gaetan@unive.it

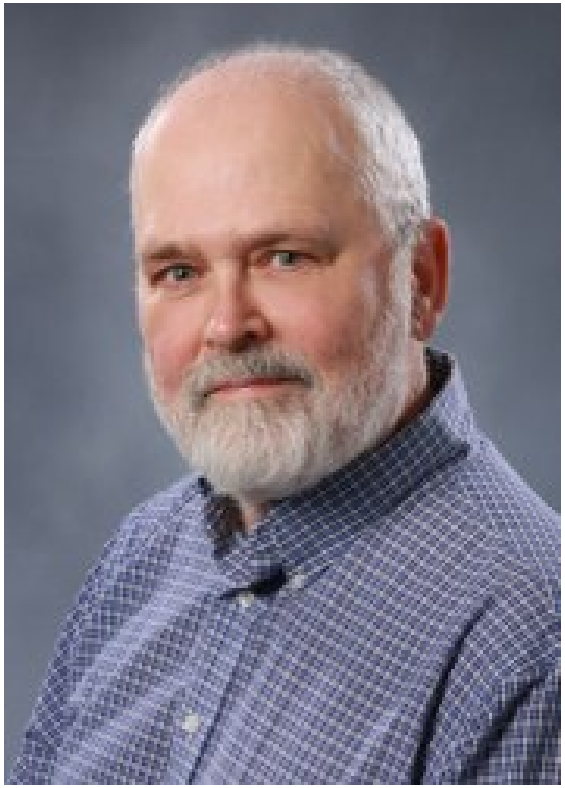
Composite likelihood for space-time data[1]

Carlo Gaetan

Paris, May 11th, 2015

[1] I thank my colleague Cristiano Varin for sharing ideas, slides and office with me.

Bruce Lindsay



March 7, 1947 - May 5, 2015

Lindsay, B.G. (1988). Composite Likelihood Methods. *Contemporary Mathematics* **80**, 221-239

General framework

- In several cases models have complex interdependencies and/or we deal with large dataset.
- Joint distribution of the data may be difficult to evaluate, or even to specify!
- Typical difficulties arise from the need to invert large matrices when we deal with large dataset and/or from approximation of intractable integrals
- Ancient Roman's principle:

  dividi et impera

  (*'diviser pour regner'*, Louis XI)
- Idea: if computing likelihoods for certain subsets of the data is possible, then one may construct a pseudolikelihood by combining such likelihood objects and use this as a surrogate for the ordinary likelihood
- Review paper: Varin et al. (2011)

Example: spatio-temporal Gaussian data I

- $s \in \mathbb{R}^d$ is a spatial location, $t \in \mathbb{R}$ is a time point
- Observations $\mathbf{z} = (z(s_1, t_1), \ldots, z(s_n, t_n))'$ from a Gaussian Random Field (GRF) $\{Z(s, t)\}$
- Weakly stationarity

$$\mathrm{cov}(Z(s, t), Z(s', t')) = C(h, u; \theta)$$

  ($h = s - s'$, spatial lag, $u = t - t'$ temporal lag).
- Likelihood computation requires inversion of $n \times n$ covariance
  - $O(n^3)$ operations, $O(n^2)$ memory
  - If n large, this may be unfeasible!

Example: spatio-temporal Gaussian data II

- Two possible strategies:
  1. simplify the model:
     - approximate with a Gaussian Markov Random Field (Lindgren et al., 2011) requiring roughly $O(n \log n)$ operations
     - approximate with low rank methods (Cressie and Johannesson, 2008) $O(n)$
  2. keep your model but simplify the fitting method: e.g. Curriero and Lele (1999) estimate θ from the pseudolikelihood

$$PL(\theta; \mathbf{z}) = \prod_{i>j} f(z(s_i, t_i) - z(s_j, t_j); \theta) w(s_i, t_j, s_j, t_j)$$

Example: spatio-temporal non-gaussian data I

- Generalized linear geostatistical models (latent Gaussian models)

$$E(Z(s,t)|U(s,t)) = g^{-1}\left(x(s,t)'\beta + U(s,t)\right)$$

  where $\{U(s,t)\}$ is a latent GRF.
- In general, the likelihood involves an n-dimensional integral

$$L(\beta,\psi,\theta;\mathbf{z}) = \int_{\mathbb{R}^n} f(\mathbf{z}|\mathbf{u};\beta,\psi)f(\mathbf{u};\theta)d\mathbf{u}$$

- Monte Carlo methods: MCEM, MCMC, etc. may be time-consuming even for moderate n
  - INLA method (Rue et al., 2009) $\Rightarrow$ see **Thomas's talk**
- Simpler: pseudolikelihood constructed from bivariate margins

$$PL(\theta;\mathbf{z}) = \prod_{i>j} f(z(s_i,t_i),z(s_j,t_j);\theta)^{w(s_i,t_j,s_j,t_j)}$$

Example: spatio-temporal extreme values I

- Observations $\mathbf{z} = (z(s_1, t_1), \ldots, z(s_n, t_n))'$ are maxima recorded values
- Classical extreme value theory says that the marginal distribution can be modeled by the GEV distribution

$$\Pr(Z(s,t) \leq z) = \exp\left[-\{1 + \xi(z-\mu)/\sigma\}^{-1/\xi}\right]$$

with $\mu \in \mathbb{R}$, $\xi \in \mathbb{R}$, $\sigma > 0$ and $\{1 + \xi(z-\mu)/\sigma\} > 0$.

- Unit Fréchet margins (i.e. $\mu = \xi = \sigma = 1$)

$$\Pr(Z(s,t) \leq z) = \exp\left[-1/z\right]$$

Example: spatio-temporal extreme values II

- A max-stable process (de Haan, 1984), is the extension for maxima of independent replications of (space-time) random fields.
- Suppose unit Fréchet margins. Finite dimesional distribution

$$\begin{aligned}\Pr(\mathbf{Z} \leq \mathbf{z}) &= \Pr(Z(s_1, t_1) \leq z_1, \ldots, Z(s_n, t_n) \leq z_n) \\ &= \exp(-V(\mathbf{z}))\end{aligned}$$

  $V(\mathbf{z})$ is a positive function such that $V(a^{-1}\mathbf{z}) = aV(\mathbf{z})$ for any $a > 0$ and $\mathbf{z} > \mathbf{0}$
- Assume a parametric model

$$\Pr(\mathbf{Z} \leq \mathbf{z}) = \exp(-V(\mathbf{z}; \theta))$$

  the likelihood corresponds to the derivative with respect to all components of $\mathbf{z}$.
  - The number of terms is the B_n Bell number, around 8.3×10^{10} for $n = 17$!

Composite likelihood: definition

Consider

1. a statistical model $\{f(\mathbf{z};\theta), \mathbf{z} \in \mathbb{R}^n, \theta \in \Theta \subseteq \mathbb{R}^p\}$;
2. a set of measurable events $\{\mathcal{A}_k; k = 1, \ldots, K\}$;
3. the associated likelihoods $\mathcal{L}_k(\theta;\mathbf{z}) = f(\mathbf{z} \in \mathcal{A}_k;\theta)$.

Then, a composite likelihood (CL) is the weighted product of the likelihoods corresponding to each single event,

$$\mathrm{CL}(\theta;\mathbf{z}) = \prod_{k=1}^{K} \mathcal{L}_k(\theta;\mathbf{z})^{w_k},$$

where $\{w_k; k = 1, \ldots, K\}$ are positive weights.

Composite conditional likelihood

- Notation $z_i = z(s_i, t_i)$, $\mathbf{z}_D = \{z_j, j \in D\}$, $D \subset \{1, \ldots, n\}$ and $\mathcal{D}$ is a set of D.

$$CCL(\theta; \mathbf{z}) = \prod_{D \in \mathcal{D}} f(\mathbf{z}_D | \mathbf{z}_{D^c}; \theta),$$

- Besag's pseudolikelihood (Besag, 1974)

$$CCL(\theta; \mathbf{z}) = \prod_{i=1}^{n} f(z(s_i) | \mathbf{z}_{\partial i}; \theta),$$

  ∂_i neighborhood of s_i
- Spatial GRF: Vecchia (1988), Stein et al. (2004)
- Limited number of space-time applications
  - Mixed states spatio-temporal auto-models on regular lattice (Hardouin and Crivelli, 2011)

Composite marginal likelihood

- More commonly used

$$CML(\theta; \mathbf{z}) = \prod_{D \in \mathcal{D}} f(\mathbf{z}_D; \theta)$$

- Independence likelihood (Smith, 1990; Chandler and Bate, 2007) where the sets D are disjoint
  - $\theta = (\theta_S, \theta_T)'$ and inference about θ_S

$$CML(\theta; \mathbf{z}) = \prod_{t=1}^{T} f(\mathbf{z}(t); \theta_S)$$

  with $\mathbf{z}(t) = (z(s_1, t), \ldots, z(s_k, t))$
- pairwise likelihood: $\prod_{i>j} f(z_i, z_j; \theta)$
- tripletwise likelihood,..., "blockwise" likelihood
- pairwise differences (for GRF): $\prod_{i>j} f(z_i - z_j; \theta)$

Hybrid methods

- hybrid pairwise likelihood: Kuk (2007)
  - optimal estimating equations for marginal parameters
  - pairwise likelihoods for estimating dependence parameters
  - no examples for space-time examples yet !
- joint composite estimating functions : Bai et al. (2012)
  - space-time example, see later.

Composite likelihood quantities I

- Composite log-likelihood

$$c\ell(\theta;\mathbf{z}) = \log CL(\theta;\mathbf{z}) = \sum_{k}^{M} \log \ell_k(\theta;\mathbf{z}) w_k$$

- Composite score

$$u(\theta;\mathbf{z}) = \sum_{k=1}^{K} \log \nabla \ell_k(\theta;\mathbf{z}) w_k = \sum_{k=1}^{K} u_k(\theta;\mathbf{z}) w_k, \Rightarrow E[u(\theta;\mathbf{Z})] = 0$$

- Maximum composite estimator $\hat{\theta}_{CL}$

$$c\ell(\hat{\theta}_{CL};\mathbf{z}) \geq c\ell(\theta;\mathbf{z}), \qquad \forall \theta \in \Theta$$

Composite likelihood quantities II

- Godambe (or sandwich) information

$$G(\theta) = H(\theta)J(\theta)^{-1}H(\theta)$$

- Sensitivity matrix: $H(\theta) = E[-\nabla u(\theta; \mathbf{Z})]$,

$$\hat{H} = -\nabla u(\hat{\theta}_{CL}; \mathbf{z})$$

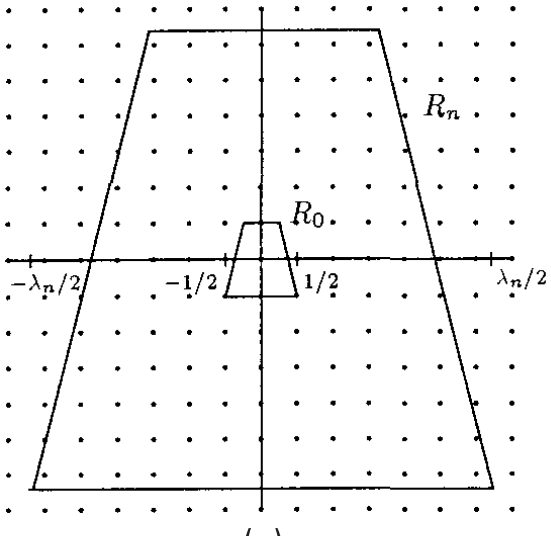
- Variability matrix: $J(\theta) = Var[u(\theta; \mathbf{Z})]$

How to estimate ?

- Misspecification $H(\theta) \neq J(\theta)$

Asymptotics I

- Two types of asymptotic frameworks
  - fixed domain
    - fill a space-time region with a sequence of observed locations
    - no results
  - increasing domain:
    - expand a a space-time region for recover more observed locations
    - regular lattice: $R_0 \subset (-\frac{1}{2}, \frac{1}{2}]^{d+1}$, $R_n = \{(s_1, t_1), \ldots, (s_n, t_n)\} = (\lambda_n R_0) \cap \mathbb{Z}^{d+1}$



Asymptotics II

- Idea: log-composite likelihood is an additive contrast function (Hardouin, 1992; Guyon, 1995)

$$U_n(\theta) = \frac{1}{K_n} \sum_{k=1}^{K_n} \log f(\mathbf{z} \in A_k; \theta) w_k$$

- mixing conditions on $\{Z(s,t)\}$
- $\hat{\theta}_{CL}$ is asymptotically Gaussian

$$G(\theta)^{1/2}(\hat{\theta}_{CL} - \theta^*) \xrightarrow{d} \mathcal{N}(0, I)$$

- θ^* is the minimizer of the composite Kullback-Leibler divergence

$$CKL(\theta; g, f) = \sum_k E_g\{\log g(\mathbf{Z} \in A_k) - \log f(\mathbf{Z} \in A_k; \theta)\} w_k$$

Asymptotics III

- consistency if all composite likelihood "blocks" correctly specified

$$\exists \theta \text{ such that } f(\mathbf{z} \in A_k; \theta) = g(\mathbf{z} \in A_k) \qquad \forall k \in K$$

- Space-time examples:
  - **GRFs**: Bevilacqua et al. (2012); Bai et al. (2012)
  - **Max-stable processes**: Davis et al. (2013)

Hypothesis testing

Null hypothesis $H_0 : \psi = \psi_0$ where $\theta = (\psi, \tau)$ and $\psi \in \mathbb{R}^q$

- Wald-type statistics (no invariance under reparameterization)
- Score-type statistics (numerical instability)
- Composite likelihood ratio statistics

$$W = 2\{c\ell(\hat{\theta}_{CL}; \mathbf{z}) - c\ell(\psi_0, \hat{\tau}_{CL}(\psi_0); \mathbf{z})\}$$

  - Non standard distribution

$$W \xrightarrow{d} \sum_{j=1}^{q} \lambda_j \chi^2_{1,j}$$

    $\lambda_1 \geq, \ldots, \geq \lambda_q$ eigenvalues of $(H^{\psi\psi})^{-1} G_{\psi\psi}$ (Kent, 1982; Guyon, 1995)
  - How to calibrate W ? Spatial example in Cattelan and Sartori (2014)

Model selection

- Model selection based on Akaike-type criterion (Takeuchi, 1976)
- composite likelihood information criterion (Varin and Vidoni, 2005)

$$CLIC = -2cl(\hat{\theta}_{CL}; \mathbf{z}) + 2\, dim(\theta)$$

- composite Bayesian information criterion (Gao and Song, 2010)

$$CBIC = -2cl(\hat{\theta}_{CL}; \mathbf{z}) + \log(n)\, dim(\theta)$$

- effective number of parameters

$$dim(\theta) = \mathrm{tr}(H(\theta)J^{-1}(\theta))$$

- CLIC has a tendency to select over-complicated models.

Model selection: example I

- 100 independent simulations from a zero mean space-time Gaussian process with covariance function $C(h, u; \theta)$
- two space-time grids: $\mathcal{S} \times \mathcal{T}$, where $\mathcal{S} = [-2, \ldots, 3]^2$ and $\mathcal{T} = \{1, \ldots, T\}$ with $T = 100, 200$,
- $Z(s_i, t_i) - Z(s_j, t_j) \sim \mathcal{N}(0, 2\gamma(s_i - s_j, t_i - t_j))$ with $\gamma(h, u) = C(0, 0; \theta) - C(h, u; \theta)$
- pairwise difference likelihood

$$PL(\theta; \mathbf{z}) = \prod_{i>j} f(z(s_i, t_i) - z(s_j, t_j); \theta)$$

Model selection: example II

model A Double exponential model (separable model)

$$C(h,u;\theta) = \sigma^2 \exp\left(-3\frac{\|h\|}{a} - 3\frac{|u|}{b}\right)$$

model B Gneiting model

$$C(h,u;\theta) = \frac{\sigma^2}{(\frac{20|u|^{2\alpha}}{b}+1)} \exp\left\{-\frac{3\|h\|}{a(\frac{20|u|^{2\alpha}}{b}+1)^{\beta/2}}\right\},$$

with $\beta = 0$ (separable model)

model C Gneiting model with $\beta \neq 0$;

Model selection: example III

		Identified					
		T=100			T=200		
	Model	**(A)**	**(B)**	**(C)**	**(A)**	**(B)**	**(C)**
True	**(A)**	84	10	6	96	2	2
	(B)	8	70	22	5	82	13
	(C)	9	19	72	3	12	85

- at least 70% of the models have been correctly identified.

Large data set: around 150000 observations

All that glisters is not gold...

The Merchant of Venice, W. Shakespeare I

- variability matrix $J(\theta) = Var[u(\theta; \mathbf{Z})]$ key ingredient for
  - evaluating uncertainty of composite likelihood estimates
  - test statistics
  - information criteria for model selection
- the simpler case is when there are T independent replicates (in time, for instance) $\mathbf{z}(t) = (z(s_1, t), \ldots, z(s_k, t))$

$$\hat{J} = \frac{1}{T} \sum_{i=1}^{T} u(\hat{\theta}_{MC}; \mathbf{z}(t)) u(\hat{\theta}_{MC}; \mathbf{z}(t))^{\top}$$

  - space-time models for extreme values (Huser and Davison, 2014)

All that glisters is not gold...

The Merchant of Venice, W. Shakespeare II

- more stable to estimate directly the variance matrix of $\hat{\theta}_{CL}$ by bootstrap but computational demanding (Bags of little bootstrap ?, Kleiner et al. (2014))
- much harder when there is no replication in space or time
- analytic expressions are possible only for few special cases (spatial GRF, Bevilacqua and Gaetan (2014)) but too computational demanding ($O(n^4)$)
- resampling methods based on idea of some form of "internal replication" (block-bootstrap, window subsampling)
  - difficult to specify tuning parameters (block dimension)
  - arduous with irregularly spaced temporal and spatial observations

Composite likelihood design I

- How to choose among many possible composite likelihoods?
- for example, which is preferable between
  - the pairwise marginal likelihood

$$c\ell_m(\theta; \mathbf{z}) = \sum_{i>j} w_{ij} \log f(z_i, z_j; \theta)$$

  - the conditional likelihood

$$c\ell_c(\theta; \mathbf{z}) = \sum_{i \neq j} w_{ij} \log f(z_i | z_j; \theta)$$

  - the pairwise difference likelihood (for GRF)

$$c\ell_d(\theta; \mathbf{z}) = \sum_{i>j} w_{ij} \log f(z_i - z_j; \theta)$$

- Same computational burden: $O(n^2)$
- In large settings: should all possible pairs be included?

Weighted composite likelihoods

- When efficiency is low, unequal weighting of likelihood components can be used to obtain some improvement
- Optimal weights computationally impractical
- It sensible to assume $w_{ij} = w_{ji}$ for $c\ell_m$ and $c\ell_d$
- If $w_{ij} = w_{ji}$ and $w_{ii} = 1$

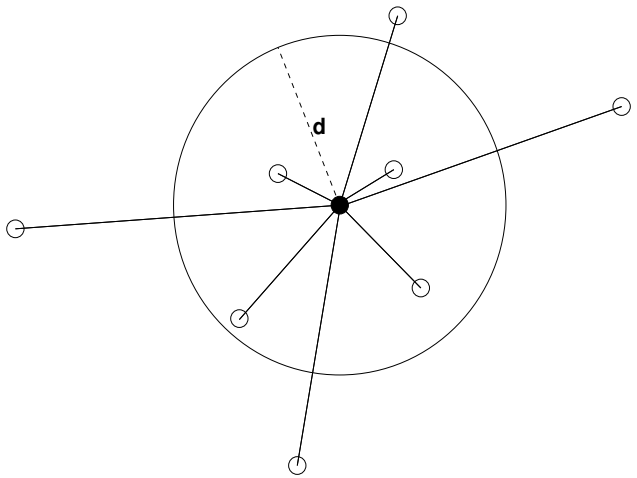
$$c\ell_c(\theta; \mathbf{z}) = 2c\ell_m(\theta; \mathbf{z}) - (n-1)\sum_{i=1}^{n} \log f(z_i; \theta)$$

  - When the marginal parameters are known, marginal and the conditional pairwise likelihood have the same efficiency.
- in space-time setting underweight observations that are far apart in time and/or space

'Simple' weighted composite likelihoods I

- weights based on the distances.

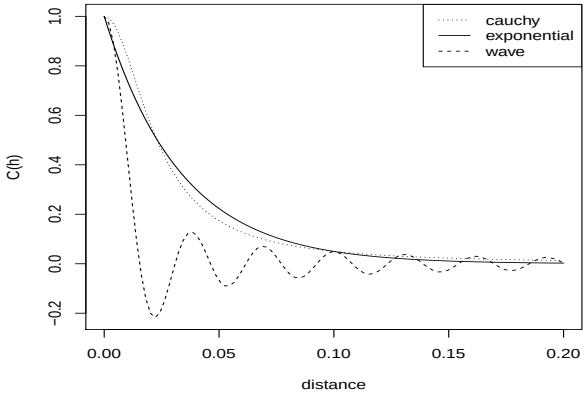
$$w_{ij}(d) = \begin{cases} 1 & \|s_i - s_j\| \leq d_s, |t_i - t_j| \leq d_t, \quad d = (d_s, d_t)' \\ 0 & otherwise \end{cases}$$

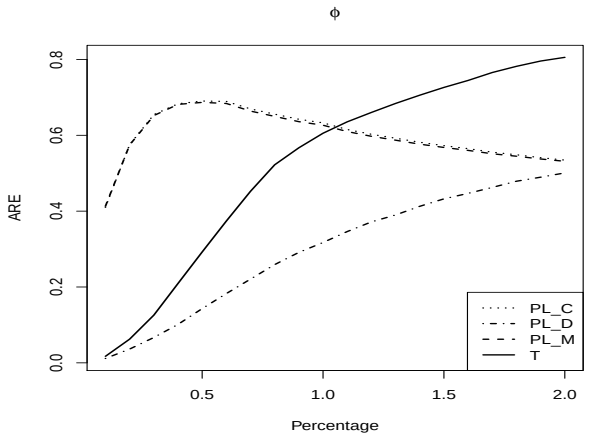


- Isotropy ?

'Simple' weighted composite likelihoods II

- Spatial setting results (Bevilacqua and Gaetan, 2014)





		Exponential model			Cauchy model			Wave model		
		μ	ϕ	σ^2	μ	ϕ	σ^2	μ	ϕ	σ^2
ML	bias	0.0037	-0.0015	-0.0067	0.0044	-0.0010	-0.0041	0.0012	-0.0001	-0.0022
	rmse	0.0843	0.0114	0.0792	0.0973	0.0089	0.0758	0.0409	0.0027	0.0648
TAP	bias	0.0039	-0.0013	-0.0074	0.0048	-0.0015	-0.0096	0.0011	-0.0026	-0.0022
	rmse	0.0851	0.0129	0.0791	0.0989	0.0094	0.0767	0.0420	0.0155	0.0675
pl_C	bias	0.0041	-0.0028	-0.0089	0.0050	-0.0034	-0.0110	0.0013	-0.0010	-0.0020
	rmse	0.0900	0.0133	0.0826	0.1033	0.0120	0.0800	0.0433	0.0086	0.0691
pl_M	bias	0.0041	-0.0028	-0.0087	0.0049	-0.0034	-0.0109	0.0013	-0.0005	-0.0017
	rmse	0.0901	0.0133	0.0827	0.1033	0.0120	0.0800	0.0435	0.0069	0.0694
pl_D	bias		0.0011	0.0088		-0.0004	0.0020		-0.0044	-0.0001
	rmse		0.0201	0.1142		0.0120	0.0891		0.0167	0.0717

'Simple' weighted composite likelihoods III

- Practical implementation (Bevilacqua et al., 2012)
  - Get a consistent estimate for θ, $\tilde{\theta}$
  - Estimate $G(\theta, d)$ by $\hat{G}(\tilde{\theta}; d)$
  - We choose the 'lag' d minimizing $[\hat{G}(\tilde{\theta}; d)]^{-1}$ in the partial order of nonnegative definite matrices or equivalently

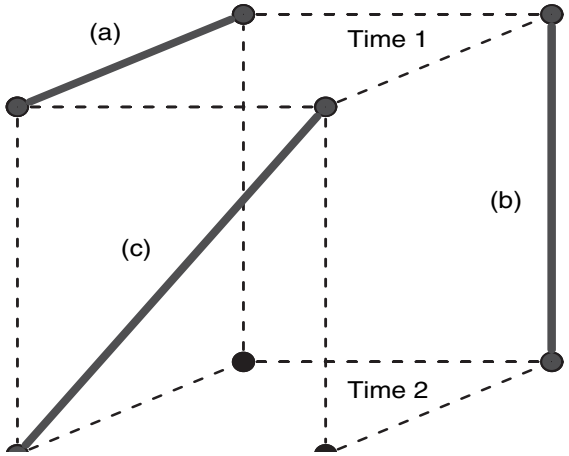
$$d^* = \underset{d \in \mathcal{D}}{\operatorname{argmin}} \operatorname{tr}[\hat{G}(\tilde{\theta}; d)]^{-1}$$

    where $\mathcal{D}$ is a set of space-time lags.
  - Maximize

$$c\ell_m(\theta; \mathbf{z}) = \sum_{i>j} w_{ij}(d^*) \log f(z_i, z_j; \theta)$$

Joint composite estimating functions I

- First partition $R = \{(s_1, t_1), \ldots, (s_n, t_n)\}$ into three subsets
  - R_S with pairs (s_i, t_i),(s_j, t_j), $i \neq j$ differing only in locations i.e. $s_i \neq s_j$, and $t_i = t_j$
  - R_T with pairs differing only in time i.e. $s_i = s_j$, and $t_i \neq t_j$
  - R_C with pairs (cross-pairs) differing in time and in space i.e. $s_i \neq s_j$, and $t_i \neq t_j$



(a) spatial pair, (b) temporal pair, (c) the spatiotemporal cross-pair

Joint composite estimating functions II

- Define $c\ell_A(\theta;\mathbf{z}) = \sum_{(i,j)\in A} w_{ij} \log f(z_i, z_j;\theta)$ we have

$$c\ell_m(\theta;\mathbf{z}) = c\ell_S(\theta;\mathbf{z}) + c\ell_T(\theta;\mathbf{z}) + c\ell_C(\theta;\mathbf{z})$$

$$\text{complete} = \text{space} + \text{time} + \text{space-time}$$

- Composite likelihood estimate $\hat{\theta}_{CL}$ solve the estimating equation

$$\begin{aligned} u(\theta;\mathbf{z}) &= 0 \\ u_S(\theta;\mathbf{z}) + u_T(\theta;\mathbf{z}) + u_C(\theta;\mathbf{z}) &= 0 \end{aligned}$$

Joint composite estimating functions III

- Instead stack the subset composite score functions into the vector $g(\theta, \mathbf{z}) = (u_S(\theta; \mathbf{z}), u_S(\theta; \mathbf{z}), u_C(\theta; \mathbf{z}))$
- weighted quadratic objective function (looks like GMM (Hansen, 1982))

$$Q(\theta; \mathbf{z}) = g(\theta, \mathbf{z})^\top W(\theta) g(\theta, \mathbf{z})$$

  where $W(\theta) = DVar[g(\theta, \mathbf{Z})]D$ and $D = diag(\sqrt{|R_S|}, \sqrt{|R_T|}, \sqrt{|R_C|}) \otimes I$ (adjustement for different size)
- the joint composite estimating function (JCEF) estimator (Bai et al., 2012)

$$\hat{\theta}_{JCEF} = \mathrm{argmax}_\theta Q(\theta; \mathbf{z})$$

Joint composite estimating functions IV

- More efficient than 'simple' weighted composite likelihood for a fixed d
- More computational demanding
  - Require at each time an evaluation of $Var[g(\theta, \mathbf{Z})]$
  - $Var[g(\theta, \mathbf{Z})]$ can be derived analytically for GRF, given the large number of possible pairs, computing it on the basis of analytic formulae is not practically feasible.
  - estimation of this matrix is typically achieved by subsampling techniques
- Model selection and hypothesis testing ?

Big data: block composite likelihoods I

- Estimation and prediction for spatial data (Eidsvik et al., 2014)
- A framework that allows parallel computing
- Similar approaches: Vecchia (1988); Stein et al. (2004); Caragea and Smith (2006).
- Partition a region D into M blocks $D_1, \ldots, D_M$, denote $\mathbf{z}_{D_k} = \{z_i, i \in D_k\}$
- Provide that the number of locations in $D_k \cup D_l$ is no so large for evaluating $f(\mathbf{z}_{D_k}, \mathbf{z}_{D_l}; \theta)$...

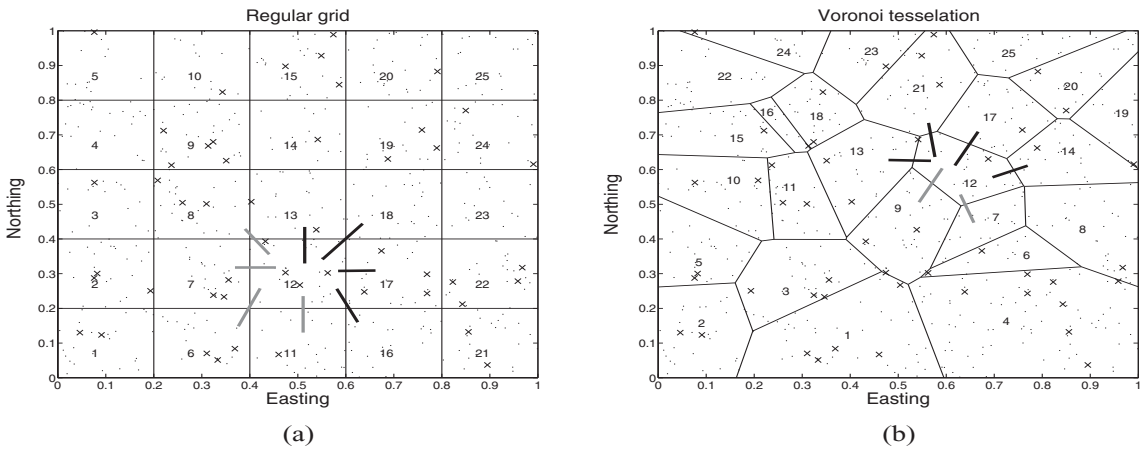
$$c\ell_B(\theta; \mathbf{z}) = \sum_{l>k} \log f(\mathbf{z}_{D_k}, \mathbf{z}_{D_l}; \theta)$$

(log) block composite likelihood (pairwise composite block-likelihood)

Big data: block composite likelihoods II

- for $M = 1$ or $M = 2$ full likelihood, $M = n$ pairwise likelihood
- M: trade-off between computational and statistical efficiency.
- N_k the neighbors of block k, $N_k^{\rightarrow} = \{l > k\} \cap \{l \in N_k\}$

$$c\ell_B(\theta; \mathbf{z}) = \sum_{k=1}^{M-1} \sum_{l \in N_k^{\rightarrow}} \log f(\mathbf{z}_{D_k}, \mathbf{z}_{D_l}; \theta)$$



Source: Eidsvik et al. (2014)

Big data: block composite likelihoods III

Recommendation for creating blocks

- check the spatial dependence and possible anisotropies with empirical variogram

Computational efficiency

- $n_k = c$ (fixed number of observations in a block)
- $O(c^3 M|N_k^{\rightarrow}|)$ i.e. $O(n)$

Concurrent approach:

- Fixed-rank kriging (Cressie and Johannesson, 2008): O(n)
- Gaussian Markov random fields (Lindgren et al., 2011): $O(n^{3/2})$ for two-dimensional spatial data and $O(n^2)$ for three spatial dimensions

Software development (in R)

- is R a convenient framework for development of composite likelihood software?
- coding in R: (block) pairwise (marginal or conditional) likelihood needs parallellizable bivariate functions in way to avoid slow loops
- Graphical processing unit
- more efficient (?) coding composite likelihoods and derived quantities in low-level languages such as Fortran or C and then call into R
- efficient numerical algorithms for low-dimension integration (for non Gaussian data)
- what is already available in R: CompRandFld package for GRF

Challenges as final remarks

- "Design" issues: which terms should be included and how they have to be combined ?
  - Preliminary analysis
- Big data: parallel inference ?
- Precise estimation of uncertainty of composite likelihood estimates
- Calibration: how to calibrate test statistics ?
- Composite likelihood in hierarchical model ?
  - Composite expectation-maximization algorithm
- Can we use composite likelihoods ideas in prediction?
  - Composite kriging ?

Thanks !

Merci !

Grazie !

Some references I

Bai, Y., Song, P. X.-K. and Raghunathan, T. E. (2012) Joint composite estimating functions in spatiotemporal models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **74**, 799–824.

Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society B*, **36**, 192–236.

Bevilacqua, M. and Gaetan, C. (2014) Comparing composite likelihood methods based on pairs for spatial Gaussian random fields. *Statistics and Computing*, 1–16.

Bevilacqua, M., Gaetan, C., Mateu, J. and Porcu, E. (2012) Estimating space and space-time covariance functions for large data sets: a weighted composite likelihood approach. *Journal of the American Statistical Association*, **107**, 268–280.

Caragea, P. and Smith, R. (2006) Approximate likelihoods for spatial processes. *Tech. rep.*, Department of Statistics, Iowa State University.

Cattelan, M. and Sartori, N. (2014) Empirical and Simulated Adjustments of Composite Likelihood Ratio Statistics. *arXiv:1403.7093 [stat]*.

Chandler, R. E. and Bate, S. (2007) Inference for clustered data using the independence loglikelihood. *Biometrika*, **94**, 167 –183.

Cressie, N. and Johannesson, G. (2008) Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society, Series B*, **70**, 209–226.

Curriero, F. and Lele, S. (1999) A composite likelihood approach to semivariogram estimation. *Journal of Agricultural, Biological and Environmental Statistics*, **4**, 9–28.

Davis, R. A., Klüppelberg, C. and Steinkohl, C. (2013) Statistical inference for max-stable processes in space and time. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **75**, 791–819.

de Haan, L. (1984) A spectral representation for max-stable processes. *The Annals of Probability*, **12**, 1194–1204.

Eidsvik, J., Shaby, B., Reich, B., Wheeler, M. and Niemi, J. (2014) Estimation and prediction in spatial models with block composite likelihoods. *Journal of Computational and Graphical Statistics*, **23**, 295–315.

Some references II

Gao, X. and Song, P. (2010) Composite likelihood bayesian information criteria for model selection in high dimensional data. *Journal of the American Statistical Association*, **105**, 1531–1540.

Guyon, X. (1995) *Random Fields on a Network*. New York: Springer.

Hansen, L. (1982) Large sample properties of generalized method of moments estimators. *Econometrica*, **50**, 1029–1054.

Hardouin, C. (1992) *Quelques résultats nouveaux en statistique des processus: contraste fort, régressions à résidus à longue portée et estimation par log-périodogramme*. Ph.D. thesis, Université Paris VII, Paris.

Hardouin, C. and Crivelli, T. (2011) Mixed-state spatio-temporal auto-models. *Statistics*, **45**.

Huser, R. and Davison, A. C. (2014) Space-time modelling of extreme events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **76**, 439–461.

Kent, J. T. (1982) Robust properties of likelihood ratio tests. *Biometrika*, **69**, 19–27.

Kleiner, A., Talwalkar, A., Sarkar, P. and Jordan, M. I. (2014) A scalable bootstrap for massive data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **76**, 795–816.

Kuk, A. (2007) A hybrid pairwise likelihood method. *Biometrika*, **94**, 939–952.

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*, **73**, 423–498.

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**, 319–392.

Smith, R. L. (1990) Regional estimation from spatially dependent data. *Preprint*. University of North Carolina.

Stein, M., Chi, Z. and Welty, L. (2004) Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society B*, **66**, 275–296.

Takeuchi, K. (1976) Distribution of information statistics and criteria for adequacy of models. *Mathematical Science*, **153**, 12–18. In Japanese.

Some references III

Varin, C., Reid, N. and Firth, D. (2011) An overview of composite likelihood methods. *Statistica Sinica*, **21**, 5–42.

Varin, C. and Vidoni, P. (2005) A note on composite likelihood inference and model selection. *Biometrika*, **52**, 519–528.

Vecchia, A. (1988) Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society B*, **50**, 297–312.