Alexandra M. Schmidt and Hedibert F. Lopes

Dynamic models

Contents

1

Dynamic models

Alexandra M. Schmidt

Universidade Federal do Rio de Janeiro, Brazil

Hedibert F. Lopes

Insper, Brazil

CONTENTS

1.1 Introduction

Dynamic models are, in a broad sense, probabilistic models that describe a set of observable measurements conditionally on a set of (latent or hidden) state-space variables whose time (and/or space) dynamics are driven by a set of time-invariant quantities (or parameters). This inherently hierarchical de-

scription renders dynamic models to the status of one of the most popular statistical structures in many areas of science, ranging from neuroscience to marketing to oceanography, from financial markets to target-tracking to signal process, and from climatology to text and sentiment analysis. Borrowing the notation from [3], also used by [11], the general structure of a dynamic model can be written as

Measurements model:	`[data	state-space, parameters]`
State-space dynamics:	`[state-space	parameters]`
Parameters prior:	`[parameters]`

In time series analysis, dynamic models are simultaneously an alternative to and an extension of Box&Jenkins' ARIMA models (See, for instance, [44] and Craigmile's chapter). The observed process, in general, is decomposed as the sum of different time series components such as deterministic or stochastic trends, seasonality and covariate effects. Moreover, different from ARIMA models, in dynamic models the parameters might change with time in order to accommodate *local* structures of the temporal process under study. More importantly, dynamic models deal rather naturally with non-stationary time series and structural changes.

In this chapter we review various aspects of dynamic models. We start by describing univariate normal dynamic linear models and their main properties. Dynamic models for non-normal processes are also discussed. Section 1.3 extends the univariate DLM to the multivariate and matrix-variate normal cases. Therein, we note that multivariate DLM naturally accommodate time series of observations made at different spatial locations. Dimension reduction through dynamic factor models is also discussed. Finally, Section 1.4 discusses use of the state-space framework to approximate convolution models, stochastic partial differential and integro-difference equations. For further reading on DLMs the readers are referred to [58], [44], [14], and [45].

1.2 Univariate Normal Dynamic Linear Models (NDLM)

Assume that $\{y_t, t \in T\}$ is a stochastic process observed at discrete time, such that $T = \{0, \pm 1, \pm 2, \pm 3, \cdots\}$. Let D_t be the set of information available at time t. Thus D_t includes the observation y_t, covariates observed at time t (if any), and all previous information D_{t-1}. A normal dynamic linear model (NDLM) decomposes an univariate time series y_t as the sum of two components, an overall time-varying trend, $\boldsymbol{F}_t'\boldsymbol{\theta}_t$, and an error component, ϵ_t, that follows a zero mean normal distribution with variance V_t. The (latent) p-dimensional vector $\boldsymbol{\theta}_t$, is known as the *state vector* at time t, and it evolves smoothly with time. More specifically, a NDLM is specified by the following

equations:

$$\begin{aligned} y_t &= \boldsymbol{F}_t'\boldsymbol{\theta}_t + \epsilon_t \qquad \epsilon_t \sim N(0, V_t) && (1.1)\\ \boldsymbol{\theta}_t &= \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \qquad \boldsymbol{\omega}_t \sim N(0, \boldsymbol{W}_t), && (1.2)\\ \boldsymbol{\theta}_0 &\sim N(\boldsymbol{m}_0, \boldsymbol{C}_0) && (1.3) \end{aligned}$$

where $\boldsymbol{\theta}_0$ is the initial information, $\boldsymbol{m}_0$ and $\boldsymbol{C}_0$ are known p-dimensional mean vector and $p \times p$ covariance matrix, respectively. $\boldsymbol{F}_t$ is a p-dimensional column vector of covariates, $\boldsymbol{G}_t$ is a $p \times p$ matrix, known as the *evolution* matrix, and $\boldsymbol{W}_t$ is a $p \times p$ covariance matrix describing the covariance structure among the components of $\boldsymbol{\theta}_t$. Equations (1.1) and (1.2) are known as the *observation* and *system*, or *state*, equations, respectively.

It is further assumed that ϵ_t and $\boldsymbol{\omega}_t$ are independent and mutually independent, and independent of $\boldsymbol{\theta}_0$. Note that, given $\boldsymbol{\theta}_t$, y_t is conditionally independent from past observations.

A NDLM is completely specified through the quadruple $\{\boldsymbol{F}_t, \boldsymbol{G}_t, V_t, \boldsymbol{W}_t\}$. If it is further assumed that $V_t = V$ and $\mathbf{W}_t = \mathbf{W}$, $\forall t \in T$ this is known as the constant model. Classical examples are the local level and local trend models, the trigonometric seasonal model and the time-varying parameter AR model, all illustrated below.

Example: *(First order polynomial model or time-varying level model)* The simplest NDLM is defined by the quadruple $\{1, 1, V, W\}$, such that $y_t = \mu_t + \epsilon_t$, and $\mu_t = \mu_{t-1} + \omega_t$, with $\epsilon_t \sim N(0, V)$, and $\omega_t \sim N(0, W)$. The ratio W/V, known as the signal-to-noise ratio, plays an important role in determining the behavior of the time varying level μ_t. As $W/V \to 0$, the first order model tends to a constant mean model, and as $W/V \to \infty$ the model reduces to a pure random walk. This model is suitable for short term forecast as it can be shown that the forecast function is constant.

Example: *(Linear trend)* A linear trend model can be defined by assuming that $y_t = \mu_t + \epsilon_t$, $\mu_t = \mu_{t-1} + \beta_{t-1} + \omega_{1t}$ and $\beta_t = \beta_{t-1} + \omega_{2t}$, where β_t is the slope of the local level μ_t. In this case, $\boldsymbol{\theta}_t = (\mu_t, \beta_t)'$ and the NDLM is defined by the quadruple $\{\boldsymbol{F}, \boldsymbol{G}, V, \boldsymbol{W}\}$, where $\boldsymbol{F}' = (1\ 0)$,

$$\boldsymbol{G} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \text{ and } \boldsymbol{W} = \begin{pmatrix} W_{11} & W_{12} \\ W_{12} & W_{22} \end{pmatrix},$$

for some known values of W_{11}, W_{12}, W_{22} and V. If $W_{22} = 0$ the model results in a constant trend model.

Example: *(Fourier representation of seasonality)* A seasonal structure can be accommodated by a Fourier representation of seasonality. Assuming the period of seasonality is p, then $\boldsymbol{F} = (1\ 0)$,

$$\boldsymbol{G} = \begin{pmatrix} \cos 2\pi/p & \sin 2\pi/p \\ -\sin 2\pi/p & \cos 2\pi/p \end{pmatrix},$$

and $\boldsymbol{W} = \text{diag}(W_1, W_2)$.

Example: *(Time-varying auto-regressive process of order p, TV-AR(p))* Assume that $y_t = x_t + \epsilon_t$, where $x_t = \sum_{j=1}^{p} \phi_{tj} x_{t-j} + \omega_t$. The model is a NDLM where $\boldsymbol{\theta}_t = (\phi_{t1}, \cdots, \phi_{tp})$, $\boldsymbol{F} = (1, 0, \ldots, 0)'$,

$$\boldsymbol{G}_t = \begin{pmatrix} \phi_{t1} & \phi_{t2} & \phi_{t3} & \cdots & \phi_{t,p-1} & \phi_{tp} \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

and $\boldsymbol{\omega}_t = (\omega_t, 0, \cdots, 0)'$. If the components of $\boldsymbol{\theta}_t$ are fixed across time and $V = 0$, the time-series x_t follows a standard AR(p) process. Loosely speaking, the whole class of well-known autoregressive integrated moving average (ARIMA) models can be thought of as special cases of DLMs. See Prado and West (2010) [45] for further details.

1.2.1 Forward learning: the Kalman filter

In a Bayesian NDLM, from Equations (1.1)–(1.3), the posterior distribution of any parameters at time t is based on all available information up to time t, D_t. One aspect of the sequential Bayesian learning dynamic models is that the posterior distribution for the parameters at time $t-1$, $p(\boldsymbol{\theta}_{t-1}|D_{t-1})$, must be propagated through the systems equation (1.2) in order to become the prior distribution of the parameters at time t, $p(\boldsymbol{\theta}_t|D_{t-1})$. This prior at time t is then combined with the likelihood $p(y_t|\boldsymbol{\theta}_t)$ from the observation equation (1.1) to produce both predictive $p(y_t|D_{t-1})$ and posterior $p(\boldsymbol{\theta}_t|D_t)$. Schematically, the sequential Bayesian inference in dynamic models cycles through three main steps: Evolution, Prediction and Updating, as follows:

$$p(\boldsymbol{\theta}_{t-1}|D_{t-1}) \overset{\text{Evolution}}{\Longrightarrow} p(\boldsymbol{\theta}_t|D_{t-1}) \overset{\text{Updating}}{\Longrightarrow} p(\boldsymbol{\theta}_t|D_t)$$

$$\Big\Downarrow \text{Prediction}$$

$$p(y_t|D_{t-1})$$

Below, following Theorem 4.3 in [58], we specify each of the distributions for each of the steps described above. The resultant distributions for each step depend on the distribution of the initial information $\boldsymbol{\theta}_0|D_0 \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0)$ and the knowledge or not of the observational variance V_t and the variance of the evolution noise, $\boldsymbol{W}_t$. We modify slightly the definition of the NDLM in equations (1.1) and (1.2) to accommodate the more general case when V is

unknown but follows an inverse gamma prior distribution, and the distribution of the noise component follows a normal distribution conditional on the value of V, that is, $\boldsymbol{\omega}_t \mid V \sim N(0, V\boldsymbol{W}_t^*)$.

Define the time $t = 0$ normal-gamma prior distribution for $p(\boldsymbol{\theta}_0, V|D_0)$, i.e. $\boldsymbol{\theta}_0|D_0, V \sim N(\boldsymbol{m}_0, V\boldsymbol{C}_0^*)$ and $V|D_0 \sim IG(n_0/2, n_0S_0/2)$. The derivation is done by induction, by assuming that $\boldsymbol{\theta}_{t-1}|V, D_{t-1} \sim N(\boldsymbol{m}_{t-1}, V\boldsymbol{C}_{t-1}^*)$ and $V|D_{t-1} \sim IG(n_{t-1}/2, n_{t-1}S_{t-1}/2)$. It can be shown that the steps above are given by [58, pages 109-110]:

- *Evolving the state:* The prior distribution of $\boldsymbol{\theta}_t$ is given by $\boldsymbol{\theta}_t|V, D_{t-1} \sim N(\boldsymbol{a}_t, V\boldsymbol{R}_t^*)$ where $\boldsymbol{a}_t = \boldsymbol{G}_t\boldsymbol{m}_{t-1}$ and $\boldsymbol{R}_t^* = \boldsymbol{G}_t\boldsymbol{C}_{t-1}^*\boldsymbol{G}_t' + \boldsymbol{W}_t^*$. Therefore, $\boldsymbol{\theta}_t|D_{t-1} \sim t_{n_{t-1}}(\boldsymbol{a}_t, S_{t-1}\boldsymbol{R}_t^*)$.

- *Prediction the observation:* The one-step forecast is given by $y_t|V, D_{t-1} \sim N(f_t, VQ_t^*)$ where $f_t = \boldsymbol{F}_t'\boldsymbol{a}_t$ and $Q_t^* = 1+\boldsymbol{F}_t'\boldsymbol{R}_t^*\boldsymbol{F}_t$. Therefore, $y_t|D_{t-1} \sim t_{n_{t-1}}(f_t, S_{t-1}Q_t^*)$.

- *Updating the variance:* $V|D_t \sim Ga(n_t/2, n_tS_t/2)$, where $n_t = n_{t-1} + 1$, and $n_tS_t = n_{t-1}S_{t-1} + (y_t - f_t)^2/Q_t^*$.

- *Updating the state:* The posterior distribution of $\boldsymbol{\theta}_t$ is given by $\boldsymbol{\theta}_t|V, D_t \sim N(\boldsymbol{m}_t, V\boldsymbol{C}_t^*)$, with $\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{A}_t(y_t - f_t)$ and $\boldsymbol{C}_t^* = \boldsymbol{R}_t^* - \boldsymbol{A}_t\boldsymbol{A}_t'Q_t^*$, where $\boldsymbol{A}_t = \boldsymbol{R}_t\boldsymbol{F}_tQ_t^{-1}$. Therefore, $\boldsymbol{\theta}_t|D_t \sim t_{n_{t-1}}(\boldsymbol{m}_t, S_t\boldsymbol{C}_t^*)$.

These recursions are commonly known as *Kalman recursions* or simply *the Kalman filter algorithm* (see, for instance, [58]). If interest lies on forecasting the process k steps ahead, based on the information up to time D_t, the forecast distribution, $y_{t+k}|D_t$ can be obtained through the Kalman filter. It can be shown that $y_{t+k}|D_t \sim t_{n_t}(f_t(k), Q_t(k))$, with $f_t(k) = \boldsymbol{F}_{t+k}'\boldsymbol{a}_t(k)$, $Q_t(k) = \boldsymbol{F}_{t+k}'\boldsymbol{R}_t(k)\boldsymbol{F}_{t+k} + S_{t+k}$, $\boldsymbol{a}_t(0) = \boldsymbol{m}_t$, and $\boldsymbol{R}_t(0) = \boldsymbol{C}_t$.

1.2.2 Backward learning: the Kalman smoother

The Kalman filter is one of the most popular algorithms for sequential update of hidden/latent states in dynamic systems. However, in the above form, it only provides posterior distribution of a given state $\boldsymbol{\theta}_t$ conditionally on the past observations, $y_1, \ldots, y_t$, $p(\boldsymbol{\theta}_t|D_t, V)$. For simplicity, we are omitting the dependence on the state variances $\boldsymbol{W}_1^*, \ldots, \boldsymbol{W}_n^*$. In many instances, however, one may want to obtain the posterior distribution of the states given the whole set of observations, $y_1, \ldots, y_n$, i.e. $p(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n|V, D_n)$; for instance, to understand the dynamics driving observations as opposed to simply forecasting its future values.

Full joint distribution of states

This joint posterior can be rewritten as

$$p(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n | V, D_n) = p(\boldsymbol{\theta}_n | V, D_n) \prod_{t=n-1}^{1} p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, V, D_t), \tag{1.4}$$

with the simplifications given by the Markov property of Equations (1.1)–(1.3). Bayes' theorem leads to

$$\begin{aligned} p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, V, D_t) &\propto f_N(\boldsymbol{\theta}_{t+1}; \boldsymbol{G}_t \boldsymbol{\theta}_t, V\boldsymbol{W}_t^*) f_N(\boldsymbol{\theta}_t; \boldsymbol{m}_t, V\boldsymbol{C}_t^*) \\ &\propto f_N(\boldsymbol{\theta}_{t+1}; \widetilde{\boldsymbol{m}}_t, V\widetilde{\boldsymbol{C}}_t^*), \end{aligned} \tag{1.5}$$

where $\widetilde{\boldsymbol{m}}_t = \widetilde{\boldsymbol{C}}_t^*(\boldsymbol{G}_t' \boldsymbol{W}_t^{-1} \boldsymbol{\theta}_{t+1} + \boldsymbol{C}_t^{*-1} \boldsymbol{m}_t)$ and $\widetilde{\boldsymbol{C}}_t^* = (\boldsymbol{G}_t' \boldsymbol{W}_t^{*-1} \boldsymbol{G}_t + \boldsymbol{C}_t^{*-1})^{-1}$, for $t = n-1, n-2, \ldots, 1$. From the Kalman filter, $V|D_n \sim IG(n_n/2, n_n S_n/2)$, so it follows that $\boldsymbol{\theta}_n | D_n \sim t_{n_n}(\widetilde{\boldsymbol{m}}_n, S_n \widetilde{\boldsymbol{C}}_n^*)$ and

$$\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t+1}, D_t \sim t_{n_n}(\widetilde{\boldsymbol{m}}_t, S_n \widetilde{\boldsymbol{C}}_t^*), \tag{1.6}$$

for $t = n-1, n-2, \ldots, 1$.

Marginal distributions of the states

Similarly, it can be shown that the marginal distribution of $\boldsymbol{\theta}_t$ is

$$\boldsymbol{\theta}_t | V, D_n \sim t_{n_n}(\overline{\boldsymbol{m}}_t, S_n \overline{\boldsymbol{C}}_t^*) \tag{1.7}$$

where $\overline{\boldsymbol{m}}_t = \boldsymbol{m}_t + \boldsymbol{C}_t^* \boldsymbol{G}_{t+1}' \boldsymbol{R}_{t+1}^{*-1} (\overline{\boldsymbol{m}}_{t+1} - \boldsymbol{a}_{t+1})$ and $\overline{\boldsymbol{C}}_t^* = \boldsymbol{C}_t^* - \boldsymbol{C}_t^* \boldsymbol{G}_{t+1}' \boldsymbol{R}_{t+1}^{*-1} (\boldsymbol{R}_{t+1}^* - \overline{\boldsymbol{C}}_{t+1}^*) \boldsymbol{R}_{t+1}^{*-1} \boldsymbol{G}_{t+1} \boldsymbol{C}_t^*$, for $t = n-1, n-2, \ldots, 1$.

Full conditional distributions of the states

Let $\boldsymbol{\theta}_{-t} = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_{t+1}, \ldots, \boldsymbol{\theta}_n\}$ and $t = 2, \ldots, n-1$, it follows that the full conditional distribution of $\boldsymbol{\theta}_t$ is

$$\begin{aligned} p(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{-t}, V, D_n) &\propto f_N(y_t; \boldsymbol{F}_t' \boldsymbol{\theta}_t, V)\, f_N(\boldsymbol{\theta}_{t+1}; \boldsymbol{G}_{t+1} \boldsymbol{\theta}_t, V\boldsymbol{W}_{t+1}^*) \\ &\times f_N(\boldsymbol{\theta}_t; \boldsymbol{G}_t \boldsymbol{\theta}_{t-1}, V\boldsymbol{W}_t^*) = f_N(\boldsymbol{\theta}_t; \boldsymbol{b}_t, V\boldsymbol{B}_t^*) \end{aligned} \tag{1.8}$$

where $\boldsymbol{b}_t = \boldsymbol{B}_t^*(\boldsymbol{F}_t y_t + \boldsymbol{G}_{t+1}' \boldsymbol{W}_{t+1}^{*-1} \boldsymbol{\theta}_{t+1} + \boldsymbol{W}_t^{*-1} \boldsymbol{G}_t \boldsymbol{\theta}_{t-1})$ and $\boldsymbol{B}_t^* = (\boldsymbol{F}_t \boldsymbol{F}_t' + \boldsymbol{G}_{t+1}' \boldsymbol{W}_{t+1}^{*-1} \boldsymbol{G}_{t+1} + \boldsymbol{W}_t^{*-1})^{-1}$. The endpoint parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_n$ also have full conditional distributions $N(\boldsymbol{b}_1, V\boldsymbol{B}_1^*)$ and $N(\boldsymbol{b}_n, V\boldsymbol{B}_n^*)$, respectively, where $\boldsymbol{b}_1 = \boldsymbol{B}_1^*(\boldsymbol{F}_1 y_1 + \boldsymbol{G}_2' \boldsymbol{W}_2^{*-1} \boldsymbol{\theta}_2 + \boldsymbol{R}^{*-1} \boldsymbol{a}_1)$, $\boldsymbol{B}_1^* = (\boldsymbol{F}_1 \boldsymbol{F}_1' + \boldsymbol{G}_2' \boldsymbol{W}_2^{*-1} \boldsymbol{G}_2 + \boldsymbol{R}^{-1})^{*-1}$, $\boldsymbol{b}_n = \boldsymbol{B}_n^*(\boldsymbol{F}_n y_n + \boldsymbol{W}_n^{*-1} \boldsymbol{G}_n \boldsymbol{\theta}_{n-1})$ and $\boldsymbol{B}_n^* = (\boldsymbol{F}_n \boldsymbol{F}_n' + \boldsymbol{W}_n^{*-1})^{-1}$. Again,

$$\boldsymbol{\theta}_t | \boldsymbol{\theta}_{-t}, D_t \sim t_{n_n}(\boldsymbol{b}_t, S_n \boldsymbol{B}_t^*) \qquad \text{for all } t. \tag{1.9}$$

Full posterior inference

When the evolution variances $\boldsymbol{W}_1^*, \ldots, \boldsymbol{W}_n^*$ are unknown, closed-form analytical full posterior inference is infeasible and numerical or Monte Carlo approximations are needed. Numerical integration, in fact, is only realistically feasible for very low dimensional settings. Markov Chain Monte Carlo methods have become the norm over the last quarter of century for state space modelers. In particular, the full joint of Equation (1.4) can be combined with full conditional distributions for V and $\boldsymbol{W}_1^*, \ldots, \boldsymbol{W}_n^*$. This is the well known *forward filtering, backward sampling (FFBS)* algorithm of [16] and [8]. The FFBS algorithm is commonly used for posterior inference in Gaussian and conditionally Gaussian DLMs. The main steps needed to use DLM's using the software `R` [48] are in the package `dlm`, detailed in [43].

1.2.3 Integrated likelihood

Another very important result of the sequential Bayesian depicted above is the derivation of the marginal likelihood of $\boldsymbol{W}_1^*, \ldots, \boldsymbol{W}_T^*$ given $y_1, \ldots, y_T$. Without loss of generality, assume that $\boldsymbol{W}_t = \boldsymbol{W}$ for all $t = 1, \ldots, n$, where n is the sample size, and that $p(\boldsymbol{W})$ denotes the prior distribution of $\boldsymbol{W}$. In this case,

$$p(y_1, \ldots, y_n|\boldsymbol{W}) = \prod_{t=1}^{n} p(y_t|D_{t-1}, \boldsymbol{W}) \tag{1.10}$$

where $y_t|D_{t-1} \sim t_{n_{t-1}}(f_t, S_{t-1}Q_t^*)$. Therefore, the posterior distribution of $\boldsymbol{W}$, $p(\boldsymbol{W}|D_n) \propto p(D_n|\boldsymbol{W})p(\boldsymbol{W})$, can be combined with $p(V|\boldsymbol{W}, D_n)$ to produce the joint posterior of $(V, \boldsymbol{W})$:

$$\begin{aligned} p(V, \boldsymbol{W}|D_n) &\propto p(D_n|\boldsymbol{W})p(V|\boldsymbol{W}, D_n)p(\boldsymbol{W}) \\ &= \left[\prod_{t=1}^{n} p(y_t; f_t, S_{t-1}Q_t^*, n_{t-1})\right] p(V; n_n/2, n_n S_n/2)p(\boldsymbol{W}). \end{aligned}$$

In words, Gaussianity and linearity, leads to straightforward posterior for V and $\boldsymbol{W}$ by integrating out all state-space vectors $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_n$. If M independent Monte Carlo draws $\{(V, \boldsymbol{W})^{(1)}, \ldots, (V, \boldsymbol{W})^{(M)}\}$ from $p(V, \boldsymbol{W}|D_n)$ are obtained, then M independent Monte Carlo draws from $p(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n|D)$ are easily obtained by repeating the FFBS of Equation (1.4) (or Equation (1.5)) M times. This leads to M independent Monte Carlo draws from $p(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n, V, \boldsymbol{W}|D)$, i.e. no MCMC is needed.

1.2.4 Some properties of NDLMs

Superposition of models

More generally, in the DLM framework, time series are modeled through a collection of components DLMs. This is possible because of a theorem proved

in [58]. Consider m time series y_{it} generated by NDLMs identified by the quadruples $M_i = \{\boldsymbol{F}_{it}, \boldsymbol{G}_{it}, V_{it}, \boldsymbol{W}_{it}\}$, $i = 1, 2, \ldots, m$. Under model M_i, the state vector $\boldsymbol{\theta}_{it}$ is of dimension p_i, and the observation and evolution error series are respectively ϵ_{it} and $\boldsymbol{\omega}_{it}$. The state vectors are distinct and, for all distinct $i \neq j$, the series ϵ_{it} and $\boldsymbol{\omega}_{it}$ are mutually independent of the series ϵ_{jt} and $\boldsymbol{\omega}_{jt}$. Then the series $y_t = \sum_{i=1}^{m} y_{it}$ follows a p-dimensional DLM $\{\boldsymbol{F}_t, \boldsymbol{G}_t, V_t, \boldsymbol{W}_t\}$, where $p = p_1 + \cdots + p_m$, the state vector $\boldsymbol{\theta}_t$ is given by $\boldsymbol{\theta}_t = (\boldsymbol{\theta}'_{1t}, \ldots, \boldsymbol{\theta}'_{mt})'$, $\boldsymbol{F}_t = \left(\boldsymbol{F}'_{1t}, \ldots, \boldsymbol{F}'_{mt}\right)'$, $\boldsymbol{G}_t = \text{block diag}(\boldsymbol{G}_{1t}, \ldots, \boldsymbol{G}_{mt})$, $\boldsymbol{W}_t = \text{block diag}(\boldsymbol{W}_{1t}, \ldots, \boldsymbol{W}_{mt})$, and $V_t = \sum_{i=1}^{m} V_{it}$.

Example: *(Dynamic regression)* If at each time t a pair of observations (y_t, x_t) is available, and there is a local linear relationship between y_t and x_t, a simple dynamic regression is defined by making $\boldsymbol{F}'_t = (1 \; x_t)$, $\boldsymbol{G} = \boldsymbol{I}_2$, the 2-dimensional identity matrix, and $\boldsymbol{W} = \text{diag}(W_1, W_2)$.

Example: *(Time varying level plus an annual seasonal component)* $\boldsymbol{F}'_t = (1\ 1\ 0)$, $\boldsymbol{W} = \text{diag}(W_1, W_2, W_3)$, and

$$G = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos 2\pi/12 & \sin 2\pi/12 \\ 0 & -\sin 2\pi/12 & \cos 2\pi/12 \end{pmatrix}.$$

Discounting factors

The sequential updating steps described in Section 1.2.1 follow normal or Student-t distributions because we assume that $\boldsymbol{W}_t$ are known for all time t. In practice, this is rarely true. One way to avoid estimating the elements of the covariance matrix $\boldsymbol{W}_t$ is through the use of discounting factors. As detailed in [58], let $\boldsymbol{P}_t = \boldsymbol{G}_t \boldsymbol{C}_{t-1} \boldsymbol{G}'_t = Var[\boldsymbol{G}_t \boldsymbol{\theta}_t | D_{t-1}]$, which can be viewed as the prior variance in a NDLM with $\boldsymbol{W}_t = \boldsymbol{0}$, that is, with no evolution error term. The usual NDLM leads to $Var[\boldsymbol{G}_t \boldsymbol{\theta}_t | D_{t-1}] = \boldsymbol{P}_t + \boldsymbol{W}_t$. The idea of discounting factor is introduced by making $\boldsymbol{W}_t = \boldsymbol{P}_t(1-\delta)/\delta$, for some $\delta \in (0, 1]$. Then, $\boldsymbol{R}_t = \boldsymbol{P}_t/\delta$, resulting in an increase in the prior variance of $\boldsymbol{\theta}_t$, at time $t-1$. Note that, for given values of δ and $\boldsymbol{C}_0$ the series $\boldsymbol{W}_1, \ldots, \boldsymbol{W}_n$ is identified. If a NDLM has different components following the superposition theorem, a possible strategy is to specify different values of the discount factor for the different components. Typically, the value of δ is used in the range $[0.9, 0.99]$ for polynomial, regression, and seasonal components [58].

Intervention and monitoring

Because of the sequential learning structure of NDLMs (Section 1.2.1), they naturally accommodate changes in the observed time series. For example, if an observation is missing at time t, then $D_t = D_{t-1}$, and $\boldsymbol{\theta}_t | D_t \sim N(\mathbf{m}_t, \boldsymbol{C}_t)$, with $\mathbf{m}_t = \mathbf{a}_t$ and $\boldsymbol{C}_t = \boldsymbol{R}_t$. On the other hand, if a change has occurred and it is difficult to attribute it to a particular component, the prior variance matrix

$\boldsymbol{R}_t$ can be changed to reflect the increased uncertainty about all parameters, without a change in the prior mean $\mathbf{a}_t$ that would anticipate the direction of the change.

The adequacy of NDLMs can be monitored online through the comparison between the observed value and the one-step ahead forecast distribution. In particular, Bayes' factors [13] can be used, and are usually based on the predictive densities of the forecast errors, $e_t = y_t - f_t$. See [58, Chapter 11] for more details both on intervention and monitoring in NDLMs.

1.2.5 Dynamic generalized linear models (DGLM)

The DLM was extended by [59] to the case wherein observations belong to the exponential family. Assume that observations y_t are generated from a dynamic exponential family (EF), defined as

$$p(y_t|\eta_t, \phi) = \exp\{\phi[y_t\eta_t - a(\eta_t)]\}b(y_t, \phi), \tag{1.11}$$

where η_t is the canonical parameter, ϕ is a scale parameter, usually time invariant. We denote the distribution of y_t as $EF(\eta_t, \phi)$. Let $\mu_t = E(y_t|\eta_t, \phi)$ and $g(.)$ be a known link function, at least twice differentiable, which relates the linear predictor with the canonical parameter, that is

$$g(\mu_t) = \boldsymbol{F}(\psi_1)'\boldsymbol{\theta}_t, \tag{1.12}$$

where $\boldsymbol{F}$, and $\boldsymbol{\theta}_t$ are vectors of dimension p, and ψ_1 denotes unknown quantities involved in the definition of $\boldsymbol{F}(\psi_1)$. Following the parameterization of the exponential family in equation (1.11) it follows that the mean and variance of y_t are, respectively, given by $E(y_t|\eta_t, \phi) = \mu_t = \dot{a}(\eta_t)$ and $V(y_t|\eta_t, \phi) = V_t = \phi^{-1}\ddot{a}(\eta_t)$.

Usually, in practice, assuming $g(.)$ as the identity function provides good results [59]. Some examples, where the link function is suggested by the definition of the canonical function, include the log-linear Poisson and logistic-linear Bernoulli models. The state parameters, $\boldsymbol{\theta}_t$, evolve through time via a Markovian structure, that is

$$\boldsymbol{\theta}_t \quad = \boldsymbol{G}(\boldsymbol{\psi}_2)\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \boldsymbol{W}). \tag{1.13}$$

The hyperparameter vector $\boldsymbol{\psi}_2$ represents possible unknown quantities in $\boldsymbol{G}(\cdot)$, the $p \times p$ evolution matrix. Lastly, $\boldsymbol{\omega}_t$ is the disturbance associated with the system evolution with covariance structure $\boldsymbol{W}$, commonly a diagonal matrix which can vary over time. The initial information of the model is denoted by $\boldsymbol{\theta}_0$, and its prior distribution is defined through a p-variate normal distribution, that is, $\boldsymbol{\theta}_0|D_0 \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0)$, where D_0 denotes the initial information set. Independence is assumed among the components of $\boldsymbol{\theta}_0$ and $\boldsymbol{\omega}_t$, $\forall\, t$.

Posterior inference

West, Harrison and Migon (1985) [59] specify only the first and second moments of $\boldsymbol{\omega}_t$. They perform inference procedure taking advantage of conjugate

prior and posterior distributions in the exponential family, and used linear Bayes' estimation to obtain estimates of the moments of $\boldsymbol{\theta}_t|D_t$.This approach is appealing as no assumption is made about the shape of the distribution of the disturbance component $\boldsymbol{\omega}_t$. However, we are limited to learn only about the first two moments of the posterior distribution of $\boldsymbol{\theta}_t|D_t$. The assumption of normality of $\boldsymbol{\omega}_t$ allows us to write down a likelihood function for $\boldsymbol{\theta}_t$ and all the other parameters in the model. Inference procedure can be performed using Markov chain Monte Carlo algorithms. However, care must be taken when proposing an algorithm to obtain samples from the posterior distribution of the state vectors $\boldsymbol{\theta}_t$. [59], [30] and [15], among others, proposed different methods to obtain approximations of the posterior distribution of parameters belonging to DGLMs. Estimation methods based on MCMC can be seen at [54], [17], [21], and [38].

The Poisson case

Time series of counts are often encountered in ecological and environmental problems. It is well known that the Poisson distribution assumes that the mean and the variance are equal, which is hardly true in practice. Usually, the variance is much greater than the mean. When this happens it is said that the data are overdispersed. To account for overdispersion it is quite common to use a mixture between the Poisson and gamma distributions, which marginally results on a negative binomial distribution. Another mixture which is quite used is between Poisson and log-normal distributions.

Assume that y_t represents the number of realizations of the process under study, for discrete time $t \in \{0, 1, 2, \ldots\}$. Usually, it is assumed that y_t follows a Poisson distribution with mean λ_t. More generally, one can assume that $y_t|\lambda_t, \delta_t \sim Poi(\lambda_t\delta_t)$ with $\log \lambda_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t$ and $\boldsymbol{\theta}_t$ following the evolution in time as in Equation (1.13). Note that, when $\delta_t = 1\ \forall t$, a dynamic Poisson model arises. From the evolution equation for $\boldsymbol{\theta}_t$ we have $\log \lambda_t = \boldsymbol{F}_t'\boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{F}_t'\boldsymbol{\omega}_t$, which is equivalent to write $\lambda_t = \exp\{\boldsymbol{F}_t'\boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{F}_t'\boldsymbol{\omega}_t\}$, and as $\boldsymbol{\omega}_t$ is normally distributed, and is assumed to be independent of $\boldsymbol{\theta}_{t-1}$, we have that $\delta_t^* = \exp\{\boldsymbol{F}_t'\boldsymbol{\omega}_t\}$ follows a log-normal distribution, whose associated normal has zero mean and variance $\boldsymbol{F}_t'\boldsymbol{W}\boldsymbol{F}_t$. Therefore, this generalized dynamic Poisson model not only accounts for the temporal structure in the time series but is also able to capture overdispersion, as for every time t the marginal (with respect to $\boldsymbol{\omega}_t$) variance of y_t is greater than its mean.

If it is believed that there is some extra variability in the data, one alternative is to assume a continuous mixture such that δ_t follows a continuous distribution assuming positive values, and δ_t is independent of λ_t, $\forall t$. If δ_t follows a Gamma distribution with mean α/β and variance α/β^2, denoted by $\delta_t|\alpha, \beta \sim Ga(\alpha, \beta)$, then the distribution of $y_t|\lambda_t, \alpha, \beta$, obtained through,

$$\int_{\delta_t} p(y_t|\lambda_t, \delta_t)p(\delta_t|\alpha, \beta)\, d\delta_t = p(y_t|\lambda_t, \alpha, \beta), \tag{1.14}$$

follows a negative binomial distribution with mean and variance given, respectively, by $E(y_t|\lambda_t, \alpha, \beta) = \lambda_t\alpha/\beta = \varphi_t$ and $V(y_t|\lambda_t, \alpha, \beta) = \varphi_t + \lambda_t^2\alpha/\beta^2$.

Another possibility is to consider a log-normal mixture, that is, to assume $\delta_t|a_\delta, V \sim LN(a_\delta, V)$, where $LN(a_\delta, V)$ stands for the log-normal distribution whose associated normal has mean a_δ and variance V. By using the properties of conditional expectation, it can be shown that $E(y_t|\lambda_t, a_\delta, V) = \lambda_t \exp\{a_\delta + V/2\} = \phi_t$ and $V(y_t|\lambda_t, a_\delta, V) = \phi_t + \lambda_t^2 \exp\{2a_\delta + V\}(e^V - 1)$.

Notice that both mixtures capture overdispersion as the variances in both cases are the means plus positive quantities. It is worth mentioning that writing y_t as following a Poisson distribution with mean $\lambda_t^* = \lambda_t\delta_t$ is equivalent to assume $\log\lambda_t^* = \log\lambda_t + \log\delta_t$ which is equal to $\boldsymbol{F}_t'\boldsymbol{\mu}_t + \log\delta_t$. Therefore, $\log\delta_t$ can be viewed as a random effect present to capture any extra variation. [58], page 556, discuss this model as an alternative to capture possible extra variations.

Zero-inflated dynamic Poisson model

Another common issue with the modeling of time-varying counts might be the presence of a great amount of zeros. In this case, one might use a zero inflated version of the Poisson dynamic model. Following [1],let x_t be a random variable representing the presence ($x_t = 1$) or absence ($x_t = 0$) of the process being observed. Assume that $x_t|\theta$ follows a Bernoulli distribution with probability of success θ. Let $\pi(y_t|\lambda_t, \delta_t)$ be a model for the process being observed *given* it is *present* ($x_t = 1$). In particular, here $\pi(\cdot|\lambda_t, \delta_t)$ is the probability distribution function of the Poisson distribution with mean $\lambda_t\delta_t$. By definition, $P(y_t|\lambda_t, \delta_t, x_t = 1) = \pi(y_t|\lambda_t, \delta_t)$ and $P(x_t = 0|\lambda_t, \delta_t, x_t = 0) = 1$. The joint density function of x_t and y_t is given by

$$p(y_t, x_t|\lambda_t, \delta_t, \theta) = \{\theta\,\pi(y_t|\lambda_t, \delta_t)\}^{x_t}\,(1-\theta)^{1-x_t}\,. \tag{1.15}$$

Therefore, the marginal distribution of y_t, with respect to x_t, is given by $p(y_t|\lambda_t, \delta_t, \theta) = \theta\,\pi(y_t|\lambda_t, \delta_t) + (1-\theta)\delta_0(y_t)$, where $\delta_0(y_t)$ is an indicator function, such that $\delta_0(y_t) = 1$ if $y_t = 0$, and zero otherwise. Clearly, $p(y_t = 0|\lambda_t, \delta_t, \theta) > \pi(y_t = 0|\lambda_t, \delta_t)$.

Notice that x_t is a latent variable, its inclusion in the model results that $P(y_t = 0|\lambda_t, \delta_t, X_t = 1) = \pi(y_t = 0|\lambda_t, \delta_t)$, $P(y_t > 0|\lambda_t, \delta_t, x_t = 0) = 0$, $P(x_t = 1|y_t > 0, \theta, \lambda_t, \delta_t) = 1$ and,

$$p_t = P(x_t = 1|y_t = 0, \theta, \lambda_t, \delta_t) = \frac{\theta\,\pi(y_t = 0|\lambda_t, \delta_t)}{\theta\,\pi(y_t = 0|\lambda_t, \delta_t) + (1-\theta)}, \tag{1.16}$$

and marginally, as previously defined, x_t follows a Bernoulli distribution with probability of success θ, denoted by $x_t|\theta \sim Ber(\theta)$. Notice that p_t provides an estimate of the probability of presence of the process, at time t, given that no cases were observed. [53] makes comparison of these class of models for artificial datasets and weekly number of cases of dengue fever between 2001 and 2002, in the city of Rio de Janeiro, Brazil.

1.3 Multivariate Dynamic Linear Models

The multivariate extension of the dynamic linear models introduced in Section 1.2 leads to the broad classes of multivariate and matrix variate Gaussian DLMs, both of which we briefly revisit here. We then focus on two quite popular and general classes of multi-dimensional DLMs: spatio-temporal models and dynamic factor models. Here, we closely follow the notation of [58, Chapter 16] and [45, Chapter 10].

1.3.1 Multivariate NDLMs

For a p-dimensional vector $\boldsymbol{y}_t = (y_{t1}, \ldots, y_{tp})'$, a multivariate normal DLM can be written as

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{F}_t'\boldsymbol{\theta}_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim N(0, \boldsymbol{V}_t), & (1.17)\\ \boldsymbol{\theta}_t &= \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(0, \boldsymbol{W}_t), & (1.18) \end{aligned}$$

where $\boldsymbol{\theta}_0 \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0)$. In what follows, for simplicity, it is assumed that both sequences of observation and evolution covariances matrices $\boldsymbol{V}_t$ and $\boldsymbol{W}_t$ are known and of dimensions $(p \times p)$ and $(k \times k)$, respectively. This assumption is relaxed later on when dealing with the special cases. Therefore, the model is completely specified by the quadruple $\{\boldsymbol{F}_t, \boldsymbol{G}_t, \boldsymbol{V}_t, \boldsymbol{W}_t\}$, where $\boldsymbol{F}_t$s are $(k \times p)$ matrices and $\boldsymbol{G}_t$s are $(k \times k)$ matrices. It follows directly from Section 1.2.1 that, for $t = 1, 2, \ldots$,

$$\begin{aligned} (\boldsymbol{\theta}_t | D_{t-1}) &\sim N(\boldsymbol{a}_t, \boldsymbol{R}_t),\\ (\boldsymbol{y}_t | D_{t-1}) &\sim N(\boldsymbol{f}_t, \boldsymbol{Q}_t),\\ (\boldsymbol{\theta}_t | D_t) &\sim N(\boldsymbol{m}_t, \boldsymbol{C}_t), \end{aligned}$$

where $\boldsymbol{a}_t = \boldsymbol{G}_t\boldsymbol{m}_{t-1}$, $\boldsymbol{R}_t = \boldsymbol{G}_t\boldsymbol{C}_{t-1}\boldsymbol{G}_t' + \boldsymbol{W}_t$, $\boldsymbol{f}_t = \boldsymbol{F}_t'\boldsymbol{a}_t$, $\boldsymbol{Q}_t = \boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t + \boldsymbol{V}_t$, with $\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{A}_t(\boldsymbol{y}_t - \boldsymbol{f}_t)$ and $\boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{A}_t\boldsymbol{Q}_t\boldsymbol{A}_t'$, with $\boldsymbol{A}_t = \boldsymbol{R}_t\boldsymbol{F}_t\boldsymbol{Q}_t^{-1}$.

On the one hand, in this parametrization the i-th time series $\{y_{ti}\}$ has its own set of regressors, represented the i-th column of the matrices $\boldsymbol{F}_t$, but shares the dynamic regression coefficients, $\boldsymbol{\theta}_t$, with all other $p-1$ time series. On the other hand, [47] proposed the *common components* multivariate DLM, where each time series $\{y_{ti}\}$ has its own dynamic regression coefficients, $\boldsymbol{\theta}_{ti}$, but shares the set of regressors in $\boldsymbol{F}_t$.

1.3.2 Multivariate common-component NDLMs

In this case, the model is written as p component univariate DLMs,

$$\begin{aligned} y_{ti} &= \boldsymbol{F}_t\boldsymbol{\theta}_{ti} + \epsilon_{ti}, \qquad \epsilon_{ti} \sim N(0, v_{ii}v_t), & (1.19)\\ \boldsymbol{\theta}_{ti} &= \boldsymbol{G}_t\boldsymbol{\theta}_{t-1,i} + \boldsymbol{\omega}_{ti}, \quad \boldsymbol{\omega}_{ti} \sim N(0, v_{ii}\boldsymbol{W}_t), & (1.20) \end{aligned}$$

so y_{ti} follows a DLM characterized by the quadruple $\{\boldsymbol{F}_t, \boldsymbol{G}_t, v_{ii}v_t, v_{ii}\boldsymbol{W}_t\}$. Besides $\boldsymbol{F}_t$, the univariate component models are also linked though (potentially) nonzero covariances on $\boldsymbol{V}$, i.e. $\text{Cov}(\epsilon_{ti}, \epsilon_{tj}) = v_t v_{ij}$ and $\text{Cov}(\boldsymbol{\omega}_{ti}, \boldsymbol{\omega}_{tj}) = v_{ij}\boldsymbol{W}_t$, for $i \neq j$.

Time-varying parameter vector autoregressions (TVP-VAR)

A quite popular special case are time-varying parameter vector autoregressive models where $\boldsymbol{F}_t$ contains several lags of all time time series in $\boldsymbol{y}_t$ (See, for instance, [46], [39] and [31]). More precisely, a standard time-varying parameter VAR(k) can be written as

$$\boldsymbol{y}_t = \sum_{l=1}^{k} \boldsymbol{B}_{lt}\boldsymbol{y}_{t-l} + \boldsymbol{\epsilon}_t, \tag{1.21}$$

where $\boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{V}_t)$, $\boldsymbol{y}_t$ is a p-dimensional vector of ecological or epidemiological measurements, for instance, and $\boldsymbol{B}_{lt}$ is the time t and lag l autoregressive coefficient matrix. Let $\boldsymbol{B}_{lt} = (\boldsymbol{b}_{lt,1}, \ldots, \boldsymbol{b}_{lt,p})'$, so $\boldsymbol{b}'_{lt,i}$ is the p-dimensional vector corresponding to the i^{th} row of $\boldsymbol{B}_{lt}$. Now let $\boldsymbol{\beta}_{it} = (\boldsymbol{b}'_{1t,i}, \boldsymbol{b}'_{2t,i}, \ldots, \boldsymbol{b}'_{kt,i})'$ be the pk-dimensional vector of coefficients corresponding to the i^{th} equation. Therefore, the above TVP-VAR(k) model can be rewritten as

$$\boldsymbol{y}_t = \boldsymbol{F}'_t\boldsymbol{\theta}_t + \boldsymbol{\epsilon}_t, \tag{1.22}$$

where $\boldsymbol{F}'_t = (\boldsymbol{I}_p \otimes \boldsymbol{x}'_t)$, $\boldsymbol{x}_t = (\boldsymbol{y}'_{t-1}, \boldsymbol{y}'_{t-2}, \ldots, \boldsymbol{y}'_{t-k})'$ is a pk-dimensional vector of lagged observations and $\boldsymbol{\theta}_t = (\boldsymbol{\beta}'_{1t}, \ldots, \boldsymbol{\beta}'_{pt})'$ is a p^2k-dimensional vector of autoregressive coefficients. In other words, Equation (1.21) has been rewritten as an observation equation for a multivariate NDLM as in Equation (1.17).

1.3.3 Matrix-variate NDLMs

The model is an extension of the previous common-component model to allow for $q \times p$ matrices of observations $\boldsymbol{y}_t$:

$$\begin{aligned} \boldsymbol{y}_t &= (\boldsymbol{I}_q \otimes \boldsymbol{F}'_t)\boldsymbol{\theta}_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim N(0, \boldsymbol{U}_t, \boldsymbol{V}_t), & (1.23)\\ \boldsymbol{\theta}_t &= (\boldsymbol{I}_q \otimes \boldsymbol{G}_t)\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(0, \boldsymbol{U}_t \otimes \boldsymbol{W}_t, \boldsymbol{V}_t), & (1.24) \end{aligned}$$

where the matrix-variate normal $\boldsymbol{\epsilon} \sim N(0, \boldsymbol{U}, \boldsymbol{V})$ basically means that the i^{th} row of $\boldsymbol{\epsilon}$ is $N(0, u_{ii}\boldsymbol{V})$, while its j^{th} column is $N(0, v_{jj}\boldsymbol{U})$. For matrix-variate normal distributions see [12]. See [32] for the matrix-variate extension of the dynamic hierarchical models of [18]. Dynamic matrix-variate graphical models are introduced by [9]. Sequential updating, forecasting and retrospective smoothing follow directly from the previous derivations. Similarly, when the prior for $\boldsymbol{V}$ is inverse Wishart, sequential updating follows directly from Section 1.2.1. See [45, Chapter 10] for additional details.

1.3.4 Hierarchical dynamic linear models (HDLM)

In certain applications, a natural hierarchical order arises in the dynamics of time series and their interdependences. For example, monthly levels of a set of pollutants might be contemporaneous related and their conditional expectations functions of regional characteristics, such as climate, industrial developments and other demographic aspects. A hierarchy, in this context, might arise to connect regressions across regions by pooling their coefficients, either via a common mean vector or through yet another set of explanatory variables.

Here we briefly review Gamerman and Migon's (1993) [18] seminal paper on dynamic hierarchical linear models (DHLM) for time series of cross-sectional data, which generalizes the hierarchical model introduced by Lindley and Smith (1972)[33]. See, for instance, [2] and [11] for additional examples of hierarchical models in spatial and spatio-temporal models.

Their basic DHM is composed of three equations: observational, structural and system, with the structural equations accounting for the levels of hierarchy. For example, a normal DHLM (NDHLM) with three hierarchical levels can be written as

$$\begin{aligned}
\boldsymbol{y}_t &= \boldsymbol{F}'_{1t}\boldsymbol{\theta}_{1t} + \boldsymbol{\epsilon}_{1t} \qquad \boldsymbol{\epsilon}_{1t} \sim N(\boldsymbol{0}, \boldsymbol{V}_{1t}) && (1.25)\\
\boldsymbol{\theta}_{1t} &= \boldsymbol{F}'_{2t}\boldsymbol{\theta}_{2t} + \boldsymbol{\epsilon}_{2t} \qquad \boldsymbol{\epsilon}_{2t} \sim N(\boldsymbol{0}, \boldsymbol{V}_{2t}) && (1.26)\\
\boldsymbol{\theta}_{2t} &= \boldsymbol{F}'_{3t}\boldsymbol{\theta}_{3t} + \boldsymbol{\epsilon}_{3t} \qquad \boldsymbol{\epsilon}_{3t} \sim N(\boldsymbol{0}, \boldsymbol{V}_{3t}) && (1.27)\\
\boldsymbol{\theta}_{3t} &= \boldsymbol{G}_t\boldsymbol{\theta}_{3,t-1} + \boldsymbol{\omega}_t \qquad \boldsymbol{\omega}_t \sim N(\boldsymbol{0}, \boldsymbol{W}_t). && (1.28)
\end{aligned}$$

The model can be easily reduced to a two-level NDHLM by making $\boldsymbol{F}_{3t} = \boldsymbol{I}$ and $\boldsymbol{V}_{3t} = \boldsymbol{0}$. If, in addition, $\boldsymbol{F}_{2t} = \boldsymbol{I}$ and $\boldsymbol{V}_{2t} = \boldsymbol{0}$, the above NDHLM reduces to a standard NDLM. A standard normal hierarchical linear model (NHLM) is obtained when $\boldsymbol{G}_t = \boldsymbol{I}$ and $\boldsymbol{W}_t = \boldsymbol{0}$. As pointed out by Gamerman and Migon (1993), a distinct feature of NDHLM is that $\dim(\boldsymbol{\theta}_{1t}) > \dim(\boldsymbol{\theta}_{2t}) > \dim(\boldsymbol{\theta}_{3t})$.

Example: *(Cross-section of random samples of linear growing exchangeable means)* For $\boldsymbol{y}_t = (y_{t1}, \ldots, y_{tn})'$, $\boldsymbol{\theta}_{1t} = (\theta_{1t,1}, \ldots, \theta_{1t,n})'$, $\boldsymbol{\theta}_{2t} = (\mu_t, \beta_t)'$ and $\boldsymbol{F}' = (\boldsymbol{1}_n \ \boldsymbol{0}_n)$, the model is written as a two-stage NDHLM:

$$\begin{aligned}
\boldsymbol{y}_t|\boldsymbol{\theta}_{1t} &\sim N(\boldsymbol{\theta}_t, \sigma^2\boldsymbol{I}_n)\\
\boldsymbol{\theta}_{1t}|\boldsymbol{\theta}_{2t} &\sim N(\boldsymbol{F}'\boldsymbol{\theta}_{2t}, \tau^2\boldsymbol{I}_n)\\
\boldsymbol{\theta}_{2t}|\boldsymbol{\theta}_{2,t-1} &\sim N(\boldsymbol{G}\boldsymbol{\theta}_{2,t-1}, \boldsymbol{W}_t),
\end{aligned}$$

where $\boldsymbol{G}$ follows the same structure of the linear growth introduced at the beginning of the chapter.

1.3.5 Spatio-temporal models

Environmental and ecological processes usually vary across space and time simultaneously. Spatio-temporal observations are naturally accommodated in

the DLM framework. We start the discussion by considering that locations vary continuously across the region of interest. Next we discuss the problem where the spatial region under study is divided into a finite number of regular or irregular subregions (areal level), resulting in observations for each subregion. Spatial interpolation is the main interest in the former, while in the latter the main interest is on spatial smoothing.

Geostatistical setup: observed locations varying continously in space

Let $\{y_t(\boldsymbol{s}), \boldsymbol{s} \in B \subset \mathbb{R}^d, t \in T\}$ be a spatial random field at discrete time t, and usually, $d = 1, 2$ or 3. If a partial realization of the spatial random field is available at each time t, the model specification in equation (1.17) is useful to describe the behaviour of the spatial process for each time t.

Following [11], let the observed data $y_t(\boldsymbol{s})$, be a noisy version of the process of interest, $Z_t(\boldsymbol{s})$, that is, assume

$$\begin{aligned} y_t(\boldsymbol{s}) &= z_t(\boldsymbol{s}) + v_t(\boldsymbol{s}) && (1.29)\\ z_t(\boldsymbol{s}) &= \boldsymbol{F}_t'\boldsymbol{\theta}_t + \epsilon_t^*(\boldsymbol{s}) && (1.30)\\ \boldsymbol{\theta}_t &= \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, && (1.31) \end{aligned}$$

where $\epsilon_t^*(\cdot)$ is assumed to follow a zero mean Gaussian process with some covariance matrix $\boldsymbol{\Sigma}_t$, while $v_t(\boldsymbol{s})$ represents measurement error assumed to follow an independent zero mean normal distribution with variance τ^2. It is assumed further that $\epsilon_t^*(\boldsymbol{s})$ are independent across time, and independent of $v_t(\boldsymbol{s})$ and $\boldsymbol{\omega}_t$, for all t and $\boldsymbol{s} \in D$.

If we substitute equation (1.30) into (1.29), we obtain $y_t(\boldsymbol{s}) = \boldsymbol{F}_t'\boldsymbol{\theta}_t + \epsilon_t^*(\boldsymbol{s}) + v_t(\boldsymbol{s})$. If we stack the observations of the process at time t onto a vector $\boldsymbol{y}_t = (y_t(\boldsymbol{s}_1), y_t(\mathbf{s_2}), \ldots, y_t(\boldsymbol{s}_n))'$, and assume a similar structure for $\boldsymbol{\epsilon}_t^*$ and $\boldsymbol{v}_t$, respectively, we can write $\boldsymbol{y}_t|\boldsymbol{F}_t, \boldsymbol{\theta}_t, \boldsymbol{v}_t, \tau^2 \sim N(\boldsymbol{F}_t'\boldsymbol{\theta}_t + \boldsymbol{v}_t, \tau^2\boldsymbol{I}_n)$. In other words, *conditional on* $\boldsymbol{v}_t$, the elements of $\boldsymbol{y}_t$ are independent across space, for each time t. We can marginalize the distribution of $\boldsymbol{y}_t$ with respect to $\boldsymbol{v}_t$. As $\boldsymbol{v}_t$ follows a zero mean multivariate normal distribution, then the distribution of $p(\boldsymbol{y}_t|\boldsymbol{\theta}_t, \boldsymbol{F}_t, \tau^2, \boldsymbol{\Sigma}_t) = \int_{\boldsymbol{v}_t} p(\boldsymbol{y}_t|\boldsymbol{F}_t, \boldsymbol{\theta}_t, \boldsymbol{v}_t, \tau^2)\ p(\boldsymbol{v}_t|\boldsymbol{\Sigma}_t)d\boldsymbol{v}_t$ follows a multivariate normal distribution with mean $\boldsymbol{F}_t'\boldsymbol{\theta}_t$ and covariance matrix $\boldsymbol{V}_t = \boldsymbol{\Sigma}_t + \tau^2\boldsymbol{I}_n$.

Note that, if spatio-temporal covariates are present in the columns of $\boldsymbol{F}_t$ the mean of the process is also varying across space for each time t. On the other hand, if spatio-temporal covariates are not available, then the temporal trend is fixed across space, and $\boldsymbol{\epsilon}_t$ captures, at each location, deviations from this overall temporal structure. The main issue is the specification of the covariance matrix $\boldsymbol{\Sigma}_t$. The simplest alternative to account for spatial correlation at each time t is to model each element of the covariance matrix $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}\ \forall t$, as a function of a common variance times a correlation function that depends on the Euclidean distance among the monitoring locations, that is, $\boldsymbol{\Sigma}_{ij} = \sigma^2\rho(d, \boldsymbol{\phi})$, where $d = ||\mathbf{s}_i - \mathbf{s}_j||$ is the Euclidean distance. The parameters in $\boldsymbol{\phi}$ are typically scalars or low dimensional vectors. For example, ϕ is

univariate when the correlation function is exponential $\rho(d,\phi) = \exp\{-d/\phi\}$ or spherical $\rho(d,\phi) = (1 - 1.5(d/\phi) + 0.5(d/\phi)^3)1_{\{d\leq\phi\}}$, and bivariate when the correlation function it is power exponential $\rho(d,\boldsymbol{\phi}) = \exp\{-(d/\phi_1)^{\phi_2}\}$ or Matérn $\rho(d,\boldsymbol{\phi}) = 2^{1-\phi_2}\Gamma(\phi_2)^{-1}(d/\phi_1)^{\phi_2}\mathcal{K}_{\phi_2}(d/\phi_1)$, where $\mathcal{K}_{\phi_2}(\cdot)$ is the modified Bessel function of the second kind and order ϕ_2. In both bivariate cases, $\phi_1 > 0$ is a range parameter controlling the speed of correlation decay between locations, while ϕ_2 is a smoothness parameter that controls the differentiability of the underlying process (see [2] for details). If it is believed that the variance of $\boldsymbol{v}_t$ changes with time, one alternative is to allow σ^2 to vary smoothly with time, for example, by making $\log\sigma_t^2 = \log\sigma_{t-1}^2 + \epsilon_{1t}$, with ϵ_{1t} following a zero mean normal distribution, with possibly unknown but constant variance W_σ, and initial information $\log\sigma_0^2 \sim N(0,b)$, for some known constant b. This will lead to a covariance matrix $\boldsymbol{V}_t$ that changes with time. When inference procedure is performed using MCMC methods, the algorithm must be adapted to include steps to sample from the posterior full conditional distributions of the elements of $\boldsymbol{\phi}$, and the variances $\sigma_0^2, \sigma_1^2, \ldots, \sigma_T^2$. This can be achieved, for example, through individual Metropolis-Hastings steps for each of these parameters.

An alternative in the modelling of the covariance matrix is to assign an inverse Wishart prior distribution for $\boldsymbol{\Sigma}_t$, which results on a nonstationary (free form) covariance matrix. The discount ideas discussed in Section 1.2.4 were extended by [34] for variance matrices discounting. [51] on the other hand, propose to define $\boldsymbol{\Sigma}_t$ as a function of covariates. Care must be taken when using covariate information in the covariance structure as it is necessary to guarantee that the resultant covariance structure is positive definite.

Spatial interpolation and temporal prediction

Usually, in spatio-temporal settings the aim is to perform spatial interpolation of the process for unmonitoring locations and temporal prediction. Initially, assume we want to predict the process for a vector of dimension k, of ungauged locations at observed time $t \in T$, say $\tilde{\boldsymbol{y}}_t = (y_t(\tilde{\boldsymbol{s}}_1), \ldots, y_t(\tilde{\boldsymbol{s}}_k))'$. Assuming that $\boldsymbol{y} = (y_1(\boldsymbol{s}_1), \ldots, y_1(\boldsymbol{s}_n), \ldots, y_T(\boldsymbol{s}_1), \ldots, y_T(\boldsymbol{s}_n))'$ represents the vector of the time series observed at n monitoring locations, we need to obtain the predictive posterior distribution, $p(\tilde{\boldsymbol{y}}_t|\boldsymbol{y})$, which is given by

$$p(\tilde{\boldsymbol{y}}_t|\boldsymbol{y}) = \int_{\boldsymbol{\theta}} p(\tilde{\boldsymbol{y}}_t|\boldsymbol{y},\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}, \tag{1.32}$$

where $\boldsymbol{\theta}$ is a parameter vector comprising all the unknowns in the model.

Spatial interpolations are obtained by considering the distribution of $(\boldsymbol{y}_t, \tilde{\boldsymbol{y}}_t)$ conditional on the parameters, and the initial information $\boldsymbol{\theta}_0$. This distribution is given by

$$\begin{pmatrix} \boldsymbol{y}_t \\ \tilde{\boldsymbol{y}}_t \end{pmatrix} |\boldsymbol{\theta}, V, W \sim N\left(\begin{pmatrix} \boldsymbol{F}_t'\boldsymbol{\theta}_t \\ \tilde{\boldsymbol{F}}_t'\boldsymbol{\theta}_t \end{pmatrix}; \begin{pmatrix} \mathbf{V}_y & \mathbf{V}_{y\tilde{y}} \\ \mathbf{V}_{\tilde{y}y} & \mathbf{V}_{\tilde{y}\tilde{y}} \end{pmatrix} \right),$$

where $\widetilde{\boldsymbol{F}}_t$ corresponds to the regression matrix for ungauged locations. Similar notation is used to split the covariance matrix into four blocks. We then have that

$$\tilde{\boldsymbol{y}}_t|\boldsymbol{y}_t,\boldsymbol{\theta} \sim N\left(\tilde{\boldsymbol{\mu}}_t + \mathbf{V}_{\tilde{y}y}(\mathbf{V}_y)^{-1}(\boldsymbol{y}_t - \boldsymbol{\mu}_t), (\mathbf{V}_{\tilde{y}\tilde{y}} - \mathbf{V}_{\tilde{y}y}(\mathbf{V}_y)^{-1}\mathbf{V}_{\tilde{y}y})\right),$$

where $\boldsymbol{\mu}_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t$ and $\tilde{\boldsymbol{\mu}}_t = \widetilde{\boldsymbol{F}}_t'\boldsymbol{\theta}_t$ for all t. Once samples from the posterior distribution of the parameter vector are available, samples from the posterior predictive distribution of $\tilde{\boldsymbol{y}}_t$ are easily obtained by sampling from the conditional distribution above.

Temporal k steps ahead prediction at monitoring locations is given by the following posterior predictive distribution

$$\begin{aligned} p(\boldsymbol{y}_{T+k}|D_T) &= \int \prod_{h=1}^{k} p(\boldsymbol{y}_{T+h}|\boldsymbol{F}_{T+h},\boldsymbol{\theta}_{T+h},\boldsymbol{V})p(\boldsymbol{\theta}_{T+h}|\boldsymbol{\theta}_{T+h-1},\boldsymbol{W}) \\ &\times \; p(\boldsymbol{\theta}|D_T)p(\boldsymbol{W})p(\boldsymbol{V})d\tilde{\boldsymbol{\theta}}d\boldsymbol{\theta}d\boldsymbol{W}d\boldsymbol{V}, \end{aligned}$$

where $D_T = \{\boldsymbol{y}_1,\ldots,\boldsymbol{y}_T\}$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_T)$ and $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\theta}_{T+1},\ldots,\boldsymbol{\theta}_{T+k})$. The integral above can be approximated by

$$p(\boldsymbol{y}_{T+k}|D_T) \approx \frac{1}{M}\sum_{m=1}^{M} p(\boldsymbol{y}_{T+k}|\boldsymbol{F}_{T+k},\boldsymbol{\theta}_{T+k}^{(m)},\boldsymbol{V}^{(m)}).$$

Here the superscript (m) denotes samples from the posterior of $\boldsymbol{\theta}_{T+1},\ldots,\boldsymbol{\theta}_{T+k}$, $\boldsymbol{V}$. Samples from the distribution above can be obtained by propagating $\boldsymbol{\theta}_{T+k}$ following the system equation in (1.18), and through the samples from the posterior distribution of the parameter vector.

The use of multivariate DLM to model spatio-temporal processes has been widely used in the literature. For example, [28] propose a spatio-temporal model for hourly ozone levels in Mexico City. The model is formulated within the state-space framework, and allows for uncertainty on any missing values of ozone concentrations and covariates. [19] extend the multivariate DLM to allow the coefficients $\boldsymbol{\theta}_t$ to vary smoothly with space. They extend this proposal to model multivariate spatio-temporal processes. [40] extend the class of DLM to allow the error term $\boldsymbol{\epsilon}_t$ to follow autoregressive processes with spatially varying coefficients.

Spatio-temporal models for non-normal data

For many environmental processes that vary across space and time, the assumption of normality of the observations is unreasonable. The univariate generalized dynamic model described in Section 1.2.5 is naturally extended to accommodate spatio-temporal observations. Commonly, the spatio-temporal process is built upon the assumption of conditional independence given the parameter vector defining the family of distribution under interest. More specifically, let $y_t(\boldsymbol{s})$ denote a spatio-temporal process under study, and we assume

that

$$\begin{aligned} y_t(\boldsymbol{s})|\eta_t(\boldsymbol{s}),\phi &\sim EF(\eta_t(\boldsymbol{s}),\phi) \\ g(\mu_t(\boldsymbol{s})) &= \boldsymbol{F}_t'(\boldsymbol{s})\boldsymbol{\theta}_t + \epsilon_t(\boldsymbol{s}), \quad \epsilon_t(\boldsymbol{s}) \sim N(\mathbf{0}, \boldsymbol{\Sigma}_t) \\ \boldsymbol{\theta}_t &= \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \boldsymbol{W}) \\ \boldsymbol{\omega}_0|D_0 &\sim N(\mathbf{m}_0, \boldsymbol{C}_0), \end{aligned} \tag{1.33}$$

where, again, $\eta_t(\boldsymbol{s})$ is the canonical parameter of the exponential family, ϕ is a scale parameter, usually time invariant, $\mu_t(\boldsymbol{s}) = E(y_t(\boldsymbol{s})|\eta_t(\boldsymbol{s}),\phi)$, $g(.)$ is a known link function, at least twice differentiable, which relates the linear predictor with the canonical parameter, and $\boldsymbol{F}_t'(\boldsymbol{s})$ is a p-dimensional vector with the covariates observed at location $\boldsymbol{s}$.

Example: This example shows the resultant properties of the model in equations (1.33) for the Poisson case, where $y_t(\boldsymbol{s}) \mid \lambda_t(\mathbf{s}) \sim Poisson(\lambda_t(\boldsymbol{s}))$. This can be viewed as a Poisson-lognormal mixture model. Using the results based on conditional expectation, it can be shown that marginal expectation, variance and covariance structures for $y_t(\boldsymbol{s})$ are given by:

$$\begin{aligned} E(y_t(\boldsymbol{s})) &= \beta_t(\boldsymbol{s}) = \exp\{\boldsymbol{F}_t(\boldsymbol{s})'\boldsymbol{\theta}_t + 0.5\boldsymbol{\Sigma}(\boldsymbol{s},\boldsymbol{s})\} \\ V(y_t(\boldsymbol{s})) &= \beta_t(\boldsymbol{s}) + \beta_t^2(\boldsymbol{s})[\exp\{\boldsymbol{\Sigma}(\boldsymbol{s},\boldsymbol{s})\} - 1] \\ C(y_t(\boldsymbol{s}_i), y_t(\boldsymbol{s}_j)) &= \beta_t(\boldsymbol{s}_i)\beta_t(\boldsymbol{s}_j)[\exp\{\boldsymbol{\Sigma}(\boldsymbol{s}_i,\boldsymbol{s}_j)\} - 1], \end{aligned} \tag{1.34}$$

which is able to capture overdispersion, as $V(y_t(\boldsymbol{s})) > E(y_t(\boldsymbol{s}))$. If $\boldsymbol{F}_t(\boldsymbol{s})$ has covariates that vary with $\boldsymbol{s}$, regardless of the covariance structure $\boldsymbol{\Sigma}$, $y_t(\boldsymbol{s})$ is a nonstationary process in space.

Note that in the case of spatio-temporal generalized dynamic models, the MCMC algorithm for the state vector $\boldsymbol{\theta}_t$ can take advantage of the FFBS algorithm to sample from the posterior full conditional of the state vectors, because of the presence of the latent component $\epsilon_t(\boldsymbol{s})$ in equation (1.33).

A practical example of generalized spatio-temporal dynamic linear model is proposed by [49], who propose a joint model for rainfall-runoff, two of the most important processes in hydrology. The challenge is to account for the different observational spatial scales of rainfall and runoff. This is done by using generalized dynamic linear models, and the effect of rainfall on runoff is modelled through a transfer function model [59]. Schmidt *et al.* (2010) [52] explore a Poisson-lognormal mixture spatio-temporal model for outbreaks of malaria in the state of Amazonas, Brazil. They compare models that assume different structures for the covariance matrix $\boldsymbol{\Sigma}$, including an inverse-Wishart prior for $\boldsymbol{\Sigma}$.

Areal level spatio-temporal processes

Refer to Sudipto's chapter? Is there areal data there? In this scenario, for each time t, spatial observations are made at a (regular or irregular) partition of the

region of interest D. Now, both temporal and spatial indices, vary discretely. This process is denoted as Y_{ti} for $t = 1, 2, \cdots$ and $i = 1, 2, \cdots, n$.

Usually, for each time t, observations are available for all subregions and, commonly, interest lies on spatial smoothing and temporal prediction. In this set up, it is common practice to capture spatial effects through Markov random fields. The idea of the Markov random field is to specify the joint distribution of the spatial effects through local specifications. [5] introduced the conditionally autoregressive (CAR) models. Let ϵ_i be a latent spatial effect at location i, and define the conditional distribution

$$\epsilon_i | \epsilon_j, j \neq i \sim N\left(\sum_j w_{ij}\epsilon_j / w_{i+}\,,\, \tau^2/w_{i+}\right) \quad \text{for } i = 1, 2, \cdots, n\,.$$

Through Brook's Lema, it can be shown that the joint distribution of the spatial effects is proportional to $p(\epsilon_1, \cdots, \epsilon_n) \propto \exp\left\{-\frac{1}{2\tau^2}\boldsymbol{\epsilon}'(\mathbf{D}_w - \boldsymbol{W})\boldsymbol{\epsilon}\right\}$, where $\mathbf{D}_w$ is diagonal with elements $(\mathbf{D}_w)_{ii} = w_{i+}$. The matrix $\boldsymbol{W}$ is a proximity matrix defining the neighbourhood structure. In practice, it is common to assume $\boldsymbol{W}$ as following a first order neighbouring structure, that is, $W_{ij} = 1$ if locations i and j are neighobors, and 0 otherwise. The joint distribution of $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_n)'$ defined in this way is improper because $(\mathbf{D}_w - \boldsymbol{W})\mathbf{1}_n = 0$. For this reason, this distribution cannot be used as a model for data. One way to make the distribution of $\boldsymbol{\epsilon}$ proper is to introduce a parameter ρ in $\mathbf{D}_w - \rho\boldsymbol{W}$ with $\rho \in (1/\lambda_{(1)}, 1/\lambda_{(n)})$ where $\lambda_{(1)}$ and $\lambda_{(n)}$ are, respectively, the smallest and biggest eigenvalues of $\mathbf{D}_w^{-1/2}\boldsymbol{W}\mathbf{D}_w^{1/2}$. See [2, pages 80-84] for details.

These models are commonly used in epidemiological studies (refer to Francesca's chapter?). For example, [41] propose a Poisson spatio-temporal model for the number of cases of malaria observed at municipalities of the state of Pará, Brazil as a function of rainfall. They propose a dynamic regression model similar to equation (1.33), and compare models that assume ϵ_{it} to be independent across space for each time t, vis-a-vis a CAR prior distribution for $\boldsymbol{\epsilon}_t$. They also explore models that capture a temporal correlation among the $\boldsymbol{\epsilon}_t$ by allowing the logarithm of the variance of the conditional autoregressive process to vary smoothly with time.

Vivar and Ferreira (2002) [57] extend the multivariate DLM to accommodate spatial structures for areal data processes. This is done by assuming proper Gaussian Markov random fields as the distribution of the innovations of both observational and evolution equations, that is, $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\omega}_t$ in equation (1.33). They discuss inference procedure under the Bayesian paradigm and they propose the forward information filter backward sampler algorithm, a modification of the forward filter backward sampler algorithm.

1.3.6 Dynamic factor models (DFM)

Here we briefly revisit two classes of dynamic factor models. In the first case both factor loadings, factor variances and idiosyncratic variances are all time-

varying and follows standard dynamic linear structures. In the second case, the unobservable factor scores themselves follow standard vector ARMA structures.

Standard normal factor analysis is the backbone of dynamic factor models. In this context, data on p related variables are considered to arise through random sampling from a zero-mean multivariate normal distribution where $\boldsymbol{\Omega}$ denotes an $p \times p$ non-singular variance matrix. For any specified positive integer $q \leq p$, the standard $q-$factor model relates each $\boldsymbol{y}_t$ to an underlying $q-$vector of random variables $\boldsymbol{f}_t$, the common factors, via

$$\boldsymbol{y}_t = \boldsymbol{\beta}\boldsymbol{f}_t + \boldsymbol{\epsilon}_t, \tag{1.35}$$

where (i) the factors $\boldsymbol{f}_t$ are independent with $\boldsymbol{f}_t \sim N(\boldsymbol{0}, \boldsymbol{I}_q)$, (ii) the $\boldsymbol{\epsilon}_t$ are independent normal $p-$vectors with $\boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$, and $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \cdots, \sigma_p^2)$, (iii) $\boldsymbol{\epsilon}_t$ and $\boldsymbol{f}_s$ are independent for all t and s, and (iv) $\boldsymbol{\beta}$ is the $p \times q$ factor loadings matrix.

Under this model, the variance-covariance structure of the data distribution is constrained with $\boldsymbol{\Omega} = V(\boldsymbol{y}_t|\boldsymbol{\Omega}) = V(\boldsymbol{y}_t|\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \boldsymbol{\Omega} = \boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Sigma}$. Conditional on the common factors, observable variables are uncorrelated. In other words, the common factors explain all the dependence structure among the p variables. For any elements y_{it} and y_{jt} of $\boldsymbol{y}_t$ and conditionally on $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, we have the characterising moments, (i) $\text{var}(y_{it}|\boldsymbol{f}) = \sigma_i^2$, (ii) $\text{cov}(y_{it}, y_{jt}|\boldsymbol{f}) = 0$, (iii) $\text{var}(y_{it}) = \sum_{l=1}^{q} \beta_{il}^2 + \sigma_i^2$, and (iv) $\text{cov}(y_{it}, y_{jt}) = \sum_{l=1}^{q} \beta_{il}\beta_{jl}$.

In practical problems, especially with larger values of p, the number of factors q will often be small relative to p, so most of the variance-covariance structure is explained by a few number of common factors. The *uniqueness,* or *idiosyncratic variance*, σ_i^2, measures the residual variability in the i^{th} component of $\boldsymbol{y}_t$ once the contribution by the factors is accounted for. Modern MCMC-based posterior inference in standard factor analysis appears in, among others, [22] and [37]. Extensions to time-varying loadings and time-varying idiosyncratic and factor variances can be found in Lopes and Carvalho (2007) [35], and the references therein.

Spatial dynamic factor models

Lopes *et al.* (2008) [36] propose a new class of nonseparable and nonstationary space-time models that resembles a standard dynamic factor model (see [42]):

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{\beta}\boldsymbol{f}_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim N(0, \boldsymbol{\Sigma}) \tag{1.36}$$

$$\boldsymbol{f}_t = \boldsymbol{\Gamma}\boldsymbol{f}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim N(0, \boldsymbol{\Lambda}) \tag{1.37}$$

where $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{nt})'$ is the n-dimensional vector of observations (locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ and times $t = 1, \ldots, T$), $\boldsymbol{\mu}$ is the mean level of the space-time process, $\boldsymbol{f}_t$ is an m-dimensional vector of common factors, for $m < n$ (m is potentially several orders of magnitude smaller than n) and $\boldsymbol{\beta} = (\boldsymbol{\beta}_{(1)}, \ldots, \boldsymbol{\beta}_{(m)})$ is the $n \times m$ matrix of factor loadings. The matrix $\boldsymbol{\Gamma}$ characterizes the evolution dynamics of the common factors, while $\boldsymbol{\Sigma}$ and $\boldsymbol{\Lambda}$ are observational and

evolutional variances. For simplicity, it is assumed that $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ and $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_m)$. The dynamic evolution of the factors is characterized by $\boldsymbol{\Gamma} = \text{diag}(\gamma_1, \ldots, \gamma_m)$, which can be easily extended to non-diagonal cases.

Equations (1.36) and (1.37) define the first level of the proposed dynamic factor model. Similar to standard factor analysis, it is assumed that the m conditionally independent common factors $\boldsymbol{f}_t$ capture all time-varying covariance structure in $\boldsymbol{y}_t$. The conditional spatial dependencies are modeled by the columns of the factor loadings matrix $\boldsymbol{\beta}$. More specifically, the j^{th} column of $\boldsymbol{\beta}$, denoted by $\boldsymbol{\beta}_{(j)} = (\beta_{(j)}(\boldsymbol{s}_1), \ldots, \beta_{(j)}(\boldsymbol{s}_n))'$, for $j = 1, \ldots, m$, is modeled as a conditionally independent, distance-based Gaussian process or a Gaussian random field (GRF), i.e.

$$\boldsymbol{\beta}_{(j)} \sim GRF(\boldsymbol{\beta}_0, \tau_j^2 \rho_{\boldsymbol{\phi}_j}(\cdot)) \equiv N(\mu_j^{\beta^*}, \tau_j^2 \boldsymbol{R}_{\boldsymbol{\phi}_j}), \tag{1.38}$$

where $\boldsymbol{\beta}_0$ is a n-dimensional mean vector. The (l,k)-element of $\boldsymbol{R}_{\phi_j}$ is given by $r_{lk} = \rho(||s_l - s_k||, \boldsymbol{\phi}_j)$, $l, k = 1, \ldots, N$, for suitably defined correlation functions $\rho(\cdot, \boldsymbol{\phi}_j)$, $j = 1, \ldots, m$. The proposed model could, in principle, accommodate nonparametric formulations for the spatial dependence, such as the ones introduced by [20], for instance. The spatial dynamic factor model is defined by equations (1.36)–(1.38).

1.4 Further aspects of spatio-temporal modeling

The spatio-temporal DLM framework previously discussed can be extended to accommodate more complex processes that result on more flexible space-time covariance structures and/or allow to account for physical information. In the following subsections we briefly discuss spatio-temporal models based on process convolution approaches, stochastic partial differential (SPDE) and integro-difference (IDE) equations. The discussion about SPDEs and IDEs follow closely Chapters 6 and 7 of [11].

1.4.1 Process convolution based approaches

A Gaussian process can be obtained through a constructive approach by convolving a continuous white noise process $\theta(\boldsymbol{s})$, $\boldsymbol{s} \in D \subset \mathbb{R}^2$ with a smoothing kernel $k(\boldsymbol{s})$. [26] allow the smoothing kernel to vary with location resulting on a nonstationary Gaussian process. On the other hand, [24] extend the spatial convolution approach to model spatio-temporal processes, and [25] extend this spatio-temporal convolution approach to allow the spatial kernel to evolve smoothly with time. This proposal is as follows. Let $Z_t(\boldsymbol{s})$ be a stochastic pro-

cess defined in continuous space D and discrete time t, such that

$$Z_t(\boldsymbol{s}) = \int_D k(\mathbf{u} - \boldsymbol{s})\theta(\mathbf{u}, t)d\mathbf{u}, \tag{1.39}$$

where $k(\boldsymbol{s})$ is a smoothing kernel. To reduce the dimension of the problem, consider $\theta(\cdot, t)$ to be nonzero at m spatial locations, $\boldsymbol{l}_1, \cdots, \boldsymbol{l}_m$, and now

$$Z_t(\boldsymbol{s}) = \sum_{j=1}^{m} k(\boldsymbol{l}_j - \boldsymbol{s})\theta_{jt}, \tag{1.40}$$

where each sequence $\{\theta\}_{jt}$ follow a Gaussian random walk, such that $\theta_{jt} = \theta_{j\,t-1} + \omega_{jt}$ with $\omega_{jt} \sim N(0, W)$. Note that $\mathcal{L} = \{\boldsymbol{l}_1, \cdots, \boldsymbol{l}_m\}$ plays the role of the spatial support that approximates the continuous process in D, and reduces the spatial dimension of the problem. If we add an independent measurement error component in equation (1.40) together with the random walk assumption for the θ_{jt} we have a DLM:

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{F}'\boldsymbol{\theta}_t + \boldsymbol{v}_t \\ \boldsymbol{\theta}_t &= \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \end{aligned}$$

where $\boldsymbol{y}_t = (Z_t(\boldsymbol{s}_1), \ldots, Z_t(\boldsymbol{s}_n))'$, $\boldsymbol{F}_{ji} = k(\boldsymbol{l}_j - \boldsymbol{s}_i)$ and $\boldsymbol{\theta}_t = (\theta_{1t}, \ldots, \theta_{mt})'$.

Higdon (2002) [25] uses this approach to model daily concentrations of ozone across 30 days. [55] use a similar approach by describing the mean of the process as a locally-weighted mixture of linear regressions, with the coefficients evolving smoothly with time. [50] consider a class of models for spatio-temporal processes based on convolving independent processes with a discrete kernel that is represented by a lower triangular matrix. This was inspired by coregionalization models for multivariate spatial data. [7] extend the dynamic process convolution model to describe multivariate spatio-temporal processes. This convolution approach can also be used to model non-normal data. For example, [27] use the dynamic process convolution to model the location parameter of a generalized extreme value distribution to describe extreme levels of ozone at some monitoring stations in Mexico City.

1.4.2 Models based on stochastic partial differential equations

Commonly, when modeling spatio-temporal processes, scientific theory is available, e.g. through the specification of deterministic models, to describe a wide class of spatio-temporal processes in ecology, environment and epidemiology. This theory might serve as prior information to describe the process under study. [64] mention that one should make use of the information in the developed science but should also recognize that there are uncertainties associated with that knowledge. For example, [23] uses stochastic partial differential equations to develop general classes of covariance functions for spatial

processes. [6] proposes a non-separable space-time covariance matrix based on a physical dispersion model, that is suitable to describe processes that spread or disperse over time, such as air pollution. [4] discusses the combination of physical reasoning and observational data through the use of Bayesian hierarchical models.

The DLM framework previously discussed can be used to accommodate the use of partial differential equations (PDEs) to account for scientific knowledge of the process. Moreover, it accounts for the uncertainty about the specification of the dynamics of the process, as the PDE is substituted by a stochastic PDE. Broadly speaking, this is done by allowing the state evolution to be defined through a partial differential equation model (e.g. reaction-diffusion processes), and the spatial process is approximated through a spatial grid and partial differential equations are solved using finite-difference methods [56]. [62] mentions that focus has been given on reaction-diffusion processes modeled via PDEs, integrodifference equations and discrete-time contact models. The choice among these models depend on whether one is considering discrete time and/or space [62].

Consider the stochastic process $\{Z(s,t), s \in D_s \subset \mathbb{R}, t \in D_t \subset \mathbb{R}\}$. [23] considers models for two-dimensional processes based on the general second-order linear SPDE which leads to three types of models: parabolic, elliptic and hyperbolic forms. Following [11], we focus on the parabolic SPDE [23, 29] which is governed by

$$\frac{\partial Z(s,t)}{\partial t} - \beta \frac{\partial^2 Z(s,t)}{\partial s^2} + \alpha Z(s,t) = \delta(s,t), \tag{1.41}$$

where $\alpha > 0$, $\beta > 0$, and $\delta(s,t)$ is a zero mean random error process. Although these models are difficult to deal with analytically, [23] shows that $Z(s,t)$ has a stationary covariance function and derives an analytical expression for the corresponding spatio-temporal correlation function. Equation (1.41) can be approximated through finite differences,

$$Z(s; t+\Delta_t) \approx \delta_1 Z(s,t) + \delta_2 Z(s+\Delta_s; t) + \delta_2 Z(s-\Delta_s; t), \tag{1.42}$$

by assuming that the first-order partial derivative in time is approximated with a forward difference, the partial second-order in one-dimensional space is approximated with a centered difference, with $\delta_1 = (1 - \alpha\Delta_t - 2\beta\Delta_t/\Delta_s^2)$, $\delta_2 = \beta\Delta_t/\Delta_s^2$, and $\alpha\Delta_t < 1$ and $2\beta\Delta_t/\Delta_s^2 < 1$.

According to equation (1.42), the future value of Z at location s is related to current values at s and neighboring spatial locations. If we discretize space by defining $D_s = \{s_0, \cdots, s_{n+1}\}$, where $s_j = s_0 + j\Delta_s$ and $j = 0, 1, \cdots, n+1$, for each $k = 1, \cdots, n$, we can write

$$Z(s_k; t+\Delta_t) = \delta_1 Z(s_k; t) + \delta_2 Z(s_{k+1}; t) + \delta_2 Z(s_{k-1}; t).$$

Defining $\mathbf{Z}_t = (Z(s_1;t), \cdots, Z(s_n;t))'$, and $\mathbf{Z}_t^{(b)} = (Z(s_0;t), Z(s_{n+1};t))'$, we have

$$\mathbf{Z}_{t+\Delta_t} = \boldsymbol{G}(\boldsymbol{\delta})\mathbf{Z}_t + \boldsymbol{G}^{(b)}(\delta_2)\mathbf{Z}_t^{(b)}, t = 0, 1, \cdots \tag{1.43}$$

where $\boldsymbol{G}(\boldsymbol{\delta})$ is a $n \times n$, tridiagonal propagator matrix, and $\boldsymbol{G}^{(b)}(\delta_2)$ is a $n \times 2$ boundary propagator matrix. Let $D_t = \{0, 1, 2, \ldots\}$, then from a given initial condition $\mathbf{Z}_0$ and boundary conditions $\{\mathbf{Z}_t^{(b)} : t = 0, 1, \ldots, \}$ we can obtain a numerical solution to the difference equation in (1.42), which in turn is an approximation to the SPDE in (1.41). If a mean-zero, random error component $\boldsymbol{\epsilon}_{t+1}$, is added to the right-hand side of equation (1.43) we obtain a stochastic spatio-temporal difference equation corresponding to the parabolic SPDE in equation (1.41). We assume that $\{\boldsymbol{\epsilon}_t\}$ is i.i.d. Whittle (1986) [60] calls this a *diffusion-injection difference equation*. As the boundary term is given, this stochastic difference equation can be seen as a multivariate first-order autoregressive model.

The idea above can be considered to model counts that vary across space and time. [62] proposes a Poisson-lognormal mixture model to estimate the abundance of House Finches over the eastern USA. The proposed model is similar to the one shown in equations (1.33) with an additional component in the $\log(\lambda_t(\boldsymbol{s}))$ that follows a reaction-diffusion model, which is approximated through first-order forward differences in time and centered differences in space.

1.4.3 Models based on integro-difference equations

As described in [11, Section 7.2] integro-difference equations (IDEs) might be useful in describing spatio-temporal processes that have long-range dependence, requiring a more general form of the propagator matrix $\boldsymbol{G}$. When time is discrete and space continuous, in the case of linear dynamics, the IDE is given by

$$Z_t(\boldsymbol{s}) = \int_{D_s} g(\boldsymbol{s}, \mathbf{x}; \boldsymbol{\delta_s}) Z_{t-1}(\mathbf{x}) d\mathbf{x}, \quad \boldsymbol{s} \in D_s \tag{1.44}$$

where D_s is the spatial domain of interest. The function $g(\boldsymbol{s}, \mathbf{x}; \boldsymbol{\delta})$ is called a *redistribution kernel* which can have a different shape for each spatial location $\boldsymbol{s}$, controlled by parameters $\boldsymbol{\delta}_s$. If we add a random noise term, $\omega_t(\boldsymbol{s})$, which is independent in time but may be correlated in space and is independent of $Z_{t-1}(\boldsymbol{s})$, such that

$$Z_t(\boldsymbol{s}) = \int_{D_s} g(\boldsymbol{s}, \mathbf{x}; \boldsymbol{\delta}_s) Z_{t-1}(\mathbf{x}) d\mathbf{x} + \omega_t(\boldsymbol{s}),$$

a stochastic IDE results.

As proposed by [63], typically one has to discretize the integral equation to reformulate the model. Considering a discretized grid, the kernel defined for locations $\boldsymbol{s}_i$, $g(\boldsymbol{s}_i, \mathbf{x}, \boldsymbol{\delta}_s)$ corresponds to the $i - th$ row of the propagator matrix $\boldsymbol{G}(\boldsymbol{\delta})$, such that

$$\mathbf{Z}_t = \boldsymbol{G}(\boldsymbol{\delta}) \mathbf{Z}_{t-1} + \boldsymbol{\omega}_t$$

where $\boldsymbol{\omega}_t \sim N(\mathbf{0}, \boldsymbol{R})$.

This discretized version of the model is suitable to achieve state-space dimension reduction and can be put onto a state-space framework [63, 10]. [11] mentions that IDE dynamical models can accommodate complicated spatio-temporal dynamics with relatively few parameters. This IDE approach is naturally extented to accommodate non-normal processes. [61] assumes a conditional spatio-temporal Poisson intensity process for cloud intensity data, and uses the kernel-based integro-difference approach as a latent component in the mean of the logarithm of the Poisson intensity process.

The processes described in the previous subsections can be incorporated as components in hierarchical models to accommodate complex structures. Typically, inference procedure is performed under the Bayesian paradigm. Applications of such models can be seen in Chapters Nychka & Wikle, and M. Berliner & C. Wikle See also [11] and, references therein, for more details.

Acknowledgements

A. M. Schmidt is grateful for partial financial support from CNPq and FAPERJ, Brazil. The research of H. F. Lopes is partially financed by Insper, CNPq and FAPESP, Brazil.

Bibliography

[1] D. K. Agarwal, A. E. Gelfand, and S. Citron-Pousty. Zero inflated models with application to spatial count data. *Environmental and Ecological Statistics*, 9:341–355, 2002.

[2] S. Banerjee, B. P. Carlin, and A. E. Gelfand. *Hierarchical modeling and analysis for spatial data*. CRC Press, 2014.

[3] L. M. Berliner. Hierarchical Bayesian time series models. In K. Hanson and R. Silver, editors, *Maximum Entropy and Bayesian Methods*, pages 15–22. Kluwer Acad., Norwell, Mass, 1996.

[4] L. M. Berliner. Physical-statistical modeling in geophysics. *Journal of Geophysical Research: Atmospheres*, 108(D24), 2003.

[5] J. Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 192–236, 1974.

[6] P. E. Brown, G. O. Roberts, K. F. Kåresen, and S. Tonellato. Blur-generated non-separable space–time models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):847–860, 2000.

[7] C. A. Calder. Dynamic factor process convolution models for multivariate space–time data with application to air quality assessment. *Environmental and Ecological Statistics*, 14(3):229–247, 2007.

[8] C. K. Carter and R. Kohn. On Gibbs sampling for state space models. *Biometrika*, 81:541–553, 1994.

[9] C. M. Carvalho and M. West. Dynamic matrix-variate graphical models. *Bayesian Analysis*, 2:69–98, 2007.

[10] N. Cressie and C. K. Wikle. Space-time Kalman filter. *Encyclopedia of Environmetrics*, 2002.

[11] N. Cressie and C. K. Wikle. *Statistics for spatio-temporal data*. John Wiley & Sons, 2015.

[12] A.P. Dawid. Some matrix-variate distribution theory: Notational considerations and a Bayesian application. *Biometrika*, 68:265–274, 1981.

[13] T.J. DiCiccio, R.A. Kass, A.E. Raftery, and L. Wasserman. Computing Bayes' factors by combining simulation and asymptotic approximations. *Journal of the American Statistical Association*, 92:903–915, 1997.

[14] J. Durbin and S.J. Koopman. *Time Series Analysis by State Space Methods*. Oxford Statistical Science Series. Clarendon Press, 2001.

[15] L. Fahrmeir. Posterior mode estimation by extended Kalman filtering for multivariate dynamic linear models. *J Am Stat Assoc*, 87:501–509, 1992.

[16] S. Frühwirth-Schnater. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15(2):183–202, 1994.

[17] D. Gamerman. Markov chain Monte Carlo for dynamic generalised linear models. *Biometrika*, 85(1):215–227, 1998.

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

[19] A. E. Gelfand, S. Banerjee, and D. Gamerman. Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics*, 16(5):465–479, 2005.

[20] A. E. Gelfand, A. Kottas, and S. N. MacEachern. Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association*, 100:1021–35, 2005.

[21] J. Geweke and H. Tanizaki. Bayesian estimation of state space models using Metropolis-Hastings algorithm within Gibbs sampling. *Comput Stat Data An*, 37:151–170, 2001.

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

[23] V Heine. Models for two-dimensional stationary stochastic processes. *Biometrika*, 42(1-2):170–178, 1955.

[24] D. Higdon. A process-convolution approach to modelling temperatures in the north atlantic ocean. *Environmental and Ecological Statistics*, 5(2):173–190, 1998.

[25] D. Higdon. Space and space-time modeling using process convolutions. In Anderson C., Barnett V., Chatwin P.C., and El-Shaarawi A.H., editors, *Quantitative methods for current environmental issues*, pages 37–56. Springer Verlag, London, 2002.

[26] D. Higdon, J. Swall, and J. Kern. Non-stationary spatial modeling. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian statistics 6: Proceedings of the Sixth Valencia International Meeting*, pages 761–768. Oxford University Press, 1999.

[27] Gabriel Huerta and Bruno Sansó. Time-varying models for extreme values. *Environmental and Ecological Statistics*, 14(3):285–299, 2007.

[28] Gabriel Huerta, Bruno Sansó, and Jonathan R Stroud. A spatiotemporal model for mexico city ozone levels. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53(2):231–248, 2004.

[29] R. H. Jones and Y. Zhang. Models for continuous stationary space-time processes. In *Modelling longitudinal and spatially correlated data*, pages 289–298. Springer, 1997.

[30] G. Kitagawa. Non-Gaussian state-space modeling of non-stationary time series. *J Am Stat Assoc*, 82:1032–1041, 1987.

[31] G. Koop and D. Korobilis. Large time-varying parameter VARs. *Journal of Econometrics*, 177:184–198, 2013.

[32] F. Landim and D. Gamerman. Dynamic hierarchical models: an extension to matrix-variate observations. *Computational Statistics & Data Analysis*, 35:11–42, 2000.

[33] D.V. Lindley and A.F.M. Smith. Bayes estimates for the linear model. *Journal of the Royal Statistical Society. Series B*, pages 1–41, 1972.

[34] F. Liu. *Bayesian Time Series: Analysis Methods Using Simulation Based Computation*. PhD thesis, Institute of Statistics and Decision Sciences, Duke University, Durham, North Carolina, USA., 2000.

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

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

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

[38] H. S. Migon, A. M. Schmidt, R. E. R. Ravines, and J. B. M. Pereira. An efficient sampling scheme for dynamic generalized models. *Computational Statistics*, 28(5):2267–2293, 2013.

[39] J. Nakajima, M. Kasuya, and T. Watanabe. Bayesian analysis of time-varying parameter vector autoregressive model for the Japanese economy and monetary policy. *Journal of The Japanese and International Economies*, 25:225–245, 2011.

[40] A. A. Nobre, B. Sansó, and A. M. Schmidt. Spatially varying autoregressive processes. *Technometrics*, 53(3):310–321, 2011.

[41] A. A. Nobre, A. M. Schmidt, and H. F. Lopes. Spatio-temporal models for mapping the incidence of malaria in pará. *Environmetrics*, 16(3):291–304, 2005.

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

[43] G. Petris, S. Petrone, and P. Campagnoli. *Dynamic linear models*. Springer, 2009.

[44] A. Pole, M. West, and J. Harrison. *Applied Bayesian forecasting and time series analysis*. CRC Press, 1994.

[45] R. Prado and M. West. *Time series: modeling, computation, and inference*. CRC Press, 2010.

[46] G. E. Primiceri. Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, 72:821–852, 2005.

[47] J.M. Quintana and M. West. Multivariate time series analysis: new techniques applied to international exchange rate data. *The Statistician*, 36:275–281, 1987.

[48] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.

[49] R. R. Ravines, A. M. Schmidt, H. S. Migon, and C. D. Rennó. A joint model for rainfallrunoff: The case of Rio Grande basin. *Journal of Hydrology*, 353(1):189 – 200, 2008.

[50] B. Sansó, A. M. Schmidt, and A. A. Nobre. Bayesian spatio-temporal models based on discrete convolutions. *Canadian Journal of Statistics*, 36(2):239–258, 2008.

[51] A. M. Schmidt, P. Guttorp, and A. O'Hagan. Considering covariates in the covariance structure of spatial processes. *Environmetrics*, 22(4):487–500, 2011.

[52] A. M. Schmidt, J. A. Hoeting, J. B. M. Pereira, and P. P. Vieira. Mapping malaria in the Amazon rainforest: a spatio-temporal mixture model. In A. O'Hagan and M. West, editors, *The Oxford Handbook of Applied Bayesian Analysis*, pages 90–117. Oxford University Press, 2010.

[53] A. M. Schmidt and J. B. M. Pereira. Modelling time series of counts in epidemiology. *International Statistical Review*, 79(1):48–69, 2011.

[54] N. Shephard and M. Pitt. Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, 84:653–667, 1997.

[55] J. R Stroud, P. Müller, and B. Sansó. Dynamic models for spatiotemporal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(4):673–689, 2001.

[56] J. R. Stroud, M. L. Stein, B. M. Lesht, D. J. Schwab, and D. Beletsky. An ensemble Kalman filter and smoother for satellite data assimilation. *Journal of the American Statistical Association*, 105(491):978–990, 2010.

[57] J. C. Vivar and M. A. R. Ferreira. Spatiotemporal models for Gaussian areal data. *Journal of Computational and Graphical Statistics*, pages 658–674, 2012.

[58] M. West and J. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer Series in Statistics, second edition, 1997.

[59] M. West, J. Harrison, and H. S. Migon. Dynamic generalized linear models and Bayesian forecasting. *J Am Stat Assoc*, 80(389):73–83, 1985.

[60] P. Whittle. *Systems in stochastic equilibrium*. John Wiley & Sons, Inc., 1986.

[61] C. K. Wikle. A kernel-based spectral model for non-Gaussian spatio-temporal processes. *Statistical Modelling*, 2(4):299–314, 2002.

[62] C. K. Wikle. Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology*, 84(6):1382–1394, 2003.

[63] C. K. Wikle and N. Cressie. A dimension-reduced approach to space-time Kalman filtering. *Biometrika*, 86(4):815–829, 1999.

[64] C. K Wikle and M. B. Hooten. A general science-based framework for dynamical spatio-temporal models. *Test*, 19(3):417–451, 2010.