Bayesian Analysis (2006) **1**, Number 4, pp. 947–968

Multi-Scale and Hidden Resolution Time Series Models

Marco A. R. Ferreira*, Mike West†, Herbert K. H. Lee‡, and David M. Higdon§

Abstract. We introduce a class of multi-scale models for time series. The novel framework couples standard linear models at different levels of resolution via stochastic links across scales. Jeffrey's rule of conditioning is used to revise the implied distributions and ensure that the probability distributions at the different levels are strictly compatible. This results in a new class of models for time series with three key characteristics: this class exhibits a variety of autocorrelation structures based on a parsimonious parameterization, it has the ability to combine information across levels of resolution, and it also has the capacity to emulate long memory processes. The potential applications of such multi-scale models include problems in which it is of interest to develop consistent stochastic models across time-scales and levels of resolution, in order to coherently combine and integrate information arising at different levels of resolution. Bayesian estimation based on MCMC analysis and forecasting based on simulation are developed. One application to the analysis of the flow of a river illustrates the new class of models and its utility.

Keywords: Autoregressive models; Bayesian inference; Combination of multi-resolution information; Jeffrey's rule of conditioning; Multi-scale stochastic models; Multi-scale time series models.

1 Introduction

We are interested in multi-scale time series models for three main reasons: first, we desire a method for consistent modeling of time series at different levels of resolution (e.g., daily and monthly aggregates of financial or meteorological data); second, we may need to combine information coherently across time scales; and third, we want to model processes that have relevant dynamics at different resolution levels. It is important to consider the scale of the process, because the temporal scale in which observations are made may lead to different conclusions with respect to the dynamics of the process under study. This concern dates back to the surprising finding of Working (1960), who showed that if a random walk process (a process with white noise first differences) is observed at a coarser resolution, where the coarsening is by non-overlapping averages,

*Departamento de Métodos Estatísticos, Universidade Federal do Rio de Janeiro, Brazil, http://www.dme.ufrj.br/marco/

†Institute of Statistics and Decision Sciences, Duke University, Durham, NC, http://www.isds.duke.edu/~mw/

‡Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA, http://www.ams.ucsc.edu/~herbie/

§Division of Statistical Sciences, Los Alamos National Laboratory, Los Alamos, NM, http://www.stat.lanl.gov/people/dhigdon.shtml

the first differences of this coarser process are correlated. Since then, several authors have studied the impact of the analysis of observations at a coarser scale than the scale of the definition of the process of interest (e.g. Telser 1967; Amemiya and Wu 1972; Palm and Nijman 1984; Drost and Nijman 1993; Schmidt and Gamerman 1997; Hwang 2000; Bollerslev and Wright 2000). Those processes may be defined on continuous or discrete time and the coarser observations may be obtained by subsampling or by non-overlapping averages. While this approach may lead to consistent modeling across time resolutions and the coherent combination of multi-scale information, it does assume that the only relevant dynamics are at the fine scale of resolution. As a point of departure from this previous body of work, we propose a new class of multi-scale time series models built from coarse to fine scales of resolution. While maintaining the capacity of consistent modeling of time series at different levels of resolution and of coherent combination of multi-scale information, our new approach allows the existence of relevant dynamics at multiple resolution levels. This approach results in a novel class of models for time series that exhibits a variety of autocorrelation structures at the fine level based on a very parsimonious parameterization and, in particular, has the capacity to emulate long memory processes.

There are two main manners to relate long memory and levels of aggregation/resolution of time series. One of them is due to Granger (1980), who has shown that the sum of many autoregressive processes of order one with Beta distributed coefficients may be fractionally integrated. This result has been used by Ding and Granger (1996) and by Byers et al. (1997) in order to motivate long memory models for volatility of speculative returns and for opinion poll series, respectively. The other manner, shown by Wornell (1990), is the construction of approximate long memory processes through the use of an inverse discrete wavelet transform (Mallat 1999; Vidakovic 1999; Müller and Vidakovic 1999). Variations of this construction have been used by several authors in order to perform fast simulations of long memory processes (Wornell 1993; McCoy and Walden 1996; Percival and Walden 2000; Craigmile 2003, 2005). The latter manner is related to the multi-scale class of models presented in this paper; we build our models in a cascade way from coarse to fine levels of resolution. In contrast to that previous work, we explicitly address cases when data *can be observed* at several different time scales, and at each scale there is a progressive correlated stochastic process. Thus, unlike current wavelet based time series modeling, our multi-scale class of models can be used in extrapolation and forecasting.

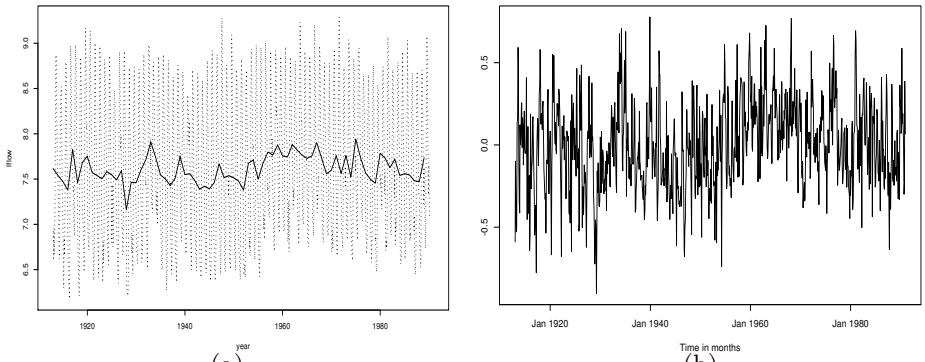
There are at least three types of problems that can be modeled within a multi-scale framework. In the first type of problem, data are observed at different scales and a multi-scale time series model (MSTSM) is used to integrate the information from the different scales. In the second case, data are observed only at the finest scale and the multi-scale model is used to induce a particular process at this finest scale, resulting in what we call Hidden Resolution Models (HRM). Finally, the multi-scale model can be used as a prior for an underlying multi-scale process, as in Ferreira et al. (2003). In this paper, we focus mainly on problems of the first two types.

Section 2 presents a motivating example for the construction of MSTSMs and HRMs. In Section 3.1 we introduce the general framework for MSTSMs and HRMs with one

coarse level and one fine level. Section 3.2 presents an example of a MSTSM and the corresponding HRMwith autoregressive processes of order one as building blocks. In Section 3.3 we derive, for the general case, the implied distribution at the fine level—that is, the Hidden Resolution Model. In Section 4 we present several properties of HRMs. Section 5 describes extensions to incorporate periodicities into the multi-scale framework, relevant for certain applications. In Section 6 we discuss issues of inference and prediction. Section 7 provides an application of HRMs to the analysis of the flow of the Fraser River in Canada. We conclude and point to future directions in Section 8.

2 Motivating example

We present here a preliminary analysis of the flow of the Fraser River in Canada as a motivating example for the construction of multi-scale time series models.

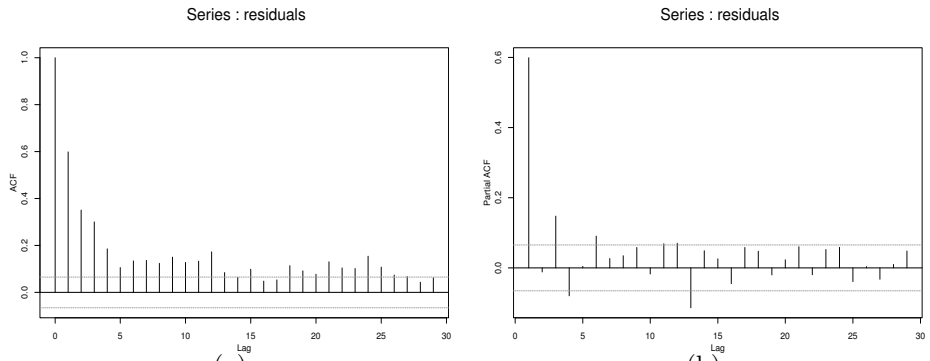


(a) (b)

Figure 1: Fraser River. (a) Log of the mean monthly flows (dotted) and annual averages (solid). (b) Monthly residuals after extracting the seasonality and the mean.

Figure 1a presents the plot of the log of the mean monthly flows of the Fraser River from January of 1913 to December of 1990, as well as the series of annual averages. From the figure we note the presence of seasonality in the monthly series. Moreover, it is evident from Figure 1a that there is strong dependence between the annual averages. Thus, a simple ARMA process with seasonality for the monthly series would be inappropriate. In particular, such a model would probably give poor medium term (12 to 36 months ahead) forecasts.

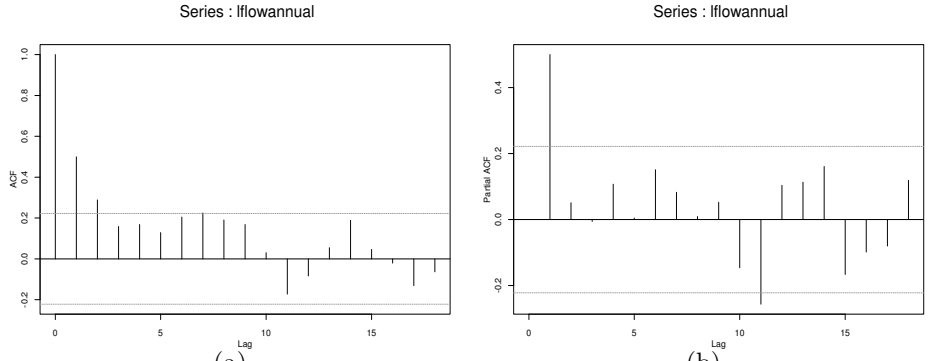
An exploratory analysis found that the seasonality can be well explained by the first, fourth and fifth harmonics. Figure 1b shows the plot of the monthly residuals after extracting the overall mean and the seasonality. Figure 2 shows the autocorrelation and partial autocorrelation functions of the monthly residuals, suggesting a long memory type process. In contrast, Figure 3 shows the autocorrelation and partial autocorrelation functions of the annual series, strongly suggesting an AR(1) process for the annual level



(a) (b)

Figure 2: Fraser River. (a) Autocorrelation and (b) partial autocorrelation functions of monthly residuals, suggesting a long memory process.

of aggregation. Thus, it seems that there is a strong annual level dynamic that is reasonably well explained by an AR(1) process and that leads to a long memory type of behavior at the monthly level. This is probably the result of large time scale climate dynamics that impact water and snow precipitation and, thus, the flow of the rivers.



(a) (b)

Figure 3: Fraser River. (a) Autocorrelation and (b) partial autocorrelation functions of the annual series.

In order to accommodate relevant dynamics at several time resolution levels, we construct in the following sections multi-scale time series models from coarser to finer resolution levels.

3 Model construction

3.1 General framework

We begin with a univariate time series x_t, $t = 1, 2, \ldots$, following a specified model. Writing $x_{1:n_x} = (x_1, \ldots, x_{n_x})$ for any integer $n_x > 0$, we denote by $p(x_{1:n_x})$ the density of the joint distribution of $x_{1:n_x}$. For example, we may assume that x_t follows a standard linear stationary time series model, such as an AR(1) process; in that case, $p(x_{1:n_x})$ is the implied joint stationary distribution. In other examples, $p(x_{1:n_x})$ may be a posterior predictive distribution arising from a dynamic model conditioned on past observations (as in West and Harrison 1997). We will refer to x_t as the *fine level* process.

For a specified positive integer m, which we call the *coarsening window*, define the *coarse level* aggregate process y_s on indices $s = 1, 2, \ldots$ by $y_s = m^{-1} \sum_{i=1}^{m} x_{(s-1)m+i} + u_s$, where the u_s, $s = 1, 2, \ldots$, are mutually uncorrelated zero-mean, normally distributed noise terms with $u_s \sim N(u_s|0, \tau)$ for some between-level variance τ. The y values are the averages of non-overlapping groups of m consecutive x values, such as weekly averages ($m = 7$) of daily data or annual averages ($m = 12$) of monthly data, subject to the addition of noise or error terms u. Thus, the marginal distribution of $y_{1:n_y}$ is $p(y_{1:n_y}) = \int p(y_{1:n_y}|x_{1:n_x})p(x_{1:n_x})dx_{1:n_x}$. As a result of these assumptions, the implied distribution of the process at the coarse level is more complex than the distribution of the process at the fine level.

The main 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, that supersedes the prior information on which $p(x_{1:n_x})$ and $p(y_{1:n_y}|x_{1:n_x})$ are based. 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 $q(y_{1:n_y})$. This update can be seen as thinking of $p(y)$ as the prior, $q(y) = q(y|G)$ as the posterior, and $l(y_{1:n_y}|G)$ as an implicit likelihood function defined 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 $p(y_{1:n_y}|x_{1:n_x})$ may be, and generally will be, inconsistent. We restore consistency by the use of Jeffrey's rule of conditioning to revise the fine scale model $p(x_{1:n_x})$ and thus make the model consistent across the different levels of resolution. The use of Jeffrey's rule assumes that conditional on $y_{1:n_y}$, the fine level $x_{1:n_x}$ is independent of the new piece of information G. Diaconis and Zabell (1982) discuss the use and roles of Jeffrey's rule in Bayesian analysis. Loschi et al. (2002) use Jeffrey's rule for inference 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 our case, Jeffrey's rule of conditioning tells us to update the fine level distribution by the formula: $q(x_{1:n_x}) = \int p(x_{1:n_x}|y_{1:n_y})q(y_{1:n_y})dy_{1:n_y}$.

This construction leads to two possible models, depending on whether $y_{1:n_y}$ is observed or not. If the two levels of resolution $x_{1:n_x}$ and $y_{1:n_y}$ are observable, the main interest is to coherently combine and integrate information arising at different levels of resolution. In this case, we call the joint model $p(x_{1:n_x}|y_{1:n_y})q(y_{1:n_y})$ a Multi-scale Time Series Model (MSTSM). Conversely, if the coarse level $y_{1:n_y}$ is an unobservable

latent process then the main interest is to induce rich processes at the fine level. In this case, the implied process at the fine level $q(x_{1:n_x})$ is what we call a Hidden Resolution Model(HRM).

In Section 3.2 we present an example using autoregressive processes of order 1 as building blocks for the construction of MSTSMs and HRMs.

3.2 Example – AR(1) building blocks

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

$$x_t = \phi_x x_{t-1} + \epsilon_t, \tag{1}$$

where ϵ_t, $t = 1, 2, \ldots$, 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. Thus, for all $n_x > 0$, $p(x_{1:n_x})$ is the implied n_x-dimensional stationary distribution

$$p(x_{1:n_x}) = N(x_{1:n_x}|0, V_x), \tag{2}$$

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

The link between levels is described by $y_{1:n_y}|x_{1:n_x} \sim N(Ax_{1:n_x}, U)$, where $n_x = n_y m$, and A is a coarsening $n_y \times n_x$ matrix. For example, if the coarsening operation is by non-overlapping arithmetic averages then 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. Moreover, if the y_s are conditionally independent with constant variance τ then the link equation is:

$$p(y_{1:n_y}|x_{1:n_x}) = \prod_{s=1}^{n_y} N(y_s|m^{-1}\sum_{i=1}^{m} x_{(s-1)m+i}, \tau) = N(y_{1:n_y}|Ax_{1:n_x}, U), \tag{3}$$

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_xA'. Here, we use the parameterization $\tau = \lambda(AV_xA')_{11}$. 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; thus, it is much easier to establish a prior for λ than for τ.

Assume that we receive a new piece of information G about the coarse level $y_{1:n_y}$— information that partially supersedes the information contained in $p(x_{1:n_x})$ and $p(y_{1:n_y}|x_{1:n_x})$. More specifically, assume that this new piece of information substitutes the implied coarse level model $p(y_{1:n_y}) = \int p(x_{1:n_x})p(y_{1:n_y}|x_{1:n_x})dx_{1:n_x})$ by the revised model $q(y_{1:n_y})$. Moreover, assume that this revised model is a simple standard stationary linear AR(1) process

$$y_s = \phi_y y_{s-1} + \eta_s, \tag{4}$$

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

n_y-dimensional stationary distribution

$$q(y_{1:n_y}) = N(y_{1:n_y}|0, Q_y), \tag{5}$$

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

Obviously, the densities $p(y_{1:n_y})$ and $q(y_{1:n_y})$ are not compatible. However, viewing the information G as superseding the information on which the model $p(x_{1:n_x})$ is based, we must adopt (5) and update $p(x_{1:n_x})$ accordingly. This is done using Jeffrey's rule of conditioning to revise the distribution of $x_{1:n_x}$. The application of Jeffrey's rule implies the assumption that the revised conditional distribution of $x_{1:n_x}$ given $y_{1:n_y}$, denoted by $q(x_{1:n_x}|y_{1:n_y})$, is equal to the conditional distribution of $x_{1:n_x}$ given $y_{1:n_y}$ implied by Equations (2) and (3), denoted by $p(x_{1:n_x}|y_{1:n_y})$. This condition means that given $y_{1:n_y}$, $x_{1:n_x}$ is independent of the new information G that led to the revision of beliefs about $y_{1:n_y}$.

In Section 3.3 we show that the resulting distribution $q(x_{1:n_x})$—the Hidden Resolution Model—is also zero-mean normal $q(x_{1:n_x}) = N(x_{1:n_x}|0, Q_x)$ with covariance matrix $Q_x = V_x - B(W - Q_y)B'$, where $W = AV_xA' + U$ and $B = V_xA'W^{-1}$.

3.3 Implied fine level distribution

We derive in this section the revised distribution at the fine level, that is, the Hidden Resolution Model. We start by using Bayes Theorem in order to obtain the conditional distribution of $x_{1:n_x}$ given $y_{1:n_y}$:

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

Therefore, by linear regression: $x_{1:n_x}|y_{1:n_y} \sim N(By_{1:n_y}, V_x - BWB')$, where $B = V_xA'W^{-1}$ and $W = AV_xA' + U$. As the distribution $p(x_{1:n_x})$ has mean zero, the expected value of $x_{1:n_x}$ given $y_{1:n_y}$ is shrunk from the corresponding value of $y_{1:n_y}$ towards the zero vector.

As $q > 0$ and $p > 0$, we can use the generalized Jeffrey's rule (Diaconis and Zabell 1982) to derive the Hidden Resolution Model:

$$\begin{aligned} q(x_{1:n_x}) &= \int p(x_{1:n_x}|y_{1:n_y})q(y_{1:n_y})dy_{1:n_y} \\ &\propto \int N(x_{1:n_x}|By_{1:n_y}, V_x - BWB') \\ &\quad N(y_{1:n_y}|0, Q_y)dy_{1:n_y}. \end{aligned} \tag{6}$$

Therefore, the resulting distribution is

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

with variance matrix

$$Q_x = V_x - B(W - Q_y)B', \tag{8}$$

where $W = AV_xA' + U$ and $B = V_xA'W^{-1}$.

We call the marginal model at the fine level $q(x_{1:n_x})$ a Hidden Resolution Model and the joint model at both levels $p(x_{1:n_x}|y_{1:n_y})q(y_{1:n_y})$ a Multi-scale Time Series Model.

In general, an HRM does not represent any recognizable model in univariate form, but in the particular case when the HRM is built with AR(1) blocks, link equation given by (3) and $\phi_x = 0$ then it is easy to show that:

$$\begin{aligned}
Var(x_t) &= \sigma_x^2 + \frac{m^{-2}\sigma_x^4}{(m^{-1}\sigma_x^2+\tau)^2}\frac{\sigma_y^2}{1-\phi_y^2} - \frac{m^{-2}\sigma_x^4}{m^{-1}\sigma_x^2+\tau}, \\
Cov(x_t, x_{t+j}) &= \frac{m^{-2}\sigma_x^4}{(m^{-1}\sigma_x^2+\tau)^2}\frac{\sigma_y^2}{1-\phi_y^2} - \frac{m^{-2}\sigma_x^4}{m^{-1}\sigma_x^2+\tau}, \quad j = 1, \ldots, m-1, \\
Cov(x_t, x_{t+j}) &= \frac{m^{-2}\sigma_x^4}{(m^{-1}\sigma_x^2+\tau)^2}\frac{\sigma_y^2}{1-\phi_y^2}\phi_y^{[j/m]}, \quad j = m, m+1, \ldots,
\end{aligned}$$

where $[z]$ denotes the integer part of z. Therefore, $\phi_x = 0$ implies a step function for the autocorrelations, with step size equal to m.

From a general viewpoint, the construction of MSTSMs and HRMs has three ingredients: the initial process 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 derive the Hidden Resolution Model. Therefore, the construction put forward in this section is quite general and can be used with any well-behaved types of processes as building blocks. Here, "well-behaved" means processes with well defined conditional mean vectors and positive definite covariance matrices in the time frame of interest. This includes stationary processes and also some nonstationary processes—for example nonstationary Markovian processes with known expected value and variance in the beginning of the time frame. Conversely, nonstationary Markovian processes with unknown initial conditions cannot be used because, in this case, the mean vector and covariance matrix in the time frame of interest are not well defined and (7) and (8) cannot be computed. When dealing with stationary processes, a natural choice for building blocks are stationary and invertible ARMA processes. In Section 4 we study properties of HRMs with AR(1) building blocks; properties of HRMs constructed with more general building blocks can be studied in the same lines of Section 4 and will not be considered here. Readers interested in multi-scale models with more than two levels of resolution are referred to Ferreira (2002).

4 Properties of the Hidden Resolution Model

We discuss here several properties of HRMs with AR(1) building blocks by analyzing plots of several possible autocorrelation functions. As a byproduct of the study of the autocorrelation functions, the parameters of HRMs are given suitable interpretations.

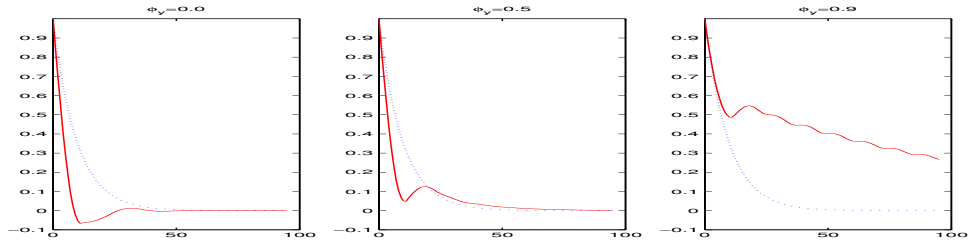


Figure 4: AR(1) building blocks. Autocorrelation functions varying ϕ_y (the autoregressive coefficient on the coarse scale). Parameters kept constant: $\phi_x = 0.9$, $\sigma_x^2 = 1$, $\sigma_y^2 = 1$, $\lambda = 0.1$ and $m = 12$. HRM (dashed line); AR(1) (dotted line).

An interesting feature of HRMs is the emulation of long memory processes (for reference on long memory processes, see Beran, 1994 and Brockwell and Davis, 1991). As can be seen in Figure 4, the parameter ϕ_y controls the rate of decay of the autocorrelation function, which shows some persistence when the value of ϕ_y increases. The main advantage of HRMs over actual long memory models is interpretability—the long memory type of behavior is explicitly modeled as a result of high autocorrelation in the hidden coarse level of the hierarchy.

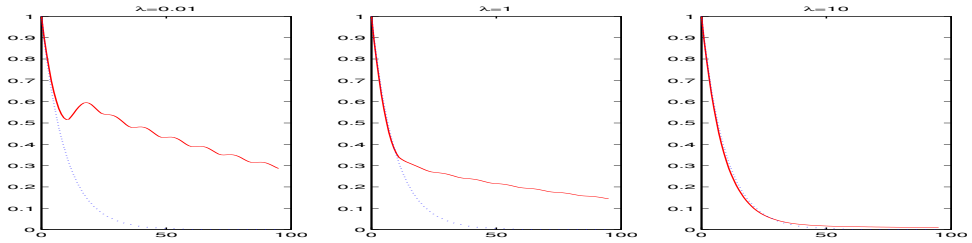


Figure 5: AR(1) building blocks. Autocorrelation functions varying λ (the lack of agreement between coarse and fine levels). Parameters kept constant: $\phi_x = 0.9$, $\sigma_x^2 = 1$, $\phi_y = 0.9$, $\sigma_y^2 = 1$ and $m = 12$. HRM (dashed line); AR(1) (dotted line).

The parameter λ controls how much the coarse level influences the fine level. In that respect, it can be shown that the limit of the autocorrelation function when λ approaches infinity is the autocorrelation function of the original autoregressive process. Figure 5 illustrates this, the autocorrelation functions of the HRM and of the original autoregressive process being 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 $\lambda = 0.01$.

As λ approaches infinity, the likelihood function approaches a constant equal to the likelihood function of an AR(1) process; thus, the use of an improper prior for λ would lead to an improper posterior distribution. Nonetheless, the autocorrelation function when $\lambda = 10$ is very close to the autocorrelation function when $\lambda \to \infty$. Thus, in

order to guarantee posterior propriety we propose the use of prior distributions for λ truncated to the interval $(0, 10)$.

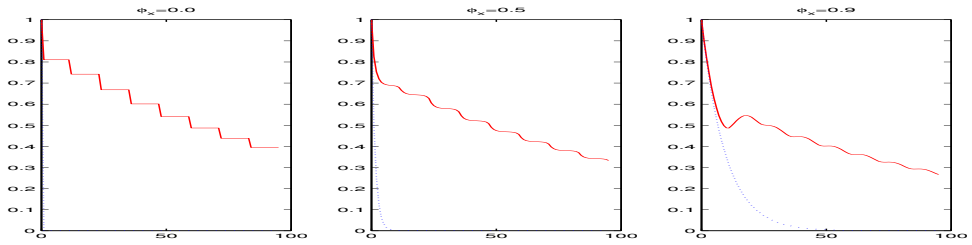


Figure 6: AR(1) building blocks. Autocorrelation functions varying ϕ_x (the autoregressive coefficient on the fine scale). Parameters kept constant: $\sigma_x^2 = 1$, $\phi_y = 0.9$, $\sigma_y^2 = 1$, $\lambda = 0.1$ and $m = 12$. HRM (dashed line); AR(1) (dotted line).

Another interesting feature of Hidden Resolution Models is the existence of a blocking effect that depends on m and ϕ_x. As discussed in Section 3.3, when $\phi_x = 0.0$ the autocorrelations are constant by blocks of length m. Additionally, as can be seen in Figure 6, 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.

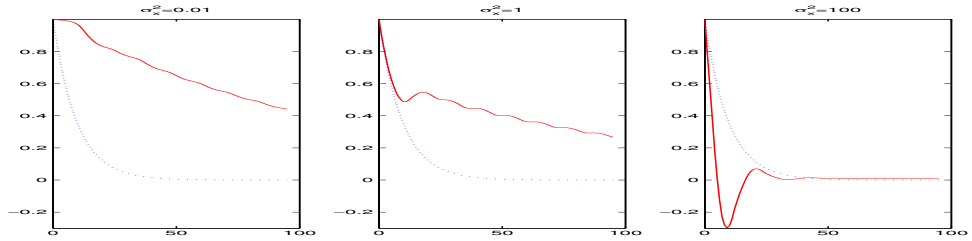


Figure 7: AR(1) building blocks. Autocorrelation functions varying σ_x^2 (the variance of the error on the fine scale). Parameters kept constant: $\phi_x = 0.9$, $\phi_y = 0.9$, $\sigma_y^2 = 1$, $\lambda = 0.1$ and $m = 12$. HRM (dashed line); AR(1) (dotted line).

Finally, it is easy to show that the autocorrelation function depends on σ_x^2 and σ_y^2 only through σ_x^2/σ_y^2. Thus, we explore the autocorrelation function keeping σ_y^2 constant and varying σ_x^2. Figure 7 depicts the behaviors of the autocorrelation function when $\phi_x = 0.9$, $\phi_y = 0.9$, $\sigma_y^2 = 1$, $\lambda = 0.1$, $m = 12$ and σ_x^2 assumes values 0.01, 1 and 100. For these values of ϕ_x, ϕ_y and λ, when σ_x^2/σ_y^2 is very small, the fine level process closely follows the coarse level process and the HRM emulates long memory behavior. Conversely, when σ_x^2/σ_y^2 is large the fine level varies widely but its non-overlapping averages are close to the coarse level; as a consequence, the autocorrelation function has an oscillatory behavior between positive and negative values.

5 Incorporating periodicities

Here we briefly discuss how to incorporate periodicities in MSTSMs and HRMs at the fine level by the inclusion of regressors corresponding to harmonics with cycle length equal to the coarsening window. As harmonics are sine and cosine functions with mean over the cycle length equal to zero, the coarsening operation eliminates the periodic pattern.

The construction of the model is the same as in Section 3.2, except that we substitute Equation (2) by $x_{1:n_x} \sim N(x_{1:n_x}|\mu_x, V_x)$, where $\mu_x = Z\beta$ is a periodic vector with period equal to the coarsening window m, V_x is the covariance matrix of the fine level process, and Z is a design matrix corresponding to the harmonics. Using developments analogous to those of Section 3, it is easy to show that the updated marginal distribution of $x_{1:n_x}$ is

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

where $B = V_x A' W^{-1}$ and $W = AV_x A' + U$ are the same as in the Section 3.2.

Note that all calculations in this section are also valid in the setting of regression analysis as long as $AZ = 0$, and the general case $AZ \neq 0$ can be accommodated with minor modifications in the model construction.

Section 6.2 discusses estimation and forecasting in the presence of periodicities. Section 7 presents an application of HRMs to time series in the presence of periodicities.

6 Inference and prediction

Due to nonlinearities, the posterior distribution is rather complicated and Bayesian analysis of MSTSMs and HRMs cannot be performed analytically. In order to explore the posterior distribution, we propose an algorithm based on Markov chain Monte Carlo (MCMC) techniques to simulate a sample from the posterior distribution. This sample is used to estimate summaries of the posterior distribution such as posterior means, standard deviations and credible intervals. Moreover, as we discuss in Section 6.1, this posterior sample may easily be used in a simulation based prediction procedure.

The MCMC algorithm used for inference and the simulation-based prediction procedure have to be tailored to the particular MSTSMs or HRMs being considered. In order to illustrate the construction of these algorithms, we consider in this section MSTSMs and HRMs with AR(1) building blocks and link equation given by (3). The model parameters are the autoregressive parameters for the fine and coarse levels, ϕ_y and ϕ_x, the fine and coarse process variances, σ_y^2 and σ_x^2, and the between-levels parameter λ. We assume independence a priori of the parameters and the following marginal priors: $\phi_y \sim TrN_{(-1,1)}(m_{\phi_y}, S_{\phi_y})$, $\sigma_y^2 \sim IG(\nu_{\sigma_y}/2, \nu_{\sigma_y} s_{\sigma_y}/2)$, $\phi_x \sim TrN_{(-1,1)}(m_{\phi_x}, S_{\phi_x})$, $\sigma_x^2 \sim IG(\nu_{\sigma_x}/2, \nu_{\sigma_x} s_{\sigma_x}/2)$, $\lambda \sim TrIG_{(0,10)}(\nu_\lambda/2, \nu_\lambda s_\lambda/2)$, where $TrN_{(a,b)}$ and $TrIG_{(a,b)}$ denote respectively the normal and inverse gamma distributions truncated to the interval (a, b). The priors for ϕ_y and ϕ_x are truncated to the interval $(-1, 1)$ to guarantee stationarity. The prior for λ is a truncated inverse

gamma, and the analysis may be fairly sensitive to the choice of the upper limit of the truncation interval when ν_λ and $\nu_\lambda s_\lambda$ are small. Thus, this upper limit has to be carefully chosen. As discussed in Section 4, values of λ greater than 10 lead to models practically indistinguishable from an AR(1) model. As a consequence, an upper limit much larger than 10 simply assigns more prior probability to the AR(1) model, while an upper limit much smaller than 10 may exclude the possibility of models close to the AR(1) model. Moreover, the analysis is reasonably insensitive to small variations of the upper limit around 10. Thus, we recommend a prior for λ truncated to the interval $(0, 10)$.

Thus the posterior distribution is proportional to:

$$p(x_{1:n_x}|y_{1:n_y}, \phi_x, \sigma_x^2, \lambda)q(y_{1:n_y}|\phi_y, \sigma_y^2)p(\phi_y)p(\sigma_y^2)p(\phi_x)p(\sigma_x^2)p(\lambda). \tag{9}$$

A major aspect of the estimation for MSTSMs and HRMs is that, conditional on the coarse level, the parameters corresponding to coarse and fine levels are independent. Thus, when analyzing HRMs the inclusion of the simulation of the hidden coarse level dramatically facilitates the implementation of a Gibbs sampler to explore the posterior distribution.

From the joint distribution given by Equation (9), it is easy to verify that conditional on the hidden coarse level, the simulation of the parameters corresponding to different levels can be done separately. As the coarse level process is generally simple, it is possible to use techniques already available in the literature to simulate the coarse level parameters. For example, if the coarse level follows an ARMA process then the coarse level parameters can be simulated with the procedure proposed by Chib and Greenberg (1994).

The simulation of the parameters of the fine level and of the link equation is not so trivial because, in general, the full conditional distributions are not available in closed form for sampling. To overcome this problem, we simulate these parameters using Metropolis-Hastings proposals. The following theorem simplifies and accelerates the computations for the simulation of $(\phi_x, \sigma_x^2, \lambda)$.

Theorem 6.1

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

where $p(y_{1:n_y}|\phi_x, \sigma_x^2, \lambda) = N(y_{1:n_y}|0, A'V_xA+U)$ *and* $p(y_{1:n_y}|x_{1:n_x}, \phi_x, \sigma_x^2, \lambda) = N(y_{1:n_y}|Ax_{1:n_x}, U)$.

Proof. See Appendix.

Note that $p(x_{1:n_x}|\phi_x, \sigma_x^2, \lambda)$ and $p(y_{1:n_y}|x_{1:n_x}, \phi_x, \sigma_x^2, \lambda)$ are easy to compute. Moreover, as the unrevised model $p(x_{1:n_x}|\phi_x, \sigma_x^2, \lambda)p(y_{1:n_y}|x_{1:n_x}, \phi_x, \sigma_x^2, \lambda)$ can be cast within a state space framework (for details, see Ferreira 2002), the computation of $p(y_{1:n_y}|\phi_x, \sigma_x^2, \lambda)$ can be performed very efficiently by the Kalman filter (for reference on the Kalman

filter, see West and Harrison 1997). Nonetheless, it is important to note that the more complex revised model $p(x_{1:n_x}|y_{1:n_y}, \phi_x, \sigma_x^2, \lambda)q(y_{1:n_y}|\phi_y, \sigma_y^2)$ can not be cast within a state space framework.

The simulation of proposals for ϕ_x and σ_x^2 is performed through Metropolis-Hastings steps. After the simulation of each proposal, Theorem 6.1 is used to compute the proposal acceptance probability. More specifically, the proposal for σ_x^2 is simulated from $U(\sigma_x^2|\sigma_x^{2\ (old)}/\delta_{\sigma_x},\ \sigma_x^{2\ (old)}\delta_{\sigma_x})$, where δ_{σ_x} has to be tuned to yield a reasonable acceptance rate.

The proposal for ϕ_x is simulated from $U(\phi_x|max(-1.0, \phi_x^{(old)} - \delta_{\phi_x}), min(1.0,\ \phi_x^{(old)} + \delta_{\phi_x}))$, where δ_{ϕ_x} has to be tuned to yield a reasonable acceptance rate.

In the case of MSTSMs, the simulation of λ is easily performed. Conversely, in the case of HRMs, the hidden coarse level $y_{1:n_y}$ and the parameter λ are highly correlated a posteriori. Thus, we simulate $y_{1:n_y}$ and λ jointly in order to improve the mixing of the Gibbs sampler. More specifically, we first simulate a proposal $\lambda^{(prop)}$ for λ from $U(\lambda|\max(0, \lambda^{(old)} - \delta_\lambda), \min(1, \lambda^{(old)} + \delta_\lambda))$. After that, we simulate a proposal for $y_{1:n_y}$ from its full conditional distribution conditional on $\lambda^{(new)}$. The joint proposal is accepted or rejected with the appropriate Metropolis-Hastings acceptance probability.

6.1 Prediction

In order to perform forecasting, it is sufficient to obtain a sample from the predictive distribution of the future observations at the different levels of resolution. Point forecasts and prediction intervals can be derived from this sample. In traditional time series analysis, depending on the decision problem at hand the analyst defines on which time scale to analyze the time series, finds a reasonable model and performs forecasts at that scale; in general, that model will not be reliable for performing forecasts at other time scales. In contrast, MSTSMs and HRMs are able to forecast at all levels of time resolution that are included in the model. As we show in the application of Section 7, the forecasts at the coarse level guide the forecasts at the fine level, reducing the prediction error.

We use a two-stage procedure to simulate each realization of the sample from the predictive distribution. First, we simulate a future realization of the hidden coarse level conditional on the past. After that, we simulate a realization of the fine level conditional on both the past and the realization of the coarse level.

The following theorem is very useful for the simulation of predictions of the coarse level. 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.

Theorem 6.2 $q(y_{n_y}|y_{1:(n_y-1)}, x_{1:(n_x-m)}) = q(y_{n_y}|y_{1:(n_y-1)}, x_{n_x-m})$.

Proof. See Appendix.

In practice, the dependence on the past coarse level decreases fast with time lag.

The following theorems simplify forecasting at the fine level conditional on the future coarse level:

Theorem 6.3

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

Proof. See Appendix.

Theorem 6.4

$$p(x_{(n_x+1):(n_x+ml)}|y_{1:(n_y+l)}, x_{1:n_x}) = p(x_{(n_x+1):(n_x+ml)}|y_{(n_y+1):(n_y+l)}, x_{n_x}).$$

Proof. This follows from Theorem 6.3 by induction.

Theorem 6.4 states that conditional on the last observation x_{n_x} at the fine level and on the future observations $y_{(n_y+1):(n_y+l)}$ at the coarse level, the future observations at the fine level are independent of the observations at the fine level up to time $n_x - 1$ and of the observations at the coarse level up to time n_y.

6.2 Estimation and forecasting in the presence of periodicities

The procedures for estimation and forecasting when there are periodicities at the fine level are analogous to the case without periodicities.

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

The full conditional distributions for ϕ_y, σ^2_y and λ are the same as in the case without periodicities. The full conditionals for ϕ_x and σ^2_x are analogous to the case without periodicities, but substituting $x_{1:n_x}$ by $x^*_{1:n_x} = x_{1:n_x} - Z\beta$.

The forecasting procedure in the case with periodicities is analogous to the case without them. More specifically, predictions are made for $x^*_{(n_x+1):(n_x+l)}$ and the periodicities are added to these predictions.

7 Application: Flow of the Fraser River

We return here to the analysis of the flow of the Fraser River in Canada as an example of the application of the HRM with seasonality. First, we start a standard time series

analysis for the monthly data using harmonics to account for the seasonality. The analysis suggests a long range dependence behavior for the monthly data, but a simple AR(1) process for the annual data. That gives us the motivation to perform an analysis using the HRM with seasonality.

As introduced in Section 2, Figure 1a shows a plot of the log of the mean monthly flows of the Fraser River from January of 1913 to December of 1990, as well as the series of annual averages. Seasonality is clearly present in the monthly series, and Figure 1a shows dependence between the annual averages. That leads us to conclude that a simple ARMA process with seasonality would be inappropriate for the monthly series, particularly for predictions 12 to 36 months out. As an alternative, we model the river flow with the HRM.

The seasonality can be well explained by the first, fourth and fifth harmonics, which will be included in the model as discussed in Section 5. Figure 1b shows the plot of the monthly residuals after extracting the overall mean and the seasonality.

In looking at the monthly residuals, their autocorrelation and partial autocorrelation functions (shown in Figure 2) suggest a long memory type process. In contrast, Figure 3 shows the autocorrelation and partial autocorrelation functions of the annual series, strongly suggesting an AR(1) process for the annual data. In order to capture this behavior, we use the HRM with the annual level of aggregation as the coarse level.

In the analysis of the HRM, we assumed prior hyperparameters equal to $m_{\phi_y} = 0$, $S_{\phi_y} = 1000$, $\nu_{\sigma_y} = 0.0001$, $\nu_{\sigma_y} s_{\sigma_y} = 0.0001$, $m_{\phi_x} = 0, S_{\phi_x} = 1000$, $\nu_{\sigma_x} = 0.0001$, $\nu_{\sigma_x} s_{\sigma_x} = 0.0001$, $\nu_\lambda = 12$, $\nu_\lambda s_\lambda = 5$. The Gibbs sampler described in Section 6 was used to generate a total of 5000 iterations. The tuning parameters of the Metropolis-Hastings proposals were set to provide reasonable acceptance probabilities, equal to 40%, 50% and 46% for ϕ_x, τ and λ respectively. Convergence was reached within 1000 iterations; the remaining 4000 iterations were used to perform the statistical analysis.

Table 1 presents posterior means and standard deviations for the parameters of the HRM. The magnitude of the posterior mean of ϕ_y shows that the series exhibits persistence through the years. In addition, the posterior mean of τ is 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.

Figure 8 shows the observed monthly flow, the forecasts and 95% predictive intervals for the years 1988, 1989 and 1990 using only the observations until 1987. As we can see in the figure, the model performs very well in terms of predictive capacity, with a mean square prediction error equal to 0.0469. In comparison, the AR(1) model and the long memory ARFIMA(1,d,0) model have mean squared prediction errors equal to 0.0549 and 0.0541, respectively. This demonstrates for this example the superiority of the Hidden Resolution Model.

It is interesting to investigate the likelihood function for the Fraser River example. Figure 9 shows different representations of the likelihood function for τ and σ_y^2 when the other parameters are kept constant at their posterior means. The behavior of the likelihood is very interesting in two aspects. First, when τ approaches infinity and σ_y^2 is

	Mean	Standard deviation
ϕ_y	0.6562	0.1331
σ_y^2	0.0193	0.0075
ϕ_x	0.5958	0.0371
σ_x^2	0.0449	0.0023
τ	0.0106	0.0049
λ	0.5365	0.2369
β_1	-0.8422	0.0177
β_2	-0.4612	0.0177
β_3	0.3391	0.0116
β_4	-0.0565	0.0114
β_5	-0.1014	0.0085
β_6	0.0670	0.0086

Table 1: Fraser River. Posterior summaries for the parameters of the HRM.

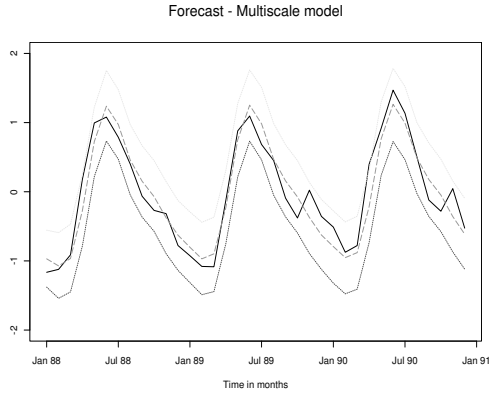


Figure 8: Fraser River. Observed flow (solid), forecast (dashed) and predictive interval (dotted) for the years 1988, 1989 and 1990 by month.

kept constant, the likelihood of the HRM does not vanish, but becomes constant equal to the likelihood of the original model $p(x)$. Second, when τ and σ_y^2 approach infinity such that σ_y^2/τ^2 is constant, the likelihood function again does not vanish, but approaches a constant value greater than zero. As a result, the analysis may heavily depend on the prior, and in this situation, naive vague priors lead to catastrophic results. This reinforces the necessity for parameterizing the model in terms of $\lambda = (AV_xA')_{1,1}/\tau$, since it is much easier to interpret and assign a sensible prior for λ.

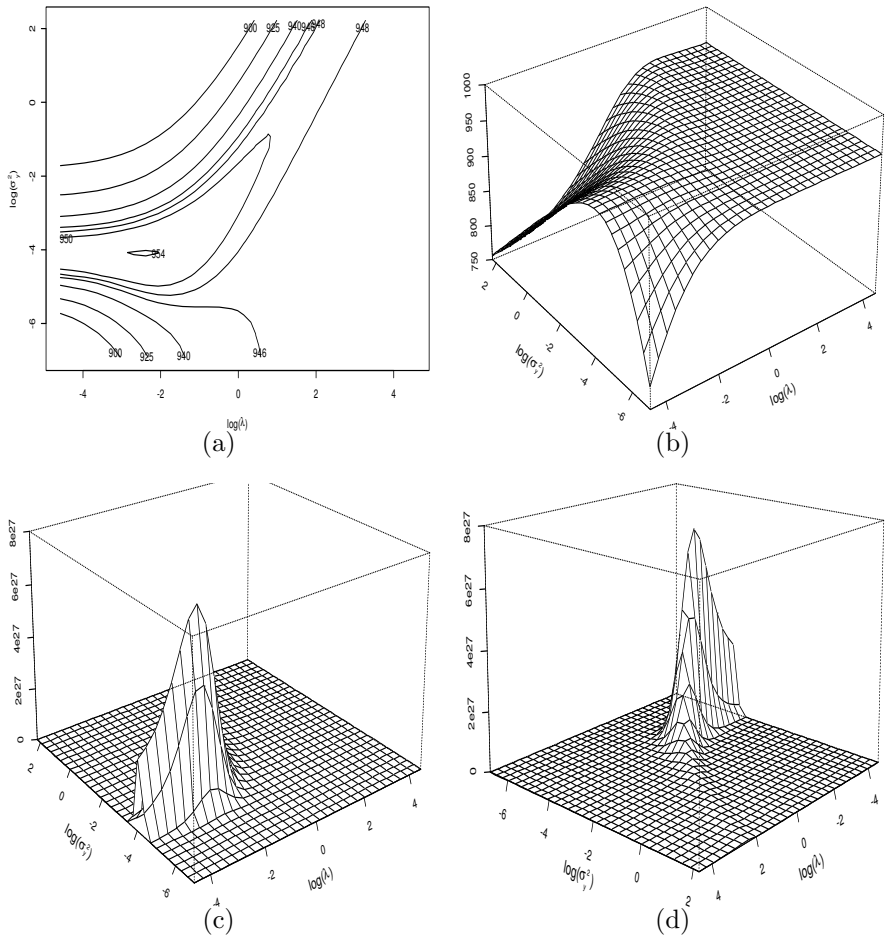


Figure 9: Fraser River. Likelihood function for τ and σ_y^2: (a) Log-likelihood contour plot; (b) Log likelihood; (c) Likelihood; (d) Likelihood from another perspective.

8 Discussion and future directions

We constructed time series models with rich autocorrelation structures by coupling processes evolving at different levels of resolution through time. Throughout this paper, we assumed knowledge of the right scale for the underlying coarse level process. One possible extension would be to accommodate an unknown scale of resolution for the coarse level. The estimation procedure for such an approach would likely use reversible jump MCMC methods (Green 1995).

Along the lines of this paper, we have been working on the construction of multi-scale models with an arbitrary number of levels. That construction assumes arrival of

information at the different resolution levels and conditional independence between the different resolution levels in order to build multi-scale processes from coarser to finer levels. Based on that work, we may be able to use our construction with a wavelet basis to obtain highly complex models that take advantage of the computational power associated with wavelets. Even though wavelets are useful for smoothing noisy data with discontinuities, their current implementations are not good for the forecasting of time series. We believe that by using stochastically dependent processes within each wavelet resolution level and by coupling those processes with the type of multi-scale framework developed here, we may obtain rich classes of stochastic processes for time series that are also useful for forecasting.

We have used this class of multi-scale models as a prior in a hydrology application (Ferreira et al. 2003)—more specifically, 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 multiple sources and available at different scales of resolution in the inference about the permeability field. In addition, we will report elsewhere on the extension of multi-scale models to the two dimensional case. This extension is of potential use for geologic applications (Lee et al. 2002) and climate modeling (Xie and Arkin 1996; Brandt and Zaslavsky 1997).

Appendix

Proof of Theorem 6.1

Since the conditional distribution of $x_{1:n_x}$ given $(y_{1:n_y}, \phi_x, \sigma_x^2, \lambda)$ is not revised by Jeffrey's rule, we have

$$\begin{aligned} q(x_{1:n_x}|y_{1:n_y}, \phi_x, \sigma_x^2, \lambda) &= p(x_{1:n_x}|y_{1:n_y}, \phi_x, \sigma_x^2, \lambda) \\ &= \frac{p(x_{1:n_x}|\phi_x, \sigma_x^2, \lambda)p(y_{1:n_y}|x_{1:n_x}, \phi_x, \sigma_x^2, \lambda)}{p(y_{1:n_y}|\phi_x, \sigma_x^2, \lambda)}. \end{aligned}$$

The result follows on application of Bayes' theorem.

Proof of theorem 6.2

Using the facts that $q(x_{1:ms}|y_{1:s}) = p(x_{1:ms}|y_{1:s}) = p(x_{1:ms})p(y_{1:s}|x_{1:ms})/p(y_{1:s})$ and $q(y_{1:s}) = q(y_1)\prod_{i=1}^{s} q(y_i|y_{i-1})$, we have

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

Moreover, as $p(x_{1:ms}) = p(x_1)\prod_{t=1}^{ms} p(x_t|x_{t-1})$ and $p(y_{1:s}|x_{1:ms}) = \prod_{i=1}^{s} p(y_i|x_{(mi-m+1):mi})$, we have

$$\begin{aligned} q(y_s|y_{1:s-1}, x_{1:m(s-1)}) &\propto \frac{q(y_s|y_{s-1})}{p(y_{1:s})} \int p(x_{(ms-m+1):ms}|x_{m(s-1)}) \times \\ &\quad p(y_s|x_{(ms-m+1):ms}) dx_{(ms-m+1):ms} \\ &\propto \frac{q(y_s|y_{s-1})}{p(y_s|y_{1:(s-1)})} p(y_s|x_{m(s-1)}). \end{aligned}$$

Therefore, $q(y_s|y_{1:s-1}, x_{1:m(s-1)}) = q(y_s|y_{1:s-1}, x_{m(s-1)})$.

Proof of theorem 6.3

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. Thus:

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

Therefore:

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

Acknowledgments

This work was partially supported by National Science Foundation grant DMS 9873275. Marco A. R. Ferreira was also partially supported by CNPq grant 402010/2003-5. Herbert K. H. Lee was also partially supported by National Science Foundation grants DMS 0233710 and DMS 0504851. The authors would like to thank the editors and two anonymous referees for their helpful comments and suggestions.

References

Amemiya, T. and R. Y. Wu. 1972. The effect of aggregation on prediction in the autoregressive model. *Journal of the American Statistical Association* 67(339): 628–632. 948

Bollerslev, T. and J. H. Wright. 2000. Semiparametric estimation of long-memory volatility dependencies: The role of high-frequency data. *Journal of Econometrics* 98: 81–106. 948

Brandt, A. and L. Y. Zaslavsky. 1997. Multiscale Algorithm for Atmospheric Data Assimilation. *SIAM Journal on Scientific Computing* 18: 949–956. 964

Byers, D., J. Davidson, and D. Peel. 1997. Modelling political popularity: an analysis of long-range dependence in opinion poll series. *Journal of the Royal Statistical Society, Series A* 160: 471–490. 948

Chib, S. and E. Greenberg. 1994. Bayes inference in regression models with ARMA(p,q) errors. *Journal of Econometrics* 64: 183–206. 958

Craigmile, P. F. 2003. Simulating a class of stationary Gaussian processes using the Davies-Harte algorithm, with application to long memory processes. *Journal of Time Series Analysis* 24(5): 505–511. 948

—. 2005. Approximate wavelet based simulation of long memory processes. *Journal of Statistical Computation and Simulation* 75: 363–380. To appear. 948

Diaconis, P. and S. L. Zabell. 1982. Updating Subjective Probability. *Journal of the American Statistical Association* 77(380): 822–830. 951, 953

Ding, Z. and C. W. J. Granger. 1996. Modeling volatility persistence of speculative returns: A new approach. *Journal of Econometrics* 73: 185–215. 948

Drost, F. C. and T. E. Nijman. 1993. Temporal aggregation of GARCH processes. *Econometrica* 61(4): 909–927. 948

Ferreira, M. A. R. 2002. *Bayesian Multi-scale Modelling*. Ph.D. thesis, Duke University. 954, 958

Ferreira, M. A. R., Z. Bi, M. West, H. Lee, and D. Higdon. 2003. Multi-Scale Modelling of 1-D Permeability Fields. In *Bayesian Statistics 7*, eds. J. Bernardo, M.J.Bayarri, J.O.Berger, A.P.Dawid, D.Heckerman, A.F.M.Smith, and M.West, 519–527. Oxford University Press. 948, 964

Granger, C. W. J. 1980. Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics* 14: 227–238. 948

Green, P. J. 1995. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika* 82: 711–732. 963

Hwang, S. 2000. The effects of systematic sampling and temporal aggregation on discrete time long memory processes and their finite sample properties. *Econometric Theory* 16: 347–372. 948

Jeffrey, R. C. 1992. *Probability and the Art of Judgement*. New York: Cambridge University Press. 951

Lee, H., D. Higdon, Z. Bi, M. Ferreira, and M. West. 2002. Markov Random Field Models for High-Dimensional Parameters in Simulations of Fluid Flow in Porous Media. *Technometrics* 44(3): 230–241. 964

Loschi, R. H., P. L. Iglesias, and R. B. Arellano-Valle. 2002. Conditioning on uncertain event: Extensions to Bayesian inference. *Test* 11: 1–29. 951

Mallat, S. G. 1999. *A wavelet tour of signal processing*. 2nd ed. San Diego: Academic Press. 948

McCoy, E. and A. Walden. 1996. Wavelet analysis and synthesis of stationary long-memory processes. *Journal of Computational and Graphical Statistics* 5(1): 26–56. 948

Müller, P. and B. Vidakovic, eds. 1999. *Bayesian Inference in Wavelet Based Models*. New York: Springer-Verlag. 948

Palm, F. C. and T. E. Nijman. 1984. Missing observations in the dynamic regression model. *Econometrica* 52(6): 1415–1435. 948

Percival, D. and A. Walden. 2000. *Wavelet methods for time series analysis*. Cambridge: Cambridge University Press. 948

Schmidt, A. M. and D. Gamerman. 1997. Temporal aggregation in dynamic linear models. *Journal of Forecasting* 16: 293–310. 948

Shafer, G. 1981. Jeffrey's rule of conditioning. *Philosophy of Science* 48: 337–362. 951

Telser, L. G. 1967. Discrete samples and moving sums in stationary stochastic processes. *Journal of the American Statistical Association* 62(318): 484–499. 948

Vidakovic, B. 1999. *Statistical modeling by wavelets*. Wiley. 948

West, M. and J. Harrison. 1997. *Bayesian Forecasting and Dynamic Models (2nd Edition)*. Springer-Verlag: New York. 951, 959

Working, H. 1960. Note on the correlation of first differences of averages in a random chain. *Econometrica* 28: 916–918. 947

Wornell, G. W. 1990. A Karhunen-Loéve-like expansion for 1/f processes via wavelets. *IEEE Transactions on Information Theory* 36(4): 859–861. 948

—. 1993. Wavelet-based representations for the 1/f family of fractal processes. *IEEE* 81(10): 1428–1450. 948

Xie, P. and P. A. Arkin. 1996. Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions. *Journal of Climate* 9: 840–858. 964