

A Class of Multi-Scale Time Series Models

Marco A. R. Ferreira^{1,2}, Mike West¹, Herbert K. H. Lee¹ and David Higdon^{1,3}

¹*ISDS - Duke University*, ²*DME - UFRJ*, ³*LANL*.

Summary. We introduce a class of multi-scale models for time series. The novel framework couples 'simple' standard Markov models for the time series stochastic process at different levels of aggregation, and links them via 'error' models to induce a new and rich class of structured linear models reconciling modeling and information at different levels of resolution. Jeffrey's rule of conditioning is used to revise the implied distributions and ensure that the probability distributions at different levels are strictly compatible. Our construction has several interesting characteristics: a variety of autocorrelation functions resulting from just a few parameters; the ability to combine information from different scales; and the capacity to emulate long memory processes. There are at least three uses for our multi-scale framework: to integrate the information from data observed at different scales; to induce a particular process when the data is observed only at the finest scale; as a prior for an underlying multi-scale process. Bayesian estimation based on MCMC analysis and forecasting based on simulation are developed. Two interesting applications are presented: in the first application, we illustrate some basic concepts of our multi-scale class of models through the analysis of a series of potential hydroelectric energy; In the second application we use our multi-scale framework to model a series of land temperatures of the northern hemisphere.

1. Introduction

In this paper, we introduce a class of multi-scale models for time series. The construction of these multi-scale models is somewhat subtle: We start with an autoregressive or ARMA process for each level, each level is connected with the immediately finer level through a linear function plus Gaussian noise, and then we use Jeffrey's rule of conditioning (see Jeffrey 1992) to revise the implied distributions and ensure that the probability distributions of the different levels are compatible. The resulting multi-scale class of models has several interesting characteristics: a variety of autocorrelation functions that result from a class of models with just a few parameters; the ability to combine information from different scales; and the ability to emulate long memory processes.

The area of multi-scale modeling has flourished in the last decade. Much of the work has been related to wavelet shrinkage, as for example the work of Vannucci and Corradi (1999), Vidakovic (1998) and Cai and Brown (1998). Another area of intense research has been the use of multi-scale models on dyadic and quad trees in a variety of situations, including: Gaussian processes (Chou *et al.* 1994); Poisson processes (Kolaczyk 1999); as prior for discrete labels in the problem of segmentation of images (Comer and Delp 1999, Laferté *et al.* 2000). All these research have focused mainly on smoothing and classification, and it is well known that these models cannot be successfully used in time series analysis for forecasting. In contrast, our multi-scale class of models is well suited for forecasting because it couples processes evolving at different scales through time.

Concerns about the scale of resolution in which the time series is observed trace back at least to Working (1960), which considered the effect of scaling up a random walk by using a non-overlapping average. More recently, Schmidt and Gamerman (1997) studied

the resulting process of the aggregated time series when the original time series follows a dynamic linear model. Bollerslev and Wright (2000) studied the effect of using data sampled at different frequencies when the finest scale follows a fractionally integrated stochastic volatility model. Thus, a common feature of these and most of previous works is that the model is defined for the finest scale and it is studied the effect of scaling up the series. Conversely, a major characteristic of our approach is that our class of multi-scale time series models is really defined from coarser to finer scales.

The construction of the model from coarser to finer scales has several interesting implications, being adequate for a wide variety of real life situations. This construction is highly appropriate when relevant features of the data can be perceived only at particular scales, more specifically, in cases when the coarser scales have a much simpler representation than the finer scales. In these cases, our multi-scale models take advantage of the simpler representation at coarser levels and allow a very simple representation of the whole process. For example, it can emulate long-range dependence at the finest level by coupling at the coarse level an underlying autoregressive process with high coefficients.

Moreover, this construction can be used to incorporate information observed at different scales of resolution. In particular, when part of the likelihood function is calculated with the help of a computer model that can be run at different levels of resolution, our construction allows the implementation of algorithms that use efficiently the runs at the different scales in order to perform inference on an underlying multi-scale process. One example of such situation can be found in Ferreira et al. (2002) where the proposed multi-scale class of models is used as prior for the permeability field in the context of subsurface hydrology.

The organization of the paper is as follows. In section 2 we use a 2-level model as an example to introduce the basic ideas of the construction of our class of models. In section 3 some properties of the multi-scale models are discussed. We generalize our class of models to an arbitrary number of levels in section 4. In section 5 we introduce our proposed estimation procedure, and the forecasting procedure is discussed in section 6. In section 7 we incorporate seasonality into the multi-scale framework and we discuss the necessary modifications in the estimation and forecasting procedures. Two applications are presented in section 8: in the first application, we illustrate some basic concepts of our multi-scale class of models through the analysis of a series of potential hydroelectric energy; In the second application we use our multi-scale framework to model a series of land temperatures of the northern hemisphere. We conclude with a brief discussion in section 9.

2. Multi-Scale Model Construction: A 2–Level Model

In this section, we present a two-level model in order to introduce the basic ideas of multi-scale modeling.

2.1. General Framework

Begin with a univariate time series x_t , ($t = 1, 2, \dots$), following a specified model. At a general level, write $x_{1:n_x} = \{x_1, \dots, x_{n_x}\}$ for any integer $n_x > 0$, and denote by $p(x_{1:n_x})$ the density of the joint distribution of $x_{1:n_x}$. For example, we may suppose that the x_t follow a standard linear stationary time series model, such as an AR(1) model, that is completely specified; in that case, $p(x_{1:n_x})$ is the implied stationary distribution for the set of n_x values. In other examples, $p(x_{1:n_x})$ could be a posterior predictive distribution arising from a dynamic model conditioned on past observations (as in West and Harrison 1997).

For a specified positive integer m , define the noisy m -block average process y_s on indices $s = 1, 2, \dots$ by

$$y_s = m^{-1} \sum_{i=1}^m x_{(s-1)m+i} + \xi_s. \quad (1)$$

The y values are the averages of non-overlapping groups of m consecutive x values plus noise, such as weekly averages ($m = 7$) of daily data or annual averages ($m = 12$) of monthly data. Refer to the x_t as the *fine level* process, and the y_s as the *coarse level* aggregate process. Then, we can compute the marginal distribution for $y_{1:n_y}$ as $p(y_{1:n_y}) = \int p(y_{1:n_y}|x_{1:n_x})p(x_{1:n_x})dx_{1:n_x}$. The implied process at the coarse level is usually complex.

Conversely, the basic idea underlying our multi-scale construction is to impose a simple process at the coarse level. This is interpreted as a new piece of information G received after $p(x_{1:n_x})$ and $p(y_{1:n_y}|x_{1:n_x})$ are defined. To be specific, suppose that the additional information G relevant to y has as a consequence the revision of the distribution of $y_{1:n_y}$ to $q(y_{1:n_y})$. That is equivalent to the existence of an implicit likelihood function for $y_{1:n_y}$ denoted by $l(y_{1:n_y}|G)$ and computed as $l(y_{1:n_y}|G) \propto q(y_{1:n_y})/p(y_{1:n_y})$.

It is extremely important to note that $p(x_{1:n_x})$, $q(y_{1:n_y})$ and $p(y_{1:n_y}|x_{1:n_x})$ are generally inconsistent. In order to overcome this problem, we use Jeffrey's rule of conditioning to revise the fine scale model $p(x_{1:n_x})$. In this way, we make the models consistent across the different levels of resolution. Diaconis and Zabell (1982) show that the measure defined by Jeffrey's rule is the closest measure to the original one for several common definitions of closeness. Loschi et al. (2002) is the only reference that we know of that uses Jeffrey's rule in a real application, in the context of financial time series. Good references on Jeffrey's rule of conditioning are Jeffrey (1992), Diaconis and Zabell (1982) and Shafer (1981).

In subsection 2.2 we present an example using autoregressive processes of order 1 as building blocks for the construction of a multi-scale time series model.

2.2. An Example

Suppose that the initial fine level model is a standard stationary linear AR(1) model

$$x_t = \phi_x x_{t-1} + \epsilon_t \quad (2)$$

where ϵ_t , ($t = 1, 2, \dots$), is a sequence of mutually uncorrelated zero-mean, normally distributed innovations with $\epsilon_t \sim N(0, \sigma_x^2)$ for some variance σ_x^2 . Then, for all $n_x > 0$, $p(x_{1:n_x})$ is the implied n_x -dimensional stationary distribution

$$p(x_{1:n_x}) = N(x_{1:n_x}|0, V_x) \quad (3)$$

where 0 is the vector of n_x zeroes and V_x is the n_x -square variance matrix with i, j element $\sigma_x^2 \phi_x^{|i-j|} / (1 - \phi_x^2)$.

Suppose that the link between levels is described very simply via the densities $p(y_{1:n_y}|x_{1:n_x})$ (for all n_y) implied by assuming that the y_s are conditionally independent and normally distributed with means $m^{-1} \sum_{i=1}^m x_{(s-1)m+i}$ and constant variance τ .

Fixing $n_x = n_y m$, define the $n_y \times n_x$ matrix A such that $E[y_{1:n_y}] = Ax_{1:n_x}$. Thus A is a sparse matrix whose non-zero elements are all $1/m$; in row i , the non-zero elements are those in columns $(i-1)m+1$ to im . Then, the link equation is defined as follows:

$$p(y_{1:n_y}|x_{1:n_x}) = \prod_{s=1}^{n_y} N(y_s | m^{-1} \sum_{i=1}^m x_{(s-1)m+i}, \tau) = N(y_{1:n_y} | Ax_{1:n_x}, U) \quad (4)$$

where $U = \tau I$ with I as the n_y -square identity matrix and *between-levels* variance τ . It is useful to parameterize τ as a function of $AV_x A'$. Here, we use the parameterization $\tau = \lambda(AV_x A')_{11}$. In particular, when the fine level follows equation (2) and the link equation is given by equation (4), $\tau = \lambda\sigma_x^2(m - m\phi_x^2 - 2\phi_x + 2\phi_x^{m+1})/[m^2(1 - \phi_x)^2(1 - \phi_x^2)]$. The parameter λ has a natural interpretation in terms of the relative increase in uncertainty at the coarse level due to the lack of agreement with the fine level. Because of this natural interpretation, it is much easier to establish a prior for λ than for τ .

Equations (2) and (4) imply an ARMA(1,1) process at the coarse level (Amemiya and Wu 1972). More generally, Amemiya and Wu (1972) proved that if equation (2) denotes an AR(p) process at the fine level, then equations (2) and (4) imply an ARMA(p,q) process at the coarse level, where q is the largest integer satisfying $q(m+1) < (p+1)m+1$.

Suppose that we receive additional information G about the coarse level $y_{1:n_y}$ — the revised model $q(y_{1:n_y})$ is a standard stationary linear AR(1) model

$$y_s = \phi_y y_{s-1} + \eta_s \quad (5)$$

where η_s , ($s = 1, 2, \dots$), is a sequence of mutually uncorrelated zero-mean, normally distributed innovations with variance σ_y^2 . Then, for all $n_y > 0$, $q(y_{1:n_y})$ is the implied n_y -dimensional stationary distribution

$$q(y_{1:n_y}) = N(y_{1:n_y}|0, Q_y) \quad (6)$$

where 0 is the vector of n_y zeroes and Q_y is the n_y -square variance matrix with i, j element $\sigma_y^2 \phi_y^{|i-j|} / (1 - \phi_y^2)$.

Obviously, the densities $p(y_{1:n_y})$ and $q(y_{1:n_y})$ are not compatible. Our approach is to impose $q(y_{1:n_y})$ and to use Jeffrey's rule of conditioning to revise the distribution of $x_{1:n_x}$. It can be shown (see Appendix A) that the resulting distribution $q(x_{1:n_x})$ is also zero-mean normal,

$$q(x_{1:n_x}) = N(x_{1:n_x}|0, Q_x) \quad (7)$$

with variance matrix

$$Q_x = V_x - B(W - Q_y)B' \quad (8)$$

where $W = AV_x A' + U$ and $B = V_x A' W^{-1}$.

The definition of the likelihood function will depend on if the coarse level is observed or not. If the coarse level $y_{1:n_y}$ is observed then the likelihood function is given by $p(x_{1:n_x}|y_{1:n_y})q(y_{1:n_y})$. Computations for the likelihood when $y_{1:n_y}$ is observed are very efficient because the use of simple building blocks for the multi-scale model implies a nice factorization of $p(x_{1:n_x}|y_{1:n_y})$. Conversely, if the coarse level is an unobservable latent process, then the likelihood function is given by equation (7). We avoid the expensive direct computation of Q_x by reparameterizing the model and introducing in the Gibbs sampler the generation of the coarse level $y_{1:n_y}$. In this way the computations for estimation and forecasting become very efficient.

2.3. Two level models with ARMA building blocks

In subsection 2.2, it was presented in detail the construction of a two level model with autoregressive processes of order one as building blocks. In this subsection we discuss the implications of using more general autoregressive moving average (ARMA) processes as building blocks.

From a general viewpoint, the construction of a two level model has three ingredients: the initial processes at the fine level, the link equation and the revised process at the coarse level. With these three ingredients, Jeffrey's rule can be used to update the process at the fine level. Therefore, the construction put forward in subsection 2.2, in particular equations (7) and (8), is quite general and can be used with any well behaved types of processes as building blocks. In particular, a natural choice for building blocks are stationary and invertible ARMA processes.

The main implications of the use of ARMA processes as building blocks are related to the properties of the updated fine level process and to the particular methods necessary to estimate the parameters of the model and perform forecasts for future observations of the fine level series. Some properties of the two level model with AR(1) building blocks are discussed in section 3. The two level model with AR(1) building blocks is very rich in its own, generating a wealthy of different behaviors for the autocorrelation functions. Properties of two level models with more general ARMA building blocks can be studied in the same lines of section 3. With respect to estimation and forecasting, the general idea is the same as the idea presented in sections 5 and 6 for the two level model with AR(1) building blocks: we can take advantage of the factorization of $p(x_{1:n_x} | y_{1:n_y})$ and of the simple form of $q(y_{1:n_y})$ to speed up the computations. Thus, in this paper we focus on AR(1) building blocks, the generalization to ARMA building blocks being relatively straightforward.

3. Some properties of the multi-scale model

In this section, we discuss some very interesting properties of the multi-scale time series class of models. More specifically, we investigate for a two-level model the behavior of the autocorrelation function at the fine level. This investigation is performed in two ways: by visual inspection of the behavior of the autocorrelation function for different values of the parameters; and by analytic study of the limit behavior of the autocorrelation function when the parameters approach certain critical values. As a byproduct of the study of the autocorrelation function, the parameters of the multi-scale model are given suitable interpretations.

An interesting feature of the multi-scale modeling is the emulation of long memory processes (for reference on long memory processes, see Beran 1994). As can be seen in figure 1, the parameter ϕ_y controls the rate of decay of the autocorrelation function, which shows some persistence when the value of ϕ_y increases. It is possible to show that the autocorrelation function at lag j for the limiting case $\lambda = 0$ and $\phi_x = 0$ is proportional to $\phi_y^{\lfloor (j-1)/m \rfloor}$ for all $j \geq m + 1$, where $\lfloor z \rfloor$ denotes the integer part of z . Note that the covariance decay of the multi-scale model is similar to the covariance decay of an AR(1) process with coefficient $\phi_y^{1/m}$. Therefore, the multi-scale class of models that we introduced does not have the long-memory property. Nonetheless, it can still resemble long memory processes. The main advantage of our class of models over actual long memory models is interpretability, that is, the long memory type of behavior is explicitly modeled as a result of a high autoregressive coefficient in the coarser level of the hierarchy.

The values of λ control how much the generator mechanism of the coarse level influences the behavior of the fine level. Thus as can be seen in figure 2, the decay of the autocorrelation function also depends on λ . In that respect, it can be shown analytically that the limit of the autocorrelation function when λ approaches infinity is the autocorrelation function of the

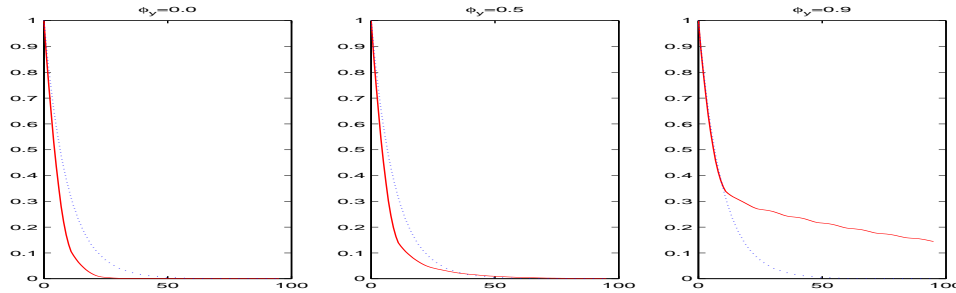


Fig. 1. Autocorrelation functions varying ϕ_y . Parameters kept constant: $\phi_x = 0.9$, $\sigma_x^2 = 1$, $\sigma_y^2 = 1$, $\lambda = 1$ and $m = 12$. Dashed line corresponds to multi-scale model and dotted line corresponds to autoregressive model.

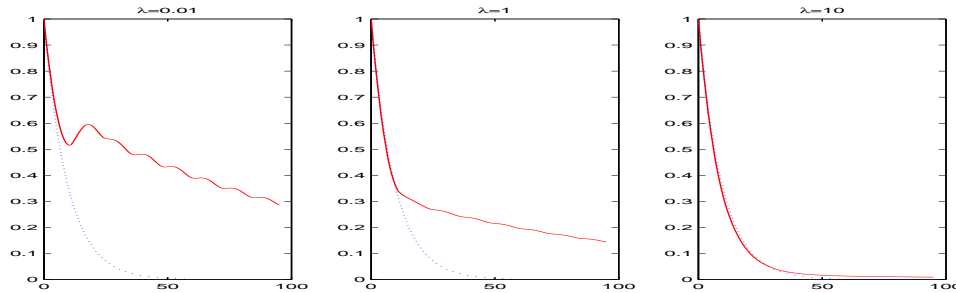


Fig. 2. Autocorrelation functions varying λ . Parameters kept constant: $\phi_x = 0.9$, $\sigma_x^2 = 1$, $\phi_y = 0.9$, $\sigma_y^2 = 1$ and $m = 12$. Dashed line corresponds to multi-scale model and dotted line corresponds to autoregressive model.

original autoregressive process with coefficient ϕ_x . Actually, as can be seen in figure 2, the autocorrelation function of the multi-scale process and of the original autoregressive process are already very close when $\lambda = 10$. When λ gets smaller, the model departs progressively more from the AR(1) model, as we can observe from the autocorrelation functions for $\lambda = 1$ and $\lambda = 0.01$.

The asymptotic behavior of the likelihood function when λ approaches infinity is a key feature of the model and has to be considered very carefully. When λ approaches infinity the likelihood function approaches a constant equal to the likelihood function of an AR(1) process. Thus, an improper prior for λ will lead to an improper posterior. In addition, as shown in figure 2, $\lambda = 10$ and $\lambda \rightarrow \infty$ provide almost the same autocorrelation function for the fine level. Therefore, a solution is to define a prior for lambda with positive mass between zero and ten and mass zero outside of this interval.

Another interesting feature is the existence of a blocking effect that depends on m and ϕ_x . More specifically, when $\phi_x = 0.0$ the autocorrelations are constant by blocks of length m . Additionally, as can be seen in figure 3, when ϕ_x increases the blocking effect is progressively reduced. Therefore, ϕ_x can be interpreted as the parameter that controls the smoothness of the autocorrelation function.

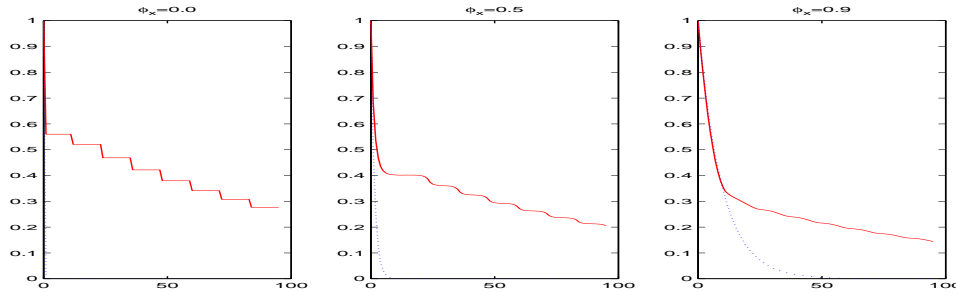


Fig. 3. Autocorrelation functions varying ϕ_x . Parameters kept constant: $\sigma_x^2 = 1$, $\phi_y = 0.9$, $\sigma_y^2 = 1$, $\lambda = 1$ and $m = 12$. Dashed line corresponds to multi-scale model and dotted line corresponds to autoregressive model.

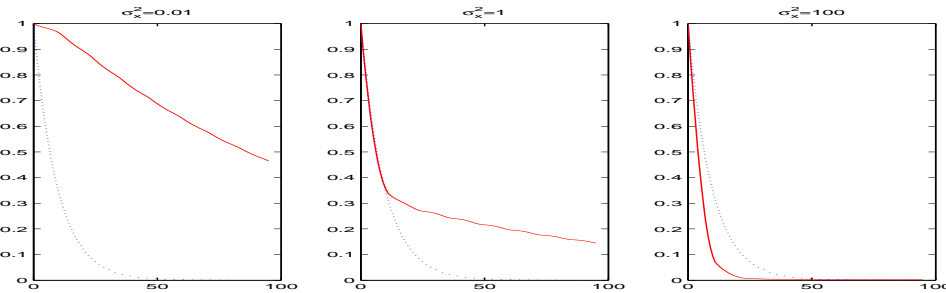


Fig. 4. Autocorrelation functions varying σ_x^2 . Parameters kept constant: $\phi_x = 0.9$, $\phi_y = 0.9$, $\sigma_y^2 = 1$, $\lambda = 1$ and $m = 12$. Dashed line corresponds to multi-scale model and dotted line corresponds to autoregressive model.

The decay of the autocorrelation function is also controlled by the parameters σ_x^2 and σ_y^2 . It is interesting to note that the decay can be faster than the decay of the autocorrelation function of the original AR(1) process. It is possible to show that the autocorrelation function depends on σ_x^2 and σ_y^2 only through σ_x^2/σ_y^2 . However, because the variance of x_t depends on σ_x^2 as well as the ratio σ_x^2/σ_y^2 , there is enough information in $x_{1:n_x}$ to separately estimate σ_x^2 and σ_y^2 .

4. Multi-Scale Model with Several Levels

In this section, we generalize the multi-scale model for any arbitrary number of levels. Instead of using Jeffrey's rule to construct the model, we use a sufficiently richer sample space; Skyrms (1980) has shown that in this way it is possible to obtain the same effect of Jeffrey's rule. This approach is equivalent to carefully conditioning each equation on different information. Besides allowing us to generalize the model for several levels in a straightforward way, conditioning each equation on different information provides a different interpretation for the multi-scale model, enlightening some of its interesting aspects.

4.1. Notation

The i -th level is denoted by $x_{i,1:n_i}$, $i = 1, 2, \dots, L$ with $x_{1,1:n_1}$ being the coarsest level and increases in i meaning finer levels. We can interpret the “movement” from coarser to finer levels as an increase in the resolution in a similar interpretation as in the wavelet framework. However, in our multi-scale framework each level can have a meaningful practical interpretation.

The information on which we are conditioning each equation is of extreme importance, mainly because the equations would be incompatible if they were conditioned on the same information. Here, we denote by I_i the knowledge about the behavior on the i -th level, and by G_i the accumulated knowledge from the coarsest level up to the i -th level $x_{i,1:n_i}$. More specifically, $I_1 = G_1$ denotes the knowledge about the generator mechanism of the coarsest level $x_{1,1:n_1}$ and the accumulated knowledge up to level $x_{1,1:n_1}$ as well. In addition $G_2 = I_1 \cap I_2$. In general, we recursively define $G_i = G_{i-1} \cap I_i$.

4.2. Model

In this subsection we define the model in a hierarchical way; It is very important to note that conditional on all available information the true dependence direction in the model is from coarser to finer levels.

We assume that the level $x_{i,1:n_i}$ given I_i follows a zero mean Gaussian process:

$$p(x_{i,1:n_i} | I_i) = N(x_{i,1:n_i} | 0, V_i). \quad (9)$$

In addition, we assume that there is a linear transformation mapping the level $x_{i,1:n_i}$ to the immediate coarser level $x_{i-1,1:n_{i-1}}$, plus noise. For example, each element of $x_{i-1,1:n_{i-1}}$ could be written as an average of m elements of $x_{i,1:n_i}$ plus noise. Moreover, we suppose that $x_{i-1,1:n_{i-1}}$ is independent of I_i given $x_{i,1:n_i}$. We can express this as follows:

$$p(x_{i-1,1:n_{i-1}} | I_i, x_{i,1:n_i}) = p(x_{i-1,1:n_{i-1}} | x_{i,1:n_i}) = N(x_{i-1,1:n_{i-1}} | A_i x_{i,1:n_i}, U_i) \quad (10)$$

Now, we assume that we receive information G_{i-1} , from another source, about the generator mechanism of the coarser level. More specifically, we assume that given G_{i-1} the coarser level $x_{i-1,1:n_{i-1}}$ follows a zero mean Gaussian process with covariance matrix Q_{i-1} .

In order to obtain the distribution of $x_{i,1:n_i}$ given $G_i = G_{i-1} \cap I_i$, we assume two hypothesis:

- (a) $p(x_{i,1:n_i} | G_i, x_{i-1,1:n_{i-1}}) = p(x_{i,1:n_i} | G_{i-1}, I_i, x_{i-1,1:n_{i-1}}) = p(x_{i,1:n_i} | I_i, x_{i-1,1:n_{i-1}})$
- (b) $p(x_{i-1,1:n_{i-1}} | G_i) = p(x_{i-1,1:n_{i-1}} | G_{i-1}, I_i) = p(x_{i-1,1:n_{i-1}} | G_{i-1})$

Hypothesis (a) states that given the generator mechanism of the finer level, the information G_{i-1} on the generator mechanism of the coarser level passes to the the finer level $x_{i,1:n_i}$ only through the realized coarser level $x_{i-1,1:n_{i-1}}$. Hypothesis (b) states that the coarser level $x_{i-1,1:n_{i-1}}$ is independent of the generator mechanism of the finer level given its own generator mechanism. Hypothesis (b) is equivalent to the use of Jeffrey’s rule, meaning that equations (9) and (10) are relevant only to define $p(x_{i,1:n_i} | I_i, x_{i-1,1:n_{i-1}})$. The distribution of $x_{i,1:n_i} | I_i, x_{i-1,1:n_{i-1}}$ is derived in Appendix B.

Analogously to the 2-level case, it can be shown (see Appendix C) that the resulting distribution $p(x_{i,1:n_i} | G_i)$ is

$$p(x_{i,1:n_i} | G_i) = N(x_{i,1:n_i} | 0, Q_i) \quad (11)$$

with covariance matrix

$$Q_i = V_i - B_i(W_i - Q_{i-1})B_i' \quad (12)$$

where $W_i = A_i V_i A_i' + U_i$ and $B_i = V_i A_i' W_i^{-1}$.

Finally, the whole model is defined recursively by first defining $p(x_{1:1:n_1}|G_1)$ and then using equations (11) and (12) to define $p(x_{i,1:n_i}|G_i), i = 2, \dots, L$.

5. Estimation

In this section we discuss the estimation of the parameters of the multi-scale time series model. The posterior distribution of the parameters is rather complicated due to nonlinearities. In order to explore the posterior distribution, we propose the use of an algorithm based on Markov Chain Monte Carlo (MCMC) techniques to generate a sample from the posterior. This sample is then used to estimate summaries of the posterior distribution such as posterior means, standard deviations and credible intervals.

When all the levels are observed, the parameters corresponding to different levels can be estimated separately because they are independent conditional on the levels. If some of the levels are unknown, then it is necessary to include their simulation in the Gibbs sampler.

Because the estimation of the parameters of different levels can be performed separately, it is straightforward to generalize the estimation procedure from the two-level model to models with an arbitrary number of levels. Therefore, in this section we consider estimation only for the two level model.

We assume the following priors for the parameters of the model: $\phi_y \sim N(m_{\phi_y}, S_{\phi_y})$, $\sigma_y^2 \sim IG(\nu_{\sigma_y}/2, \nu_{\sigma_y} s_{\sigma_y}/2)$, $\phi_x \sim TrN_{(-1.0, 1.0)}(m_{\phi_x}, S_{\phi_x})$, $\sigma_x^2 \sim IG(\nu_{\sigma_x}/2, \nu_{\sigma_x} s_{\sigma_x}/2)$ and $\lambda \sim TrIG_{(0, 10)}(\nu_{\lambda}/2, \nu_{\lambda} s_{\lambda}/2)$, where $TrN_{(a, b)}$ denotes the normal distribution truncated to the interval (a, b) and $TrIG_{(c, d)}$ denotes the inverse gamma distribution truncated to the interval (c, d) .

The estimation procedure depends on whether the coarse level is observed or not observed. When the coarse level is observed the estimation procedure is simpler. If the coarse level is not observed we can still use part of the estimation procedure by reparameterizing the model by the inclusion of a latent coarse level. Thus, in any of the two cases the joint distribution of the random quantities involved is:

$$p(x_{1:n_x} | y_{1:n_y}, \phi_x, \sigma_x^2, \lambda) q(y_{1:n_y} | \phi_y, \sigma_y^2) p(\phi_y) p(\sigma_y^2) p(\phi_x) p(\sigma_x^2) p(\lambda).$$

From the joint distribution, it is easy to verify that conditional on the coarse level the generation of the parameters corresponding to different levels can be done separately. More specifically, the general full conditional distributions for the parameters of the different levels are the following:

$$p(\phi_y, \sigma_y^2 | y_{1:n_y}) \propto p(\phi_y) p(\sigma_y^2) q(y_{1:n_y} | \phi_y, \sigma_y^2) \quad (13)$$

$$p(\phi_x, \sigma_x^2, \lambda | x_{1:n_x}, y_{1:n_y}) \propto p(\phi_x) p(\sigma_x^2) p(\lambda) p(x_{1:n_x} | y_{1:n_y}, \phi_x, \sigma_x^2, \lambda). \quad (14)$$

Moreover, if the coarse level follows an ARMA process, we can generate very easily the parameters correspondent to the coarse level, for example by using the procedures proposed by Chib and Greenberg (1994).

The estimation of the parameters of the fine level and link equation is not so easy. In general, the full conditional distributions are not available in closed form for sampling. Here, we propose the use of Metropolis-Hastings steps for the generation of these parameters.

The generation of ϕ_x , σ_x^2 and λ can be efficiently performed because the conditional distribution of $x_{1:n_x}$ given $y_{1:n_y}$ factorizes very nicely: $p(x_{1:n_x}|y_{1:n_y}, \phi_x, \sigma_x^2, \lambda) = p(x_{1:n_x}|\phi_x, \sigma_x^2) p(y_{1:n_y}|x_{1:n_x}, \phi_x, \sigma_x^2, \lambda) / p(y_{1:n_y}|\phi_x, \sigma_x^2, \lambda)$. Moreover, the computation of $p(y_{1:n_y}|\phi_x, \sigma_x^2, \lambda)$ can be done very efficiently by the Kalman filter (for reference on the Kalman filter, see West and Harrison 1997). After the generation of each proposal for ϕ_x , σ_x^2 and λ , the factorization of $p(x_{1:n_x}|y_{1:n_y}, \phi_x, \sigma_x^2, \lambda)$ is used to compute the probability of acceptance of the proposal.

The proposal for σ_x^2 is generated from $U(\sigma_x^{2(oid)}/\delta_{\sigma_x}, \sigma_x^{2(oid)}\delta_{\sigma_x})$, where δ_{σ_x} has to be tuned to yield a reasonable acceptance rate.

We generate a proposal for ϕ_x from $U[\max(-1.0, \phi_x^{(oid)} - \delta_{\phi_x}), \min(1.0, \phi_x^{(oid)} + \delta_{\phi_x})]$, where δ_{ϕ_x} has to be tuned to yield a reasonable acceptance rate. Note that the proposal is not symmetric and therefore this is not a Metropolis step.

The generation of λ depends on whether the coarse level is observed or not. If the coarse level is observed, then a proposal for λ is generated from $U(\max(0, \lambda^{(oid)} - \delta_\lambda), \min(1, \lambda^{(oid)} + \delta_\lambda))$, where δ_λ has to be tuned to yield a reasonable acceptance rate.

Conversely, if the coarse level is not observed then there is a high posterior correlation between the coarse level $y_{1:n_y}$ and λ . Thus, it is good practice to generate $y_{1:n_y}$ and λ jointly. We perform this by first generating a proposal $\lambda^{(new)}$ for λ from $U(\max(0, \lambda^{(oid)} - \delta_\lambda), \min(1, \lambda^{(oid)} + \delta_\lambda))$. Then, we generate a proposal for $y_{1:n_y}$ from its full conditional on $\lambda^{(new)}$. The full conditional of $y_{1:n_y}$ is $N(y_{1:n_y}|C_y^{-1}U^{-1}Ax_{1:n_x}, C_y)$, where $C_y = Q_y^{-1} + U^{-1} - W^{-1}$. The joint proposal is accepted or rejected with the appropriate probability.

As illustrated in section 8, the MCMC-based approach presented here works very well in practice.

6. Forecasting future observations

In this section, we discuss the issue of forecasting future observations at the coarse and fine levels. In order to perform the forecasts, it is sufficient to obtain a sample from the predictive distribution of the future observations at the different levels. Point and interval forecasts can then be derived from this sample.

In order to generate each realization of the sample from the predictive distribution, we use a two-stage procedure. First, we generate a future realization of the coarse level conditional on the past. After that, we generate a realization of the fine level conditional on the past and on the realization of the coarse level. As stated by the theorems presented in this section, we sample from those distributions in a very efficient way.

Some additional notation is necessary. Let $r = x_{n_x-m}(\phi_x, \phi_x^2, \dots, \phi_x^m)'$ and $R_{ij} = \sigma_x^2 \phi_x^{|i-j|} (1 - \phi_x^{2 \min(i,j)}) / (1 - \phi_x^2)$ respectively be the m -step ahead predictive mean vector and covariance matrix for the original AR(1) process at the fine level. Moreover, let us denote by $p(y_s|y_{1:s-1}) = N(y_s|p_s, P_s)$ the *original* distribution of y_s given the values of y up to time $s-1$. It is important to note that $p(y_s|y_{1:s-1})$ can be efficiently computed by the Kalman filter.

The generation of predictions for the coarse level is facilitated by theorem 6.1. The theorem states that the one-step ahead predictive distribution at the coarse level depends

on the whole past at the coarse level but depends only on the last observation at the fine level:

$$\text{THEOREM 6.1. } q(y_{n_y}|y_{1:(n_y-1)}, x_{1:(n_x-m)}) = q(y_{n_y}|y_{1:(n_y-1)}, x_{n_x-m}).$$

In practice, the dependence on the past coarse level falls reasonably fast with time. Theorem 6.1 is proved in appendix D.

The following corollary holds when AR(1) processes are used as building blocks for the two level model:

$$\text{COROLLARY 6.1. } \textit{Assume conditions of subsection 2.2. Then } q(y_{n_y}|y_{1:(n_y-1)}, x_{1:(n_x-m)}) = N(y_{n_y}|f_y, F_y) \textit{ where } F_y = [\sigma_y^{-2} + (m^{-2}\bar{\Gamma}'R\bar{\Gamma} + \tau)^{-1} - P_{n_y}^{-1}]^{-1} \textit{ and } f_y = F_y[\sigma_y^{-2}\phi_y y_{n_y-1} + m^{-1}(m^{-2}\bar{\Gamma}'R\bar{\Gamma} + \tau)^{-1}\bar{\Gamma}'r - P_{n_y}^{-1}p_{n_y}].$$

The following theorem simplifies the task of forecasting the fine level when the future coarse level is known:

$$\text{THEOREM 6.2. } p(x_{(n_x-m+1):n_x}|y_{1:n_y}, x_{1:(n_x-m)}) = p(x_{(n_x-m+1):n_x}|y_{n_y}, x_{n_x-m}).$$

Theorem 6.2 is proved in appendix E.

The following corollary holds when AR(1) processes are used as building blocks for the two level model:

$$\text{COROLLARY 6.2. } \textit{Assume conditions of subsection 2.2. Then } x_{(n_x-m+1):n_x}|y_{n_y}, x_{1:(n_x-m)} \sim N(f_x, F_x), \textit{ where } F_x = R - m^{-2}R\bar{\Gamma}(m^{-2}\bar{\Gamma}'R\bar{\Gamma} + \tau)^{-1}\bar{\Gamma}'R \textit{ and } f_x = r + m^{-1}R\bar{\Gamma}(m^{-2}\bar{\Gamma}'R\bar{\Gamma} + \tau)^{-1}(y_{n_y} - \bar{\Gamma}'r).$$

7. Incorporating seasonality into the multi-scale model

In this section, we discuss how to incorporate seasonality in the multi-scale model at the fine scale. We assume that the coarsening operation eliminates the seasonal pattern. In particular, the seasonality is modeled by the inclusion at the fine scale of regressors corresponding to harmonics. Harmonics are sine and cosine functions with mean over a specified period equal to zero, implying trivially the vanishing of the seasonal pattern at the coarse scale.

The construction of the model is the same as in section 2, except that we substitute equation (3) by:

$$x_{1:n_x} \sim N(x_{1:n_x}|\mu_x, V_x)$$

where μ_x is a periodic vector with period m , and V_x is the covariance matrix of an AR(1) process. Moreover, we assume that $A\mu_x = 0$, that is, we assume that the seasonal effects vanish with the coarsening, which is typical in time series analysis. In this work, we choose to define $\mu_x = Z\beta$, where Z is a design matrix corresponding to the harmonics. As a consequence of the very definition of harmonics, $AZ = 0$. As in the case without seasonality, we assume the following link equation:

$$p(y_{1:n_y}|x_{1:n_x}) = N(y_{1:n_y}|Ax_{1:n_x}, U).$$

Let us update the distribution of $y_{1:n_y}$ to $q(y_{1:n_y}) = N(y_{1:n_y}|0, Q_y)$ where Q_y is the covariance matrix of an AR(1) process. Then, the updated marginal distribution of $x_{1:n_x}$ is

$$q(x_{1:n_x}) = N(x_{1:n_x}|Z\beta, V_x - B(W - Q_y)B')$$

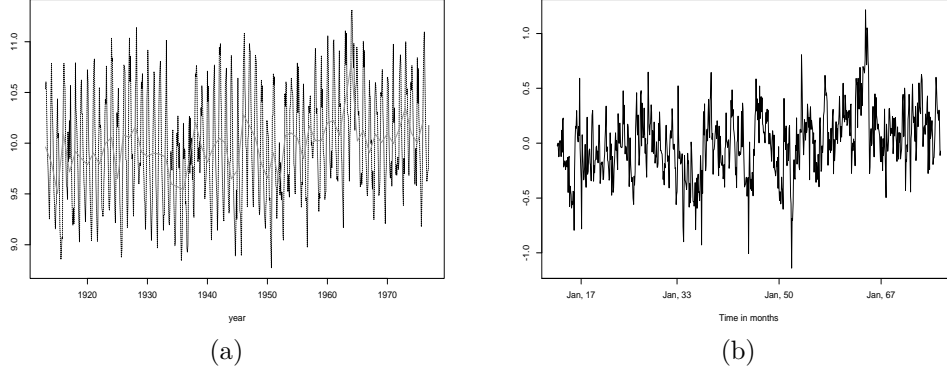


Fig. 5. Potential hydroelectric energy. (a) Log of the monthly potential (dotted) and annual averages (solid). (b) Monthly residuals after extracting the seasonality and the mean

where $B = V_x A' W^{-1}$ and $W = A V_x A' + U$ are the same as in the subsection 2.2.

In the next subsection we discuss the issue of estimation and forecasting. In subsection 8.1 we present an example of the application of this model to the analysis of potential hydroelectric energy.

7.1. Estimation and forecasting

As in the case without seasonality, we use an MCMC algorithm to generate a sample from the posterior distribution of the parameters.

The full conditional for β is $N(m_\beta^*, C_\beta^*)$ where $C_\beta^* = (C_\beta^{-1} + Z' V x^{-1} Z)^{-1}$ and $m_\beta^* = C_\beta^* (C_\beta^{-1} m_\beta + Z' V x^{-1} x_{1:n_x})$.

The full conditionals for ϕ_y , σ_y^2 and τ are the same as in the case without seasonality. The full conditionals for ϕ_x and σ_x^2 are analogous to the case without seasonality, but substituting $x_{1:n_x}$ by $x_{1:n_x}^* = x_{1:n_x} - Z\beta$.

The forecasting procedure in the case with seasonality is analogous to the non-seasonality case. More specifically, predictions are made for the deseasonalized series $x_{(n_x+1):(n_x+l)}^*$, and the seasonality is added to these predictions.

Therefore, the use of harmonics allows a straightforward extension of the multi-scale class of models to the case with seasonality. When compared with the non-seasonality case, the tasks of estimation and forecasting can be performed with minor modifications.

8. Applications

8.1. Analysis of potential hydroelectric energy

In this subsection, we present the analysis of the series of potential hydroelectric energy of the rivers of the southeast region of Brazil as an example of the application of the two-level multi-scale model with known coarse level and seasonality in the fine level.

Figure 5 (a) presents the plot of the log of the monthly potential hydroelectric energy of the southeast region of Brazil from January of 1936 to December of 2000, as well as

Table 1. Potential hydroelectric energy - Posterior summaries for the parameters

	Multi-scale model		AR(1) model	
	Mean	Standard deviation	Mean	Standard deviation
ϕ_y	0.3690	0.1310	-	-
σ_y^2	0.0992	0.0412	-	-
ϕ_x	0.4914	0.0402	0.7190	0.0256
σ_x^2	0.0371	0.0022	0.0435	0.0022
τ	0.0057	0.0030	-	-
λ	0.5278	0.2594	-	-
β_1	0.2525	0.0161	0.2530	0.0212
β_2	0.5591	0.0161	0.5593	0.0207
β_3	0.0432	0.0118	0.0428	0.0124
β_4	0.0799	0.0114	0.0801	0.0120

the series of annual averages. The presence of seasonality is quite obvious in the monthly series. After an exploratory analysis, we verified that the first and fourth harmonics are enough to explain most part of the seasonal behavior. Figure 5 (b) shows the plot of the monthly residuals after extracting the seasonality and the overall mean. The autocorrelation and partial autocorrelation functions of the monthly residuals suggest a long memory type process.

In addition, it is evident from figure 5 that there is strong dependence between years. That would probably cause a simple ARMA model with seasonality for the monthly series to give poor medium term forecasts. Alternatively, we model the potential hydroelectric energy series with the two-level multi-scale time series class introduced in section 2. More specifically, the coarse level is the annual mean of the log of the mean monthly flow, and the fine level is the log of the mean monthly flow. The seasonality is modeled using harmonics as proposed in section 7.

We used the MCMC algorithm described in section 5, with 5000 iterations, reaching convergence as soon as after 1000 iterations. So, we considered a burn-in of 1000 iterations and kept 4000 iterations of the Gibbs sampling. The tuning parameters were adjusted in order to yield reasonable acceptance rates for ϕ_x , σ_x^2 and λ , equal respectively to 46%, 53% and 43%.

In order to check the predictive capacity of the model, we fit an AR(1) model, the $p(x)$ model. We used the data until 1997 to fit both models, and then we computed the three years ahead forecasts. The multi-scale model performed much better than the AR(1) model in terms of mean squared prediction error, equal to 0.0548 and 0.0621 for the multi-scale and the AR(1) model respectively, a decrease of 11.8%. Furthermore, the mean squared prediction error is likely to decrease if we have a model for the annual series with better predictive capacity. For example, we could have a model for the coarse level based on the knowledge of the physical process of the flow of the rivers in the southeast region of Brazil. Just as an indication for the potential increase in the predictive capacity, the mean squared prediction error would decrease by around 45.4% to 0.0339 if the future coarse level was known.

Finally, figure 7 shows the observed potential hydroelectric energy and the predictive mean based on the multi-scale model for the years 1998, 1999 and 2000 using only the observations until 1997, as well as the 95% predictive interval. As we can see in figure 7, the model performs very well in terms of predictive capacity.

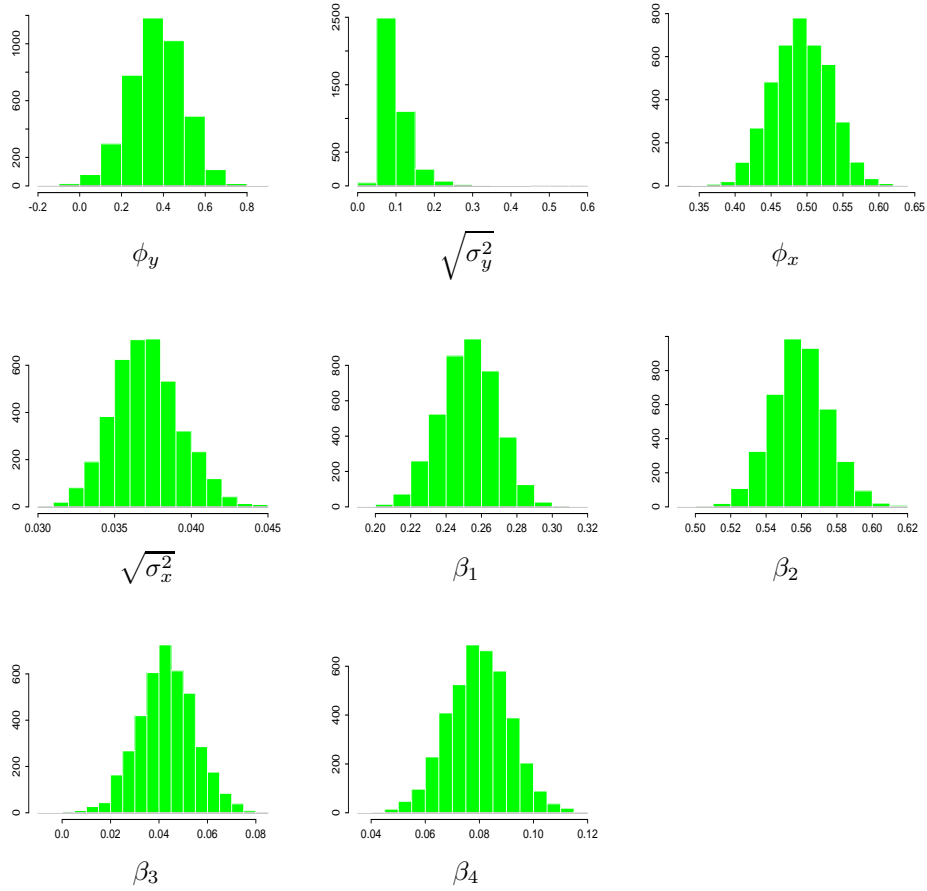


Fig. 6. Potential hydroelectric energy. Histogram of parameters

8.2. Analysis of temperatures of the Northern Hemisphere

In this subsection, we present the analysis of the series of annual land only average temperatures of the Northern Hemisphere as an example of an application of the two-level multi-scale model with unknown coarse level. Figure 8 shows the plot of annual land only average temperatures of the Northern Hemisphere from 1851 to 1986, from the report of the Intergovernmental Panel on Climate Change Houghton et al. (1990). Similar datasets were analyzed by Smith (1993) and Petris and West (1998) using long range dependence models: Smith (1993) analyzed the series of monthly overall average temperatures of the Northern Hemisphere, and Petris and West (1998) analyzed the monthly overall average temperatures of the Southern Hemisphere. Here we use our two-level multi-scale model in order to emulate long range dependence, assuming that the fine level is the annual time scale, and that there is an *unobservable* coarse level with coarsening window $m = 4$.

In order to fit the multi-scale model, the MCMC algorithm described in section 5 was used with 5000 iterations, and convergence was obtained within 1000 iterations. After convergence, the acceptance rates for ϕ_x , σ_x^2 and λ were equal to 0.60, 0.46 and 0.54,

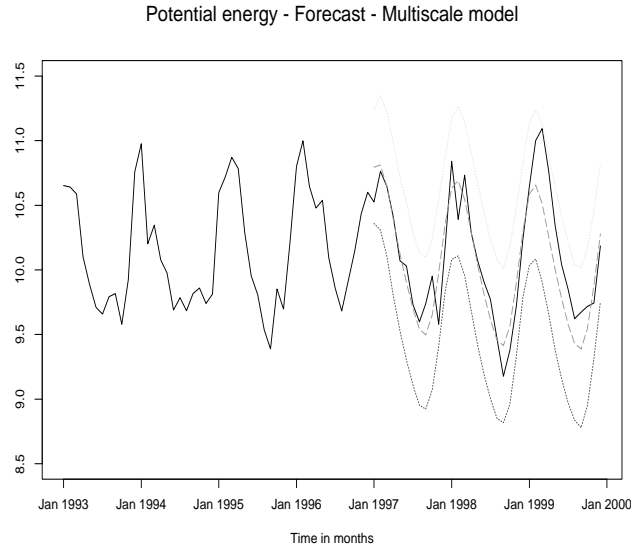


Fig. 7. Observed potential hydroelectric energy (solid), forecast (dashed) and predictive interval (dotted).

Table 2. Northern hemisphere temperatures - Posterior summaries for the parameters

	Multi-scale model		AR(1) model	
	Mean	Standard deviation	Mean	Standard deviation
ϕ_y	0.8407	0.0955	-	-
σ_y^2	0.0235	0.0164	-	-
ϕ_x	0.1014	0.1284	0.4843	0.0816
σ_x^2	0.0352	0.0058	0.0465	0.0063
τ	0.0064	0.0042	-	-
λ	0.5201	0.2540	-	-

respectively.

Figure 9 shows the posterior histograms of the parameters of the multi-scale model. The posterior distribution of ϕ_y is very skewed, mainly because of the restrictions $-1 < \phi_y < 1$. The posterior distributions of σ_y^2 , ϕ_x , σ_x^2 , τ and λ are well behaved.

Table 8.2 presents posterior means and standard deviations for the parameters of both the multi-scale and the AR(1) models. There is a big difference between the posterior means for ϕ_x in the two models. This is expected when the multi-scale model is sufficiently different from the AR(1) model such that ϕ_x has different interpretations in each of the models. The impact of the multi-scale model can also be noticed by the magnitude of the posterior mean of ϕ_y , that is reasonably close to one. In addition, the posterior mean for τ is very small imposing a high degree of agreement between coarse and fine levels. Therefore, the introduction of an underlying coarse level seems to have a big impact in the process at the fine level.

In order to check the predictive capacity of the model, we compared it with autoregressive models. We used the data until 1970 to fit both models, and then we computed the forecasts

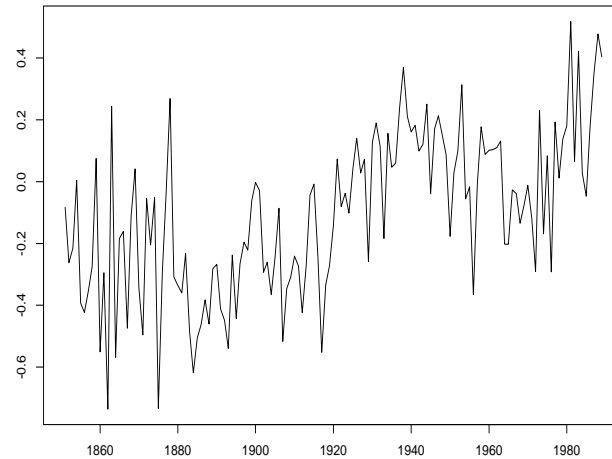


Fig. 8. Annual average land temperatures of the Northern Hemisphere.

from 1971 to 1986. When compared with the original AR(1) model, the multi-scale model reduces the mean squared prediction error from 0.0936 to 0.0754, a reduction of about 19.4%. When compared with the best autoregressive model in this application, the AR(5), the reduction is of about 2.8%. In addition, our multi-scale model is more parsimonious, it has one parameter less than the AR(5) model, and therefore it is preferable. Finally, figure 10 shows the observed annual average temperature and the forecast based on the multi-scale model for the period from 1971 to 1986 as well as the 95% predictive interval. The forecast was computed as the predictive mean. The figure suggests that maybe it would be adequate to consider a linear trend in the model, which will be the subject of future investigation.

9. Discussion and future directions

In this paper, we introduced a new class of Gaussian multi-scale time series models. We used Jeffrey's rule in order to build the two level model, and we generalized that idea using a richer sample space to build models with an arbitrary number of levels. We derived several important properties of this class of models and incorporated seasonality through the use of harmonics. We introduced MCMC-based estimation and forecasting procedures. Finally, we presented three applications in order to illustrate the potential applicability of our class of models.

Our multi-scale time series class of models is of potential use in problems that have information available at different levels of resolution. For example, some problems in ecology have information at daily, monthly and annual basis, and researchers have analyzed them only at a particular arbitrary scale. In these cases, our class of models would provide a natural framework of analysis.

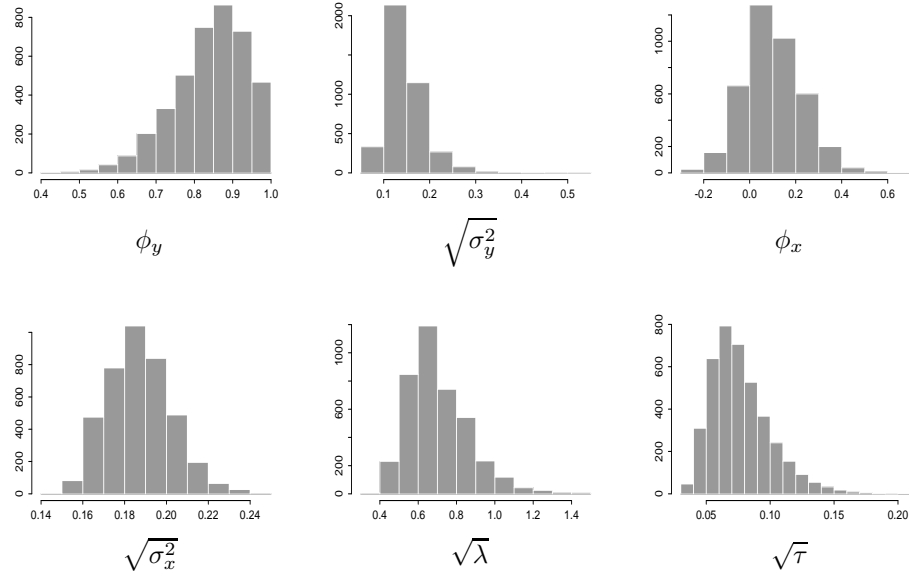


Fig. 9. Northern land temperatures. Histograms of ϕ_y , $\sqrt{\sigma_y^2}$, ϕ_x , $\sqrt{\sigma_x^2}$, $\sqrt{\lambda}$ and $\sqrt{\tau}$

In our class of models, there is a multitude of possible future directions to be pursued. For example, a more detailed study of different classes of processes at the several levels of resolution would be very welcome. This issue only reinforces the necessity of investigation on model selection. Another very important open question when the coarse level is unknown is what is the aggregation that implies the simplest model at the coarse level.

We have worked on the use of this class of multi-scale models as a prior in a problem of hydrology (Ferreira *et al.* 2002). More specifically, we use a multi-scale model as a prior for a permeability field in the problem of 1-D fluid flow through porous media. This allows the incorporation of information obtained from from multiple sources and available at different scales of resolution in the inference about the permeability field.

In addition, we are working on the extension of this class of multi-scale models to the two dimensional case. These will lead to a new class of multi-scale random field models. For that case, theoretical results that increase the velocity of the computations are of great interest. These extensions will probably be very useful, for example, for petroleum reservoir problems (Lee *et al.* 2002) and climate modeling (Xie and Arkin 1996, Brandt and Zaslavsky 1997).

10. Acknowledgements

This work was supported by National Science Foundation grant DMS 9873275. Marco A. R. Ferreira's graduate studies were partially supported by a grant from CNPq, Brazil. We would like to thank Ajax Monteiro, IPEA, for the data on the potential hydroelectric energy of the brazilian southeast region.

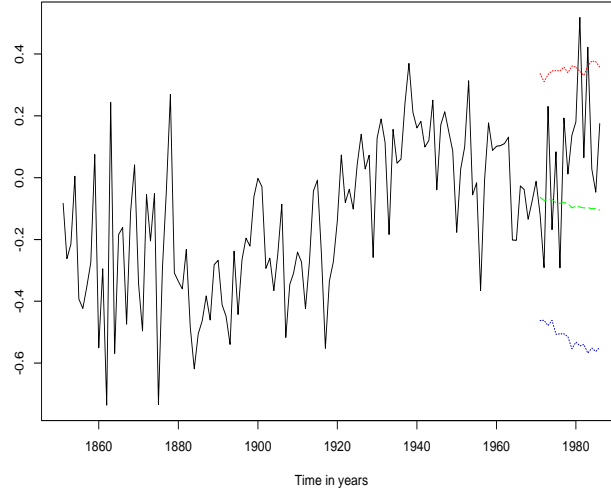


Fig. 10. Northern land temperatures. Observed series (solid), forecast (dashed) and predictive interval (dotted).

A. Deriving the marginal of the fine level in the two level model

By Bayes Theorem:

$$\begin{aligned} p(x_{1:n_x} | y_{1:n_y}) &\propto p(x_{1:n_x}) p(y_{1:n_y} | x_{1:n_x}) \\ &= N(x_{1:n_x} | 0, V_x) N(y_{1:n_y} | Ax_{1:n_x}, U) \end{aligned}$$

Therefore, by linear regression:

$$x_{1:n_x} | y_{1:n_y} \sim N(By_{1:n_y}, V_x - BWB')$$

where $B = V_x A' W^{-1}$ and $W = AV_x A' + U$. So, as $q > 0$ and $p > 0$, we can use the generalized Jeffrey's rule of Diaconis and Zabel (1982) to derive the marginal distribution of $x_{i,1:n_x}$:

$$\begin{aligned} q(x_{i,1:n_x}) &= \int p(x_{i,1:n_x} | y_{1:n_y}) q(y_{1:n_y}) dy_{1:n_y} \\ &\propto \int N(x_{1:n_x} | By_{1:n_y}, V_x - BWB') \\ &\quad N(y_{1:n_y} | 0, Q_y) dy_{1:n_y} \end{aligned}$$

Therefore, $q(x_{1:n_x}) = N(0, Q_x)$ where $Q_x = V_x - B(W - Q_y)B'$.

B. Distribution of $x_{i,1:n_i}$ given I_i and $x_{i-1,1:n_{i-1}}$

Using the fact that $x_{i-1,1:n_{i-1}}$ is independent of I_i given $x_{i,1:n_i}$, we have:

$$\begin{aligned} p(x_{i,1:n_i} | I_i, x_{i-1,1:n_{i-1}}) &\propto p(x_{i,1:n_i} | I_i) p(x_{i-1,1:n_{i-1}} | x_{i,1:n_i}, I_i) \\ &= p(x_{i,1:n_i} | I_i) p(x_{i-1,1:n_{i-1}} | x_{i,1:n_i}) \\ &\propto N(x_{i,1:n_i} | 0, V_i) N(x_{i-1,1:n_{i-1}} | A_i x_{i,1:n_i}, U_i) \end{aligned} \quad (15)$$

Therefore, by Bayesian linear regression, $x_{i,1:n_i} | I_i, x_{i-1,1:n_{i-1}} \sim N(B_i x_{i-1,1:n_{i-1}}, V_i - B_i W_i B_i')$, where $B_i = V_i A_i' W_i^{-1}$ and $W_i = A_i V_i A_i' + U_i$.

C. Deriving the distribution of $x_{i,1:n_i}$ given G_i

Using conditions a and b we have:

$$\begin{aligned} p(x_{i,1:n_i} | G_i) &= \int p(x_{i,1:n_i} | G_i, x_{i-1,1:n_{i-1}}) p(x_{i-1,1:n_{i-1}} | G_i) dx_{i-1,1:n_{i-1}} \\ &= \int p(x_{i,1:n_i} | I_i, x_{i-1,1:n_{i-1}}) p(x_{i-1,1:n_{i-1}} | G_{i-1}) dx_{i-1,1:n_{i-1}} \\ &\propto \int N(x_{i,1:n_i} | B_i x_{i-1,1:n_{i-1}}, V_i - B_i W_i B_i') N(x_{i-1,1:n_{i-1}} | 0, Q_{i-1}) dx_{i-1,1:n_{i-1}} \end{aligned}$$

Therefore, $x_{i,1:n_i} | G_i \sim N(0, Q_i)$ where $Q_i = V_i - B_i(W_i - Q_{i-1})B_i'$.

D. On the predictive distribution at the coarse level

Proof of theorem 6.1:

$$\begin{aligned} q(y_s | y_{1:s-1}, x_{1:[m(s-1)]}) &\propto \int q(y_{1:s}) p(x_{1:m_s} | y_{1:s}) dx_{(m_s-m+1):m_s} \\ &\propto q(y_s | y_{s-1}) \int \frac{p(x_{1:m_s}) p(y_{1:s} | x_{1:m_s})}{p(y_{1:s})} dx_{(m_s-m+1):m_s} \\ &\propto \frac{q(y_s | y_{s-1})}{p(y_{1:s})} \int p(x_{(m_s-m+1):m_s} | x_{m(s-1)}) p(y_s | x_{(m_s-m+1):m_s}) dx_{(m_s-m+1):m_s} \\ &\propto \frac{q(y_s | y_{s-1})}{p(y_s | y_{1:(s-1)})} p(y_s | x_{m(s-1)}) \end{aligned}$$

E. On the predictive distribution at the fine level

In order to prove theorem 6.2, we just need to use the fact that $q(x_{1:n_x} | y_{1:n_y}) = p(x_{1:n_x} | y_{1:n_y})$, meaning that the conditional distribution of the fine level given the coarse level is not revised by Jeffrey's rule. Then:

$$\begin{aligned} q(x_{1:n_x} | y_{1:n_y}) &= p(x_{1:n_x} | y_{1:n_y}) \propto p(x_{1:n_x}) p(y_{1:n_y} | x_{1:n_x}) \\ &\propto p(x_1) \left[\prod_{i=2}^{n_x} p(x_i | x_{i-1}) \right] \left[\prod_{j=1}^{n_y} p(y_j | x_{(m_j-m+1):(m_j)}) \right] \end{aligned} \quad (16)$$

Therefore:

$$\begin{aligned}
 p(x_{(n_x-m+1):n_x} | y_{1:n_y}, x_{1:(n_x-m)}) &\propto p(x_{1:n_x} | y_{1:n_y}) \\
 &\propto \left[\prod_{i=n_x-m+1}^{n_x} p(x_i | x_{i-1}) \right] p(y_{n_y} | x_{(n_x-m+1):n_x}) \\
 &\propto p(x_{(n_x-m+1):n_x} | y_{n_y}, x_{n_x-m}) \tag{17}
 \end{aligned}$$

References

- Amemiya, T. and Wu, R. Y. (1972). The effect of aggregation on prediction in the autoregressive model. *Journal of the American Statistical Association*, 67(339):628–632.
- Beran, J. (1994). *Statistical Methods for Long Memory Processes*. Chapman & Hall.
- Bollerslev, T. and Wright, J. H. (2000). Semiparametric estimation of long-memory volatility dependencies: The role of high-frequency data. *Journal of Econometrics*, 98:81–106.
- Brandt, A. and Zaslavsky, L. Y. (1997). Multiscale algorithm for atmospheric data assimilation. *SIAM Journal on Scientific Computing*, 18:949–956.
- Cai, T. T. and Brown, L. D. (1998). Wavelet shrinkage for nonequispaced samples. *The Annals of Statistics*, 26:1783–1799.
- Chib, S. and Greenberg, E. (1994). Bayes inference in regression models with arma(p,q) errors. *Journal of Econometrics*, 64:183–206.
- Chou, K. C., Willsky, A. S., and Benveniste, A. (1994). Multiscale recursive estimation, data fusion, and regularization. *IEEE Transactions on Automatic Control*, 39:464–478.
- Comer, M. and Delp, E. (1999). Segmentation of textured images using a multiresolution gaussian autoregressive model. *IEEE Transactions on Image Processing*, 8:408–420.
- Diaconis, P. and Zabell, S. L. (1982). Updating subjective probability. *Journal of the American Statistical Association*, 77(380):822–830.
- Ferreira, M. A. R., Bi, Z., West, M., Lee, H., and Higdon, D. (2002). Multi-scale modeling of 1-d permeability fields. Technical report, Duke University - ISDS.
- Houghton, J. T., Jenkins, G. J., and Ephraums, J. J., editors (1990). *Climate change: the IPCC scientific assessment*. Cambridge University Press. Intergovernmental Panel on Climate Change.
- Jeffrey, R. C. (1992). *Probability and the art of judgement*. Cambridge University Press, New York.
- Kolaczyk, E. D. (1999). Bayesian multiscale models for poisson processes. *Journal of the American Statistical Association*, 94:920–933.
- Laferté, J.-M., Pérez, P., and Heitz, F. (2000). Discrete markov image modeling and inference on the quadtree. *IEEE Transactions on Image Processing*, 9(3):390–404.

- Lee, H., Higdon, D., Bi, Z., Ferreira, M. A. R., and West, M. (2002). Markov random field models for high-dimensional parameters in simulations of fluid flow in porous media. *Technometrics*. To appear.
- Loschi, R. H., Iglesias, P. L., and Arellano-Valle, R. B. (2002). Conditioning on uncertain event: Extensions to bayesian inference. *Test*, 11:1–29.
- Petris, G. and West, M. (1998). Bayesian time series modelling with long-range dependence. Technical Report 686, Carnegie Mellon University.
- Schmidt, A. M. and Gamerman, D. (1997). Temporal aggregation in dynamic linear models. *Journal of Forecasting*, 16:293–310.
- Shafer, G. (1981). Jeffrey’s rule of conditioning. *Philosophy of Science*, 48:337–362.
- Skyrms, B. (1980). *Causal Necessity*. New Haven: Yale University Press. Appendix 2.
- Smith, R. L. (1993). Long-range dependence and global warming. In Barnett, V. and Turkman, F., editors, *Statistics for the Environment*, pages 141–161. Wiley, Chichester.
- Vannucci, M. and Corradi, F. (1999). Covariance structure of wavelet coefficients: Theory and models in a Bayesian perspective. *Journal of the Royal Statistical Society, Series B, Methodological*, 61:971–986.
- Vidakovic, B. (1998). Nonlinear wavelet shrinkage with bayes rules and bayes factors. *Journal of the American Statistical Association*, 93:173–179.
- West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models (Second Edition)*. Springer-Verlag.
- Working, H. (1960). Note on the correlation of first differences of averages in a random chain. *Econometrica*, 28:916–918.
- Xie, P. and Arkin, P. A. (1996). Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions. *Journal of Climate*, 9:840–858.