# A hidden Markov approach to the analysis of space-time environmental data with linear and circular components

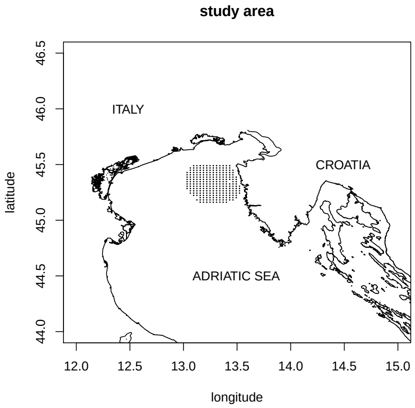
Francesco Lagona (Roma Tre)
Marco Picone (ISPRA)
Antonello Maruotti (Southampton)
Simone Cosoli (OGS)

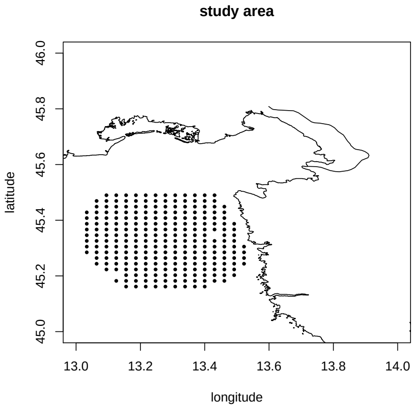
ADISTA 2014 Brussels

# Outline

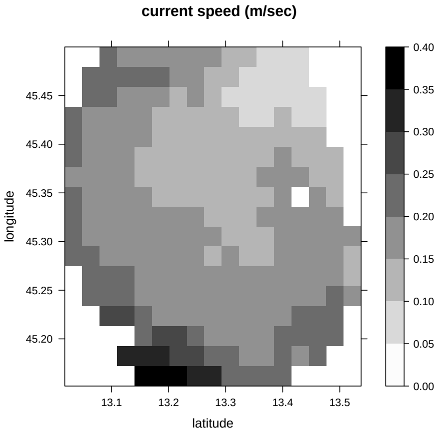
1. **Space-time linear-circular data**
2. **A hidden Markov model**
3. **Maximum likelihood estimation**
4. **Identification of sea regimes**
5. **Discussion**

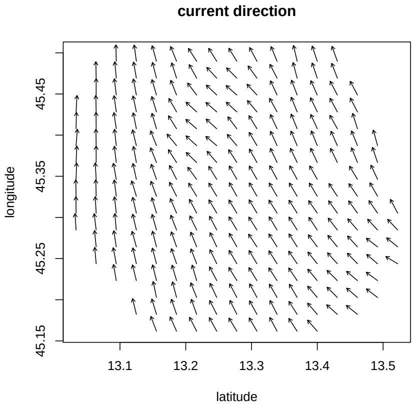
# Marine currents in the Adriatic sea



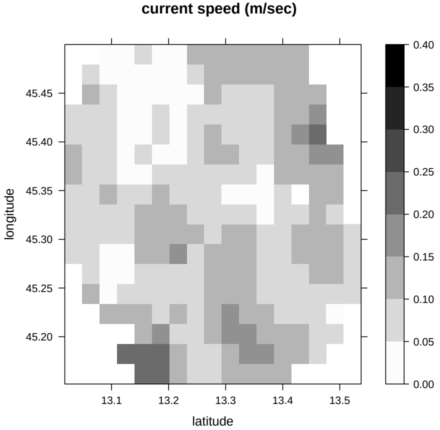



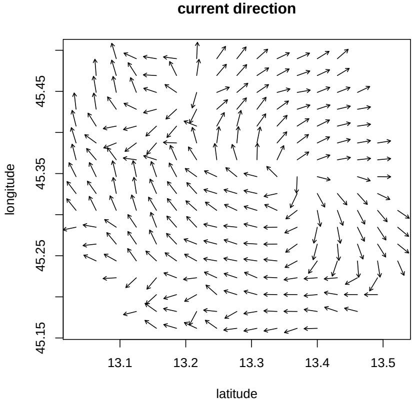
# A Sirocco event



# Calm sea

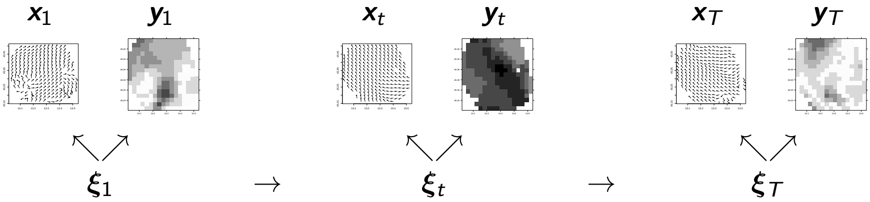

current speed (m/sec)
45.45
45.40
45.35
45.30
45.25
45.20
longitude
13.1
13.2
13.3
13.4
13.5
latitude
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00


current direction
45.45
45.35
45.25
45.15
longitude
13.1
13.2
13.3
13.4
13.5
latitude

# Data issues

- **GOAL**: identification of typical sea regimes
- **APPROACH**: mixture-based clustering
- analysis complicated by:
    - multiple correlation sources (in time, space, between variables)
    - mixed support of the data (linear and circular)
    - special nature of circular data
- literature on space-time linear-circular data is limited:
    - Modlin *et al* 2012, Environmetrics 23, 46-53
    - Wang *et al* 2014, Statistica Sinica (to appear)
    - Lagona *et al* 2014, Stochastic Environmental Research and Risk Assessment (to appear)

# A hidden Markov model (HMM)




- latent process is a Markov chain with $K$ states and parameters $\boldsymbol{\pi} = (\pi_k, \pi_{hk}, h, k = 1 \ldots K)$:

$$p(\boldsymbol{\xi}_{0:T}; \boldsymbol{\pi}) = \prod_{k=1}^{K} \pi_k^{\xi_{0k}} \prod_{t=1}^{T} \prod_{h=1}^{K} \prod_{k=1}^{K} \pi_{hk}^{\xi_{t-1,h}\xi_{tk}}$$

- observation process is a product of linear and circular random fields:

$$f(\boldsymbol{x}_{0:T}, \boldsymbol{y}_{0:T} | \boldsymbol{\xi}_{0:T}) = \prod_{t=0}^{T} \prod_{k=1}^{K} \left( f(\boldsymbol{x}_t | \boldsymbol{\theta}_k^{\mathrm{circ}}) f(\boldsymbol{y}_t | \boldsymbol{\theta}_k^{\mathrm{lin}}) \right)^{\xi_{tk}}$$

## HMM clustering

- find the MLE $\hat{\boldsymbol{\theta}}$ that maximizes the likelihood

$$L(\boldsymbol{\theta}; \boldsymbol{x}_{0:T}, \boldsymbol{y}_{0:T}) = \sum_{\boldsymbol{\xi}_{0:T}} p(\boldsymbol{\xi}_{0:T}; \boldsymbol{\pi}) f(\boldsymbol{x}_{0:T}, \boldsymbol{y}_{0:T} | \boldsymbol{\xi}_{0:T})$$

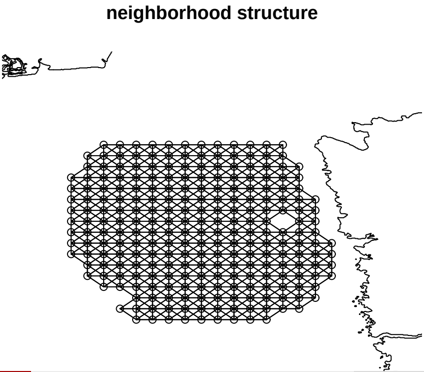
- cluster linear and circular spatial patterns according to the posterior probabilities of class membership

$$\hat{p}_{tk} = P(\xi_{tk} = 1 | \boldsymbol{x}_{0:T}, \boldsymbol{y}_{0:T}; \hat{\boldsymbol{\theta}}) = \mathbb{E}(\xi_{tk} | \boldsymbol{x}_{0:T}, \boldsymbol{y}_{0:T}; \hat{\boldsymbol{\theta}})$$

# Neighborhood structure

- each observation site $i$ is associated with a set of neighbors $N(i)$
- it is defined by a $n \times n$ connectivity matrix $\boldsymbol{C}$

$$c_{ij} = \begin{cases} 1 & j \in N(i) \\ 0 & \text{otherwise} \end{cases}$$



## Gaussian Markov field

- state-specific joint distribution

$$f(\boldsymbol{y}_t; \boldsymbol{\theta}_k^{\text{lin}}) \propto \exp\left(-\frac{1}{2\tau_k^2}(\boldsymbol{y} - \boldsymbol{\mu}_k^{\text{lin}})^{\mathsf{T}}(\boldsymbol{I} - \rho_k \boldsymbol{C})(\boldsymbol{y} - \boldsymbol{\mu}_k^{\text{lin}})\right)$$

- the conditional univariate distribution is Markov w.r.t. the neighborhood structure:

$$f(y_{it}|y_{jt}, j \neq i) \propto \exp\left(-\frac{1}{2\tau_k^2}\left(y_{it} - \mu_{ik}^{\text{lin}} - \rho_k \sum_{j \in N(i)} \left(y_{jt} - \mu_{jk}^{\text{lin}}\right)\right)^2\right)$$

## Von Mises Markov field

- state-specific joint distribution

$$f(\boldsymbol{x}_t; \boldsymbol{\theta}_k^{circ}) \propto$$

$$\exp\left(\kappa_k \sum_{i=1}^{n} \cos(x_{it} - \mu_{ik}^{\mathrm{circ}}) + \frac{\lambda_k}{2} \sum_{i=1}^{n} \sin(x_{it} - \mu_{ik}^{\mathrm{circ}}) \sum_{j \in N(i)} \sin(x_{jt} - \mu_{jk}^{\mathrm{circ}})\right)$$

- the conditional univariate distribution is Markov w.r.t. the neighborhood structure:

$$\begin{aligned} f(x_{it} \mid x_{1t} \ldots x_{i-1,t}, x_{i+1,t} \ldots x_{nt}; \boldsymbol{\theta}_k^{\mathrm{circ}}) &= f_{\mathrm{vm}}(x_{it}; \nu_{ik}, \kappa_{ik}) \\ \kappa_{ik} &= \sqrt{\kappa_k^2 + \lambda_k^2 \tilde{s}_{ik}^2} \\ \nu_{ik} &= \mu_{ik}^{\mathrm{circ}} + \arctan\left(\lambda_k \frac{\tilde{s}_{ik}}{\kappa_k}\right) \\ \tilde{s}_{itk} &= \sum_{j \in N(i)} \sin(x_{jt} - \mu_{jk}^{\mathrm{circ}}) \end{aligned}$$

# EM algorithm

- data: $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y})$
- an algorithm based on the complete-data log-likelihood

$$\log L_{\mathrm{comp}}(\boldsymbol{\theta}, \boldsymbol{\xi}_{0:T}, \boldsymbol{z}_{0:T}) = \sum_{k=1}^{K} \xi_{0k} \log \pi_k + \sum_{t=1}^{T} \sum_{h=1}^{K} \sum_{k=1}^{K} \xi_{t-1,h} \xi_{t,k} \log \pi_{hk}$$
$$+ \sum_{t=0}^{T} \sum_{k=1}^{K} \xi_{tk} \log f(\boldsymbol{x}_t; \boldsymbol{\theta}_k^{\mathrm{circ}})$$
$$+ \sum_{t=0}^{T} \sum_{k=1}^{K} \xi_{tk} \log f(\boldsymbol{y}_t; \boldsymbol{\theta}_k^{\mathrm{lin}})$$

# E step

- given the estimate $\hat{\boldsymbol{\theta}}_s$, evaluate

$$
\begin{aligned}
Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}_s) = & \sum_{k=1}^{K} \mathbb{E}(\xi_{0k}|\boldsymbol{z}_{0:T}, \hat{\boldsymbol{\theta}}_s) \log \pi_k \\
& + \sum_{t=1}^{T} \sum_{h=1}^{K} \sum_{k=1}^{K} \mathbb{E}(\xi_{t-1,h}\xi_{tk}|\boldsymbol{z}_{0:T}, \hat{\boldsymbol{\theta}}_s) \log \pi_{h,k} \\
& + \sum_{t=0}^{T} \sum_{k=1}^{K} \mathbb{E}(\xi_{tk}|\boldsymbol{z}_{0:T}, \hat{\boldsymbol{\theta}}_s) \log f(\boldsymbol{x}_t; \boldsymbol{\theta}_k^{\mathrm{circ}}) \\
& + \sum_{t=0}^{T} \sum_{k=1}^{K} \mathbb{E}(\xi_{tk}|\boldsymbol{z}_{0:T}, \hat{\boldsymbol{\theta}}_s) \log f(\boldsymbol{y}_t; \boldsymbol{\theta}_k^{\mathrm{lin}}).
\end{aligned}
$$

# M step

- M-step: maximizes $Q$ and obtain
  - the Markov chain probabilities

$$\hat{\pi}_{hk(s+1)} = \frac{\sum_{t=1}^{T} \hat{p}_{t-1,t,hk}(\hat{\boldsymbol{\theta}}_s)}{\sum_{t=1}^{T} \hat{p}_{t-1,h}(\hat{\boldsymbol{\theta}}_s)}, \qquad h, k = 1, \ldots, K.$$

  - the state-specific Gaussian MRF parameters $\boldsymbol{\theta}_k^{\mathrm{lin}}$, via the estimation of a spatial conditional autoregressive model with case weights (R library `spdep`)
  - the state-specific von Mises MRF parameters $\boldsymbol{\theta}_k^{\mathrm{circ}}$, by maximizing the pseudo-loglikelihood (Mardia *et al* 2008, Canadian Journal of Statistics 36)

$$\mathrm{pl}_k(\boldsymbol{\theta}^{\mathrm{circ}} \mid \boldsymbol{x}_t) = \sum_{i=1}^{n} \log f_{\mathrm{vm}}(x_{it}; \nu_{ik}, \kappa_{ik})$$

## large scale variation

- a linear spatial trend

$$\mu_{ik}^{\text{lin}} = \mu_{k}^{\text{lin}} + (X_{i1} - \bar{X}_1)\beta_{1k}^{\text{lin}} + (X_{i2} - \bar{X}_2)\beta_{2k}^{\text{lin}}$$

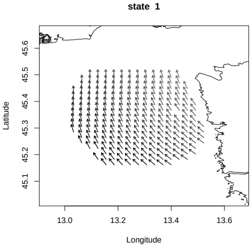
- a circular spatial trend

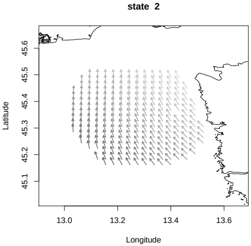
$$\mu_{ik}^{\text{circ}} = \mu_{k}^{\text{circ}} + 2\arctan\left((X_{i1} - \bar{X}_1)\beta_{1k}^{\text{circ}} + (X_{i2} - \bar{X}_2)\beta_{2k}^{\text{circ}}\right).$$

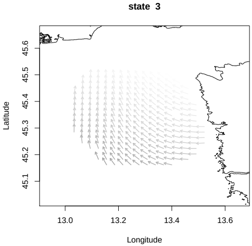
# Estimates

| | | State 1 | | State 2 | | State 3 | |
|---|---|---|---|---|---|---|---|
| | | estimate | s. e. | estimate | s.e. | estimate | s.e. |
| Direction | Mean | 2.056 | 0.012 | 1.969 | 0.289 | 2.290 | 0.254 |
| | Longitude | 0.566 | 0.029 | 0.765 | 0.046 | 1.742 | 0.061 |
| | Latitude | -0.934 | 0.010 | -0.498 | 0.029 | -0.618 | 0.031 |
| | Spatial dependence | 8.020 | 0.024 | 2.047 | 0.011 | 0.687 | 0.009 |
| | Spatial concentration | 45.988 | 2.123 | 9.847 | 1.451 | 1.940 | 0.128 |
| Log-speed | Mean | -1.670 | 0.122 | -2.078 | 0.430 | -2.399 | 0.771 |
| | Longitude | -0.469 | 0.032 | -0.714 | 0.051 | -0.441 | 0.022 |
| | Latitude | -0.804 | 0.032 | -1.285 | 0.099 | -1.579 | 0.081 |
| | Spatial dependence | 0.128 | 0.055 | 0.128 | 0.042 | 0.128 | 0.031 |
| | Spatial variance | 0.014 | 0.004 | 0.042 | 0.011 | 0.082 | 0.020 |
| Transition | | | | Destination states | | | |
| probabilities | Origin states | State 1 | | State 2 | | State 3 | |
| | | estimate | s. e. | estimate | s.e. | estimate | s.e. |
| | State 1 | 0.796 | 0.002 | 0.195 | 0.001 | 0.009 | 0.001 |
| | State 2 | 0.177 | 0.003 | 0.613 | 0.004 | 0.210 | 0.005 |
| | State 3 | 0.004 | 0.001 | 0.208 | 0.006 | 0.788 | 0.001 |

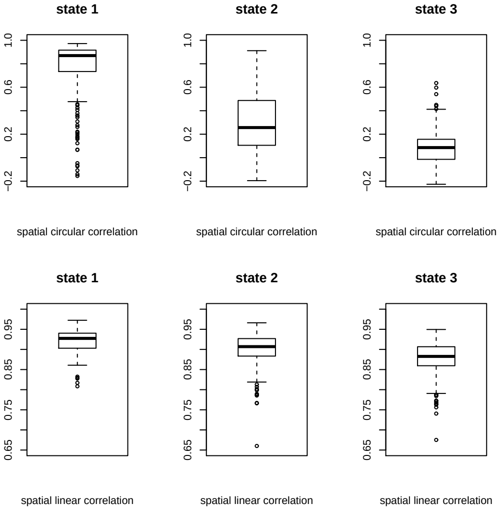
# large scale variation



# small scale variation



## advantages

- the model captures several sources of heterogeneity:
  - time-dependence, through the hidden Markov chain
  - association between spatial patterns of circular and linear measurements, through latent classes
  - large-scale spatial variation, through pairs of circular and linear spatial gradients
  - small-scale spatial variation, through pairs of circular and linear spatial auto-correlation parameters
- it is relatively easy to estimate, under a likelihood-based setting
- intuitively appealing meaning of all the parameters

## limitations

- the model depends on prior neighborhood structures and a prior number of latent classes:
  - model selection (e.g., via BIC) can be problematic
  - incorporation of model order in the EM leads to computationally intensive algorithms
- the model depends on a homogeneous Markov chain
  - reasonable in short period of time
- it depends on homogeneous random Markov fields
  - reasonable in areas of moderate size

many thanks !
questions?