}$$

$$h_{d-1}^f = h_d^0(1 + \tanh(8K_d^0)). \tag{14}$$

To determine the values of the hyperparameters of the prior at the next level, under the assumption that the hyperparameters are uniform on a plaquette, this map can be inverted to produce the backward RG map (Equation (2) which is implemented by Equations (9) and (10)), linking the posterior at scale  $d - 1$  to the prior at scale  $d$ :

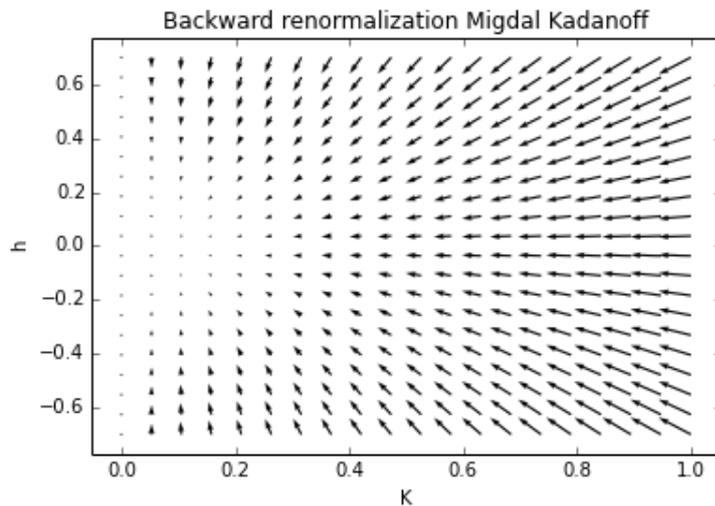
$$K_d^0 = \frac{1}{8} \cosh^{-1}(e^{2K_{d-1}^f}), \quad h_d^0 = \frac{h_{d-1}^f}{1 + \tanh(\cosh^{-1}(e^{2K_{d-1}^f}))}. \tag{15}$$

This coupling flow is shown in Figure 1. This means that if after learning at a coarse scale  $d - 1$  the knowledge is codified with a set of parameters  $\theta_{d-1}^f = (K_{d-1}^f, h_{d-1}^f)$  for a spin  $s_{d-1,I}$ , then the set  $\theta_d^0 = (K_d^0, h_d^0)$  will be used in the prior for the inference of the group of sources  $\{s_{d,i}\}_{i \in I}$  at the finer scale within block  $I$ . Even if the process is started with a spatially uniform set of  $\theta_0$ , the learning step will introduce variations that will reflect that the data  $V$  does not favor the update of probabilities of the sources at different places in the same from, and spatial differences will appear reflecting the construction of the inference. The plaquette distribution is

$$f_d(\{s_{di}\}|\theta) = \frac{1}{z_I} e^{K \sum_{i,j \in I} s_i s_j + h \sum_{i \in I} s_i}, \tag{16}$$

where the sum  $i, j \in I$  means that every pair inside the plaquette is included only once and  $i \neq j$ . The square plaquette partition function is  $z = 2 \exp(6K) \cosh(4h) + 8 \cosh(2h) + 6 \exp(-2K)$  and the local magnetization

$$m = \frac{1}{z} (2 \exp(6K) \sinh(4h) + 4 \sinh(2h)) \tag{17}$$



**Figure 1.** The Migdal Kadanoff Backward Renormalization Group flow.

### 3.2. The Learning Step

Once the parameters  $\theta_d^0 = (K_d^0, h_d^0)$  have been determined at a finer scale by Equations (15), data must be incorporated via Maximum Entropy, or equivalently through Bayes as shown by Equation (9) and the resulting posterior projected to the parametric space giving a new set  $\theta_d^f = (K_d^f, h_d^f)$ .

The fMRI model is as follows. A lattice  $\Lambda_d$  is a square lattice  $2^d \times 2^d$  sites, with  $1 \leq d \leq D$  and distance between nearest neighbors  $L/(2^d - 1)$ . Some estimate of the configuration of  $s_D$  is the object of all this exercise. The data collected, sampled at times  $\{t_k\}_{k=1 \dots T}$  is  $V_{D,i}(t_k)$  and the model

$$V_{D,i}(t_k) = H(t_k) \frac{s_{D,i} + 1}{2} + \xi_{ik} \tag{18}$$

where  $H(t)$  is the hemodynamic response function and the  $s$  variables take on values  $\pm 1$  instead of 0, 1. A plaquette in a lattice  $\Lambda_d$  is made up of the four sites of a square of smallest possible side  $L/(2^d - 1)$ . Lattice  $\Lambda_d$  can be naturally divided into plaquettes whose center are the sites of lattice  $\Lambda_{d-1}$ . The data to be used in each renormalization stage is obtained from the data on the previous lattice by renormalization

$$V_{d-1,I}(t) = \frac{1}{4} \sum_{i \in I} V_{d,i} \tag{19}$$

and the renormalized model

$$V_{d-1,I}(t) = \frac{H(t)}{8} \sum_{i \in I} s_{d,i} + \frac{H(t)}{2} + \frac{1}{4} \sum_{i \in I} \xi_i = H(t) \frac{s_{d-1,I} + 1}{2} + \xi_{d-1,I} \tag{20}$$

The choice of the likelihood is an important step of the modelling process. For the fMRI model assume a site independent Gaussian noise process,

$$Q(V_{d-1,I}(t)|\{s_{d,i}\}_{i \in I}) = \frac{1}{\mathcal{N}} \exp \left( -\frac{1}{2T\sigma_{d-1}^2} \sum_{t_k=0}^T \left( V_{d-1,I}(t_k) - \frac{H(t_k)}{4} \sum_{i \in I} \frac{s_{d-i} + 1}{2} \right)^2 \right) \quad (21)$$

where the denominator  $\mathcal{N}$  is independent of the activities. Although this is a probability of  $V$  given  $\{s_{d,i}\}$  it can be written as proportional to a member in the Ising family given by Equation (16):

$$Q(V|\{s_{d,i}\}_{i \in I}) \propto e^{\delta K(V_{d-1},H) \sum_{(i,j)} s_i s_j + \delta h(V_{d-1},H) \sum_{i \in I} s_i}, \quad (22)$$

this is an example of a conjugated prior, where the likelihood is proportional to a member of the family of the prior. This results in a great simplification since step 1.b is unnecessary. Note that the Ising interaction term allows for spatial correlations in the activity.

Then Maximum Entropy (see Equations (9) and (10)) leads to a posterior

$$f_d(\{s_{d,i}\}_{i \in I} | \theta_d^f) = \frac{1}{z_p} e^{K_I^f \sum_{(i,j) \in I} s_i s_j + h_I^f \sum_{i \in I} s_i}, \quad (23)$$

with the learning update (Equation (1)) of the parameters given by

$$K_I^f = K_I^0 + \delta K(V_{d-1}, H), \quad (24)$$

$$h_I^f = h_I^0 + \delta h(V_{d-1}, H). \quad (25)$$

Using the notation  $\langle a \rangle = \frac{1}{T} \sum_{t_k} a(t_k)$ , it follows from Equations (20) and (21), that the learning step changes the parameters by

$$\delta K = -\frac{\langle H^2 \rangle}{64\sigma_{d-1}^2}, \quad (26)$$

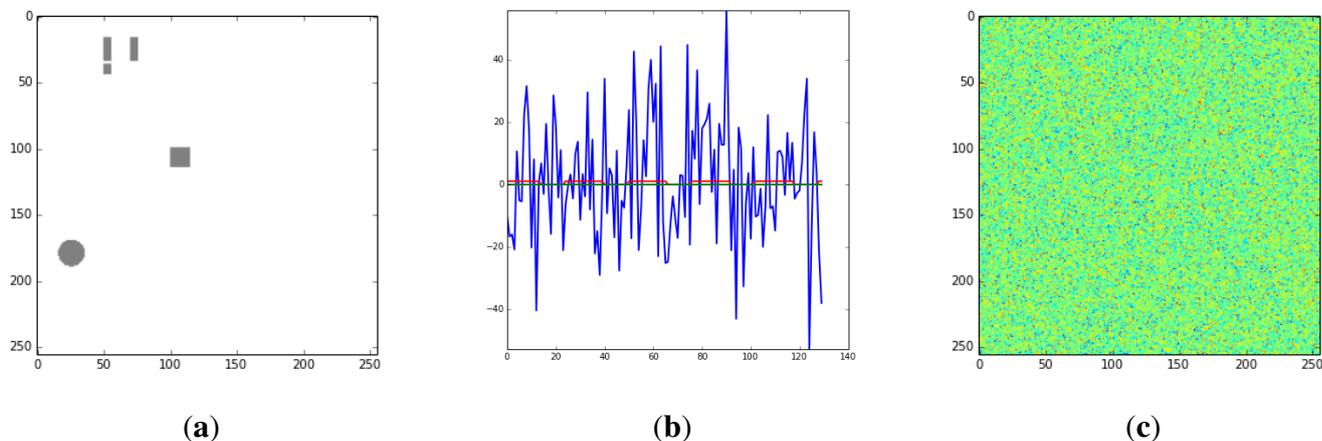
$$\delta h = \frac{2\langle V_{d-1,I}H \rangle - \langle H^2 \rangle}{16\sigma_{d-1}^2}. \quad (27)$$

The learning process is such that the effective local field increases if there is a time correlation between the renormalized data and the hemodynamic response function. The pair interactions are reduced by incorporating the data and the backward RG drives the pair coupling in the direction of the critical value. The next section has the results of a simulation of an fMRI experiment.

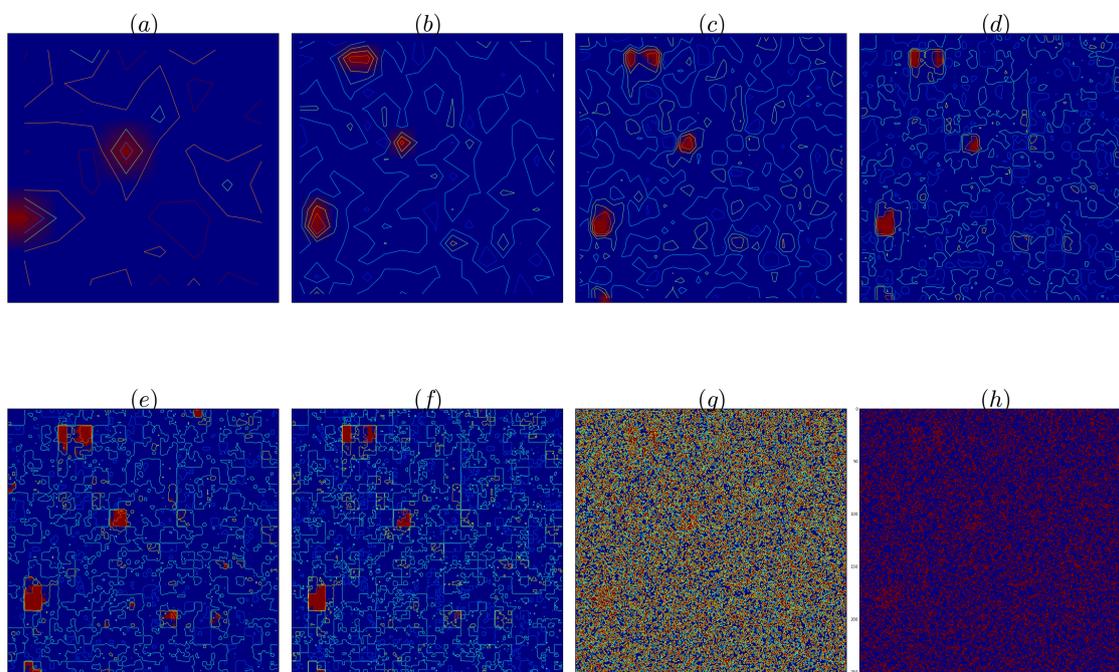
#### 4. Numerical Results

The simulated experiments consist of a sequence of active and rest periods where a number of images  $T_{act}, T_{rest}$  are collected. This is repeated  $N_{rep}$  times. Figures 2–7 essentially tell the story. A few small regions, shown in Figure 2a are active. The fMRI simulated data is built by multiplying, for each pixel, the activity and the periodic hemodynamic response function and corrupting by the addition of noise. A typical pixel time series from the active region is shown in Figure 2b and one data image from an active period is shown in Figure 2c. The likelihood in Equation (21) shows that the corrupting Gaussian noise is assumed in the reconstruction, but the experimental noises could be of other types, so Gaussian and

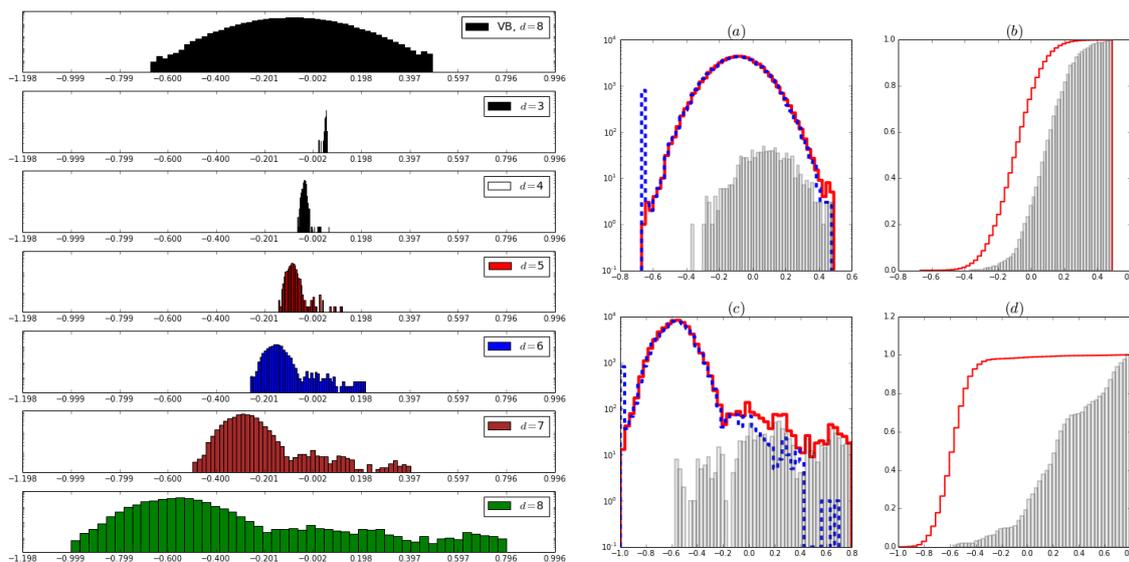
Cauchy noise are considered in the corruption of the data and Gaussian noise is always considered in this reconstruction in this paper. If some additional information about the noise becomes available, the method can be adapted accordingly.



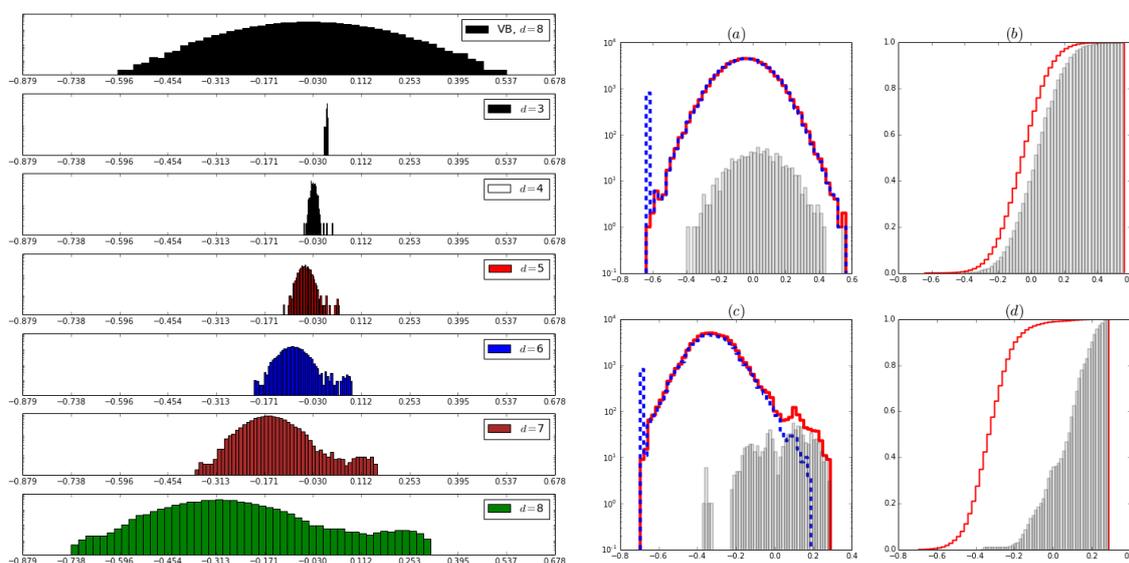
**Figure 2.** (a): Active areas, without noise. (b): An active pixel times series is a square periodic hemodynamic function with amplitude 1 or zero (red) plus Gaussian noise with  $\sigma_{data}(= 20)$  (blue). (c): A typical simulated data image during the active period. Active time  $T_{act} = 14$ , resting time  $T_{rest} = 12$ , number of active-rest repetitions  $N_{rep} = 5$ . Time units: one image per unit.



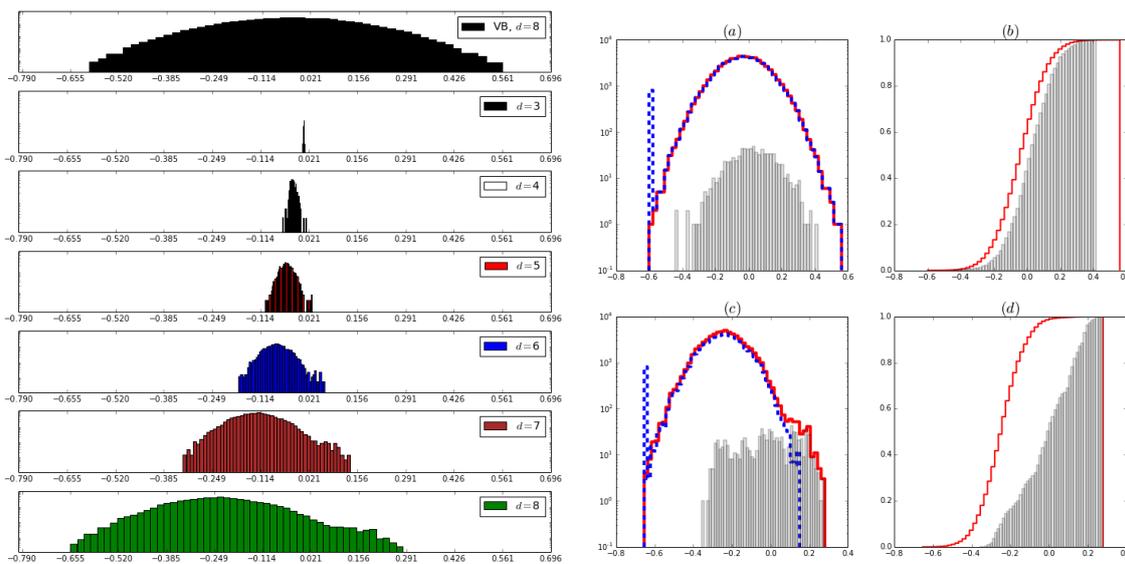
**Figure 3.** (a–e) Contour plots and Backward Renormalization hyperparameters  $h_d(i)$  for lattices  $L \times L$ ,  $L = 2^d$ ,  $d = 3 - 7$  (positive values in red, negative values in blue). (f) Final result for lattice with  $d = 8$ , which shows the threshold image of the fields. The threshold is chosen from the histogram of Figure 6. (g) Pure Bayes field contour plot and (h) field. Data corrupted with Gaussian noise with  $\sigma = 15$ .



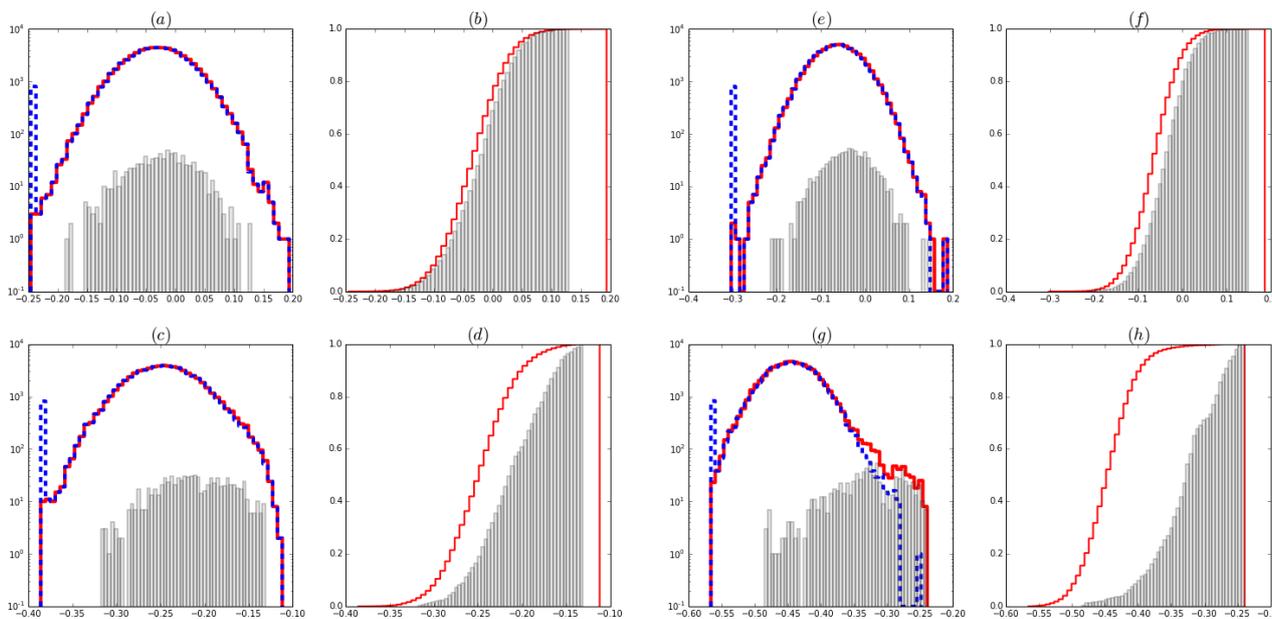
**Figure 4.** Left panel: Normalized histograms of the magnetic fields hyperparameters  $h_i$ . Vertical axis in log scale. From top to bottom, first figure shows the Gaussian like Bayes results for the full lattice. Second to seventh figures show the histograms for the BRG priors as the lattice side increases from  $2^3$  to  $2^8$ . Note that the inverse renormalization (Equations (15)) sends the distribution to the left towards lower values of the field. The learning step (Equations (27)) send most values to the left but a few to the right, hence the distributions become bimodal, separating noise from signal. Right panel, (a) and (c): The normalized histograms  $h_i$  in logarithmic scale. Steps (red): whole image. Dashed lines (blue): only from the region known to be inactive in the simulation. Bars (grey): from the region known to be active in the simulation. Right (b) and (d): cumulative histograms. Steps (red): Whole image. Bars (grey): the active regions. Top row ((a) and (b)): Bayes. Bottom row ((c) and (d)): final result of the BRG priors in  $L \times L$  lattice,  $L = 2^8$ . Data corrupted with Gaussian noise with  $\sigma = 7$



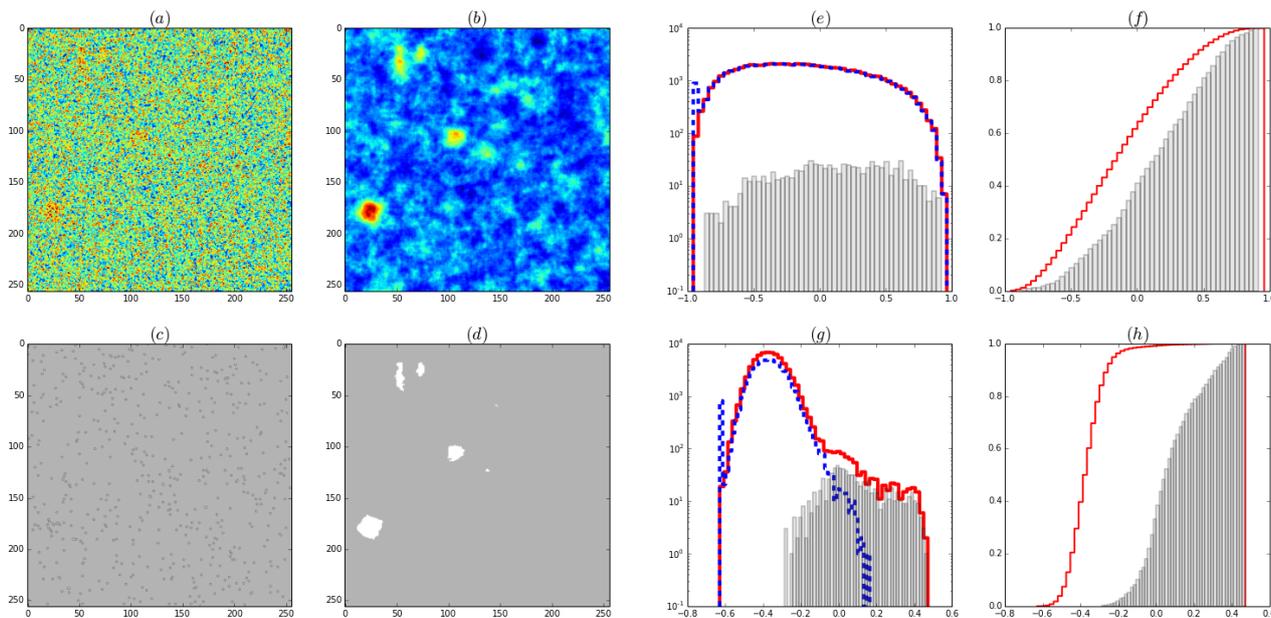
**Figure 5.** Same as Figure 4, data corrupted with Gaussian noise with  $\sigma = 15$ .



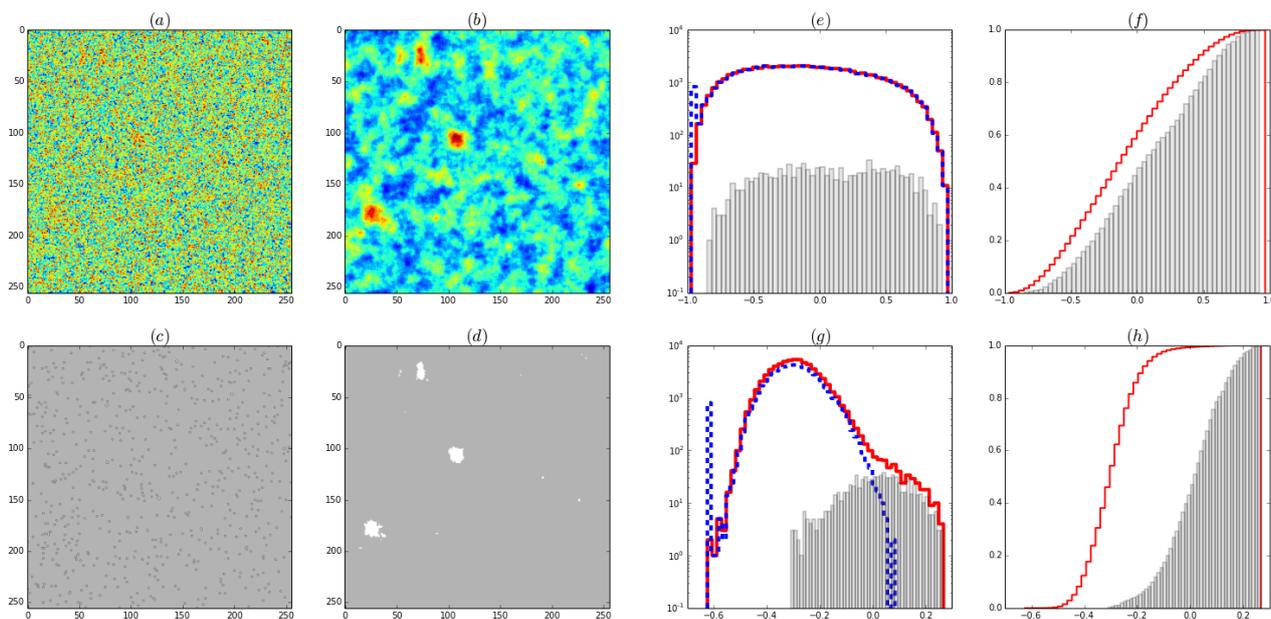
**Figure 6.** Same as Figure 4, data corrupted with Gaussian noise with  $\sigma = 20$ . The images corresponding to this histograms are shown in Figure 3.



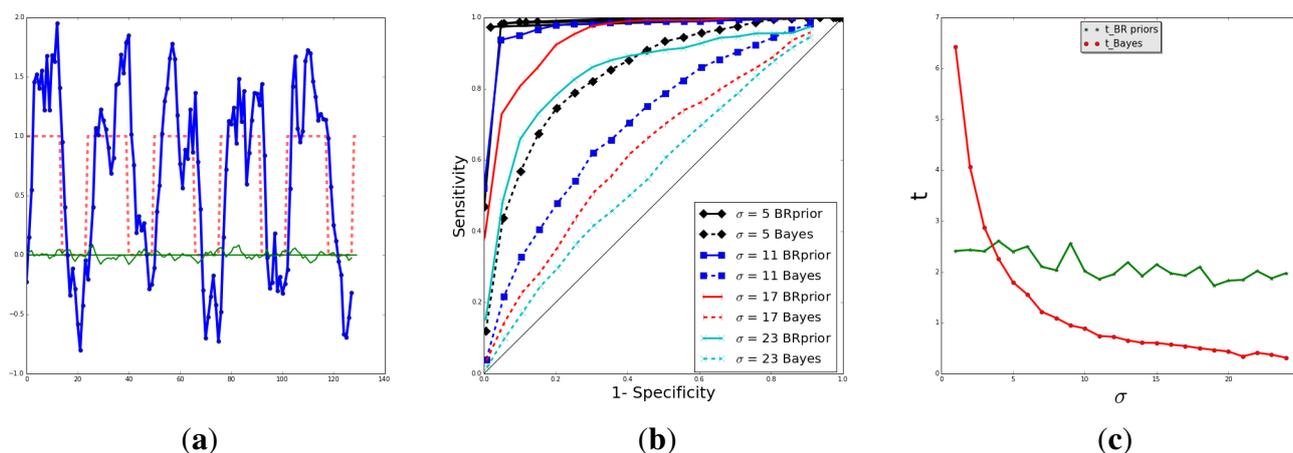
**Figure 7.** Normalized histograms of the magnetic fields  $h_i$ . Data corrupted with Cauchy noise with amplitude at half width  $\sigma_C = 20$  (figures a-d) and  $\sigma_C = 10$  (figures e-h). Images (a, b, e, f) are the histograms for the Bayes algorithm and (c,d,g,h) for the BRG priors. Note first that for the case  $\sigma_C = 10$  the histogram in figure (g) develops a shoulder, consistent with the fact that a partial identification of the active areas is possible. The other cases (a, c, e) fail to develop a shoulder and the methods can't identify the active areas. Figure (g) shows that even though the noise was Cauchy and the data analysis done under the hypothesis of Gaussian noise, the BRG priors can partially find the active areas. Continuous steps (red): full image. Dashed steps (blue) inactive regions. Bars (grey) histogram of  $h_i$  of Active regions.



**Figure 8.** Magnetization or expected value of activity for  $\sigma = 15$ . (a) Pure Bayes, (b) BRG priors: Average of  $L_s \times L_s$  shifts with  $L_s = 32$ . Thresholded magnetization for (c) Pure Bayes (d) BRG priors. (e-f) Normalized histograms of the magnetization. (e) and (f)(cumulative) Pure Bayes, (g) and (h) Normalized histograms (cumulative) BRG priors. In continuous line (red) the magnetization of the full image, in dashed line (blue) the histogram of the inactive region, in (grey) bars the histograms of the region where there is activity. The full histogram is experimentally available. The separation into signal (bars) and noise (dashed) can only be done in the simulation. The thresholds of (c) and (d) are chosen to give the same number of active sites. The threshold for figure (d) is chosen at the beginning of the shoulder in the continuous histogram in figure (g).



**Figure 9.** Same as Figure 8 but  $\sigma = 20$ .



**Figure 10.** (a) Hemodynamic response as a function of time. Dashed line (Red) the active-inactive task. Thick line (Blue)  $HRF(t)$ , the data spatial average over active region. Thin line (Green) the baseline  $B(t)$ , the data spatial average over the inactive sites. BRG priors-shifted average algorithm. Spatial averages over regions where local magnetization is higher than threshold (experimentally available from the onset of the shoulder of Figure 5 Right.(c)). Data corrupted with Gaussian noise with  $\sigma = 15$ . Range of Shift of Lattice origin =  $32 \times 32$ . The times series shown are smoothed using a moving-average with a window of 5 time steps. (b) ROC curves, 1– Specificity vs Sensitivity, obtained by changing the magnetization threshold for activity, BRG priors are shown in continuous lines, Pure Bayes in dashed lines. (c) The statistics  $t = (m_f - m_s) / \max(\text{var } m_f, \text{var } m_s)$  measures the separability of two distributions of magnetization as a function of the magnitude of the corrupting noise:  $m_f$  of the full picture and  $m_s$  of the active region.

The numerical simulations shown here correspond to a brutal corruption with noise. For example, in Figure 3, the on-off hemodynamic function is of amplitude one and the Gaussian noise variance is 15, a signal to noise ratio of  $\text{SNR} = -23\text{dB}$ . The pure Bayes method (see Figures 3g,h) fails to identify the active regions, while the backward renormalization iteration leads to a rough identification of at least some active areas. The  $h$  field histograms help to understand the joint effect of the steps in the process. Note that the Bayes histograms, shown at the top of the left panel in Figures 4–6, are almost Gaussian. The renormalization transformation shifts the whole histogram to the origin, and the learning step Equation (27) shifts most fields to the left, where  $\langle V_{d-1,l}H \rangle$  is small, and a few fields to the right, where it is large. This results in non-Gaussian distributions as shown at the bottom of the left panel in Figures 4–6. The meaning of the bumps is very interesting. Since these are simulations, the effects on the signal (active pixels) can be analyzed separately from the noise (inactive pixels) and this is clearly shown in the right hand side histograms of these figures. Since the active area is very small, the signal histogram is difficult to see. The shift in the signal histogram, relative to the whole picture field histogram is clearly seen even for the Bayes. But the shift is not large enough to escape the noise histogram umbrella. In the bottom pictures of the right panels in Figures 4–6, the signal histogram can be seen escaping the noise histogram and in breaking away making a bump in the otherwise Gaussian histograms. But this bump can be seen in the histogram of the fields of the whole image, signaling where to draw the line in

deciding what is noise and what is signal. The same behavior can be seen in Figure 7, where the noise corruption is not Gaussian but Cauchy.

#### 4.1. The Effect of the Lattice

It is due to the structure of the square renormalized lattices that the images of the active regions tend to show sharp horizontal or vertical edges. In order to get rid of this bias, an average of the magnetizations obtained with different choices of the origin of the lattice was performed. This was done by repeating the calculation  $L_s \times L_s$  times, with the origin of the lattice at a position translated by  $(x, y)$  pixels, with  $0 \leq x, y < L_s$ . For each one the magnetization given by Equation (17), is calculated and then averaged at each pixel. Figures 8 and 9 show the results for  $L = 2^8$  and  $L_s = 32$ . It can be seen that the sharp edges are smoothed by the averaging process leading to a better image.

Once the shift-averaged magnetization is available it can be used for instance to infer the form of the hemodynamic response function by doing a spatial average of the data time series over the region that has been identified as active [11]:

$$HRF(t) = \langle V \rangle_{active} = \frac{\sum_i \Theta(m_i - m_0) V_i(t)}{\sum_i \Theta(m_i - m_0)}, \quad (28)$$

$$B(t) = \langle V \rangle_{inactive} = \frac{\sum_i \Theta(-m_i + m_0) V_i(t)}{\sum_i \Theta(-m_i + m_0)}, \quad (29)$$

where  $\Theta(x) = 0$  or  $1$  if  $x < 0$  or  $x \geq 0$  respectively and the threshold  $m_0$  can be obtained from the onset of the shoulder of the histogram (see Figure 8g).  $HRF(t)$ , the spatial average over the active regions and  $B(t)$  the baseline, the spatial average over the regions identified as inactive, are shown in Figure 10a.

## 5. Discussion

The use of data to inform the prior is not new since Empirical Bayes (EB), e.g., see [12], has been around in many flavors for many decades. The idea behind the method presented here is different in at least two aspects. First, in EB the estimates of hyperparameters is made using the marginals of the data and the resulting posteriors of the sources are conditioned on estimates of the hyperparameters. In this sense a multigrid approach would be more aligned with the EB approach [12]. Second, the source variables in different lattices are related by renormalization and are not in the same footing as hyperparameters.

What are the main ingredients a problem has to satisfy so that the backward renormalization helps the inferential process? A method that builds priors that lead to better results has to introduce the use of some form of information that is not used in the pure Bayes. The new information is that the active regions are not isolated pixels but have some spatial extension. The data in the fMRI models come in two types of time series. One, the noise, has a mean value  $\mu_n = 0$  and the other has a mean value  $\mu_s \neq 0$ , the active pixels. In the simulations, Section 4,  $\mu_s \sim 1$  and the variance of the noise  $\xi$  is taken up to  $\sigma \sim 20$  and still a partial identification of the active areas is possible. The renormalization of the data by blocking leads to a decrease in the effective variance  $\sigma_{eff} = \sigma / (RT_a^{\frac{1}{2}})$ , where  $R^2$  is the typical area of the active regions and  $T_a$  the number of images where there is activity. For the pure Bayes approach the effective noise variance is only reduced by  $T_a^{\frac{1}{2}}$ . Blocking reduces the spatial resolution, which is

recovered by the Backward renormalization step and the use of a more refined version of the data. This argument suggests that the large regions of activity will be more easily detected than the small regions as indeed happens in the simulations. Also the quality of recovery of a region of size  $R$  can not be expected to be better than that with pure Bayes at a variance  $1/R$  times smaller. On the other hand, the inactive area is much larger than the active (about 7% of the total image shown in Figure 2 is active). This means that the identification of the inactive areas will be very effective translating into very few false positives. The backward renormalization step determines an effective transmission of the information to higher resolutions. Independently of the starting point, the pair interaction is driven to the critical value (in this case  $K \approx .03$ ) by the Backward renormalization step.

In Figure 10b, the receiver operating characteristic (ROC) curves are shown for different values of  $\sigma$  for both methods. The lines are obtained by measuring the specificity ( $TN/(FP + TN)$ ) and the sensitivity ( $TP/(TP + FN)$ ) when the magnetization threshold  $m_0$  is changed ( $(T, F) = (\text{true}, \text{false})$ ,  $(P, N) = (\text{positive}, \text{negative})$ ). Note that the ROC curve for the pure Bayes with  $\sigma = 5$  is approximately the same as for BRG priors with  $\sigma = 23$ , compatible with the discussion above for  $R = 4$  or  $5$ , a reasonable estimate, as seen in Figure 2.

In this simple case, the magnitude of the HRF is the same in every active region. But HRF amplitudes may vary in space in experimental cases. It is not difficult to include this variability in the model, e.g., see [11]. If this variation is permitted the pair interaction will also be driven by the data to space dependent values.

## 6. Conclusion

In this paper, the information used to construct priors is essentially that the sources are spatially localized but that single site activity is not the rule. The computational demands can be very much reduced if the starting point of an iterative process, like the Variational Bayes, is close to the final solution. The building of better priors through the building of more spatially refined lattices leads to simple problems at each new scale. The method presented here has several potential application scenarios. That a simulation of fMRI was chosen to exemplify the method is due to the simplicity and importance of the task-based or stimulus driven rest-active experimental paradigm. Applications to EEG/MEG source determination, and Connectome information flow will be presented elsewhere. We have not strived to present an extremely realistic simulation, but note that that doesn't make it simpler. Noise to signal ratios in the simulations results presented here are very high and attest to the potential of the method. Even the error of corrupting with very high Cauchy noise and analyzing under the assumption that the noise was Gaussian could be handled quite efficiently. The Renormalization Group is one of the most important developments in Physics in the second half of the last century. It is typically used to integrate out degrees of freedom associated to the finer scales degrees of freedom. But here the reverse direction was taken. Typically the RG is referred to as a semigroup. But that applies only if the issue is how to go from the dynamical variables from one lattice to another. Then only the direction towards the coarser lattices is possible. However, if only the hyperparameters are considered then the RG is invertible. The hyperparameters or coupling constants, from one scale are related to the other scales and by inverting the RG map, the posterior in one scale feeds the prior in the next refinement. The use of

the Migdal-Kadanoff renormalization group is not essential since there are several other possibilities for Ising variables. In general, the nature of the sources will dictate the renormalization group transformation useful to transport information from a coarse to a finer lattice. In addition, the choice of the type of lattice can be very important. Each application will have its own peculiarities, but in all of them, the choice of the prior at the coarsest spatial representation will be quite irrelevant.

### Acknowledgments

I thank Ariel Caticha for discussions about Entropic Inference; Said Rabbani, Selene Amaral, Leonardo Barbosa and Leonardo Tavares for discussions about multigrid priors; Felipe Alves Pereira for his help with the python code. The author received support from the Núcleo de Apoio à Pesquisa CNAIPS of the University of São Paulo.

### Conflicts of Interest

The author declares no conflict of interest.

### References

1. Wilson, K.G. The Renormalization Group and Critical Phenomena. *Rev. Mod. Phys.* **1983**, *55*, 583–600.
2. Goodman, J.; Sokal, A.D. Multigrid Monte Carlo Method. Conceptual Foundations. *Phys. Rev. D* **1989**, *40*, 2035–2071.
3. Amaral, R.; Rabbani, S.R.; Caticha, N. Multigrid Priors for a Bayesian Approach to FMRI. *NeuroImage* **2004**, *23*, 654–662.
4. Giffin, A.; Caticha, A. Updating Probabilities with Data and Moments. Presented at the 27th Interenational Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Saratoga Springs, NY, USA, 8–13 July 2007; doi:10.1063/1.2821302.
5. Bishop, C.M. *Pattern Recognition and Machine Learning*, 1st ed.; Information Science and Statistics Series; Springer: New York, NY, USA, 2007.
6. Logothetis, N.K. What We Can Do and What We Cannot Do with FMRI. *Nature* **2008**, *453*, 869–878.
7. Ashby, F.G. *Statistical Analysis of FMRI Data*; MIT Press: Hong Kong, China, 2011.
8. Migdal, A.A. Phase transitions in gauge and spin-lattice systems *Sov. Phys. JETP* **1975**, *42*, 743–746.
9. Kadanoff, L.P. Notes on Migdal's Recursion Formulas. *Ann. Phys.* **1976**, *100*, 359–394.
10. Griffiths, R.B.; Kaufman, M. Spin Systems on hierarchical lattices. Introduction and thermodynamic limit. *Phys. Rev. B* **1982**, *26*, 5022–5032.
11. Amaral, R.; Rabbani, S.R.; Caticha, N. BOLD Response Analysis by Iterated Local Multigrid Priors. *Neuroimage* **2006**, *36*, 361–369.

12. Carlin, B.P.; Louis, T.A. *Bayes and Empirical Bayes methods for Data Analysis*; Chapman and Hall/CRC: Boca Raton, Florida, USA, 1996.

© 2015 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).