Atelier RESSTE
– Estimation d'une fonction de covariance –

Thomas OPITZ
avec Nicolas Desassis

Atelier RESSTE, Avignon, 27-29 avril 2016

28/04/2016

Modèle gaussien spatiotemporel

champ gaussien spatiotemporel

$$Z_t(x) = \mu_t(x) + \sigma_t(x) Z_t^*(x), \quad x \in \mathbb{R}^2, t \geq 0$$

où

- moyenne $\mu_t(x \mid \theta_\mu)$,
- variance $\sigma_t(x \mid \theta_\sigma)$,
- $Z_t^*(x)$ champ gaussien standard ($\mathbb{E}Z_t^*(x) = 0$, $\mathbb{V}(Z_t^*(x)) = 1$), fonction de corrélation $\mathrm{Cor}((x_1, t_1), (x_2, t_2))$

Pour ajuster un modèle, nous allons faire des **hypothèses de stationnarité (espace/temps), d'isotropie, de famille paramétrique, ...**

Densité multivariée et vraisemblance

Pour une observation $\boldsymbol{x} = (x_1, \ldots, x_d)$ d'un vecteur gaussien $Z \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ de dimension d, la densité de probabilité est

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = |\Sigma|^{-1/2} (2\pi)^{-d/2} \exp\left(-0.5(\boldsymbol{x} - \boldsymbol{\mu})\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right),$$

où $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta}_\mu), \Sigma = \Sigma(\boldsymbol{\theta}_\Sigma)$ avec un vecteur de paramètres $\boldsymbol{\theta} = (\boldsymbol{\theta}_\mu, \boldsymbol{\theta}_\Sigma)$.

Vraisemblance : $\boldsymbol{\theta} \mapsto f_{\boldsymbol{\theta}}(\boldsymbol{z})$
$\rightsquigarrow$ maximum de vraisemblance $\boldsymbol{\theta}^* = \arg\max_\theta f_{\boldsymbol{\theta}}(\boldsymbol{z})$

Si $d \gg 1$, le **coût de calcul numérique** de $|\Sigma|$ et de Σ^{-1} à partir de Σ peut être **très important voire prohibitif**.

Ici : $d = \#\{\text{sites d'observation}\} \times \#\{\text{temps d'observation}\}$

Fonction de covariance C et semi-variogramme γ

Semi-variogramme spatiotemporel

$$\gamma((s_1, t_1), (s_2, t_2)) = 0.5\mathbb{E}\left(Z(s_1) - Z(s_2)\right)^2$$

si champ stationnaire :

$$\gamma((s_1, t), (s_2, t_2)) = \gamma(s_2 - s_1, t_2 - t_1) = \sigma^2 - C(s_2 - s_1, t_2 - t_1)$$

où $\sigma^2 = C((0,0),0)$

Notations :
$\Delta s = s_2 - s_1$, $\Delta t = t_2 - t_1$, $\gamma(\Delta s, \Delta t) = \gamma_{\Delta t}(\Delta s)$, $C(\Delta s, \Delta t) = C_{\Delta t}(\Delta s)$

Modèles paramétriques de covariance spatiotemporelle

- modèles spatiaux stationnaires et isotropes
  $C(x_1, x_2) = C(\Delta x) = C(\|\Delta x\|) = C(h)$
  anisotropie géométrique si $C(\Delta x) = C_{\text{iso}}(\|M\Delta x\|)$ avec matrice d'anisotropie A
- modèles s.t. séparables :

$$C_{\Delta t}(\Delta x) = C_S(\Delta x) \times C_T(\Delta t) = \theta_{\text{sill}} C_S^*(\Delta x) \times C_T^*(\Delta t)$$

  (anisotropie spatiale possible)
- product-sum (nonséparable) :

$$C_{\Delta t}(\Delta x) = kC_S(\Delta x)C_T(\Delta t) + C_S(\Delta x) + C_T(\Delta t), \quad k > 0$$

  trois paramètres de palier identifiables (C_S, C_T, global)
  (anisotropie spatiale possible)
- modèles s.t. spécifiques :
  - classe de Gneiting
  - classe Porcu
  - ...

Le modèle de Gneiting

[Gneiting, 2002]

$$C_t(\boldsymbol{x}) = \sigma^2 g_T(t)^{-1} \exp\left(-\frac{d_S(\boldsymbol{h})}{g_T(t)^{-0.5\eta K_S}}\right)$$

où

- $0 \leq \eta \leq 1$ paramètre de nonséparabilité (séparable si $\eta = 0$)
- $d_T(t) = (|t|/\psi_T)^{\kappa_T}$ variogramme puissance (temporel)
- $d_S(\boldsymbol{h}) = (\|\boldsymbol{h}\|/\psi_T)^{\kappa_S}$ variogramme puissance (spatial)
- $g_T(t) = 1 + d_T(t)$, $g_S(\boldsymbol{h}) = 1 + d_S(\boldsymbol{h})$

Certaines généralisations sont possible :

$$C_t(x) = (\psi(t) + 1)^{-\delta/2} \varphi(h/\sqrt{(\psi(t) + 1)}),$$

où $\psi(\cdot)$ variogramme, $\varphi(\cdot)$ covariance (mélange de covariance gaussienne)

Valeurs à estimer ou à fixer

Hypothèse : Champ spatiotemporel stationnaire

- moyenne $\theta_{\text{mean}} = \mu(s,t) \equiv \mu_0(o)$
- palier $\theta_{\text{sill}} = C_t(s) = C_0(o)$
- pépite ou erreur de mesure $\theta_{\text{nugget}} = C_0(o) - \lim_{t\to 0, \|x\|\to 0} C_t(x)$
- échelle θ_{scale} : $\Delta s \rightsquigarrow \Delta s/\theta_{\text{scale}}$
- anisotropie géométrique : $\Delta s \rightsquigarrow \begin{pmatrix} b & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(a) & -\sin(a) \\ \sin(a) & \cos(a) \end{pmatrix} \Delta s$
  - angle de rotation $a \in [0, \pi)$
  - "étirement" $b > 0$
- paramètres de forme de la fonction de covariance, ...

Estimation empirique de θ_{mean} et de θ_{sill} :

- $\hat{\theta}_{\text{mean}} = \frac{1}{d} \sum_{(s_i,t_i)} x_{s,t}$, où $d = \#\{(s_i, t_i)\}$
- $\hat{\theta}_{\text{sill}} = \frac{1}{d} \sum_{(s_i,t_i)} \left(x_{s,t} - \hat{\theta}_{\text{mean}} \right)^2$

Estimation de la covariance

Techniques :

- **estimation empirique** : nuée variographique
  + lissage nonparamétrique (LOESS etc.) ou classes de distance
  $\rightsquigarrow$ **variogramme empirique**
  $\rightsquigarrow$ visualisation spatiotemporelle ?
  e.g., "variogramme spatial" pour $\Delta t = 0, 1, ..., t_{\max}$

- **estimation paramétrique**
  - méthode des moments, moindres carrés pondérés
  - maximum de vraisemblance
  - maximum de vraisemblance composite, *e.g.* vraisemblance par paires

Moindres carrés pondérés

classes de distance $D_k = \{(s_i, t_i) : \mathrm{dist}((s_i, t_i), (\tilde{s}_k, \tilde{t}_k)) \leq \varepsilon_k\}$, $k = 1, \ldots, K$
(typiquement, grille spatio-temporelle des barycentres des classes)

$$\boldsymbol{\theta}^* = \arg\min_{\theta} \sum_{k=1}^{K} \omega_k \left| \gamma_{\boldsymbol{\theta},k} - \hat{\gamma}_k \right|^2$$

- $\omega_k = 1$ moindres carrés ordinaires
- $\omega_k = |D_k| / \gamma^2_{\boldsymbol{\theta},k}$: tenir compte de la variance de $\hat{\gamma}_k$
- surpondérer distances s.t. faibles, ...

Maximum de vraisemblance

$$\boldsymbol{\theta}^* = \arg\max_{\theta} f_{\boldsymbol{\theta}}(\boldsymbol{z})$$

- solution numérique itérative si contraintes paramétriques
- **tapering** [Kaufman et al., 2008] pour réduire la complexité numérique des calculs
  - utiliser $\tilde{C}_t(x) = C_t(\boldsymbol{x}) \times C_{0,t}(x)$ où $C_{0,t}$ est à support compact (*Wendland* etc.)
    $\rightsquigarrow$ matrice de variance-covariance "creuse" avec peu d'entrées non-zéro
  - peut considérablement faciliter et accélerer les calculs numériques
  - cependant, souvent la performance est décevante en pratique [Stein, 2013]

Maximum de vraisemblance composite

[Lindsay, 1988, Varin et al., 2011]

souvent basé sur les lois bivariées : **vraisemblance par paires**

$$\mathrm{PL}(\theta; \boldsymbol{z}) = \prod_{i_1, i_2} f_{\theta, i_1, i_2}(z_{i_1}, z_{i_2})$$

maximum de vraisemblance par paires : $\boldsymbol{\theta}^* = \arg\max_{\theta \in \Theta} \mathrm{PL}(\theta; \boldsymbol{z})$

- $f_{\theta, i_1, i_2} \in$ densité de
  $(Z_{t_1}(x_1), Z_{t_2}(x_2))$ ou $Z_{t_1}(x_1) - Z_{t_2}(x_2)$ ou $Z_{t_1}(x_1) \mid Z_{t_2}(x_2)$
- réduction de la complexité de calcul en ne considérant que des paires "proches" en espace et en temps
  $\rightsquigarrow$ peu de biais supplémentaire, voire réduction de biais dans certains cas
- consistance et limite normale asymptotiques
- blocs plus grands possibles,
  par exemple spatial-full et temporel-composite

Estimation de la moyenne

Comment estimer $\mu_t(x)$?

- effets saisonniers
- effets spatiaux

Il est souvent utile d'estimer $\mu_t(x)$ dans un premier temps, et ensuite estimer la fonction de covariance à partir des données centrées $Z_t(x) - \mu_t(x)$

⚠ **confusion de moyenne et dépendance** possible si approche (semi-) paramétrique

Modèles courants :

- séparation $\mu_t(x) = \mu_t \times \mu(x)$
- régression avec trends linéaires, splines
- régression locale $\mu(s) \sim x + y$, e.g. `LOESS`
- ...

Sélection de modèle

Souvent, nous devons sélectionner un modèle parmi plusieurs modèles ajustés.

Méthodes de sélection de modèle :

- comparaison visuelle ajusté vs. empirique (variogramme, fn. de covariance, ...)
- validation croisée (enlever une partie des données pour l'ajustement, prédire ces données)
- vraisemblance composite : CLIC (*Composite Likelihood Information Criterion*)
- tests d'hypothèse statistique pour modèles emboîtés, adaptés à la vraisemblance composite

Packages pour modèles gaussiens spatiotemporels

- accent sur l'**estimation ponctuelle** :
  - `gstat` : modèles s.t. séparables et product-sum, variogramme empirique s.t., moindres carrés pondérés s.t.
  - `CompRandFld` : modèles s.t. séparables et nonséparables, variogramme empirique s.t., vraisemblance et vraisemblances par paires s.t., moindres carrés pondérés s.t., tapering s.t., pas d'anisotropie géométrique
  - `RandomFields` : large gamme de modèles séparables et nonséparables, variogramme empirique s.t., vraisemblance (s.t. ?), moindres carrés pondérés (s.t. ?)
  - peu de fonctionnalité s.t. : `rgeos`, `Rgeostat`, ...
- **modélisation bayésienne** (pas considérée ici) :
  - `INLA` (`www.r-inla.org`) : modèles hiérarchiques, modèle de covariance spatiale markovienne solution d'une EDP stochastique, autorégression en t, approximations analytiques de Laplace dans la densité a posteriori [Rue et al., 2009, Lindgren et al., 2011]
  - `spate` : modèle gaussien s.t. solution d'une EDP stochastique, inférence MCMC [Sigrist et al., 2015]
  - `spBayes`, ...

Fonctionnalité de gstat

- `variogram(...)` pour calculer un variogramme empirique (s.t.)

```
variogram(formula, data, width, cutoff, tlags)
```

- `vgm(...)` pour définir un modèle paramétrique de variogramme spatial

```
vgm(psill, model, range, nugget, anis, kappa = 0.5)
```

où `model` $\in$ "Exp", "Sph", "Gau", "Mat", ...
`vgm()` rend la liste des modèles disponibles

- `vgmST(...)` pour définir un modèle paramétrique de variogramme s.t.

```
vgmST(stModel, space, time, sill, nugget)
```

où
  - `stModel` $\in$ "separable", "productSum", ...
  - `space` et `time` sont des objets `vgm(...)`

- `fit.variogram(...)` pour ajuster un modèle de variogramme spatial
- `fit.StVariogram(...)` pour ajuster un modèle de variogramme s.t.
- `krigeST(...)` pour le krigeage spatiotemporel

Fonctionnalité de RandomFields

syntaxe plutôt technique,
pas beaucoup de documentation sur l'estimation s.t.
(implémentation de l'estimation paramétrique pas encore aboutie (?))

- `RFempiricalvariogram(...)` pour variogramme empirique
- `?RMmodel` modèles de covariance et syntaxe
  Exemple : modèle de Gneiting $C_t(x) = (\psi(t)+1)^{-\delta/2}\varphi(h/\sqrt{(\psi(t)+1)})$,
  où $\psi(\cdot)$ variogramme, $\varphi(\cdot)$ covariance (mélange de covariance gaussienne)

```
model <- RMnsst(phi=RMgauss(), psi=RMfbm(alpha=1), delta=2)
x <- seq(0,10,0.25)
plot(model, dim=2)
plot(RFsimulate(model, x=x, y=x))
```

- `?RMmodelsAdvanced` modèles de covariance s.t. (entre autres)
- `RFsim(...)` pour simuler
- `RFfit(...)` pour ajuster des modèles paramétriques
- `RFinterpolate(...)` krigeage
- `?fitgauss` information sur les méthodes d'ajustement

Fonctionnalité de CompRandFld

- modèles de covariance paramétriques
  - types : ...
  - `Covariogram(...)`, `Covmatrix(...)` calcul, tapering, visualisation
  - `RFsim` simulation
- **estimation**
  - `EVariogram(...)` empirical space(-time) variogram
  - estimation de modèles paramétriques
    - `WLeastSquare(...)` moindres carrés pondérés
    - `FitComposite(...)` vraisemblance (classique ou par paires)
    - `HypoTest(...)` pour tester modèles emboîtés à partir de fits avec `FitComposite(...)`
- `Kri(...)` krigeage (tapering possible)

⚠⚠⚠

- **pas d'anisotropie géométrique spatiale**
- **support des données : produit cartésien espace $\times$ temps**
- **pas de possibilité de données manquantes**

Résumé packages R

- peu de compatibilité entre packages concernant
  - représentation des données
  - types d'objets R utilisés
  - disponibilité et paramétrisation des modèles de covariance
- estimation spatiotemporelle avec `RandomFields::RFfit(...)` – est-ce possible ?
- gestion des paramètres pas toujours claires
  e.g., ordre des paramètres dans `gstat`
- anisotropie géométrique spatiale
  - `RandomFields` : fonctionnalité très complète
  - `gstat` : estimation partielle
  - `CompRandFld` : pas implémentée
- estimation d'un palier nonstationnaire ?

Quelques rappels sur la structure des données

- données d'observation (PM10, O3, ...) sur stations espacées de façon irrégulière
  - Europe, $\approx$ 400 stations en France
  - données horaires et journalières
  - ⚠ beaucoup de données manquantes
    ⚠ censure informative ?
  - stations "background" vs. "traffic", "industry", ...
- prédictions CHIMERE journalières sur grille avec mailles $50km \times 50km$, issues d'un krigeage spatial des prédictions temporelles site-par-site
- ici, pas de covariables explicatives (météo, urbanisation, activité industrielle, ...)

Pour éviter le traitement de fortes nonstationnarités et ruptures locales, nous écartons les stations non-background.

Modélisation proposée

Modèle gaussien s.t. de correction d'erreur pour

$$\text{observation}(t) - \texttt{prédiction CHIMERE}(t)$$

- ainsi, on garde le modèle de transport CHIMERE pour la moyenne du processes

- ⚠ différents supports spatiaux
  $\rightsquigarrow$ interpolation des données CHIMERE aux sites d'observation

Questions ?


Gneiting, T. (2002).
Nonseparable, stationary covariance functions for space–time data.
Journal of the American Statistical Association, 97(458) :590–600.


Kaufman, C. G., Schervish, M. J., and Nychka, D. W. (2008).
Covariance tapering for likelihood-based estimation in large spatial data sets.
Journal of the American Statistical Association, 103(484) :1545–1555.


Lindgren, F., Rue, H., and Lindström, J. (2011).
An explicit link between Gaussian fields and Gaussian Markov random fields : the stochastic partial differential equation approach.
Journal of the Royal Statistical Society : Series B (Statistical Methodology), 73(4) :423–498.


Lindsay, B. G. (1988).
Composite likelihood methods.
Contemporary mathematics, 80(1) :221–39.


Rue, H., Martino, S., and Chopin, N. (2009).
Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations.
Journal of the royal statistical society : Series b (statistical methodology), 71(2) :319–392.


Sigrist, F., Künsch, H. R., and Stahel, W. A. (2015).
Stochastic partial differential equation based modelling of large space–time data sets.
Journal of the Royal Statistical Society : Series B (Statistical Methodology), 77(1) :3–33.


Stein, M. L. (2013).
Statistical properties of covariance tapers.
Journal of Computational and Graphical Statistics, 22(4) :866–885.


Varin, C., Reid, N., and Firth, D. (2011).
An overview of composite likelihood methods.
Statistica Sinica, pages 5–42.