

Hierarchical Statistical Models Applied to Spatio-Temporal Models

Juan Martínez

10/23/2020

Hierarchical Statistical Models

Statistical uncertainty can be introduced into a model by *thinking conditionally* and building complexity through a series of **conditional-probability** models. For instance, if most of the complex dependencies in the data are due to the **underlying process** of interest, then one should model the distribution of the data **conditioned** in that process (*e.g.*, data model). Then, a model of the process' behavior and its uncertainties (*e.g.*, process model).

There will be unknown parameters in both the statistical model for the data (conditioned on the process) and the statistical model for the process.

Also, when a **dynamic model** of one or several variables is placed within a **hierarchical model** formulation, one obtains what has been called a **state-space model**. In other words, a model for a data set that has been collected sequentially over time. These data is modeled as *noisy* observations of an underlying **state process** evolving statistically through time.

One of the challenges is to find an efficient approach to perform inference on the underlying state process of interest while accounting for the noise.

There are three main stages for **state-space** models:

- *Smoothing*: During this stage one performs inference on the **hidden state process** during a fixed time period in which we have observations throughout the time period.
- *Filtering*: This stage is about inference on the **hidden state** value at the most current time based on the current and all past data.
- *Forecasting*: This refers to inference on the **hidden state** value at any time point beyond the current time, where data are either not available or not considered in the forecast.

In this case, we are model spatial processes evolving through time.

Moreover, there might be uncertainties in the parameters. These uncertainties could be accounted statistically by assigning **prior distributions** on the parameters. The structure of this type of models are also known as **hierarchical models** (HMs), in which *data model*, a *process model*, and a *parameter model* are included.

In this framework, the main key is that at any *level* of spatio-temporal HM, it is convenient to put as much of the dependence structure as possible in the **conditional-mean** specification in order to simplify the **conditional-covariance** specification.

When the parameters are given prior distributions at the **bottom level** of the **hierarchy**, then we say that the model is a *Bayesian hierarchical model* (BHM). A BHM is often necessary for complex-modeling situations, because the parameters themselves may exhibit complex spatio-temporal structure.

A simpler approach is to estimate the parameters present in the top two levels in some way using the data or other sources if data; then we like to say that the hierarchical model is an *empirical hierarchical model* (EHM). In this framework, the modeler does not have to assign prior distributions to the parameters.

An important advantage of the **data-process-parameters** modeling paradigm in an HM is that, while marginal-dependence structure are difficult to model directly, **conditional-dependence** structure usually come naturally. For example, it is often reasonable to assume that the *data covariance matrix* (given the corresponding values of the hidden process) is simply a diagonal matrix of measurement-error variances. This allows the *process covariance matrix* to capture the “pure” spatio-temporal dependence, ideally (but not necessarily) from physical or mechanistic knowledge. Then, these two covariance matrices, the seemingly complex *marginal covariance matrix* of the data can simply obtained.

The product of the **conditional-probability** components of the HM gives the **joint probability model** for all random quantities (*i.e.*, all “unknowns”). Thus, the HM could be either a BHM or an EHM, depending on whether, respectively, a **prior distribution** is assigned to the parameters (BHM case), or if the parameters are estimated (EHM case).

In addition, the BHM approach allows the modeler to obtain the **posterior distribution** of the parameters given the data. Both **predictive** and **posterior distributions** are obtained using *Bayes’ Rule*.

References

Appendix