

Modern Bayesian Statistics
Part III: high-dimensional modeling
Example 2: Spatial hierarchical factor analysis

HEDIBERT FREITAS LOPES¹
hedibert.org

13^a aMostra de Estatística
IME-USP, October 2018

¹Professor of Statistics and Econometrics at Insper, São Paulo.

Example 2: Spatial hierarchical factor analysis

Vulnerability index for Uruguay

Uruguay has an area of 176,215 km^2 and roughly 3.3 million inhabitants, half of which live in the capital, Montevideo. Around 93% of the population lives in urban areas.



Census tracts per capital

Capital	Census tracts	Capital	Census tracts
Bella Unión	11	Durazno	35
Canelones	20	Maldonado	36
Colonia	21	Tacuarembó	38
Fray Bentos	22	Mercedes	39
Trinidad	27	Melo	43
Rocha	28	Rivera	45
Treinta y Tres	29	Paysandú	72
Florida	31	Salto	84
Minas	33	Montevideo	1031
San José	34		

Main goals

To characterize the vulnerability of the population of Uruguay to diseases transmitted through vectors (e.g. Dengue Fever, Malaria, etc.);

To help prioritizing the allocation of fundings;

We have information on $p = 11$ variables per census tracts of the $l = 19$ Departamental Capitals of the country.

Source: Census 1996 (latest Census in Uruguay)

Table: Description of the $p = 11$ variables, observed in the census tract level of the departmental capitals, to build the vulnerability index of the population of Uruguay to vector-borne diseases.

Levels	Variables
Personal characteristic	Illiteracy rate (ILL)
	Population with access to public health care (PHC)
	Male without formal jobs (UQW)
Household characteristic	Owed houses (OWH)
	Households headed by a woman (WHF)
	Households without sewage system (AHS)
	Average number of persons per household (APH)
	Households with more than two persons per room (OVC)
	Households without access to drinkable water (ADW)
	Households with air conditioner (ACO)
Households poorly built (HOQ)	

Sample correlations

	ILL	PHC	OVC	UQW	AHS	ADW	APH
PHC	0.85						
OVC	0.78	0.79					
UQW	0.67	0.65	0.68				
AHS	0.64	0.59	0.67	0.60			
ADW	0.60	0.47	0.49	0.51	0.62		
APH	0.53	0.52	0.54	0.38	0.32	0.26	
HOQ	0.45	0.36	0.43	0.40	0.63	0.57	0.23

The sample correlations between OWH or WHF or ACO and any one of the attributes are below 18% (in absolute value).

Model structure

Observational Level:

$$y_{ijk} = \mu_k + \beta_k f_{ij} + \sigma_k \varepsilon_{ijk} \quad k = 1, \dots, p,$$

where μ_k represents the overall grand mean.

Modeling f_{ij} :

$$f_{ij} = \theta_i + \tilde{f}_{ij} + \sqrt{\omega_i} u_{ij}$$

where θ_i is the common factor for capital i .

Spatial variation within capitals:

$$\tilde{f}_i \sim N(0, \tau_i^2 P_i)$$

where $P_i = (I_{n_i} + \phi M_i)^{-1}$, $M_i = D_i - W_i$, with w_{ijl} , the (j, l) component of W_i , given by $w_{ijl} = 1/d_{jl}$ if j and l are neighbors (denoted here by $j \sim l$) and zero otherwise, $d_{jl} = \|s_j - s_l\|$ is the Euclidean distance between centroids of capitals j and l ,

$D_i = \text{diag}(w_{i1+}, \dots, w_{in_i+})$ and $w_{ij+} = \sum_{l \sim j} w_{ijl}$.

Model structure (cont.)

Spatial variation between capitals:

$$\theta \sim N(\mathbf{1}_I \theta_0, \delta^2 H(\lambda)),$$

where $\theta = (\theta_1, \dots, \theta_I)$.

Although each capital i has its own vulnerability factor, the above model allows borrowing-strength across neighboring regions.

Table: *Comparing SHFM and UHF*M: Comparing the unstructured hierarchical factor (UHFM) and spatial hierarchical factor models (SHFM) for different values of ϕ . Best models appear in italic. DIC: deviance information criterion, EPD: expected posterior deviation, CRPS: continuous ranked probability score, MSE: mean square error and MAE: mean absolute error. CRPS are in tens of thousands.

Criterion	UHFM		SHFM		
	$\theta = 0$	unknown θ	$\phi = 1$	$\phi = 5$	$\phi = 7$
DIC	-21445.4	-21493.3	-21785.8	<i>-21827.4</i>	<i>-21827.0</i>
EPD	2557.4	2510.9	2453.1	2433.6	<i>2432.6</i>
CRPS	1030.7	1024.2	1014.2	<i>1010.3</i>	1010.3
MAE	2397.0	2381.8	2374.5	<i>2367.9</i>	2369.1
MSE	1222.3	1200.1	1177.2	1169.2	<i>1168.9</i>

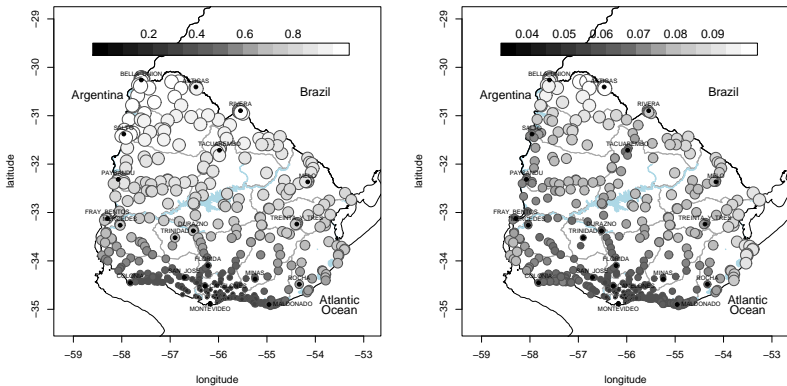


Figure: Posterior mean of θ_i and standard deviations (second column) for observed and unobserved cities under the SHFM when $\phi = 5$.

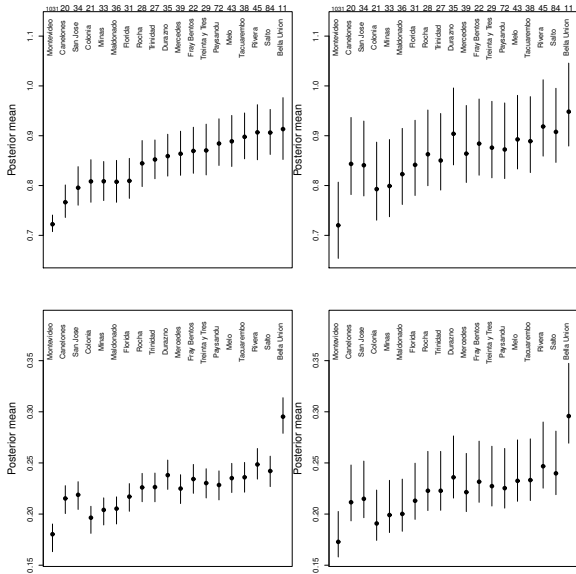


Figure: Posterior means of the θ_i and 95% CI. *Top row:* SHFM with $\phi = 5$ (left) and UHFm (right). *Bottom row:* ASFM (left) and AFM (right).

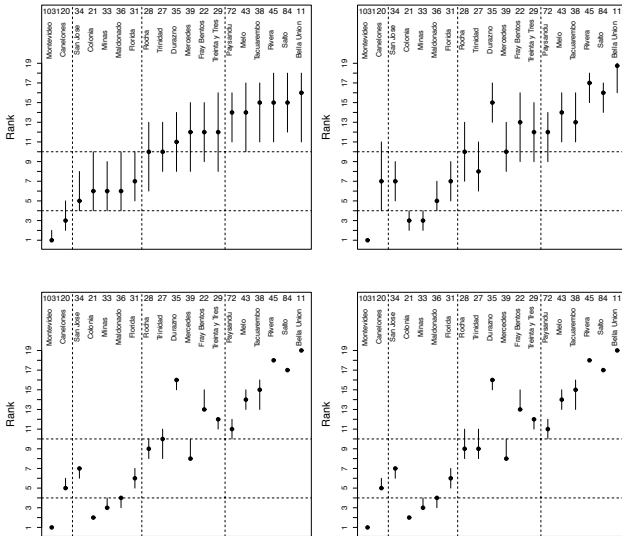
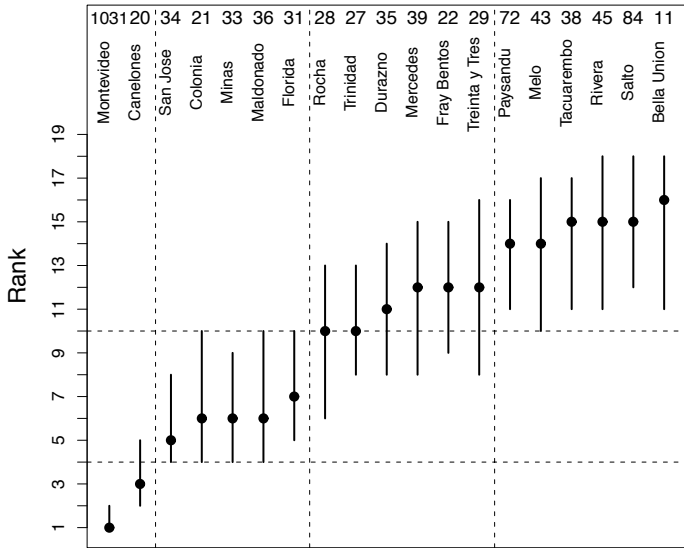
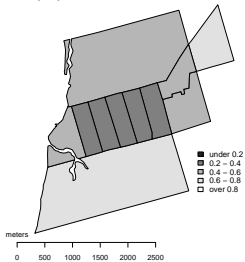


Figure: Posterior rankings of the capitals. *Top row:* SHFM with $\phi = 5$ (left) and UHFM (right). *Bottom row:* ASFM (left) and AFM (right).



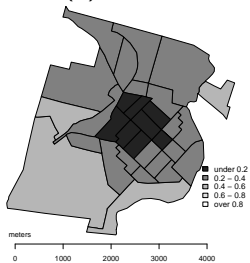
(a) Bella Unión



(b) Melo



(c) Florida



(d) Montevideo

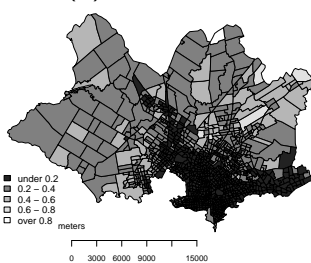


Figure: Within-city posterior vulnerability index per census

