

Conjugate Gamma Markov Random Fields for Modelling Nonstationary Sources

A. Taylan Cemgil¹ and Onur Dikmen^{2,*}

¹ Engineering Dept., University of Cambridge, CB2 1PZ, Cambridge, UK
atc27@eng.cam.ac.uk

<http://www-sigproc.eng.cam.ac.uk/~atc27>

² Dept. of Computer Eng., Bogazici University, 80815 Bebek, Istanbul, Turkey

Abstract. In modelling nonstationary sources, one possible strategy is to define a latent process of strictly positive variables to model variations in second order statistics of the underlying process. This can be achieved, for example, by passing a Gaussian process through a positive nonlinearity or defining a discrete state Markov chain where each state encodes a certain regime. However, models with such constructs turn out to be either not very flexible or non-conjugate, making inference somewhat harder. In this paper, we introduce a conjugate (inverse-) gamma Markov Random field model that allows random fluctuations on variances which are useful as priors for nonstationary time-frequency energy distributions. The main idea is to introduce auxiliary variables such that full conditional distributions and sufficient statistics are readily available as closed form expressions. This allows straightforward implementation of a Gibbs sampler or a variational algorithm. We illustrate our approach on denoising and single channel source separation.

1 Introduction

In the Bayesian framework, various signal estimation problems can be cast into posterior inference problems. For example, source separation [6,5,9,11,3,2], can be stated as

$$p(\mathbf{s}|\mathbf{x}) = \frac{1}{Z_x} \int d\theta_o d\theta_s p(\mathbf{x}|\mathbf{s}, \theta_o) p(\mathbf{s}|\theta_s) p(\theta_o) p(\theta_s) \quad (1)$$

where $\mathbf{s} \equiv s_{1:K,1:N}$ and $\mathbf{x} \equiv x_{1:K,1:M}$. Here, the task is to infer N *source signals* $s_{k,n}$ given M *observed signals* $x_{k,m}$ where $n = 1 \dots N$, $m = 1 \dots M$ at each index k where $k = 1 \dots K$. Here, k typically denotes time or a time-frequency atom in a linear transform domain. In Eq.(1), the (possibly degenerate, deterministic) conditional distribution $p(\mathbf{x}|\mathbf{s}, \theta_o)$ specifies the *observation model* where θ_o denotes

* This research is funded by EPSRC (Engineering and Physical Sciences Research Council of UK) under the grant EP/D03261X/1 entitled “Probabilistic Modelling of Musical Audio for Machine Listening” and by TUBITAK (Scientific and Technological Research Council of Turkey) under the grant entitled “Time Series Modelling for Bayesian Source Separation”.

the collection of mixing parameters such as the mixing matrix, observation noise variance, etc. The prior term $p(\mathbf{s}|\Theta_s)$, the *source model*, describes the statistical properties of the sources via their own prior parameters Θ_s . The normalisation term $Z_x = p(\mathbf{x})$ is the marginal probability (*evidence*) of the data under the model and plays a key role for model order selection (such as determining the number of sources) [8]. The hierarchical model is completed by postulating hyper-priors over the nuisance parameters Θ_s and Θ_o . Estimates of the sources can be obtained from posterior features such as marginal maximum-a-posteriori (MMAP) or minimum-mean-square-error (MMSE) estimate¹

$$\mathbf{s}^* = \underset{\mathbf{s}}{\operatorname{argmax}} p(\mathbf{s}|\mathbf{x}) \quad \langle \mathbf{s} \rangle_{p(\mathbf{s}|\mathbf{x})} = \int \mathbf{s} p(\mathbf{s}|\mathbf{x}) d\mathbf{s}$$

Unfortunately, exact calculation of these quantities is intractable for almost all relevant observation and source models, even under conditionally Gaussian and independence assumptions. Hence, approximate numerical integration techniques have to be employed.

In applications, the key object is often the source model $p(\mathbf{s}|\Theta_s)$. Indeed, many popular signal estimation algorithms can be obtained by choosing a particular source prior and applying Bayes rule. If the sources have some known structure, one can design more realistic prior models to improve the estimates. In this paper, we explore generic source models that explicitly model nonstationarity.

Perhaps the prototypical example of a nonstationary process is one where the conditional variance is slowly changing in time. In finance literature, such models are known as *stochastic volatility* models and are important to characterise non-stationary behaviour observed in financial markets [12]. In spatial statistics, similar constructions are needed in 2-D where one is interested in changing intensity over a region [15]. In audio processing, the energy content of a signal is typically time-varying hence it is natural to model audio with a process with a time varying power spectral density on a time frequency plane [10,14,4].

In the sequel, we introduce an alternative model that is useful for modelling a random walk on variances. The main idea is to introduce auxiliary variables such that full conditional distributions and sufficient statistics are readily available as closed form expressions. This allows straightforward implementation of a Gibbs sampler or a variational algorithm. Consequently we extend the model to 2-D random fields, which is useful for modelling nonstationary time-frequency energy distributions or intensity functions that need to be strictly positive. We illustrate our approach on various denoising and source separation scenarios.

2 Model

The *inverse Gamma* distribution with shape parameter a and scale parameter z is defined as

$$\mathcal{IG}(v; a, z) \equiv \exp((a + 1) \log v^{-1} - z^{-1}v^{-1} + a \log z^{-1} - \log \Gamma(a))$$

¹ Here, and elsewhere the notation $\langle f(x) \rangle_{p(x)}$ will denote the expectation of the function $f(x)$ under the distribution $p(x)$, i.e. $\langle f(x) \rangle_p \equiv \int dx f(x)p(x)$.

Here, Γ is the gamma (generalised factorial) function. The sufficient statistics of the inverse-Gamma distribution are given by $\langle v^{-1} \rangle_{\mathcal{IG}} = az$ and $\langle \log v^{-1} \rangle_{\mathcal{IG}} = \Psi(a) - \log z^{-1}$ where Ψ is the digamma function defined as $\Psi(a) \equiv d \log \Gamma(a) / da$. The inverse gamma distribution is the conjugate prior for the variance v of a Gaussian distribution $\mathcal{N}(s; \mu, v) \equiv \exp(-(s-\mu)^2 v^{-1} / 2 + \log v^{-1} / 2 - \log(2\pi) / 2)$. When the prior $p(v)$ is inverse Gamma, the posterior distribution $p(v|s)$ can be represented as an inverse Gamma distribution since the logarithm of a Gaussian is a polynomial in v^{-1} and $\log v^{-1}$. Similarly, the *Gamma* distribution with shape parameter a and scale parameter z is defined as

$$\mathcal{G}(\lambda; a, z) \equiv \exp((a - 1) \log \lambda - z^{-1} \lambda + a \log z^{-1} - \log \Gamma(a))$$

The sufficient statistics of the Gamma distribution are given by $\langle \lambda \rangle_{\mathcal{G}} = az$ and $\langle \log \lambda \rangle_{\mathcal{G}} = \Psi(a) - \log z^{-1}$. Gamma distribution is the conjugate prior for the precision parameter (inverse variance) of a Gaussian distribution as well as for the intensity parameter λ of a Poisson distribution

$$c \sim \mathcal{PO}(c; \lambda) \equiv e^{-\lambda} \lambda^c / c! = \exp(c \log \lambda - \lambda - \log \Gamma(c + 1))$$

We will exploit this property to estimate intensity functions of non-homogeneous Poisson processes.

2.1 Markov Chain Models

It is possible to define a Markov chain on inverse Gamma random variables in a straightforward way by $v_k | v_{k-1} \sim \mathcal{IG}(v_k; a, v_{k-1} / a)$. The full conditional distribution $p(v_k | v_{k-1}, v_{k+1})$ is conjugate, i.e. it is also inverse Gamma. However, by this construction it is not possible to attain positive correlation between v_k and v_{k-1} . Positive correlations can be obtained by conditioning on the reciprocal of v_{k-1} and defining $p(v_k | v_{k-1}) = \mathcal{IG}(v_k; a, (v_{k-1} a)^{-1})$; however in this case the full conditional distribution $p(v_k | v_{k-1}, v_{k+1})$ becomes non-conjugate since it has $v_k, 1/v_k$ and $\log v_k$ terms. The basic idea is to introduce latent auxiliary variables z_k between v_k and v_{k-1} such that when z_k are integrated out we restore positive correlation between v_k and v_{k-1} while retaining conjugacy. We define an *Inverse Gamma-Markov* chain (IGMC) for $k = 1 \dots K$ as follows

$$z_1 \sim \mathcal{IG}(z_1; a_z, b_z / a_z) \quad v_k | z_k \sim \mathcal{IG}(v_k; a, z_k / a) \quad z_{k+1} | v_k \sim \mathcal{IG}(z_{k+1}; a_z, v_k / a_z)$$

Here, z_k are auxiliary variables that ensure the full conditionals

$$p(v_k | z_k, z_{k+1}) \propto \exp((a + a_z + 1) \log v_k^{-1} - (a z_k^{-1} + a_z z_{k+1}^{-1}) v_k^{-1}) \tag{2}$$

and $p(z_k | v_k, v_{k-1})$ are inverse Gamma. By integrating out over the auxiliary variable z_k we obtain the effective transition kernel of the Markov chain, where it can be easily shown that

$$\begin{aligned} p(v_k | v_{k-1}) &= \int dz_k p(v_k | z_k) p(z_k | v_{k-1}) = \int dz_k \mathcal{IG}(v_k; a, z_k / a) \mathcal{IG}(z_k; a_z, v_{k-1} / a_z) \\ &= \frac{\Gamma(a + a_z)}{\Gamma(a_z) \Gamma(a)} \frac{(a_z v_{k-1}^{-1})^{a_z} (a v_k^{-1})^a}{(a_z v_{k-1}^{-1} + a v_k^{-1})^{(a_z + a)}} v_k^{-1} \end{aligned} \tag{3}$$

This distribution, which in our knowledge does not have a designated name, is a scale mixture of inverse Gamma distributions where the scaling function is also inverse Gamma. The transition kernel $p(v_k|v_{k-1})$ has positive correlation for various shape parameters a_z and a . The absolute value of a_z and a control the strength of the correlation and the ratio a_z/a controls the skewness. For $a_z/a < 1$ ($a_z/a > 1$), the probability mass is shifted towards the interval $v_k < v_{k-1}$ ($v_k > v_{k-1}$) hence, typical trajectories from a IGMC will exhibit a systematic negative (positive) drift. Using an exactly analogous construction, we define a *Gamma-Markov* chain (GMC) as $z_1 \sim \mathcal{G}(z_1; a_z, (b_z a_z)^{-1})$, $\lambda_k|z_k \sim \mathcal{G}(\lambda_k; a_\lambda, (z_k a_\lambda)^{-1})$, $z_{k+1}|\lambda_k \sim \mathcal{G}(z_{k+1}; a_z, (\lambda_k a_z)^{-1})$. The effective transition kernel has a very similar expression as in Eq.3.

Example 1, Nonstationary Gaussian Process: We define a non-stationary Gaussian process $\{y_k\}_{k=1,2,\dots}$ by drawing the variances $\{v_k\}_{k=1,2,\dots}$ from an IGMC and drawing $y_k|v_k \sim \mathcal{N}(y_k; 0, v_k)$ In Figure 1(a)-top, we show a realisation of $v_{1:K}$ from the IGMC, labelled as “true” and generate $y_{1:K}$ conditionally Figure 1(a)-bottom. Given a realisation $y_{1:K}$, we can estimated the posterior variance $\langle v_k|y_{1:K} \rangle$. In this case, inference is carried out with variational Bayes as will be detailed in section 3.

Example 2, Nonhomogeneous Poisson Process: We partition an interval \mathbb{I} on the real line into small disjoint regions R_k of area L such that $\mathbb{I} = \cup_{k=1}^K R_k$. We assume that the unknown intensity function of the process is piecewise constant and has the value λ_k on region R_k . The intensity function $\{\lambda_k\}_{k=1,2,\dots}$ is drawn according to a GMC. The number points in R_k , given the intensity function, is denoted by the Poisson random variable $c_k|\lambda_k \sim \mathcal{PO}(c_k; \lambda_k L)$ To generate a realisation from the Poisson process, we can uniformly draw c_k points in each region R_k . In Figure 1(b), we show a realisation from the model. Given the number of events in each region R_k , we can estimate the value of the intensity function on R_k by calculating $\langle \lambda_k|c_{1:K} \rangle$.

2.2 (Inverse) Gamma Markov Random Fields – (I)GMRF

We have defined the IGMC and GMC in the previous section using conditional distributions. An alternative but equivalent factorisation, that encodes the same distribution but corresponds to

$$p(\mathbf{z}, \mathbf{v}) \propto \psi(b_z^{-1}, a_z z_1^{-1}) \prod_k \phi(v_k^{-1}; a + a_z) \phi(z_k^{-1}; a + a_z) \psi(a z_k^{-1}, v_k^{-1}) \psi(a_z v_k^{-1}, z_{k+1}^{-1})$$

where we specify singleton potentials ϕ and pairwise potentials ψ as

$$\phi(\xi; \alpha) = \exp((\alpha + 1) \log \xi) \qquad \psi(\xi, \eta) = \exp(-\xi \eta)$$

Generalising this to a general undirected graph with vertex set \mathcal{V} and undirected edge set \mathcal{E} , we define an IGMRF on $\boldsymbol{\xi} = \{\xi_i\}_{i \in \mathcal{V}}$ by a set of connection weights $\mathbf{a} = \{a_{i,j}\}_{(i,j) \in \mathcal{E}}$ for $i, j \in \mathcal{V}$ and $i \neq j$

$$p(\boldsymbol{\xi}; \mathbf{a}) = \frac{1}{Z_{\mathbf{a}}} \prod_{i \in \mathcal{V}} \phi(\xi_i^{-1}; \sum_j a_{i,j}) \prod_{(i,j) \in \mathcal{E}} \psi(\xi_i^{-1}, (a_{i,j}/2) \xi_j^{-1}) \equiv \frac{1}{Z_{\mathbf{a}}} p_{\mathbf{a}}^*(\boldsymbol{\xi}) \quad (4)$$

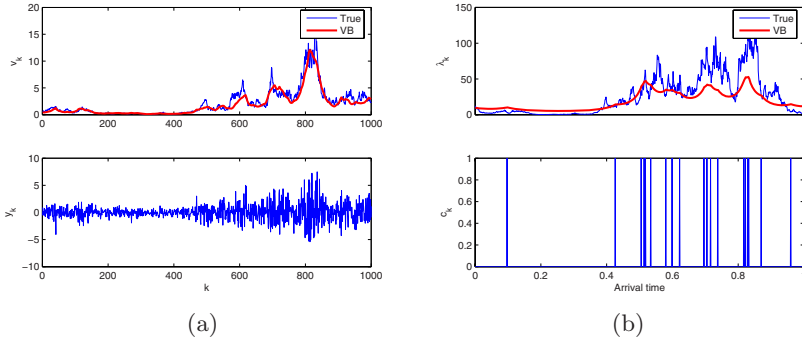


Fig. 1. Synthetic examples generated from the model. The thick line shows the result of the variational inference (a) Non-stationary Gaussian Process. (a-Top) a typical draw of $v_{1:K}$ from the IGMG $b_z = 1, a = a_z = 100$. (a-Bottom) Draw from the Gaussian process $y_{1:K}$ given $v_{1:K}$. (b) Non-homogeneous Poisson Process. (b-Top) a typical draw from the GMC with $b_z = 10, a = a_z = 100$ and frame length $\mu(R_k) = L = 0.001$. (b-Bottom) Number of events in each R_k .

where ϕ and ψ are defined above. A GMRF is defined similarly by the potentials $\phi(\xi_i; \sum_j a_{i,j})$ and $\psi(\xi_i, (a_{i,j}/2)\xi_j)$ but with $\phi(\xi; \alpha) = \exp((\alpha - 1) \log \xi)$.

3 Inference

Exact inference in random fields is in general intractable and various numerical methods have been developed, based on sampling (Monte Carlo-stochastic) or analytic approximation (Variational-deterministic). Here, we focus on a particularly simple variational algorithm (mean field - variational Bayes [1,13]) – but application of Monte Carlo methods, such as the Gibbs sampler [7] is algorithmically very similar[2]. Variational methods have been applied extensively, notably in machine learning for inference in large models. While lacking theoretical guarantees of Monte Carlo approaches, variational methods have been viable alternatives in several practical situations where only a fixed amount of CPU budget is available.

The main idea in variational Bayes is to approximate a target distribution $\mathcal{P} \equiv p^*(\boldsymbol{\xi})/Z$ (such as the IGMRF defined in Eq.(4)) with a simple distribution \mathcal{Q} . The variational distribution \mathcal{Q} is chosen such that its expectations can be obtained easily, preferably in closed form. One such distribution is a factorised one $\mathcal{Q}(\boldsymbol{\xi}) = \prod_{i \in \mathcal{V}} \mathcal{Q}_i(\xi_i)$. An intuitive interpretation of mean field method is minimising the KL divergence with respect to (the parameters of) \mathcal{Q} where $KL(\mathcal{Q}||\mathcal{P}) = \langle \log \mathcal{Q} \rangle_{\mathcal{Q}} - \langle \log p^*/Z_{\mathbf{a}} \rangle_{\mathcal{Q}}$. Using non-negativity of KL, one can obtain a lower bound on log-normalisation constant

$$\log Z_{\mathbf{a}} \geq \langle \log p^* \rangle_{\mathcal{Q}} - \langle \log \mathcal{Q} \rangle_{\mathcal{Q}} \tag{5}$$

The maximisation of this lower bound is equivalent to finding the “nearest” \mathcal{Q} to \mathcal{P} in terms of KL divergence. Whilst the solution is in general not available in

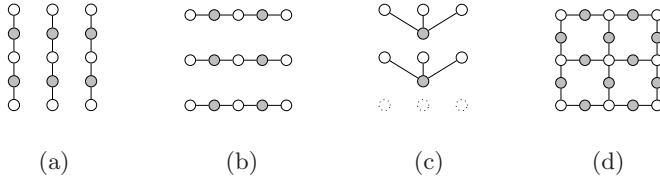


Fig. 2. Various IGMRF topologies as priors for time frequency energy distributions. White nodes (placed always on a rectangular grid) correspond to $v_{\nu,\tau}$ where the vertical and horizontal axis corresponds to the frequency band index ν and time index τ respectively. Gray nodes correspond to the auxiliary variables z . (a)-(d) **vertical, horizontal, band, grid.**

closed form, it can be easily shown, e.g. see [13], that each factor Q_i of the optimal approximating distribution should satisfy the following fixed point equation

$$Q_i \propto \exp \left(\langle \log p^* \rangle_{Q_{-i}} \right) \tag{6}$$

where $Q_{-i} \equiv Q/Q_i$, that is the joint distribution of all factors excluding Q_i . Hence, the mean field approach leads to a set of (deterministic) fixed point equations that need to be iterated until convergence. For a MRF, this fixed point expression is efficient to evaluate since it depends only on the neighbouring variables $j \in \mathcal{N}(i)$. Finally, for conjugate models the factors are available in closed form; for example IGMRF leads to the factors $Q_i^{(t)}(\xi_i) = \mathcal{IG}(\xi_i; \alpha_i^{(t)}, \beta_i^{(t)})$ with

$$\alpha_i^{(t)} = \theta_{\alpha,i} + \sum_{j \in \mathcal{N}(i)} a_{i,j} \qquad \beta_i^{-1(t)} = \theta_{\beta,i} + \sum_{j \in \mathcal{N}(i)} a_{i,j} \langle \xi_j^{-1} \rangle_{Q_j^{(t-1)}}$$

Here, $\theta_{\alpha,i}$ and $\theta_{\beta,i}$ denote the data contributions when a IGMRF is used as a prior where some of the ξ are observed via a conjugate observation model. For example, in the conditionally Gaussian observation model of section 2 we have $\theta_{\alpha,i} = 1/2$ and $\theta_{\beta,i} = y_i^2/2$. Similarly, the Poisson model with a GMRF prior has $\theta_{\alpha,i} = c_i$ and $\theta_{\beta,i} = L$.

3.1 Simulation Experiments

In Figure 1, we show the results of variational inference for two synthetic examples. In the following, we will illustrate the IGMRF model used as a prior for time-frequency energy distributions of nonstationary sources.

Linear time-frequency representations decompose a signal $y(t)$ as a linear decomposition of form $y(t) = \sum_{(\nu,\tau)} s_{(\nu,\tau)} f_{\nu,\tau}(t)$ where $s_{(\nu,\tau)}$ is the expansion coefficient corresponding to the basis function $f_{\nu,\tau}(t)$. Somewhat succinctly we can write $\mathbf{y} = F\mathbf{s}$, where the collection of basis functions is denoted by a matrix F where each column corresponds to a particular $f_{\nu,\tau}$. The well known Gabor representation or modified cosine transform (MDCT) have this form and can be

computed using fast transforms where τ corresponds to time and ν corresponds to frequency. In the sequel, we will impose a conditionally Gaussian prior on transform coefficients $\mathcal{N}(s_{(\nu,\tau)}; 0, v_{(\nu,\tau)})$ where the covariance structure will be drawn from a IGMRF.

Denoising: In the first experiment, we illustrate the denoising performance of 4 MRF topologies on a set of 5 audio clips (`speech`, `piano`, `guitar`, `perc1`, `perc2`) in 3 different noise conditions `low`, `medium`, `high`. We transform each clip via MDCT to $s_{(\nu,\tau)}^{true}$ and add independent white Gaussian noise with variance $r \sim \mathcal{IG}(r; a_r, b_r)$ to obtain $x_{(\nu,\tau)}$. Note that since MDCT is an orthonormal linear transform, we could have added noise in time domain and the noise characteristics would have remained unaltered. As inference engine, we use variational Bayes. The task of the inference algorithm is to infer the latent source coefficients $s_{(\nu,\tau)}$ by integrating out the noise variance r and the IGMRF variables ξ . The optimisation of MRF parameters \mathbf{a} is carried out by the Newton method where we maximise the lower bound in Eq.5. We assume homogeneous MRF structure where the coupling values are the same throughout the network². The signal to noise ratio of reconstructions and inference results are given in Figure 3-(a). The SNR results do not show big variations across topologies, with the grid model consistently providing good reconstruction, especially in high and medium noise conditions. We note that SNR may not be the best metric to measure perceptual quality and the reader is invited to listen to the audio examples provided online at <http://www.cmpe.boun.edu.tr/~dikmen/ICA07/>. In informal subjective listening tests, we perceive a somewhat better reconstruction quality with the grid topology.

Source Separation: In the second experiment, we illustrate the viability of the approach on a single channel separation task with $j = 1 \dots J$ sources. At each time-frequency location $k \equiv (\nu, \tau)$, the generative model is $\mathbf{v}_j \sim \text{IGMRF}_j, s_{k,j} \sim \mathcal{N}(s_{k,j}; 0; v_{k,j}), x_k = \sum_{j=1}^J s_{k,j}$. In this scenario, the reconstruction equations have a particularly simple form. Given the variance estimate $V_{k,j} = \langle 1/v_{k,j} \rangle^{-1}$ at k , we define a positive quantity, which we shall name as *responsibility* also know as filter factors, $\kappa_{k,j} = V_{k,j} / (\sum_{j'} V_{k,j'})$, where by construction $0 < \kappa_j$ for all j and $\sum_j \kappa_j \leq 1$. It turns out that the sufficient statistics can be compactly written as

$$\langle s_{k,j} \rangle = \kappa_{k,j} x_k \quad \langle s_{k,j}^2 \rangle = V_{k,j} (1 - \kappa_{k,j}) + \kappa_{k,j}^2 x_k^2$$

We illustrate this approach to separate a piano sound into its constituent components. We assume that $J = 2$ components are generated independently by two IGMRF models with vertical and horizontal topology. In figure 3-(b), we observe that the model is able to separate transients and harmonic components.

² In (a) (**vertical**), each white node is connected to two gray nodes with a_{north} or a_{south} and in (b) (**horizontal**) with a_{west} or a_{east} . The **grid** topology (d) has couplings in four directions and in (c) (**band**), we use a single a .

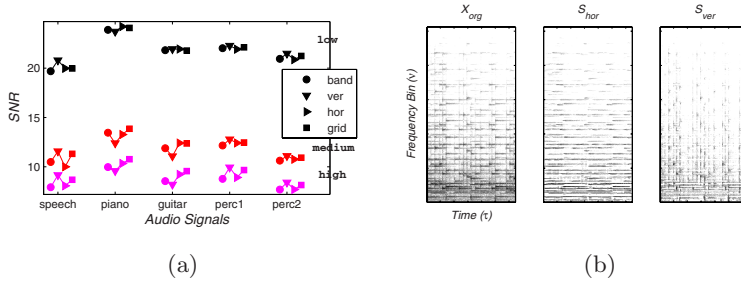


Fig. 3. (a) Signal-to-Noise ratio results for reconstructions obtained from the audio clips in *low*, *medium*, *high* noise conditions. (b) Single channel Source Separation example, left to right, log-MDCT coefficients of the original signal and reconstruction with horizontal and vertical IGMRF models.

3.2 Discussion

We have introduced a conjugate gamma Markov random field model for modelling nonstationary sources. The conjugacy makes it possible to design fast inference algorithms in a straightforward way. The simple MRF topologies considered here are quite generic, yet provide good results without any hand tuned parameters. One can envision more structured models with a larger set of parameters to capture physical reality, certainly for acoustical signals. We are currently investigating further applications such as restoration, transcription or tracking time varying intensity functions.

References

1. Attias, H.: Independent factor analysis. *Neural Computation* 11(4), 803–851 (1999)
2. Cemgil, A.T., Fevotte, C., Godsill, S.J.: Variational and Stochastic Inference for Bayesian Source Separation. *Digital Signal Processing* (in print, 2007)
3. Cichocki, A., Amari, S.I.: *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, revised version edition. Wiley, Chichester (2003)
4. Févotte, C., Daudet, L., Godsill, S.J., Torrèsani, B.: Sparse regression with structured priors: Application to audio denoising. In: *Proc. ICASSP*, Toulouse, France (May 2006)
5. Hyvärinen, A., Karhunen, J., Oja, E.: *Independent Component Analysis*. John Wiley & Sons, New York (2001)
6. Knuth, K.H.: Bayesian source separation and localization. In: *SPIE'98: Bayesian Inference for Inverse Problems*, San diego, pp. 147–158 (July 1998)
7. Liu, J.S.: *Monte Carlo strategies in scientific computing*. Springer, Heidelberg (2004)
8. MacKay, D.J.C.: *Information Theory, Inference and Learning Algorithms*. Cambridge University Press, Cambridge (2003)
9. Miskin, J., Mackay, D.: Ensemble learning for blind source separation. In: Roberts, S.J., Everson, R.M. (eds.) *Independent Component Analysis*, pp. 209–233. Cambridge University Press, Cambridge (2001)

10. Reyes-Gomez, M., Jojic, N., Ellis, D.: Deformable spectrograms. In: AI and Statistics Conference, Barbados (2005)
11. Rowe, D.B.: A Bayesian approach to blind source separation. *Journal of Interdisciplinary Mathematics* 5(1), 49–76 (2002)
12. Shepard, N. (ed.): *Stochastic Volatility, Selected Readings*. Oxford University Press, Oxford (2005)
13. Wainwright, M., Jordan, M.I.: Graphical models, exponential families, and variational inference. Technical Report 649, Department of Statistics, UC Berkeley (September 2003)
14. Wolfe, P.J., Godsill, S.J., Ng, W.J.: Bayesian variable selection and regularisation for time-frequency surface estimation. *Journal of the Royal Statistical Society* (2004)
15. Wolpert, R.L., Ickstadt, K.: Poisson/gamma random field models for spatial statistics. *Biometrika* (1998)