Predictive spatio-temporal models for spatially sparse environmental data

Xavier de Luna and Marc G. Genton

`xavier.deluna@stat.umu.se` and `genton@stat.ncsu.edu`

`http://www.stat.umu.se/egna/xdl/index.html`
`http://www4.stat.ncsu.edu/~mggenton/`

Umeå University

NC State University
DEPARTMENT OF STATISTICS

Outline

1. Introduction

Outline

1. Introduction
2. VAR models with spatial structure

Outline

1. Introduction
2. VAR models with spatial structure
3. Spatio-temporal correlation analysis

Outline

1. Introduction
2. VAR models with spatial structure
3. Spatio-temporal correlation analysis
4. Prediction performance analysis

Outline

1. Introduction
2. VAR models with spatial structure
3. Spatio-temporal correlation analysis
4. Prediction performance analysis
5. Conclusions

1. Introduction

- **Spatio-temporal data:** $z(\mathbf{s}_i, t)$
  at stations $\mathbf{s}_i, i = 1, \ldots, N$ (irregular)
  and times $t = 1, \ldots, T$ (regular)
- **Data:** sparse in space and rich in time
- **Applications:**
  • **Environmental:** e.g. air pollution levels at meteorological stations in a given region/country
  • **Economics:** e.g. socio-economic measurements made at different geographical levels, where stations are whole sub-regions or countries
- **Goal:** build a simple model for predictions in the future at the stations with as few assumptions as possible (e.g. no stationarity or isotropy in space)

Approaches to space-time modeling

- Direct joint modeling of space-time covariance
- Hierarchical Bayesian models
- Vector of time series
- Vector of spatial random fields

see survey by Kyriakidis and Journel (1999)

Many papers in the literature, for example:

- Wikle and Cressie (1999): Kalman filter
- Stroud, Müller, and Sansó (2001): state space
- Tonellato (2001): Bayesian multivar. time series
- Pfeifer and Deutsch (1980), Stoffer (1986): STARMAX
- Previous talks...

2. VAR models with spatial structure

A simple model for $\boldsymbol{z}_t = (z(\mathbf{s}_1, t), \ldots, z(\mathbf{s}_N, t))'$ is:

$$\boxed{\boldsymbol{z}_t = \boldsymbol{\beta} + \sum_{i=1}^{p} R_i \boldsymbol{z}_{t-i} + \boldsymbol{\varepsilon}_t,}$$

- $\boldsymbol{\beta} = (\beta(\mathbf{s}_1), \ldots, \beta(\mathbf{s}_N))'$ is a vector of spatial effects (spatial trend)
- R_i, $i = 1, \ldots, p$, are unknown $N \times N$ parameter matrices
- $\boldsymbol{\varepsilon}_t$, N-dimensional white noise process: $\mathrm{E}(\boldsymbol{\varepsilon}_t) = \mathbf{0}$, $\mathrm{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \Sigma_\varepsilon$, $\mathrm{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_u') = \mathbf{0}$ $u \neq t$

Vector autoregressive (VAR) model commonly used in multivariate time series analysis (e.g., Lütkepohl, 1991). However: spatial structure of data typically has major consequences on R_i (specific structure)

Remarks:

- The deterministic dynamic system is stable if:

$$\det(I - R_1 x - \ldots - R_p x^p) \neq 0, \text{ for } x \in \mathbb{C}, |x| \leq 1$$

- Stability implies time-stationarity: $\mathrm{E}(\boldsymbol{z}_t) = \boldsymbol{\mu}$, for all t, and $\mathrm{Cov}(\boldsymbol{z}_t, \boldsymbol{z}_{t-\tau}) = \Gamma_{\mathbf{z}}(\tau)$, i.e. is a function of τ only, for all t and $\tau = 0, 1, 2, \ldots$
- Covariance matrix $\Gamma_{\mathbf{z}}(\tau)$ can easily be computed from $R_1, \ldots, R_p$ and $\Sigma_{\boldsymbol{\varepsilon}}$. For instance, $p = 1$: $\mathrm{vec}(\Gamma_{\mathbf{z}}(0)) = (I_{N^2} - R_1 \otimes R_1)^{-1} \mathrm{vec}(\Sigma_{\boldsymbol{\varepsilon}})$ and $\Gamma_{\mathbf{z}}(\tau) = R_1^{\tau} \Gamma_{\mathbf{z}}(0)$
- A temporal trend needs to be modeled/removed
- No assumption of spatial stationarity/isotropy

Estimation and inference

Estimation of parameters of our model with:

- maximum likelihood (if distributional assumptions are made)
- least squares
- moments estimators (Yule-Walker type)

More about estimation and inference in, e.g., Lütkepohl (1991)

Robust estimation of the parameters with:
robust estimators of moments (Ma and Genton, 2000) together with Yule-Walker estimating equations (de Luna and Genton, 2002)

Estimation of the parameters: all stations simultaneously or station-wise
Model building must be performed separately for each station

Spatio-temporal trends

- Deterministic trend specification:

$$z(\mathbf{s},t) = \mu + g(\mathbf{s},t) + y(\mathbf{s},t)$$

where $y(\mathbf{s},t)$ is a time stationary process

- Often removed by time differencing with ∇^d:

$$\nabla z(\mathbf{s},t) = g(\mathbf{s},t) - g(\mathbf{s},t-1) + \nabla y(\mathbf{s},t)$$

where $\nabla y(\mathbf{s},t)$ is a time stationary process

- $\nabla z(\mathbf{s},t)$ can be modeled by a VAR if $\beta(\mathbf{s}) = g(\mathbf{s},t) - g(\mathbf{s},t-1)$ is a function of s only
- Will happen most often in practice, at least approximately (as soon as $g(\mathbf{s},t)$ is a polynomial function in t with coefficients possibly function of location s), e.g. $g(\mathbf{s},t) = \gamma_1(\mathbf{s}) + \gamma_2(\mathbf{s})t$

- Periodic time trends or cycles may also be tackled by differencing, e.g. observations taken monthly will typically have to be stationarized with the seasonal differencing operator
$\nabla_{12} z(\mathbf{s}, t) = z(\mathbf{s}, t) - z(\mathbf{s}, t-12)$

- Modeling a deterministic trend with a weighted sum of known basis functions, where the weights are typically estimated by regression:
Periodic functions: to account for seasonal effects along the time axis
Polynomials: to model smooth variations in space
see review article by Kyriakidis and Journel (1999)

- When other variables are observed at the same locations and time, these may also be used to model trends by regressing on them

Model building strategy

- Identify zeroes in the R_i matrices
- Each station is modeled separately: covariate selection problem, where the available predictors are the lagged values at each station
Complex model selection problem
- From the spatio-temporal structure: define an ordering to sequentially introduce the predictors in the model for $z(\mathbf{s}, t)$ say
- Such ordering is used in purely time series models: to explain x_t the lag one variable x_{t-1} is considered first, then lag two x_{t-2}, etc...

NC STATE UNIVERSITY

- To explain $z(\mathbf{s}, t)$, consider predictors in the following order:

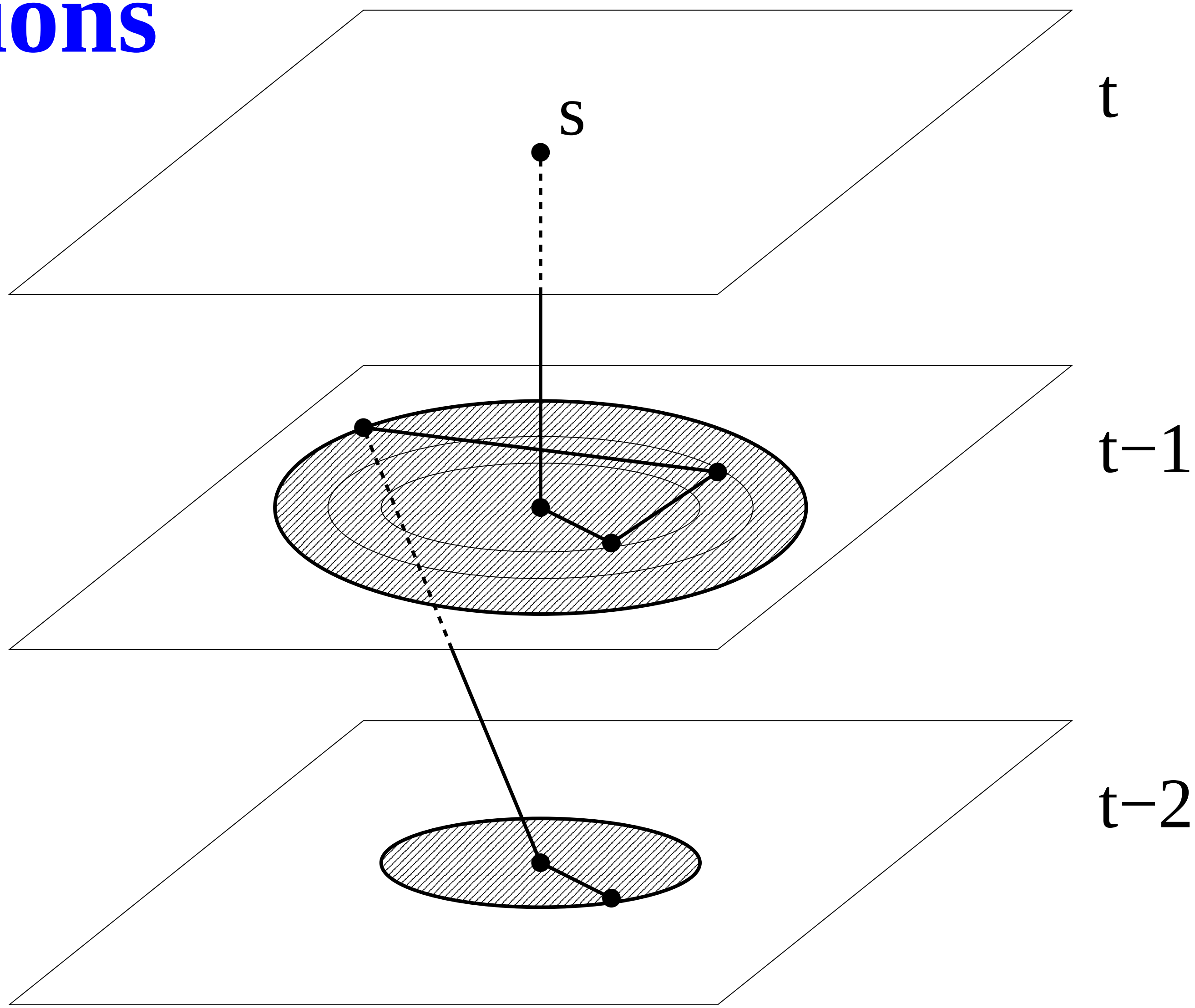
$$z(\mathbf{s}, t-1), z(\mathbf{s}(1), t-1), z(\mathbf{s}(2), t-1), \ldots$$
$$z(\mathbf{s}, t-2), z(\mathbf{s}(1), t-2), \ldots$$

where $\mathbf{s}(1), \mathbf{s}(2), \ldots, \mathbf{s}(N-1)$ is an ordering of the stations, e.g., in ascending order with respect to their distance (using a given metric) to s

- Other ordering motivated by physical knowledge about the underlying process, e.g. Irish wind speed application

- Predictors can be entered in the model sequentially: simplification of the model building stage

Spatio-temporal ordering of the stations



Several strategies to decide on the number of predictors to be used

- Popular technique in time series modeling: partial autocorrelations
Generalization to space-time: look at partial correlation along the ordering of predictors

- Three random variables: x, z, and $\boldsymbol{y}$
The partial correlation of x and z given $\boldsymbol{y}$:

$$\text{Corr}(x, z|\boldsymbol{y}) = \text{Corr}(x - P(x|\boldsymbol{y}), z - P(z|\boldsymbol{y}))$$

where $P(x|\boldsymbol{y})$ is the best linear predictor of $x|\boldsymbol{y}$
This partial correlation has the property that it is zero when x & z are independent conditional on $\boldsymbol{y}$

- Renaming:

$$\begin{aligned} x_1 &= z(\mathbf{s}(1), t-1), \ldots, x_N = z(\mathbf{s}(N), t-1) \\ x_{N+1} &= z(\mathbf{s}(1), t-2), \ldots, x_{2N} = z(\mathbf{s}(N), t-2) \\ &\ldots \end{aligned}$$

Partial correlation function (PCF) for station s:

$$\boxed{\rho_{\mathbf{s}}(h) = \mathrm{Corr}(z(\mathbf{s}, t), x_h | x_1, \ldots, x_{h-1})}$$

- PCF is useful for model selection

Define h_1 to be such that $\rho_{\mathbf{s}}(h_1) \neq 0$ and $\rho_{\mathbf{s}}(h) = 0$ for $h_1 < h \leq N$

Similarly, h_i can be defined for each time lag i, such that $\rho_{\mathbf{s}}(h_i) \neq 0$ and $\rho_{\mathbf{s}}(h) = 0$ for $h_i < h \leq iN$

- Identify the h_i with the sample PCF:

$$\hat{\rho}_{\mathbf{s}}(h) = \widehat{\mathrm{Corr}}(x - \hat{P}(x|y), z - \hat{P}(z|y))$$

- Test for $\rho_{\mathbf{s}}(h) = 0$: under joint normality

$\hat{\rho}_{\mathbf{s}}(h)\sqrt{(n-h)/(1-\hat{\rho}_{\mathbf{s}}(h)^2)}$ is t_{n-h}-distributed

see e.g. Krzanowski (1988, Sec. 14.4)

Strategy

Step 1: Identify h_1 by looking at the sample PCF $\hat{\rho}_{\mathbf{s}}(h)$, $h = 1, \ldots, N$.

Step 2: Identify h_2 by looking at the sample PCF $\hat{\rho}_{\mathbf{s}}(h)$, $h = N+1, \ldots, 2N$ when $x_{h_1+1}, \ldots, x_N$ have been put to zero

Step 3: Identify h_3 by looking at the sample PCF $\hat{\rho}_{\mathbf{s}}(h)$, $h = 2N+1, \ldots, 3N$ when $x_{h_1+1}, \ldots, x_N$,$x_{h_2+1}, \ldots, x_{2N}$ have been put to zero

Step 4: Step 3 is repeated in a similar manner for all necessary time lags in order to identify h_4, h_5, etc.

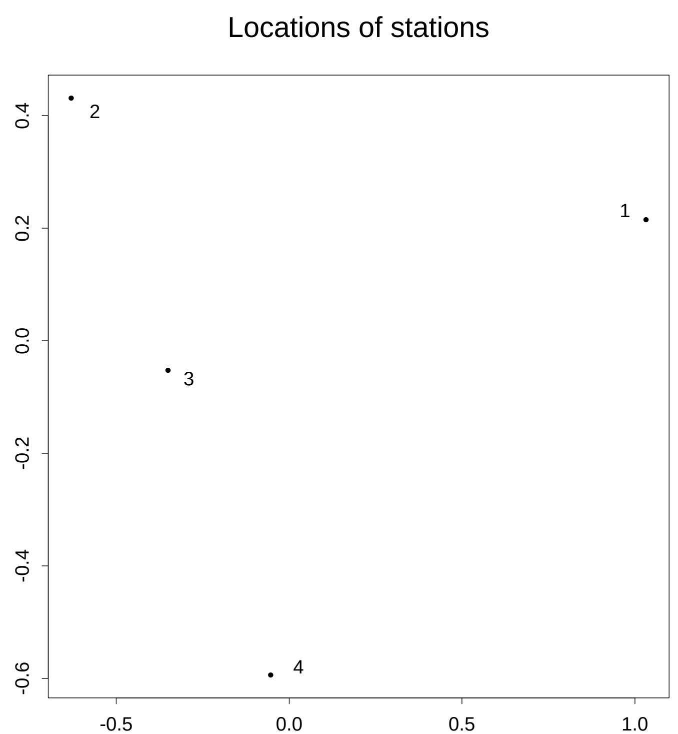
Deletion of uninteresting predictors
Use model selection criterion: AIC, BIC,...

Carbone monoxide in Venice

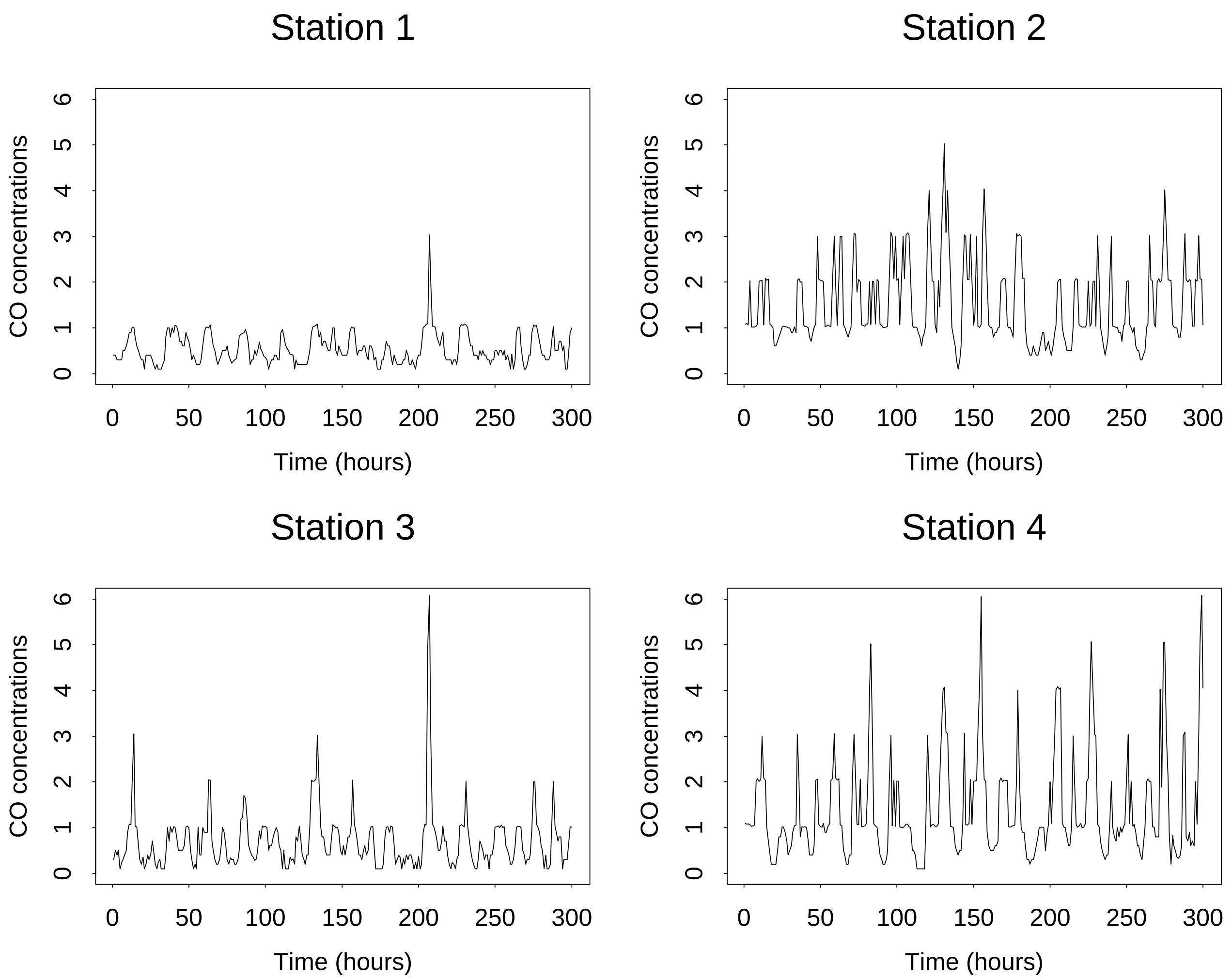
Implementation of the methodology using Splus

$T = 300$ hourly observations of atmospheric concentrations (micrograms per cubic meter) of carbon monoxide (CO) recorded in September 1995 at $N = 4$ monitoring stations located in Mestre (Venice, Italy)

Locations of the 4 stations



Time series of CO concentrations



Stations 2 and 4 are located along streets with high intensity traffic, whereas station 1 is in a garden and station 3 in a pedestrian area

Data set has been analyzed by Tonellato (2001) in a Bayesian dynamic linear model framework

Our methodology is appropriate for this application:

- Italian law requires public authorities to produce short-term forecasts of air pollutant concentrations at locations where monitoring stations are present
- With so few stations, it is not possible to model the spatial dependence adequately, therefore, the use of a spatial stationary isotropic exponential correlation function suggested by Tonellato (2001) seems questionable
- Wind speed and direction can influence air pollutant concentrations in a nonstationary and anisotropic way
- Explanatory variables such as wind and road traffic information are not available

We take the logarithm of the CO concentrations and we estimate a trend for each station by regressing on a family of daily harmonics

Fitted models with BIC:

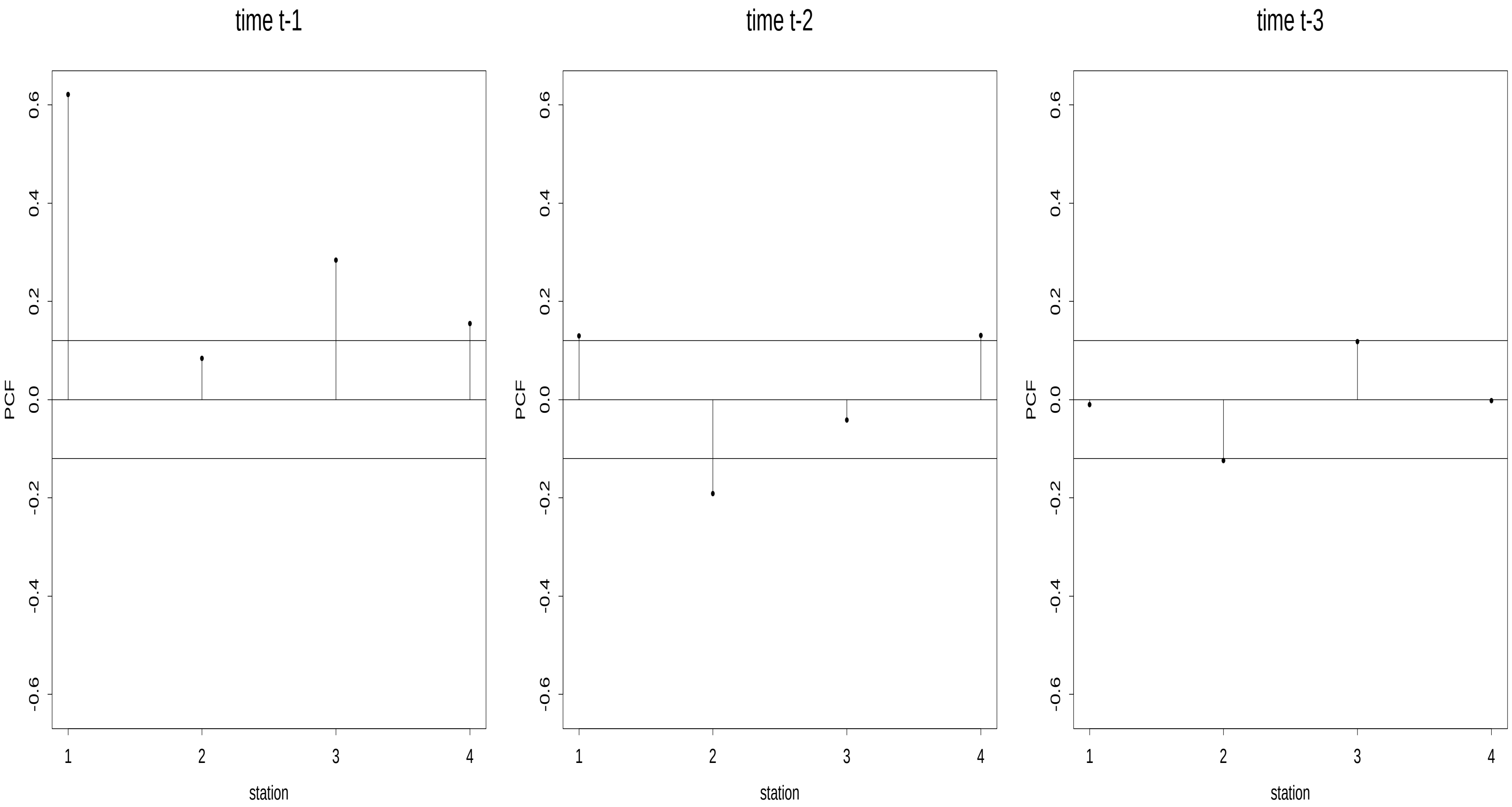
Univariate AR time series models

$\hat{R}_1$	$\hat{R}_2$	$\hat{\Sigma}_\varepsilon$
$\begin{pmatrix} 0.64 & 0 & 0 & 0 \\ 0 & 0.67 & 0 & 0 \\ 0 & 0 & 0.48 & 0 \\ 0 & 0 & 0 & 0.53 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.19 & 0 \\ 0 & 0 & 0 & 0.17 \end{pmatrix}$	$\begin{pmatrix} 0.13 & 0 & 0 & 0 \\ 0 & 0.11 & 0 & 0 \\ 0 & 0 & 0.22 & 0 \\ 0 & 0 & 0 & 0.13 \end{pmatrix}$

Spatial VAR model

$\hat{R}_1$	$\hat{R}_2$	$\hat{\Sigma}_\varepsilon$
$\begin{pmatrix} 0.59 & 0 & 0 & 0.16 \\ 0 & 0.67 & 0 & 0 \\ 0.14 & 0.15 & 0.32 & 0.25 \\ 0 & 0 & 0 & 0.53 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.15 & 0 \\ 0 & 0 & 0 & 0.17 \end{pmatrix}$	$\begin{pmatrix} 0.13 & 0.01 & 0.03 & 0.03 \\ 0.01 & 0.11 & 0.03 & 0.01 \\ 0.03 & 0.03 & 0.20 & 0.04 \\ 0.03 & 0.01 & 0.04 & 0.13 \end{pmatrix}$

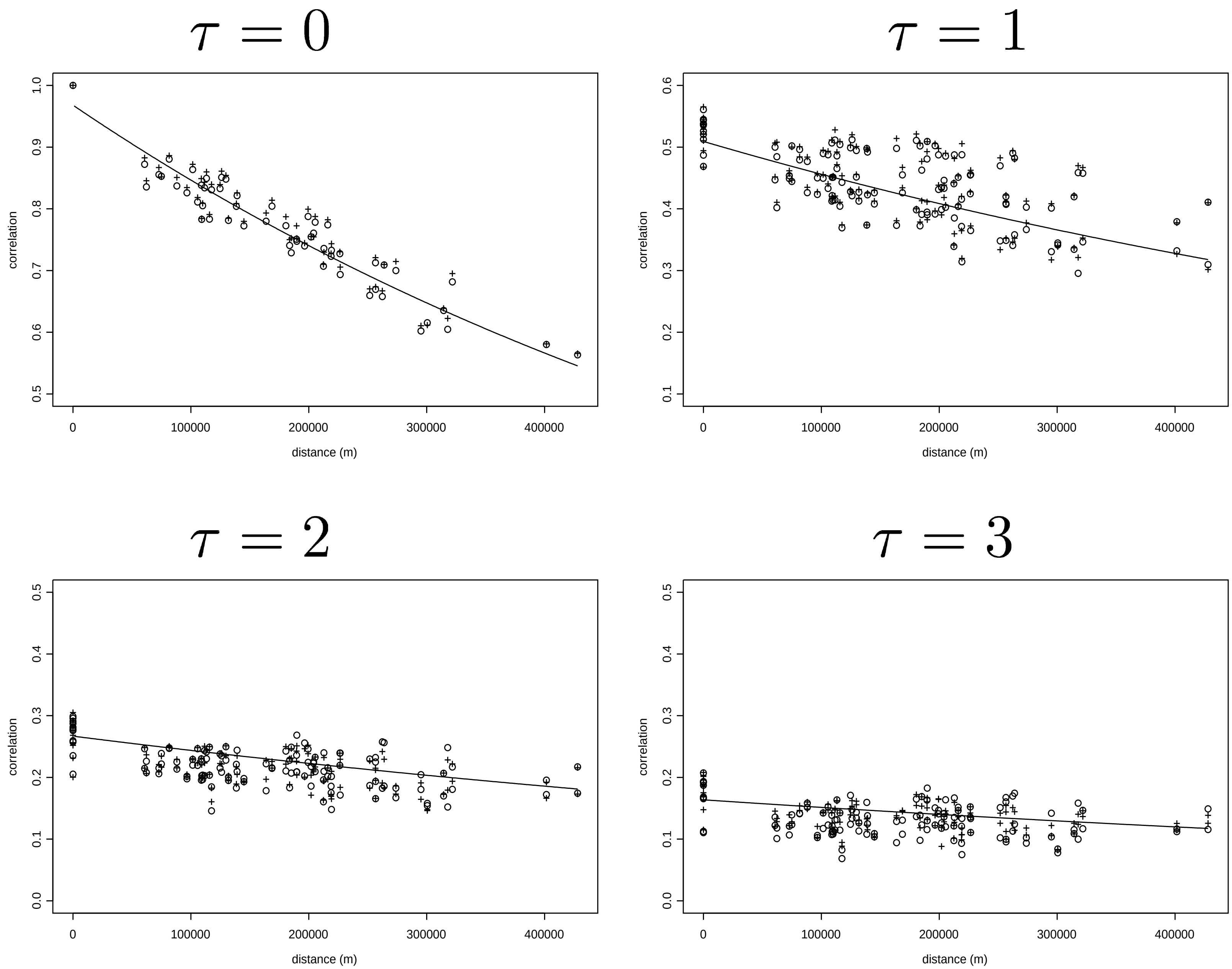
Station 3



PCF for the station 3 up to lag $t - 3$

3. Spatio-temporal correlation analysis

- Irish wind data set analyzed by Haslett and Raftery (1989)
- Daily averages of wind speeds recorded at $N = 11$ meteorological stations in Ireland 1961-1978
- Square root transformation to stabilize the variance
- Subtraction of an estimated seasonal effect and spatially varying mean from the data
- Gneiting (2002): assumption of full symmetry of the spatio-temporal correlation not realistic because wind patterns are predominantly westerly over Ireland. We incorporate this physical information by defining a special ordering of the stations
- With BIC, we fit a spatial VAR(3) model



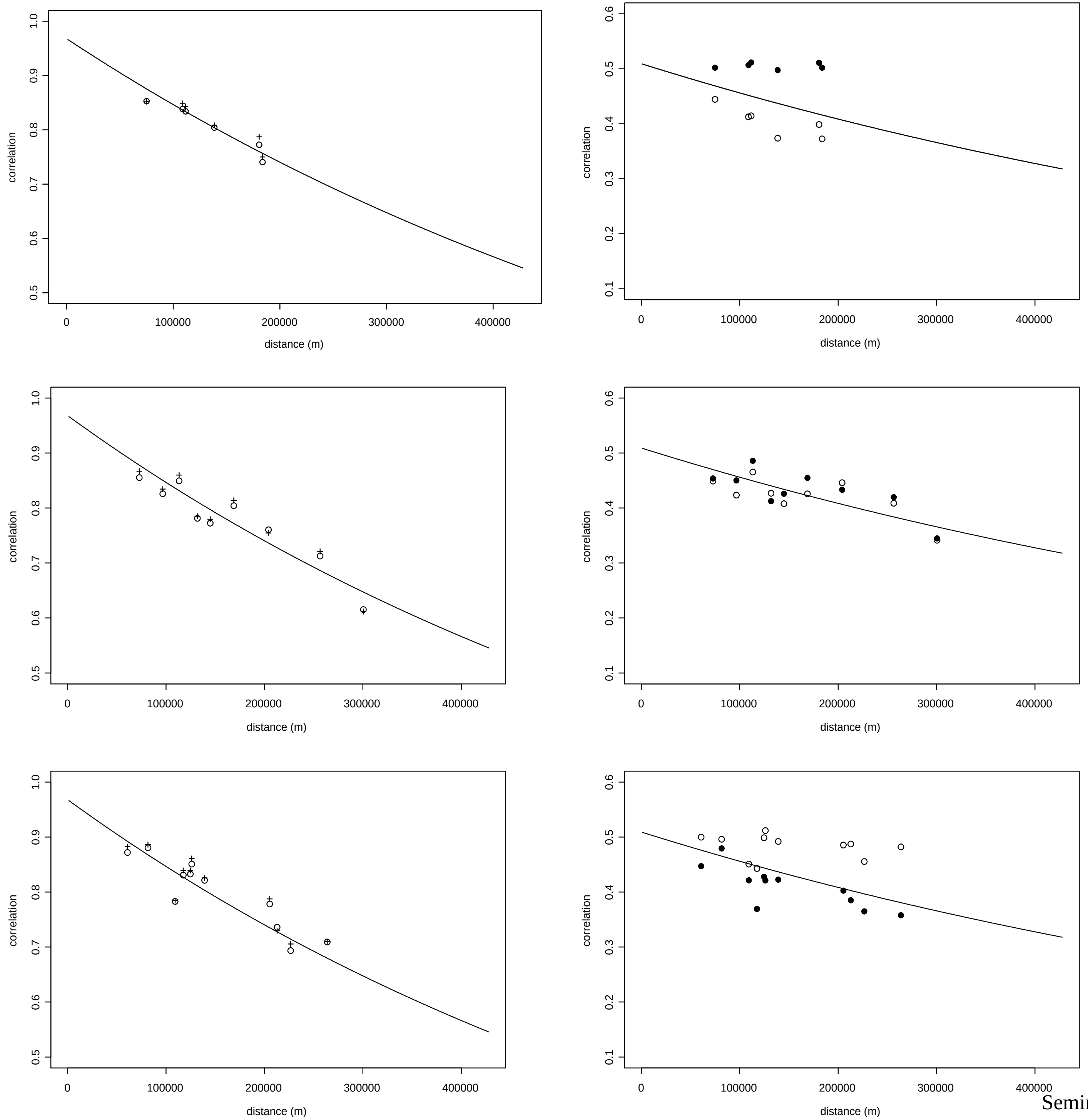
Spatial correlation: from fitted VAR(3) (open circles) and empirical (pluses)

$\tau = 0$ $\tau = 1$

WE

NS

SW-NE



4. Prediction performance analysis

- Comparison of model selection strategies through out-of-sample validation based on recursive prediction errors, see de Luna and Skouras (2003)

- For given $\boldsymbol{s}$, a model selection strategy can be evaluated with the accumulated prediction error criterion:

$$\sum_{t=M}^{T} (z(\boldsymbol{s},t) - \hat{z}^{t-1}(\boldsymbol{s},t))^2$$

where $\hat{z}^{t-1}(\boldsymbol{s},t)$ is the prediction of $z(\boldsymbol{s},t)$ obtained by applying the model selection strategy on the sub-sample $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{t-1}$

- Choosing the model strategy minimizing the above criterion will eventually identify the best strategy with probability one

RAMSE with $M = 1,000$, $T = 6,570$

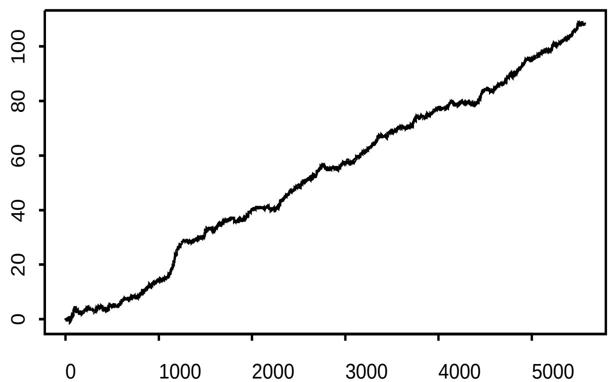
Station	Strategy					
	AR-BIC	AR-AIC	SVAR-BIC		SVAR-AIC	
			dist order	wind order	dist order	wind order
Roche's Pt. (1)	0.492	0.492	0.474	**0.473**	**0.473**	**0.473**
Valentia (2)	0.503	0.503	**0.498**	**0.498**	0.499	0.499
Kilkenny (3)	0.446	0.446	0.424	0.424	**0.423**	0.424
Shannon (4)	0.461	0.460	0.455	0.454	**0.452**	**0.452**
Birr (5)	0.481	0.480	0.470	0.470	0.469	**0.468**
Dublin (6)	0.461	0.461	0.445	0.445	**0.444**	**0.444**
Claremorris (7)	0.488	0.488	0.485	0.485	0.483	**0.482**
Mullingar (8)	0.446	0.445	**0.433**	**0.433**	**0.433**	**0.433**
Clones (9)	0.477	0.477	**0.467**	0.468	**0.467**	**0.467**
Belmullet (10)	0.490	0.490	0.491	0.490	**0.489**	**0.489**
Malin Head (11)	0.499	0.499	0.494	**0.492**	0.493	0.493

For two strategies S_1 and S_2, plot the sums:

$$\sum_{t=M}^{i} \left(\left(z(\boldsymbol{s},t) - z_{S_1}^{t-1}(\boldsymbol{s},t)\right)^2 - \left(z(\boldsymbol{s},t) - z_{S_2}^{t-1}(\boldsymbol{s},t)\right)^2 \right)$$

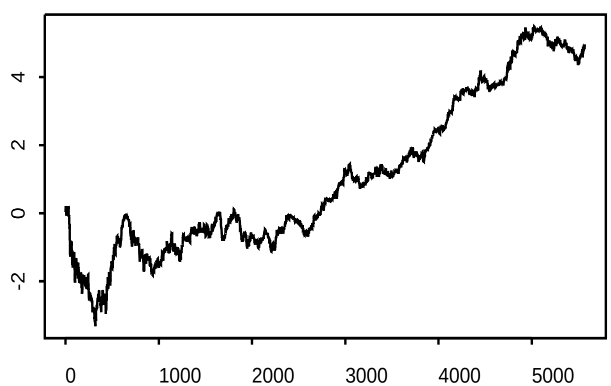
against $i = M, \ldots, T$.

AR-AIC vs SVAR-AIC (dist order)

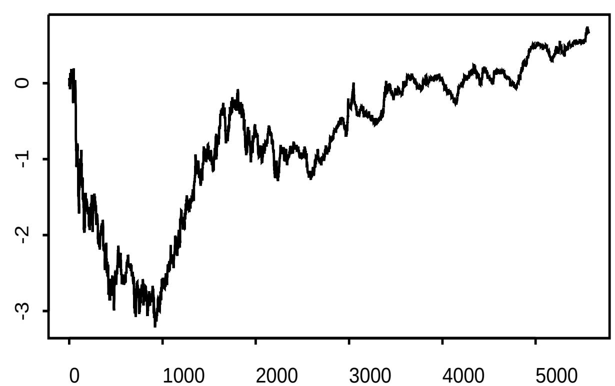


SVAR-BIC (dist order) vs SVAR-AIC (dist order)

SVAR-BIC (wind order) vs SVAR-AIC (dist order)



SVAR-AIC (wind order) vs SVAR-AIC (dist order)



5. Conclusions

- Simple predictive model for nonstationary spatio-temporal data (sparse in space, rich in time)
- Spatio-temporal trend handled by time differencing/weighted sum of known basis functions
- Estimation by least squares, Yule-Walker, or maximum likelihood
- Spatio-temporal ordering of the stations
- Model building strategy with PCF/AIC/BIC
- Each station is modeled separately
- Applications: Carbone monoxide in Venice; Irish wind speed;...
- Future research: comparison with other predictive methods

Predictive spatio-temporal models for spatially sparse environmental data

Xavier de Luna and Marc G. Genton

xavier.deluna@stat.umu.se and genton@stat.ncsu.edu

http://www.stat.umu.se/egna/xdl/index.html
http://www4.stat.ncsu.edu/∼mggenton/

Umeå University

NC State University
DEPARTMENT OF STATISTICS