Insert title of book

Insert optional subtitle of book

Dani Gamerman
Departamento de Métodos Estatísticos, Universidade Federal do Rio de Janeiro, Brazil

Esther Salazar
Department of Electrical and Computer Engineering, Duke University, USA

CLARENDON PRESS • OXFORD
2012

CONTENTS

1

HIERARCHICAL MODELING IN TIME SERIES: THE FACTOR ANALYTIC APPROACH

Dani Gamerman
Departamento de Métodos Estatísticos, Universidade Federal do Rio de Janeiro, Brazil

Esther Salazar
Department of Electrical and Computer Engineering, Duke University, USA

1.1 Introduction

This book is very timely, right after Professor Adrian Smith's contributions to Science earned him a well-deserved knighthood. And we congratulate the Editors for compiling this tribute.

Adrian Smith's contribution to Statistics spans over a very wide range of topics. One can single out his relentless effort towards making the Bayesian approach to inference applicable, in a series of computationally oriented papers with approximating methods. This effort reached its climax with his landmark paper [18], after which MCMC methods became famous and widespread.

Another line of contributions was more methodological in terms of proposing new routes for exploring more elaborate data structures. It led to another landmark paper [29]. This JRSSB discussion paper was devoted to explaining how information from different but related sources of information could be combined in a regression framework with a hierarchical structure.This paper was extended to the time-varying context in [16] with the use of dynamic models. One of us was fortunate to interact with Adrian in [17], where hierarchical and dynamic models were also used to combine information from different time series sources.

The idea of *borrowing information* from related sources is very powerful. It proved to be very useful in the last decades where complex data structures began to be tackled, as they required sophisticated modeling strategies. A vital element in such structured settings is the ability to extract from the data possible similarity patterns. This can be achieved in a number of ways, including hierarchical modeling, non-parametric components and factor analysis.

This Chapter will address the issue of combining information from a possibly large time series with a factor analytic approach. Results obtained from this exercise are a (hopefully much) smaller number of latent time series that represent the main features of the complete dataset of time series originally available. Each combination of a time series and a factor gives rise to a weight or loading that informs in which ways the different original series were combined. These loadings are useful quantities as they allow the identification of common features and allow the interpretation of the relationship or correlation structure between the different series.

These concepts will be discussed and combined in a number of forms in this Chapter. Special attention will be devoted to the exploration of these ideas in the area of spatial statistics. It will be shown that this area is not only an area of application of these ideas but is one of the main beneficiaries of these developments. Spatial Statistics is devoted to the analysis of a collection of processes that exhibit correlation due to their (geographic) location. The main goal there is to appropriately capture the spatial dependence in order to be able to extrapolate information from a few data sources to the whole region of interest. This inevitably leads to the need for parsimonious forms for representing the correlation structure. This Chapter will show how the ideas behind dynamic factor models apply in this setting via illustrative examples with real data problems.

This Chapter is organized as follows. Section 1.2 reviews the literature on factor analysis. Section 1.3 presents some basic factor model extensions for modeling high-dimensional multivariate time series. Section 1.4 describes applications of these ideas in the context of spatial analysis. Section 1.5 describes how regression ideas can be incorporated into the factor model setting. This is accomplished by enlarging the scope of the models to include explanation via covariate time series. Section 1.6 draws some concluding remarks and points at possible directions for further work.

1.2 A short review of factor analysis

Factor analysis is a useful statistical technique widely used for modeling multivariate data by a few unobserved set of variables called latent factors. More specifically, the observed variables are modeled as linear combinations of the latent factors plus an idiosyncratic error. In general, this approach is applied for the following purposes: (*i*) dimension reduction, (*ii*) identifying underlying structures, and (*iii*) modeling of sparse covariance structures. From a classical point of view, the term was first introduced by Thurstone [49] and later discussed in [50, 1, 22] among many others. In recent years, a fully Bayesian treatment of factor models became feasible due to the improvements in Bayesian computation, specially Markov chain Monte Carlo (MCMC) simulation methods. In this context, the Bayesian specifications proposed by Geweke and Zhou [21], Polasek [42], Arminger and Muthén [3] can be mentioned.

The factor model is defined as follows

$$\boldsymbol{y}_t = \boldsymbol{\beta} \boldsymbol{f}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}), \tag{1.1}$$

where $t = 1, \ldots, T$, $\boldsymbol{y}_t$ is a n-dimensional observational vector, $\boldsymbol{\beta}$ is a $n \times k$ factor loadings matrix, $\boldsymbol{f}_t \sim N(\mathbf{0}, \boldsymbol{I}_k)$ are independent k-dimensional vectors called latent factors such that $k \ll n$ and $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$. This model implies that, given the factors, each $\boldsymbol{y}_t$ have independent components that is

$\mathrm{var}(y_{it}|\boldsymbol{f}_t) = \sigma_i^2$ and $\mathrm{cov}(y_{it}, y_{jt}|\boldsymbol{f}_t) = 0$ $(i \neq j)$. Moreover, dependence among components is induced by marginalizing over the distribution of the factors so $\boldsymbol{y}_t|\boldsymbol{\beta}, \boldsymbol{\Sigma} \sim N(\mathbf{0}, \boldsymbol{\Omega})$ where $\boldsymbol{\Omega} = \boldsymbol{\beta}\boldsymbol{\beta}^T + \boldsymbol{\Sigma}$. Note also that independence of the factors and the idiosyncratic terms $\boldsymbol{\epsilon}_t$ induces independence of the observations.

Two important issues have to be mentioned at this point. The first one regards identifiability problems related to the non-unique decomposition of $\boldsymbol{\Omega}$ and the inference about the number of factors. Many ways to handle this problem can found in the literature. Basically, the idea is to impose constraints on $\boldsymbol{\beta}$, as, for example the lower constraint of Geweke and Zhou [21], Aguilar and West [2] and Lopes and West [34]. However, in some applications, identifiability of the factor loadings is not required, especially for covariance matrix estimation, variable selection and prediction (see [5] for more details). This issue will be further discussed in the next Section where structured priors for $\boldsymbol{\beta}$ can be used. On the other hand, uncertainty about the number of latent factors has been studied in different ways. The most common approach was fitting the factor model for different choices of k and then using a selection criteria like AIC or BIC for model selection. Lopes and West [34] proposed fully Bayesian inference on the number of factors through a reversible jump MCMC (RJMCMC) [23]. Their proposal was compared with other number of alternatives based on bridge sampling [35]. Another recent approach relies on zeroing a subset of factor loadings using variable selection priors such as binary indicator δ_{ij} (George and McCulloch [19]). Based on this idea, interesting applications can be mentioned. Carvalho et al. [8] and Frühwirth-Schnatter and Lopes [15] for gene expression and financial modeling, respectively are a few examples. In this Chapter, we further discuss the RJMCMC scheme as a tool for model selection.

1.3 Dynamic factor models

1.3.1 *Basic definitions*

Dynamic factor models (DFM's) were developed in a number of ways and have become a useful tool for modeling high-dimensional multivariate time series. The core idea is to explain the common dynamic structure of the multivariate time series through a set of common (time series) factors. This is achieved by the introduction of flexible temporal correlation structures for the latent factors, previously assumed to be independent. This renders the DFM capable of assessing the complexity of time series data. Models along these lines include Geweke [20], Sargent and Sims [45], Molenaar [36], Engle and Watson [10], Peña and Box [39], Forni et al. [13] and Bai and Ng [4].

Earlier approaches have been primarily concerned with modeling multivariate stationary time series considering latent factors with a time-varying mean function. In this context, [39] proposed a methodology to identify the number of latent factors in a vector of stationary times series. Specifically, temporal correlation is introduced through a k-dimensional vector that follows an autoregressive moving average process (see [36, 4] and references therein for related ideas). For the nonstationary case, Peña and Poncela [40, 41] proposed a methodology for building DFM for nonstationary time series in state space form and, more recently, Pan and Yao [38] introduced a new approach that allows nonstationary factors not necessarily driven by unit roots.

In this Section we focus on DFM for both stationary and nonstationary time series where the k-dimensional latent factor $\boldsymbol{f}_t$ (state vector) follows a general VARMA(p, q) representation

$$\boldsymbol{f}_t = \boldsymbol{\Gamma}_1 \boldsymbol{f}_{t-1} + ... + \boldsymbol{\Gamma}_p \boldsymbol{f}_{t-p} + \boldsymbol{\omega}_t + \boldsymbol{\Xi}_1 \boldsymbol{\omega}_{t-1} + ... + \boldsymbol{\Xi}_q \boldsymbol{\omega}_{t-q}, \tag{1.2}$$

where $\boldsymbol{\omega}_t \sim N(\mathbf{0}, \boldsymbol{\Lambda})$, $\forall t$. The latent nature of the factors make it difficult to precisely estimate this full model. It what follows, we will concentrate the presentation on a simplified VAR(1) version, obtained when $p = 1$ and $q = 0$. A number of features are more clearly understood in this setting and will be discussed below. This factor evolution is driven by the following transition equation

$$\boldsymbol{f}_t = \boldsymbol{\Gamma} \boldsymbol{f}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \boldsymbol{\Lambda}), \tag{1.3}$$

where $\boldsymbol{\Gamma}$ is a symmetric $k \times k$ autoregressive coefficient matrix characterizing the dynamic evolution of the common factors and $\boldsymbol{\Lambda}$ is a $k \times k$ covariance matrix with elements λ_{ij}, $i, j = 1, \ldots, k$. Note that $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}$ are not necessarily diagonal matrices so they can be defined to deal, for instance, with seasonal components and nonstationary common factors. Equations (1.1) and (1.2) or (1.3) define the dynamic factor model and, in a similar fashion to standard factor analysis, the latent factors $\boldsymbol{f}_t$ capture the time-varying correlation structure of the data.

Working within a Bayesian framework, some important issues related to model specification and posterior inference can be mentioned at this point:

Prior specification The prior for the latent factor is given in eqn (1.3) and completed by $\boldsymbol{f}_0 \sim N(m_0, C_0)$ with known hyperparameters m_0, C_0. As was mentioned before, many specifications for $\boldsymbol{\Gamma}$ can be considered. One possibility for the $\boldsymbol{\Lambda}$ matrix is a diagonal form with elements λ_i. In this case, a typical choice of prior for the λ_i's is independent Gamma distributions. Similar independence assumptions can be made for the autoregressive matrix $\boldsymbol{\Gamma}$. One possibility is to consider $\boldsymbol{\Gamma} = \text{diag}(\gamma_1, \ldots, \gamma_k)$ such that, $\gamma_j \sim N(0, a)$ independent, for $j = 1, \ldots, k$, for some large value of a if one want to represent vague prior information. If one is concerned with the possibility of unit roots and non-stationarity, the mixture prior $\gamma_j \sim \pi N_{(-1,1)}(0, a) + (1 - \pi)\delta_1(\gamma_j)$ may be assumed for the autoregressive coefficients, where $\pi \in (0, 1]$ and a are known hyperparameters, $N_{(l,u)}(\cdot, \cdot)$ denotes the normal distribution constrained to assumed values only in (l, u), $\delta_1(\gamma_j) = 1$ if $\gamma_j = 1$ and $\delta_j(\gamma_j) = 0$ if $\gamma_j \neq 1$ (see [26] for more details); for $\pi \neq 1$, the mixture prior allows the possibility that nonstationary factors be incorporated, if $\pi = 1$ we are in the stationary case.

Correlated factors can also be incorporated into the DFM. One example of that is the inclusion of h seasonal common factors to capture a possibly periodic behavior of the time series. In that case, $\boldsymbol{\Gamma}$ could be specified as $\boldsymbol{\Gamma} = \text{diag}(\boldsymbol{\Gamma}_0, \boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_h)$ where $\boldsymbol{\Gamma}_0 = \text{diag}(\gamma_{0,1}, \ldots, \gamma_{0,k})$,

$$\boldsymbol{\Gamma}_l = \begin{pmatrix} \cos(2\pi l/p) & \sin(2\pi l/p) \\ -\sin(2\pi l/p) & \cos(2\pi l/p) \end{pmatrix}, \quad l = 1, \ldots, h,$$

p is the seasonal period and $h = p/2$ is the *maximum* number of harmonics needed to capture the seasonal behavior of the time series, (see [54], Chapter 8, for more details). As a consequence $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\Lambda}_0, \boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_h)$ and each $\boldsymbol{\Lambda}_l$ is no longer diagonal with *inverted Wishart* distribution as a prior.

Factor loadings specification For the factor loadings one can take independent normal priors for each element of $\boldsymbol{\beta}$ such that $\beta_{ii} \sim N_{(0,\infty)}(0, b)$, $\beta_{ij} \sim N(0, b)$ only for $i > j$ (see [34] for more details) since identifiability constraints impose $\beta_{ij} = 0$ for $i < j$. However, in practice, one may be also interested in including conditional dependencies within the elements of $\boldsymbol{y}_t$. In order to do that, the underlying idea is to include a flexible correlation structure into the columns of $\boldsymbol{\beta}$, denoted by $\boldsymbol{\beta}_{(j)}$ $(j = 1, \ldots, k)$. In the context of spatial analysis, a number of papers have examined inducing dependencies through $\boldsymbol{\beta}_{(j)}$. For example, in [52] the columns of $\boldsymbol{\beta}$ are modeled as orthonormal basis functions and in [6, 44] smoothed deterministic kernels are used to built $\boldsymbol{\beta}$. Alternatively, Lopes et al. [32] introduced a spatial DFM where the columns of the factor loadings matrix follow independent Gaussian random fields. This idea will be discussed and illustrated in the next Section for modeling space-time data. Additional developments on factor loadings specification include, for example, Carvalho et al. [8] and Bhattacharya et al. [5] for sparse factor analysis, and Lopes and Carvalho [30] for latent time-varying loadings, among others.

Posterior inference Fully Bayesian treatment of the standard and dynamic factor models via MCMC methods is described in detail in [34] and [2], respectively. More specifically, the inference procedure is designed for two cases: known and unknown number of factors k. Considering that the number of factor k is known, the MCMC scheme described in [34] can be easily adapted where the common factor are jointly sampled via the well known forward filtering backward sampling (FFBS) scheme [7,14]. For the second case, model selection is performed by computing *posterior model probabilities* (PMP) for different choices of k. In particular, the *reversible jump* MCMC algorithm, proposed/described in [34,32] for DFM, can be used. The algorithm allows for a simple method of calculating the PMP from preliminary MCMC runs. As mentioned in the previous references, the Bayesian model search via RJMCMC penalizes over and under-parametrized factor models.

1.3.2 *Hierarchical DFM*

A common criticism in DFM is that the common latent factors are difficult to interpret. In large n settings and for multi-level datasets, the dimension reduction of the problem may involve loss of data structure. In this context, a hierarchical construction of the model to allow a progressive reduction in the dimensionality as the levels becomes higher may be desired. For example, the hierarchical construction for dynamic linear models (DLM's) proposed by Gamerman and Migon [16] provides a general framework for analysis of multivariate time series. In accordance with this idea and following the same notation introduced in Subsection 1.3.1, the *3-level hierarchical dynamic factor models* (HDFM) can be written as

$$\boldsymbol{y}_t = \boldsymbol{\beta}_1 \boldsymbol{f}_{1t} + \boldsymbol{\epsilon}_{1t}, \quad \boldsymbol{\epsilon}_{1t} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_1), \tag{1.4}$$

$$\boldsymbol{f}_{1t} = \boldsymbol{\beta}_2 \boldsymbol{f}_{2t} + \boldsymbol{\epsilon}_{2t}, \quad \boldsymbol{\epsilon}_{2t} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_2), \tag{1.5}$$

$$\boldsymbol{f}_{2t} = \boldsymbol{\beta}_3 \boldsymbol{f}_{3t} + \boldsymbol{\epsilon}_{3t}, \quad \boldsymbol{\epsilon}_{3t} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_3), \tag{1.6}$$

$$\boldsymbol{f}_{3t} = \boldsymbol{\Gamma} \boldsymbol{f}_{3,t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \boldsymbol{\Lambda}), \tag{1.7}$$

where $\boldsymbol{f}_{it}$, $i = 1, 2, 3$, are k_i-dimensional vectors satisfying $k_1 > k_2 > k_2$, $\boldsymbol{\beta}_1$ is a $n \times k_1$ matrix, $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$ are $k_1 \times k_2$ and $k_2 \times k_3$ matrices respectively, and $\boldsymbol{\Gamma}$ is a $k_3 \times k_3$ matrix. More specifically, eqn (1.4) represents the observation equation, eqns (1.5) and (1.6) the structural equations and eqn (1.7) the system equation. As mentioned in [16], the previous HDFM can be reduced to considering only two level/stages of hierarchy by setting $\boldsymbol{\beta}_3 = \boldsymbol{I}_{k_3}$ and $\boldsymbol{\Sigma}_3 = \mathbf{0}$, a zero matrix. Again, further levels are easily induced but this would rarely be required.

1.3.3 *Generalized DFM*

The DFM can also be extended to allow for non Gaussian observations. More specifically, the generalized DFM (GDFM) is a hierarchical model where the first level equation (observation equation) is given by

$$p(y_{ti}|\eta_{ti}, \psi) = \exp\{\psi(y_{ti}\eta_{ti} - b(\eta_{ti})) + c(y_{ti}, \psi)\}, \tag{1.8}$$

where η_{ti} is the natural parameter and ψ is the dispersion parameter. The natural parameter η_{ti} is related to the temporal components through the link function v such that $\eta_{ti} = v(\theta_{ti})$. Consequently, the model is completed by specifying the following two levels of hierarchy

$$\boldsymbol{\theta}_t = \boldsymbol{\mu}_t + \boldsymbol{\beta} \boldsymbol{f}_t, \tag{1.9}$$

$$\boldsymbol{f}_t = \boldsymbol{\Gamma} \boldsymbol{f}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \boldsymbol{\Lambda}), \tag{1.10}$$

where $\boldsymbol{\theta}_t = (\theta_{t1}, \ldots, \theta_{tn})^T$, $\boldsymbol{\mu}_t$ is the mean level, and $\boldsymbol{\beta}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}$ have the same specifications as the SDFM. See [53] for more details in the context of generalized DLM's.

Full Bayesian treatment for this new class is more challenging, specifically for MCMC sampling the common factor. In the previous cases, the full conditional distribution for joint sampling this component was normal and thus easily sampled from using for example the FFBS. This is no longer valid here and efficient proposal are very difficult to obtain, specially for large time series with large T. Componentwise sampling is also very inefficient. The solution here is a compromise with this component sampled in blocks. To this end, a block sampling scheme that combines techniques such as extended Kalman filter and block sampling was proposed in [31] with good performance in the applications.

1.4 Applications to spatial statistics

In this Section, we discuss some applications of the above mentioned approaches for spatial and spatio-temporal processes.

The use of factor analysis to model multivariate spatial data has been treated in a number of ways. Here, we focus on the case in which factor analysis is used to identify cluster or groups of locations/regions (spatial dependence) whose temporal behavior is driven by a set of common dynamic latent factors (temporal dependence). In previous works, either common dynamic factors or factor loadings matrices are restricted to be deterministic functions. Specifically, when the common factors are non-stochastic the space-time dynamic model proposed by [48] is obtained. On the other hand, when $\boldsymbol{\beta}$ is defined as a deterministic function the structure proposed by [52, 6] is obtained.

Lopes et al. [32] introduced a new class of models called *spatial dynamic factor model* (SDFM) derived from the standard DFM. More specifically, the temporal dependence is modeled by the latent common factors and the spatial dependence is also modeled stochastically by the columns of the factor loadings $\boldsymbol{\beta}_{(j)}$. These are assumed to follow independent Gaussian processes. The role played by the stochastic structure is to allow further flexibility to the deterministic specification, that is restrictive by definition.

The SDFM is defined by eqns (1.1) and (1.3) where $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{nt})^T$ such that y_{it} is a observation measured at time t and location $s_i \in \mathbb{R}^d$. Each column $\boldsymbol{\beta}_{(j)} = (\beta_j(s_1), \ldots, \beta_j(s_n))^T$ is defined as

$$\boldsymbol{\beta}_{(j)} \sim N(\boldsymbol{\theta}_j, \tau_j^2 \boldsymbol{R}_{\boldsymbol{\phi}_j}), \quad \text{for} \quad j = 1, .., k, \tag{1.11}$$

where $\boldsymbol{\theta}_j$ is the n-dimensional mean vector, τ_j^2 is the common variance of the spatial process, $\boldsymbol{R}_{\boldsymbol{\phi}_j}$ is the matrix correlation function with (l, m)-element given by $\{\boldsymbol{R}_{\phi_j}\}_{l,m} = \rho_{\phi_j}(\|s_l - s_m\|)$, $l, m = 1, \ldots, n$ and $\rho_\phi(\cdot)$ represents a spatial correlation function like exponential or Matérn specified by the parameter ϕ.

As an illustration, Fig. 1.1 shows a simulated SDFM with $k = 2$ common latent factors. Note that the surfaces for $\boldsymbol{y}_t$ are driven by the spatial behavior of the $\boldsymbol{\beta}_{(j)}$'s $(j = 1, 2)$ weighted by the values of the common factors. It is important to mention that the SDFM implies nonseparable forms of the covariance function when $k \geqslant 2$. In fact, if $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}$ are diagonal matrices, the covariance between two different sites at two different time indexes is given by

$$\text{cov}(y_{it}, y_{j,t+h}) = \sum_{l=1}^{k} \lambda_l \gamma_l^h (1 - \gamma_k^2)(\tau_j^2 \rho_{\phi_j} + \theta_{il}\theta_{jl}).$$

That characteristic implies that the SDFM is able to model complex space-time interactions. In contrast, when $k = 1$ spatial and temporal covariance functions are separately identified.

Example 1.1 The SDFM is used to examine the spatio-temporal variation in weekly concentration levels of nitrate (NO_3) across 22 monitoring stations located in eastern US for $T = 312$ weeks (1st week of 1998 – 52nd week of 2003).

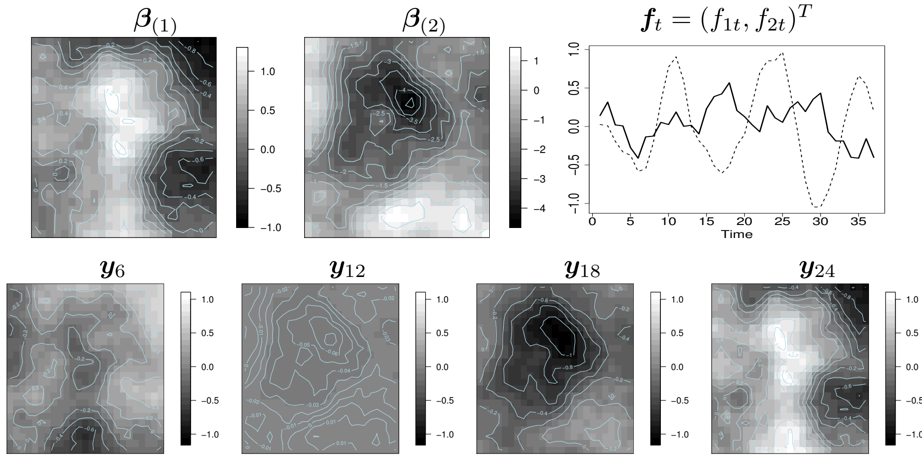


FIG. 1.1. Simulated spatial dynamic 2-factor model. *First row*: Gaussian processes for the two columns of $\boldsymbol{\beta}$ and simulated dynamic factors (time series) $\boldsymbol{f}_t = (f_{1t}, f_{2t})^T$ for $t = 1, \ldots, 36$. The first factor (dashed line) has a seasonal behavior with period $p = 12$. The second factor (solid line) follows an AR process with autoregressive parameter $\gamma_{22} = 0.9$. *Second row*: $\boldsymbol{y}_t$ processes following eqn (1.1) for $t = 6, 12, 18, 24$.

The logarithm transformation was used to normalize the data and a seasonal common factor was considered to capture the yearly periodic behavior repeated every 52 weeks (seasonal period). The SDFM with $k = 3$ regular factors and 1 seasonal factor was compared against other models and selected for fit to the data. Also, a Matérn correlation function is used to specify the spatial correlation structure of each $\boldsymbol{\beta}_{(j)}$ $(j = 1, \ldots, 4)$ with smoothness parameter equal to 1.

Fig. 1.2 shows some posterior results of the fitted model. The four maps of the factor loadings (estimated via Bayesian interpolation) show distinct spatial patterns across the study area. Note that the temporal behavior of the time series is directly related with the higher values of the interpolated surfaces (white areas). The loadings for the second factor are higher in the western region, specifically in some areas of Indiana, Ohio and Kentucky. This factor mimics, roughly, the spatial pattern of the NO_3 across time. Also, the results indicated that this factor is nonstationary. In addition, Fig 1.3, panel (a) shows a plot of observed vs. fitted values (considering the original scale). The points roughly follow a straight line indicating good predictions. Panel (b) presents interpolation results for one of the out-of-sample monitoring stations (SPD station). The NO_3 interpolated values are very close to the real values indicating the good interpolated performance of the model. Finally, panels (c) and (d) show forecast values for the 1st and 12th weeks of 2004. Note that the spatial pattern is almost preserved but with lowest values for the entire region for the 12th week. That behavior is ex-

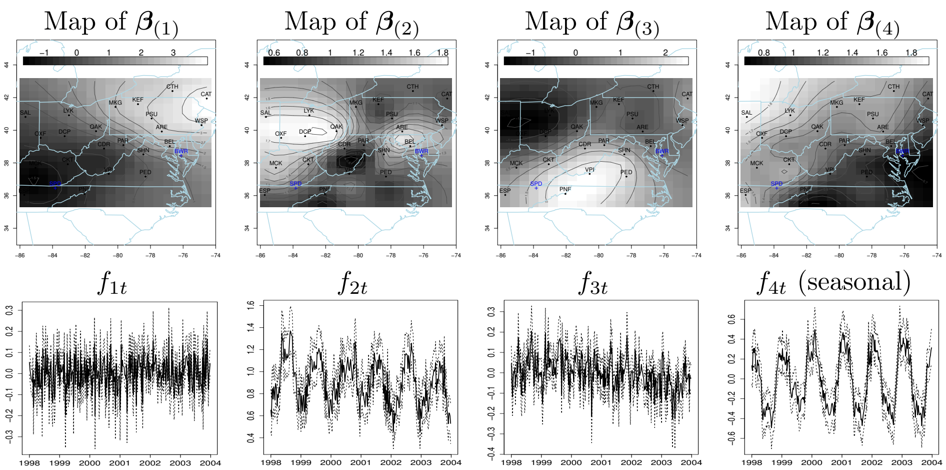
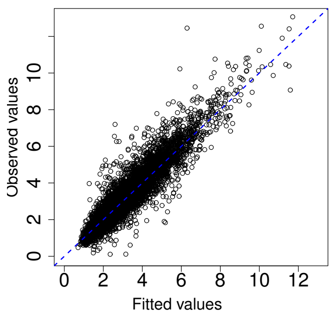


FIG. 1.2. *First row*: Bayesian interpolation of the factor loadings. Values represent the range of the posterior means. *Second row*: Posterior means of the factors. Solid lines represent the posterior means and dashed lines the 95% credible intervals.

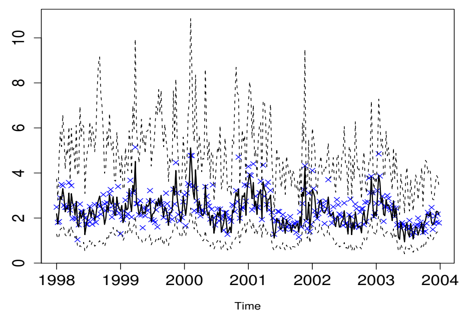
pected since high concentration levels of nitrate are expected in the first weeks of the year. ‖

The SDFM can also be extended to allow for non Gaussian observations. Lopes et al. [31] introduced a new class of spatio-temporal models for multivariate exponential family data, called the *generalized spatial dynamic factor model* (GSDFM). In this formulation, the spatial and temporal components are modeled via a latent factor analysis of the canonical transformation of the mean function. The model is given by (1.8)–(1.10) with the addition of the Gaussian prior (1.11) for the loadings. Note that this class of model also leads to a nonseparable spatio-temporal covariance structure. This characteristic is associated with the linear predictor $\boldsymbol{\theta}_t$ where, for $k > 1$, both spatial and temporal covariance structures can not be separately identified.

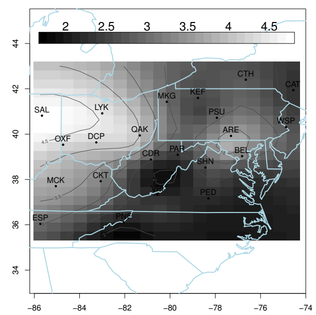
Example 1.2 We are interested in modeling daily rainfall occurrences (over 1 mm) in northern Oceania in 2001. The data contains $T = 365$ binary observations measured at 19 meteorological stations, 14 of them located in the Federated States of Micronesia and 5 in the Marshall Islands. Fig. 1.4(a) shows the study area as well as the geographic location of the stations. We aim to identify microclimates over the study region and also fit rain probability maps for the whole area and across time. The model considered is given by eqns (1.8)–(1.10) such that $y_{ti} \sim \text{Bernoulli}(p_{ti})$, logistic link function $\theta_{ti} = \log(p_{ti}/(1 - p_{ti}))$, a Matérn correlation function for the specification of the factor loadings matrix and $\boldsymbol{\theta}_{(j)} = \mu_j \mathbf{1}_{17}$. The GSDFM was fitted considering $k = 1, 2, 3, 4$ common factors. Comparison between models are based on the PMP. In this application,



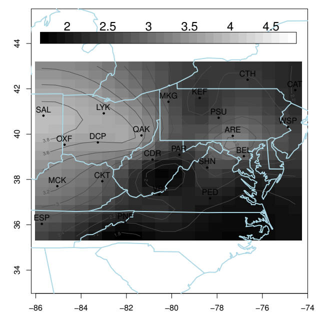
(a) Fitted vs observed values



(b) Interpolated values at SPD station



(c) 2004, 1st week

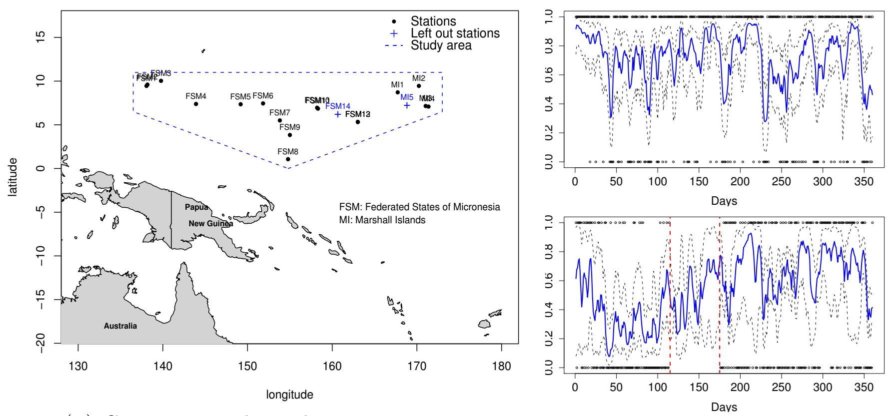


(d) 2004, 12th week

FIG. 1.3. (a) Plot of fitted vs. observed values of NO_3 for the whole period and for the 22 stations. (b) Interpolated values at SPD station left out from the sample used for fitting. Dashed lines represent the 95% credible intervals and the symbol $\times$ the observed NO_3 concentration level. (c)-(d) Forecast values for two different weeks in 2004.

the model with 3 common factor shows the best results with PMP equal to 0.46.

Fig. 1.5 shows the interpolated loading associated with the first, second and third factor as well as the estimated temporal behavior of the common factors. The spatial loadings, interpolated via Bayesian kriging, indicate a smooth variation in different directions, specially for the first factor. These findings allow the recognition of microclimates, more specifically in the eastern part (around Marshall Islands), as showed in the map for $\boldsymbol{\beta}_{(3)}$. In addition, the results indicate the presence of one nonstationary factor (2nd factor) with posterior probability $\mathrm{pr}(\gamma_2 = 1|\boldsymbol{y}_{1:T}) = 0.89$. Other interesting results of the model are the rain probability time series for observed and interpolated stations, the latter interpolated via Bayesian kriging. Fig. 1.4(b) shows those time series for two stations, FSM13 (included in the analysis) and MI5 (left out of the analysis). The posterior probabilities of rainfall occurrence seem to follow the general trend of the observed binary time series. Also, for station MI5, we tested the capability of the model in handling missing data, specifically for the days 121–170 (delimited by the ver-



(a) Stations and study area

(b) Posterior probability of rain

FIG. 1.4. (a) Location of the monitoring stations. (b) Daily posterior probability of rainfall occurrence at FSM13 station (above) and MI5 station(below). The latter shows the results of the spatial interpolation since station MI5 was left out of the analysis for interpolation purposes. Dots are rain indicators, solid lines are rain mean probabilities and dashed lines are 95% credibility intervals.

tical dashed lines). Note that the temporal behavior of the probability is mainly driven by the third factor, which is expected given the location of the station. ‖

An alternative specification for the factor loadings matrix can be considered for the SDFM. For example, the discrete process convolution approach proposed by Higdon [25], or (more recently) the spatial model considering compact support kernels as proposed in Lemos and Sansó [28]. Here, we discuss a related approach that uses an approximate Gaussian processes for $\boldsymbol{\beta}_{(j)}$'s instead of deterministic kernels. This new model specification was recently proposed by Salazar et al. [43] for comparing and blending regional climate model predictions. See the cited paper for more details related to the fully Bayesian treatment of the model.

More specifically, for each $y_t(s)$ measured at time t and location s we have that

$$y_t(s) = \mu_t(s) + \omega_t(s) + \epsilon_t(s), \quad \epsilon_t(s) \sim N(0, \sigma^2), \tag{1.12}$$

where $\mu_t(s)$ may represent a regression component and $\omega_t(s)$ is the space-time component that follows a Gaussian process. Many specifications can be considered for $\omega_t(s)$. Here we opted to use the *modified predictive process* (see [11] and the references therein), letting $\omega_t(s) = \tilde{\omega}_t(s) + \tilde{\epsilon}_t(s)$, where $\tilde{\epsilon}_t(s) \sim N(0, \tau^2 - v(s)^T \boldsymbol{H}^{-1} v(s))$ and $\tilde{\omega}_t(s)$ is represented on a set of k basis functions $B_l(s) = [v(s)^T \boldsymbol{H}^{-1}]_l$ where $v(s) = \tau^2(\rho_\phi(s, s_1^*), \ldots, \rho_\phi(s, s_k^*))$, $\{\boldsymbol{H}\}_{lm} = \tau^2 \rho_\phi(s_l^*, s_m^*)$ for $l, m = 1, \ldots, k$ and $\{s_l^*; l = 1, \ldots, k\}$ a set of *selected knots* and is given by $\tilde{\omega}_t(s) = \sum_{l=1}^k B_l(s)\gamma_{t,l} = \boldsymbol{B}(s)^T \boldsymbol{\gamma}_t$.

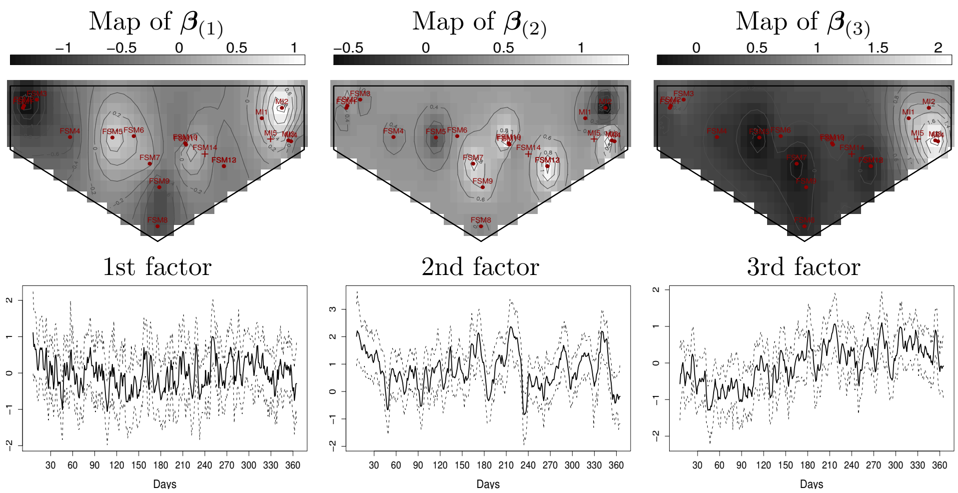


FIG. 1.5. *First row*: Bayesian interpolation of the three columns of the factor loadings matrix. Values represent the range of the posterior means. *Second row*: Daily posterior means of the first, second and third dynamic factors. Solid lines represent the posterior means and dashed lines the 95% credible intervals.

The temporal evolution of $\boldsymbol{\gamma}_t$ is specified as $\boldsymbol{\gamma}_t \sim N(\psi\boldsymbol{\gamma}_{t-1}, \boldsymbol{H})$. After a SVD decomposition of $\boldsymbol{H} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T$ and letting $\boldsymbol{\gamma}_t = \boldsymbol{P}\boldsymbol{f}_t$ we can rewrite $\tilde{\omega}_t(s)$ as $\tilde{\omega}_t(s) = \boldsymbol{B}(s)^T\boldsymbol{P}\boldsymbol{f}_t = \boldsymbol{\beta}(s)^T\boldsymbol{f}_t$ and therefore $\boldsymbol{f}_t \sim N(\psi\boldsymbol{f}_{t-1}, \boldsymbol{\Lambda})$ (with independent elements given that $\boldsymbol{\Lambda}$ is a diagonal matrix).

If we conveniently rewrite eqn (1.12) in vector notation and by considering the previous specification we have

$$\begin{aligned}\boldsymbol{y}_t &= \boldsymbol{\mu}_t + \boldsymbol{\beta}\boldsymbol{f}_t + \tilde{\boldsymbol{\epsilon}}_t + \boldsymbol{\epsilon}_t,\\ \boldsymbol{f}_t &\sim N(\psi\boldsymbol{f}_{t-1}, \boldsymbol{\Lambda}).\end{aligned}$$

The previous specification resembles the SDFM where $\boldsymbol{\beta}$ is a $n \times k$ matrix with k is the number of pre-selected knots (fixed), $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_k)$ with $\lambda_1 > \ldots > \lambda_k$, therefore $\boldsymbol{\beta}_{(1)}$ describes the main model of spatial variability, $\boldsymbol{\beta}_{(2)}$ the second, and so on.

Finally, in the following example we describe a *spatial hierarchical dynamic factor model* (SHDFM) for socio-economic multi-level measurements. The idea is to build a model-based vulnerability index that account for the different levels of hierarchy (for example, census tracts and capitals). The spatial version of this model was proposed by Lopes et al. [33] to build Uruguayan vulnerability index at different geographical resolutions.

Example 1.3 Consider the p-dimensional vector of socio-economical variables $\boldsymbol{y}_{ijt} = (y_{ijt,1}, \ldots, y_{ijt,p})$ at capital i ($i = 1, \ldots, I$), census tract j ($j = 1, \ldots, n_i$)

and time t. We aim to infer vulnerability indexes at two levels of resolution: capitals (coarse level) and census tracts (fine level). The proposed *two level SHDFM* can be written as

$$\begin{aligned}
\boldsymbol{y}_{ijt} &= \mu + \boldsymbol{\beta} f_{ijt}^{(1)} + \boldsymbol{\epsilon}_{ijt}^{(1)}, \quad \boldsymbol{\epsilon}_{ijt}^{(1)} \sim N(\mathbf{0}, \boldsymbol{\Sigma}),\\
f_{ijt}^{(1)} &= \theta_{it} + f_{ijt}^{(2)} + \epsilon_{ijt}^{(2)}, \quad \epsilon_{ijt}^{(2)} \sim N(0, \psi),\\
\boldsymbol{f}_{it}^{(2)} &= \boldsymbol{f}_{i,t-1}^{(2)} + \boldsymbol{w}_{it}, \quad \boldsymbol{w}_{it} \sim N(\mathbf{0}, \tau_i^2 \boldsymbol{P}_i)\\
\boldsymbol{\theta}_t &= \boldsymbol{\theta}_{t-1} + \boldsymbol{v}_t, \quad \boldsymbol{v}_t \sim N(\mathbf{0}, \delta^2 \boldsymbol{H})
\end{aligned}$$

where $\boldsymbol{\beta} = (1, \beta_2, \ldots, \beta_p)^T$ (that implies a 1-factor model for the first level), $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2)$. Within capital i, the one dimensional factor $f_{ijt}^{(1)}$ is decomposed as the sum of two spatial components: θ_{it} (capital-level) and $f_{ijt}^{(2)}$ (census tract-level). Note that the $f_{ijt}^{(2)}$'s are conditionally independent and the joint vector $\boldsymbol{f}_{it}^{(2)} = (f_{i1t}^{(2)}, \ldots, f_{in_it}^{(2)})^T$ follows a Markovian evolution where the system innovation $\boldsymbol{w}_{it}$ follows a *proper Gaussian Markov random field*. More specifically, $\boldsymbol{P}_i = (\boldsymbol{I}_{n_i} + \phi \boldsymbol{M}_i)^{-1}$ where $\{\boldsymbol{M}_i\}_{lk} = m_k^{(i)}$ if $l = k$ and $\{\boldsymbol{M}_i\}_{lk} = -1/d_{lk}^{(i)}$ if sites l and k are neighbors (denoted by $l \sim k$) and zero otherwise, $m_k^{(i)} = \sum_{l \sim k} 1/d_{lk}^{(i)}$ and $d_{lk}^{(i)}$ is the Euclidean distance between centroids of regions l and k (see [51] for more details about this construction). An additional assumption is that the θ_{it}'s are conditionally independent so the joint vector $\boldsymbol{\theta}_t = (\theta_{1t}, \ldots, \theta_{n_it})^T$ follows a Markovian evolution where the innovation $\boldsymbol{v}_t$ follows a zero mean Gaussian process with covariance structure $\boldsymbol{H}$ driven by the Euclidean distances between the centroids of the capitals. In this multi-level factor model, $f_{ijt}^{(1)}$ represents the vulnerability index at the census tract level of the capital i (fine level) and θ_{it} is the capital vulnerability index (coarse level). Note that the SHDFM takes full advantage of the multi-level data structure through the hierarchical specification of the common factor. ‖

1.5 Regression with dynamic factor models

The ideas so far have been restricted to a single collection of time series. Even though any time series problem can be cast in a single collection of time series, the collections usually considered have a unified framework relating them. Typically they are measurements in a variety of settings of the same quantity. As such, they behave like a (random) sample of time series.

This Section extends the scope of DFM beyond random samples by considering regression. The general idea of a regression is to explain a variable by a set of covariates. In time series context, this means explaining the behavior of a (possibly multivariate) time series by a number of related explanatory time series. Although the approach is quite general, it is better explained without much loss of generality in the context of simple regression.

So from now on, we will restrict our attention to the situation where a collection of time series forming a multivariate time series $\boldsymbol{y}_t$ of a given variable is explained by another collection of time series forming a multivariate time series $\boldsymbol{x}_t$ of another given variable. This is a well known set-up in time series, sometimes referred to as transfer response models, covered in many standard time series books.

The idea of dimensionality reduction via factor models in the regression context is also not new and is also related to the basic factor model set-up, as expected. Considering a set of multivariate observations $\boldsymbol{y}_t$ related to another collection of observations $\boldsymbol{x}_t$ at a latent level gives rise to the structural equation model (SEM)

$$\begin{aligned}
\boldsymbol{y}_t &= \boldsymbol{\beta}_y \boldsymbol{g}_t + \boldsymbol{\epsilon}_{yt}, \quad \boldsymbol{\epsilon}_{yt} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_y), \\
\boldsymbol{x}_t &= \boldsymbol{\beta}_x \boldsymbol{f}_t + \boldsymbol{\epsilon}_{xt}, \quad \boldsymbol{\epsilon}_{xt} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_x), \\
\boldsymbol{g}_t &= \boldsymbol{\Delta}_y \boldsymbol{g}_t + \boldsymbol{\Delta} \boldsymbol{f}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_{gt} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_g).
\end{aligned}$$

The loading matrices $\boldsymbol{\beta}_x$ and $\boldsymbol{\beta}_y$ play exactly the same role as in factor models. The novelty here is the introduction of the relational matrices $\boldsymbol{\Delta}$ and $\boldsymbol{\Delta}_y$, establishing a regression relation between the set of variables $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$ at a latent level. This basic SEM set-up is described in detail by [46]. The Bayesian approach to SEM is described in [37]. It is worth pointing out that the standard factor model (1.1) is recovered after suitable concatenation of observables $(\boldsymbol{x}_t, \boldsymbol{y}_t)$ and factors $(\boldsymbol{f}_t, \boldsymbol{g}_t)$, respectively.

The extension towards time series problems is not difficult to obtain following the standard recipe of the previous sections of this Chapter. Just like (1.2) establishes the dynamic of the factors for the time series settings, [9] proposed the dynamics of the two sets of factors as

$$\begin{aligned}
\boldsymbol{g}(t) &= \sum_{i=0}^{p} \boldsymbol{\Delta}_{yi} \boldsymbol{g}_{t-i} + \sum_{j=0}^{q} \boldsymbol{\Delta}_j \boldsymbol{f}_{t-j} + \boldsymbol{\varepsilon}_{gt}, \quad \boldsymbol{\varepsilon}_{gt} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_g) \\
\boldsymbol{f}(t) &= \sum_{j=1}^{s} \boldsymbol{\Delta}_{xj} \boldsymbol{f}_{t-j} + \boldsymbol{\varepsilon}_{ft}, \quad \boldsymbol{\varepsilon}_{ft} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_f).
\end{aligned}$$

He refer to this model as dynamic SEM. [12] cast the dynamic SEM in state space form and applied it to the analysis of environmental problems. Once again, the DFM given in (1.1)–(1.2) can be recovered by appropriate concatenation of observables $(\boldsymbol{x}_t, \boldsymbol{y}_t)$ and factors $(\boldsymbol{f}_t, \boldsymbol{g}_t)$. Even more so than in DFM, it is very hard to estimate this model in its full expression for typical applications. The more natural simplifications can be obtained by restricting the order p, q, s of the auto- and cross-regressions to small values, say 1 or 2. Once again, the model can be written in state-space form by appropriately enlarging the state vector according to the order of lagged dependence of the latent factors.

These ideas were applied to the Spatial Statistics context by [27]. Once again, the columns of the loading matrices were assumed to follow independent Gaussian processes in order to impose stochastic similarity between neighboring sites.

The presence of two sets of variables introduces further possibilities beyond standard DFM. In particular, relationships between the loading matrices $\boldsymbol{\beta}_x$ and $\boldsymbol{\beta}_y$ may be introduced. For example, [27] use $\boldsymbol{\beta}_x$ as a (latent) design matrix for the mean of $\boldsymbol{\beta}_y$. Illustrative examples and further discussion about model specification and evaluation in provided in [27].

1.6 Concluding remarks

This Chapter was concerned with the discussion on the use of factor models in the time series context via state space formulation. The key element of the approach is its ability to reduce the dimensionality in the multivariate time series context and at the same time shed some light on the structure of the relationship between the different time series. Our presentation has focused exclusively on the discussion about model building. As a result, a number of other issues were not addressed. We will briefly comment upon them now.

By far the most important item not yet discussed is prediction. Time series is primarily concerned with forecasting into the future. The model-based approach of state space models followed here enables easy calculation of the predictive distributions $p(\boldsymbol{y}_{T+h}|\boldsymbol{y}^T)$ where $\boldsymbol{y}^t = (\boldsymbol{y}_1, ..., \boldsymbol{y}_t)$, for all t, h. This is available approximately after obtaining the predictive distribution $p(\boldsymbol{f}_{T+h}|\boldsymbol{y}^T)$ for the latent factors and predictions can be approximated by samples. This exercise was made in examples 1.2 and 1.3. Note also that the prediction exercise is very similar to the kriging exercise required for spatial extrapolation and that was also illustrated in the examples above. Details are provided in [32].

Another important issue is generated by the large amount of possibilities rendered by these classes of models. There are a number of options provided by the choice of the number of regular and seasonal factors and the orders of the factors dynamics. There are a few options available for model selection including AIC, BIC and DIC. These are mostly based on model fit after some penalization for complexity. One may also consider the estimation of the number of factors in a RJMCMC algorithm. In this time series, we feel that model comparison should be more heavily based on predictions rather than fit. Even more so than in the other areas of Statistics, given the relevance of prediction for the time series context. Standard practice in this area is based on cross-validation, where a portion of the data is left out of the fit. This portion typically consist on the last observed points to mimic the real exercise of forecasting into the future.

The description above illustrate some of the many possibilities for the use of factor models in the time series context. The presentation of the spatial applications was entirely on data collected under continuous spatial variation. Similar ideas were applied to the context of discrete spatial variation or areal data by [47]. There are a number of extensions that can be envisaged by appropriately combining some of the model components described in this Chapter. We are currently working on some of them and will be reporting it in the near future.

Acknowledgements

D. Gamerman was supported by CNPq-Brazil and *Fundação de Amparo à Pesquisa no Estado do Rio de Janeiro* (FAPERJ foundation). E. Salazar would like to thank the support of the Department of Electrical and Computer Engineering at Duke University.

References

[1] Anderson, T. W. (1963). The use of factor analysis in the statistical analysis of multiple time series. *Psychometrika*, **28**, 1–25.

[2] Aguilar, O. and West, M. (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business and Economic Statistics*, **18**, 338–357.

[3] Arminger, G. and Muthén, B. O. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika*, **63**, 271–300.

[4] Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, **70**, 191–222.

[5] Bhattacharya, A. and Dunson, D. B. (2011). Sparse Bayesian infinite factor models. *Biometrika*, **98**, 291–306.

[6] Calder, C. (2007). Dynamic factor process convolution models for multivariate space-time data with application to air quality assessment. *Environmental and Ecological Statistics*, **14**, 229–247.

[7] Carter, C., and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, **81**, 541–553.

[8] Carvalho, C., Chang, J., Lucas, J., Nevins, J., Wang, Q. and West, M. (2008). High-dimensional sparse factor modeling: applications in gene expression genomics. *Journal of the American Statistical Association*, **103**, 1438–1456.

[9] Cziráky, D. (2004). Estimation of dynamic structural equation models with latent variables. *Metodološki zvezki*, **1**, 185–204.

[10] Engle, R. and Watson, M. (1981). A one-factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association*, **76**, 774–81.

[11] Finley, A. O., Sang, H., Banerjee, S. and Gelfand, A. E. (2009). Improving the performance of predictive process modeling for large datasets. *Computational Statistics and Data Analysis*, **53**, 2873–2884.

[12] Fontanella, L., Ippoliti, L. and Valentini, P. (2007). Environmental pollution analysis by dynamic structural equation models. *Environmetrics*, **18**, 265–283.

[13] Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000). The generalized dynamic factor model: identification and estimation. *The Review of Economics and Statistics*, **82**, 540–554.

[14] Frühwirth-Schnatter. S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, **15**, 183–202.

[15] Frühwirth-Schnatter, S. and Lopes, H. F. (2010). Parsimonious Bayesian factor analysis when the number of factors is unknown. *Technical Report.*

The University of Chicago Booth School of Business.

[16] Gamerman, D. and Migon, H. S. (1993). Dynamic hierarchical models. *Journal of the Royal Statistical Society, Series B*, **55**, 629–642.

[17] Gamerman, D. and Smith, A. F. M. (1996). Bayesian analysis of longitudinal data studies, in *Bayesian Statistics 5* (eds J. M. Bernardo et al.), Oxford University Press, Oxford, pp. 587–98.

[18] Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398–409.

[19] George, E. I. and R. McCulloch (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, **88**, 881–889.

[20] Geweke, J. (1977). The dynamic factor analysis of economic time series models. In *Latent Variables in Socio-Economic Models* (eds. D. J. Aigner and A. S. Goldberger). North Holland: Amsterdan, 365–383.

[21] Geweke, J. F. and Zhou, G. (1996). Measuring the pricing error of the arbitrage pricing theory. *The Review of Financial Studies*, **9**, 557–587.

[22] Gorsuch, R. L. (1983). *Factor Analysis.* 2nd edition, Hillsdale: Lawrence Erlbaum Associates.

[23] Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711–732.

[24] Higdon, D. (1998). A process-convolution approach to modelling temperatures in the north Atlantic Ocean. *Environmental and Ecological Statistics*, **5**, 173–190.

[25] Higdon, D. (2002). Space and space-time modeling using process convolutions. In *Quantitative Methods for Current Environmental Issues*, eds. C. Anderson, V. Barnett, P. C. Chatwin, and A. H. El-haarawi, London: Springer Verlag, pp 37–56.

[26] Huerta, G. and West, M. (1999). Priors and component structures in autoregressive time series models. *Journal of the Royal Statistical Society, Series B*, **61**, 881–899.

[27] Ippoliti, L., Valentini, P. and Gamerman, D. (2012). Space-time modelling of coupled spatio-temporal environmental variables. To appear in *Applied Statistics*.

[28] Lemos, R. T. and Sansó, B. (2009). A spatio-temporal model for mean, anomaly, and trend fields of north Atlantic sea surface temperature. *Journal of the American Statistical Association*, **104**, 5–25.

[29] Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model (with discussion). *Journal of the Royal Statistical Society, Series B*, **34**, 1–41.

[30] Lopes, H. F. and Carvalho, C. M. (2007). Factor stochastic volatility with time varying loadings and Markov switching regimes. *Journal of Statistical Planning and Inference*, **137**, 3082–3091.

[31] Lopes, H. F., Gamerman, D. and Salazar, E. (2011) Generalized spatial dynamic factor analysis. *Computational Statistics and Data Analysis*, **55**,

1319–1330.

[32] Lopes, H. F., Salazar, E. and Gamerman, D. (2008). Spatial dynamic factor analysis. *Bayesian Analysis*, **3**(4), 759–792.

[33] Lopes, H. F., Schmidt, A. M., Salazar, E., Gomez, M. and Achkar, M. (2011) Measuring the vulnerability of the Uruguayan population to vector-borne diseases via spatially hierarchical factor model. *Annals of Applied Statistics.* (To appear)

[34] Lopes, H. F. and West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica*, **14**, 41–67.

[35] Meng, X. L. and Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica*, **6**, 831–860.

[36] Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, **50**, 181–202.

[37] Palomo, J., Dunson, D. B. and Bollen, K. (2007). Bayesian structural equation modeling. *Handbook of Latent Variable and Related Models*, S-Y. Lee (editor), Elsevier.

[38] Pan, J. and Yao, Q. (2008). Modelling multiple time series via common factors. *Biometrika*, **95**, 365–379.

[39] Peña, D. and Box, G. (1987). Identifying a simplifying structure in time series. *Journal of the American Statistical Association*, **82**, 836–843.

[40] Peña, D. and Poncela, P. (2004). Forecasting with nonstationary dynamic factor models. *Journal of Econometrics*, **119**, 291–321.

[41] Peña, D. and P. Poncela (2006). Nonstationary dynamic factor analysis. *Journal of Statistical Planning and Inference*, **136**, 1237–1257.

[42] Polasek, W. (1997). Factor analysis and outliers: a Bayesian approach. Discussion paper, University of Basel.

[43] Salazar, E., Sansó, B., Finley, A., Hammerling, D., Steinsland, I., Wang, X. and Delamater, P. (2011) Comparing and blending regional climate model predictions for the American southwest. *Journal of Agricultural, Biological, and Environmental Statistics.* (To appear)

[44] Sansó, B., Schmidt, A. M., and Nobre, A. A. (2008). Bayesian spatio-temporal models based on discrete convolutions. *Canadian Journal of Statistics*, **36**, 239–258.

[45] Sargent, T. J. and Sims, C. A. (1977). Business cycle modelling without pretending to have too much a priori economic theory. In *New Methods in Business Research* (ed. C. A. Sims). Federal Reserve Bank of Minneapolis.

[46] Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling*, Chapman & Hall, Boca Raton.

[47] Strickland, C., Simpson, D., Turner, I., Denham, R. and Mengersen, K. (2010). Fast Bayesian analysis of spatial dynamic factor models for multi-temporal remotely sensed imagery. *Applied Statistics*, **60**, 109-124.

[48] Stroud, R., Müller, P., and Sansó, B. (2001). Dynamic models for spatiotemporal data. *Journal of the Royal Statistical Society, Series B*, **63**, 673-689.

[49] Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review*, **38**(5), 406–427.

[50] Thurstone, L. L. (1947). *Multiple Factor Analysis.* Chicago: University of Chicago Press.

[51] Vivar, J. C. and Ferreira, M. A. R. (2009). Spatiotemporal models for Gaussian areal data. *Journal of Computational and Graphical Statistics*, **18**(3), 658–674.

[52] Wikle, C. K. and Cressie, N. (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika*, **86**, 815–829.

[53] West, M., Harrison, P. J. and Migon, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting (with discussion). *Journal of the American Statistical Association*, **81**, 741–750.

[54] West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*, Springer, New York.