Spatial Statistics for Energy Challenges

Tilmann Gneiting

Heidelberg Institute for Theoretical Studies (HITS)
Karlsruhe Institute of Technology (KIT)

tilmann.gneiting@h-its.org

http://www.h-its.org/english/research/cst/

KAUST, March 8, 2014

Outline

Part 1: Random Fields on Euclidean Spaces

1.1 Spatial statistics: Geostatistical approaches

1.2 Random fields, covariances, and positive definiteness

1.3 Stationary and isotropic correlation functions

1.4 Examples

1.5 Sample path properties

1.6 Non-stationary covariance functions

Part 2: Random Fields on Spheres

2.1 Global data

2.2 Isotropic correlation functions on spheres

2.3 Examples

2.4 Some challenges

Part 3: Space-Time Covariance Functions

3.1 Geostatistical models for spatio-temporal data
3.2 Stationarity, separability and full symmetry
3.3 Stationary space-time covariance functions
3.4 Irish wind data

Part 4: Calibrated Probabilistic Forecasts at the Stateline Wind Energy Center

4.1 Probabilistic forecasts
4.2 The Stateline wind energy center
4.3 The regime-switching space-time (RST) method
4.4 Tools for evaluating predictive performance
4.5 Results

Part 1

Random Fields on Euclidean Spaces

1.1 Spatial statistics: Geostatistical approaches

spatial statistics is concerned with the analysis of **spatially** or **spatio-temporally referenced data** and with the study of the associated probabilistic models and **stochastic processes**

applications abound, including agriculture, astronomy, climatology, economics, **energy**, environmental science, geographical epidemiology, geology, hydrology, medical imaging, meteorology, . . .

spatial statistics can be broadly divided into three complementary parts, dealing with **point patterns**, **lattice data** and **geostatistical data**, respectively

the state of the art in **spatial statistics** is summarized in the Handbook of Spatial Statistics (Gelfand, Diggle, Fuentes and Guttorp, eds., 2010) and in Cressie and Wikle (2011)

1.1 Spatial statistics: Geostatistical approaches

geostatistical data are referenced at potentially **scattered locations** in space or space-time

interest thus centers on building statistical **models** in **continuous space** or **space-time**: given observations $Z(x_1), \ldots, Z(x_k)$ at scattered sites $x_1, \ldots, x_k$ in $\mathbb{R}^d$, we seek to fit a probability model for the process

$$\left\{ Z(x) : x \in \mathbb{R}^d \right\}$$

examples include

- meteorological variables such as **temperature** or **precipitation accumulation** observed at weather stations
- atmospheric **pollutant concentrations** observed at environmental monitoring stations
- **permeabilities** in an aquifer observed at well logs
- **ore grades** within a deposit or mine

1.1 Spatial statistics: Geostatistical approaches

major tasks in geostatistics comprise

- the **characterization** of **spatial variability**
- **spatial prediction**

  given observations $Z(x_1), \ldots, Z(x_k)$ at $x_1, \ldots, x_k \in \mathbb{R}^d$, find a point predictor for the unknown value of Z at $x_0 \in \mathbb{R}^d$ or, preferably, find a conditional distribution for $Z(x_0)$
- **geostatistical simulation**

  given observations $Z(x_1), \ldots, Z(x_k)$ at $x_1, \ldots, x_k \in \mathbb{R}^d$, sample conditional realizations of the process $\{Z(x) : x \in \mathbb{R}^d\}$

the state of the art in **geostatistics** is summarized in Stein (1999) and Chilès and Delfiner (1999, 2012)

1.2 Random fields, covariances, and positive definiteness

geostatistical techniques depend on **random function**, **random field** or **stochastic process** models that are indexed in continuous space or space-time

$$\left\{ Z(x) = Z(x, \omega) : x \in \mathbb{R}^d \right\}$$

spatial location	"chance"
$x \in \mathbb{R}^d$	$\omega \in \Omega$

simplifying assumptions on the random field model are crucial, with the most basic approach requiring **stationarity**, in that

- the **expectation** $\mathbb{E}(Z(x))$ exists and does not depend on $x \in \mathbb{R}^d$
- the **covariance** $\text{cov}(Z(x), Z(x+h))$ exists and does not depend on the location $x \in \mathbb{R}^d$

1.2 Random fields, covariances, and positive definiteness

for any **stationary random field** $\{Z(x) : x \in \mathbb{R}^d\}$, we can define the **covariance function**,

$$C : \mathbb{R}^d \to \mathbb{R}, \quad h \mapsto C(h) = \mathrm{cov}(Z(x), Z(x+h))$$

and the **correlation function**,

$$c : \mathbb{R}^d \to \mathbb{R}, \quad h \mapsto c(h) = \mathrm{corr}(Z(x), Z(x+h)) = \frac{C(h)}{C(0)}$$

Gaussian random fields are **characterized** by the first two **moments**; in particular, mean zero unit variance stationary Gaussian random fields are **characterized** by the **correlation function**

these functions are **positive definite**: given any finite system of real numbers $a_1, \ldots, a_m \in \mathbb{R}$ and locations $x_1, \ldots, x_m \in \mathbb{R}^d$,

$$\sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j C(x_i - x_j) = \mathrm{var}\left(\sum_{j=1}^{m} a_j Z(x_j) \right) \geq 0$$

1.2 Random fields, covariances, and positive definiteness

Theorem (Kolmogorov; Bochner): Let $c : \mathbb{R}^d \to \mathbb{R}$ be continuous with $c(0) = 1$. Then the following statements are equivalent:

(a) The function c is **positive definite**.

(b) There exists a mean zero unit variance **stationary Gaussian random field** with **covariance** and **correlation function** c.

(c) There exists a symmetric **probability measure** F on $\mathbb{R}^d$ with **Fourier transform**

$$c(h) = \int_{\mathbb{R}^d} e^{ih'x} \, \mathrm{d}F(x). \tag{1}$$

(d) There exists a d-variate symmetric **random vector** X with **characteristic function** $c(h) = \mathbb{E}[e^{ih'X}]$.

The probability measure F in the **spectral representation** (1) is called the **spectral measure**.

1.2 Random fields, covariances, and positive definiteness

Inversion theorem. If the correlation function $c : \mathbb{R}^d \to \mathbb{R}$ is continuous and integrable, the associated spectral measure F is absolutely continuous with **spectral density**

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-ix'h} \, c(h) \, \mathrm{d}h, \quad x \in \mathbb{R}^d.$$

Corollary. Suppose that $c : \mathbb{R}^d \to \mathbb{R}$ is continuous and integrable. Then c is a correlation function if, and only if, $c(0) = 1$ and

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-ix'h} \, c(h) \, \mathrm{d}h \geq 0$$

for almost all $x \in \mathbb{R}^d$.

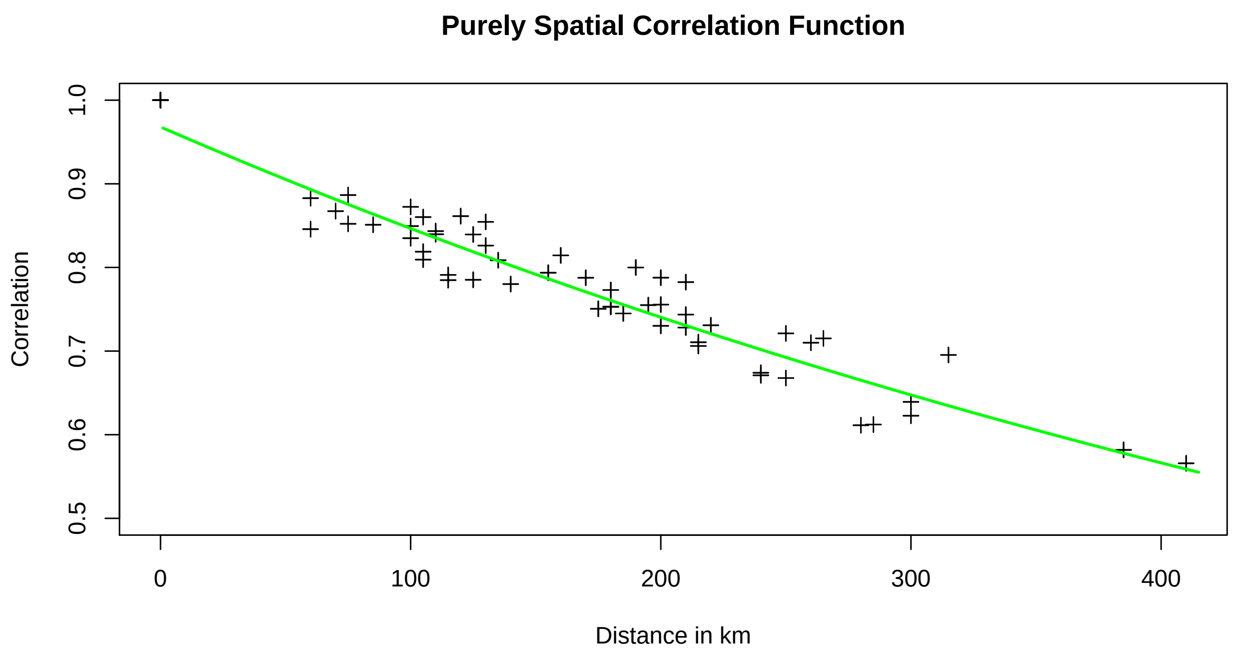
	Positive definite	Nonnegative
Probability measure	characteristic function	probability density function
Stationary random field	correlation function	spectral density function

1.3 Stationary and isotropic correlation functions

in practice, the **correlation function** is often assumed to be **isotropic**, in that

$$c(h) = \varphi(|h|), \qquad h \in \mathbb{R}^d$$

for some function $\varphi : [0, \infty) \to \mathbb{R}$, where $|\cdot|$ denotes the **Euclidean norm** or distance



Irish wind data (Haslett and Raftery 1989; Gneiting 2002)

1.3 Stationary and isotropic correlation functions

the discontinuous **nugget effect**,

$$c(h) = \begin{cases} 1, & h = 0, \\ 0, & h \neq 0, \end{cases}$$

is an isotropic correlation function in $\mathbb{R}^d$ that represents **measurement error** and/or **microscale variability**

in dimension $d \geq 2$, every measurable isotropic correlation function

$$c(h) = \varphi(|h|), \quad h \in \mathbb{R}^d$$

is a **convex combination** of a **nugget effect** and a **continuous** correlation function

thus, we may assume from now on that φ is **continuous**

1.3 Stationary and isotropic correlation functions

class Φ_d:

$$\Phi_d = \left\{ \varphi : [0,\infty) \to \mathbb{R} \;\middle|\; \begin{array}{c} \varphi \text{ is continuous, } \varphi(0) = 1, \text{ and} \\ h \mapsto \varphi(|h|) \text{ is positive definite in } \mathbb{R}^d \end{array} \right\}$$

multiple interpretations of the class **Φ_d:**

- continuous **correlation functions** of stationary and isotropic random fields in $\mathbb{R}^d$
- **characteristic functions** of isotropic random vectors in $\mathbb{R}^d$
- continuous and isotropic, standardized **positive definite functions** in $\mathbb{R}^d$

quite obviously,

$$\Phi_1 \supset \Phi_2 \supset \cdots \qquad \text{and} \qquad \Phi_d \downarrow \Phi_\infty = \bigcap_{d=1}^{\infty} \Phi_d$$

1.3 Stationary and isotropic correlation functions

Theorem (Schoenberg):

(a) The members of the class $\boldsymbol{\Phi_d}$ are of the form

$$\varphi(t) = \int_{[0,\infty)} \Omega_d(rt) \, \mathrm{d}F(r)$$

for $t \geq 0$, where F is a probability measure on $[0, \infty)$, i.e., they are **scale mixtures** of the Bessel function

$$\Omega_d(t) = \Gamma(d/2) \left(\frac{2}{t}\right)^{(d-2)/2} J_{(d-2)/2}(t).$$

(b) The members of the **class $\boldsymbol{\Phi_\infty}$** are of the form

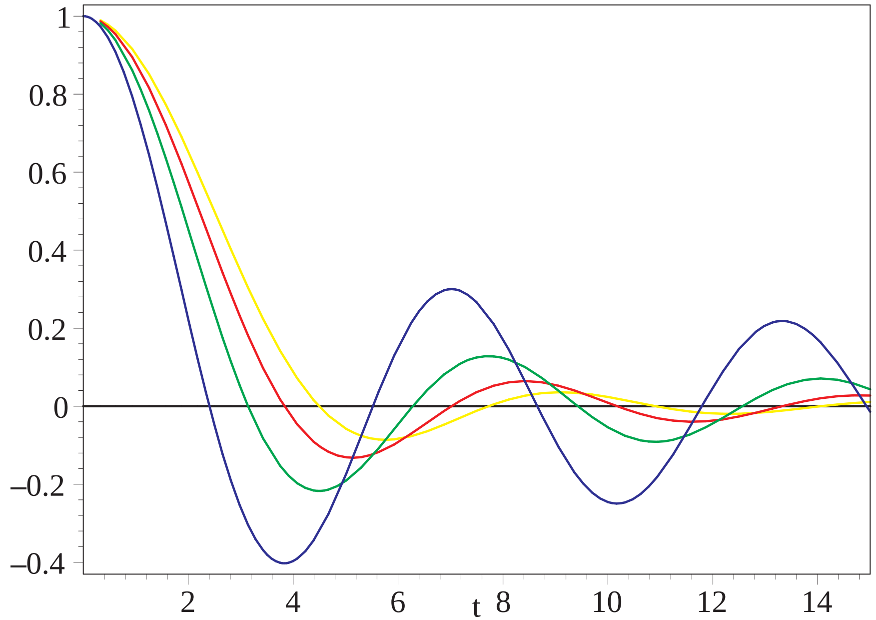
$$\varphi(t) = \int_{[0,\infty)} e^{-r^2 t^2} \, \mathrm{d}F(r)$$

for $t \geq 0$, where F is a probability measure on $[0, \infty)$, i.e., they are **scale mixtures** of the Gaussian function, $\exp(-t^2)$.

In particular, they are **nonnegative** and **decreasing** functions with **support** $[0, \infty)$.

1.3 Stationary and isotropic correlation functions

Dimension d	1	2	3	4	5	∞
$\Omega_d(t)$	$\cos t$	$J_0(t)$	$\frac{\sin t}{t}$	$2\,\frac{J_1(t)}{t}$	$3\,\frac{\sin t - t\cos t}{t^3}$	$\exp(-t^2)$
Lower bound	-1	-0.403	-0.218	-0.133	-0.086	0



$$\lim_{d\to\infty} \Omega_d(\sqrt{2d}t) = \exp(-t^2) \text{ uniformly in } t \geq 0$$

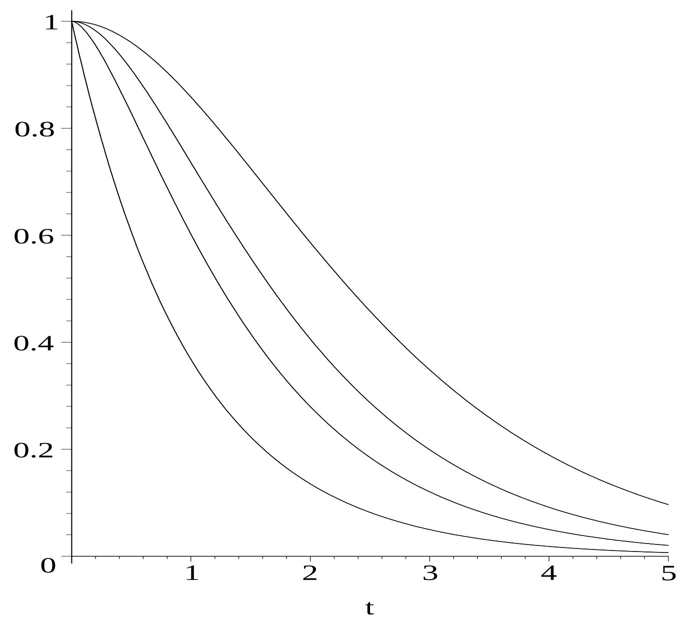
1.4 Examples

Matérn family

$$\varphi(t) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{t}{s}\right)^{\nu} K_{\nu}\left(\frac{t}{s}\right) \qquad (\nu > 0;\ s > 0)$$

member of the class Φ_∞
studied by Whittle (1954), Matérn (1960) and Tatarski (1961)

$\nu \to \infty$	$\varphi(t) = \exp(-t^2)$ when $s^{-1} = 2\sqrt{\nu}$
$\nu = \frac{5}{2}$	$\varphi(t) = \left(1 + t + t^2/3\right) \exp(-t)$
$\nu = \frac{3}{2}$	$\varphi(t) = (1 + t) \exp(-t)$
$\nu = 1$	$\varphi(t) = t\, K_1(t)$
$\nu = \frac{1}{2}$	$\varphi(t) = \exp(-t)$
$\nu \to 0$	nugget effect



1.4 Examples

Powered exponential family

$$\varphi(t) = e^{-(t/s)^\alpha} \qquad (\alpha \in (0,2];\ s > 0)$$

member of the class Φ_∞

$\alpha = 2$	Gaussian	$\varphi(t) = \exp(-t^2)$
$\alpha = 1$	exponential	$\varphi(t) = \exp(-t)$
$\alpha \to 0$	nugget effect	

Cauchy family

$$\varphi(t) = \left(1 + \left(\frac{t}{s}\right)^\alpha\right)^{-\beta/\alpha} \qquad (\alpha \in (0,2];\ \beta > 0;\ s > 0)$$

member of the class Φ_∞
studied by Gneiting and Schlather (2004)

1.4 Examples

as noted, the members of the class Φ_∞ are **nonnegative** and **decreasing** functions with **support** $[0, \infty)$

moderate **negative correlations** and **non-monotonicity** are ocassionally observed and referred to as **hole effect**

a possible **hole effect** correlation model is

$$\varphi(t) = \exp(-at) \cos(bt) \qquad \left(a > 0;\ \frac{|b|}{a} \leq \tan \frac{\pi}{2d} \right)$$

$$\text{member of the class } \Phi_d \iff \tfrac{|b|}{a} \leq \tan \tfrac{\pi}{2d}$$

1.4 Examples

covariance structures with **compact support** allow for computationally efficient prediction and simulation

parametric families of compactly supported members of the class $\mathbf{\Phi_3}$ with support parameter $s > 0$ and shape parameter $\tau \geq 0$

Family	Analytic Expression	Smoothness
Spherical	$\varphi(t) = \left(1 + \frac{1}{2}\frac{t}{s}\right)\left(1 - \frac{t}{s}\right)_+^2$	$k = 0 \quad \alpha = 1$
Askey	$\varphi(t) = \left(1 - \frac{t}{s}\right)_+^{2+\tau}$	$k = 0 \quad \alpha = 1$
Wendland	$\varphi(t) = \left(1 + (4+\tau)\frac{t}{s}\right)\left(1 - \frac{t}{s}\right)_+^{4+\tau}$	$k = 1 \quad \alpha = 2$

1.5 Sample path properties

a stationary and isotropic, standardized Gaussian random field is **characterized by** the **correlation function**,

$$c(h) = \varphi(|h|), \qquad h \in \mathbb{R}^d$$

the **sample path properties** of the **random field** thus depend on the **analytic properties** of the **function** $\boldsymbol{\varphi}$

behavior at the origin: if φ has **fractal index** $\boldsymbol{\alpha \in (0, 2]}$, in the sense that

$$1 - \varphi(t) \sim t^\alpha \quad \text{as} \quad t \downarrow 0,$$

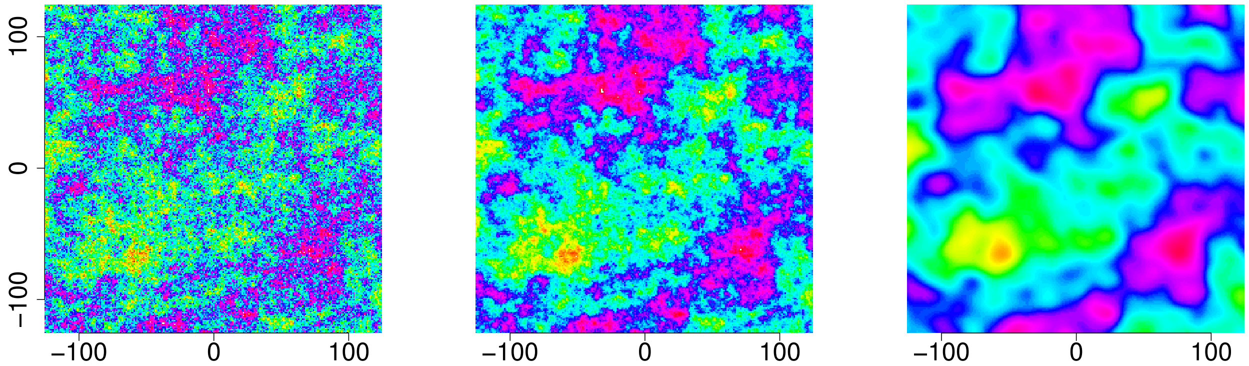
then the **graph** $\{(x, Z(x)) : x \in \mathbb{R}^d\} \subset \mathbb{R}^{d+1}$ of a **sample path** of the Gaussian random field has **fractal** or **Hausdorff dimension**

$$D = d + 1 - \frac{\alpha}{2}$$

almost surely

1.5 Sample path properties

behavior at the origin: if φ has fractal index $\boldsymbol{\alpha \in (0, 2]}$, the graph of a Gaussian sample path has fractal or Hausdorff dimension $\boldsymbol{D = d + 1 - \frac{\alpha}{2}}$ almost surely



$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$
$D = 2.75$	$D = 2.5$	$D = 2$

if the **fractal index** is $\boldsymbol{\alpha = 2}$ then $u \to \varphi(|u|)$ is at least twice differentiable on $\mathbb{R}$ and the following holds:

$u \mapsto \varphi(|u|)$ is $2k$ times differentiable $\Longleftrightarrow$ a Gaussian **sample path** is $\boldsymbol{k}$ **times differentiable** almost surely

Matérn family: $\alpha = 2\min(\nu, 1)$; $k = \lceil \nu \rceil$; **powered exponential** and **Cauchy** families: $k = 0$ if $\alpha \in (0, 2)$ and $k = \infty$ if $\alpha = 2$

1.5 Sample path properties

long-memory dependence: typically, not important in spatial prediction, but a significant consideration in estimation problems

Hurst effect in time series analysis:

$(X_t)_{t=1,2,\ldots}$ standard time series model $\Rightarrow$ $\text{sd}(\bar{X}_n) \sim n^{-1/2}$

long-memory models with **Hurst coefficient** $H \in [\frac{1}{2}, 1)$ have

$$\text{sd}(\bar{X}_n) \sim n^{-(1-H)} \qquad \text{as} \quad n \to \infty$$

empirically observed in **time series data** (Hurst 1951: Nile river) as well as in **spatial data** (Fairfield Smith 1938; Whittle 1956, 1962: agricultural field trials)

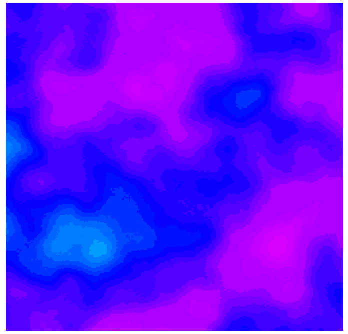
corresponds to a **power law** for the **correlation function**,

$$\varphi(t) \sim t^{-\beta} \quad \text{as} \quad t \to \infty$$

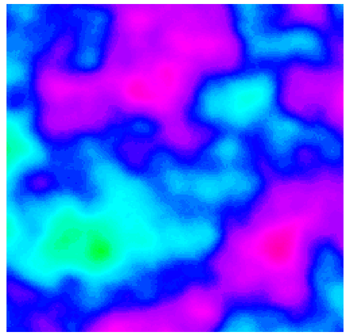
where $\beta = 2 - 2H \in (0, 1]$, i.e., $H \in [\frac{1}{2}, 1)$, with the **Cauchy family** serving as key example

1.5 Sample path properties

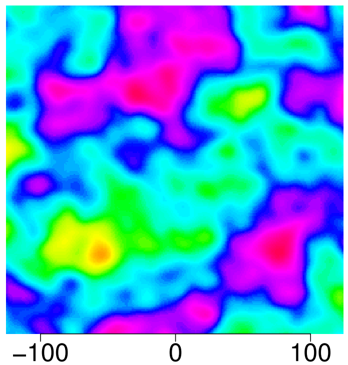
$\beta = 0.025$
$H = 0.9875$



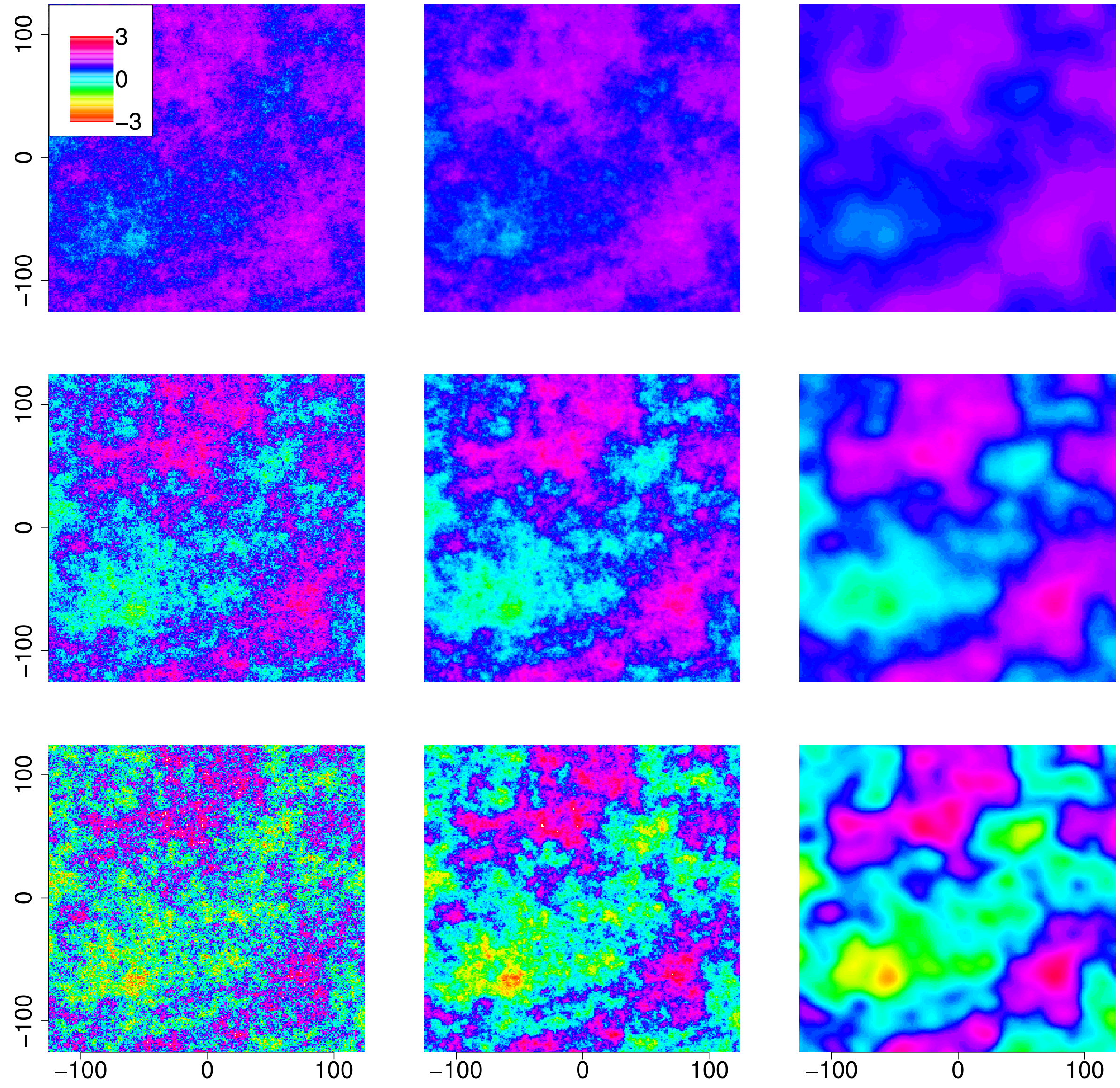
$\beta = 0.2$
$H = 0.9$



$\beta = 0.9$
$H = 0.55$



1.5 Sample path properties



1.6 Non-stationary covariance functions

stationary and **isotropic** models form the basic building blocks of more realistic and more complex, **non-stationary** models

in a general, typically Gaussian **random field**

$$\left\{ Z(x) : x \in \mathbb{R}^d \right\}$$

the covariance between $Z(x)$ and $Z(y)$ depends on both x and y, and simplifying assumptions may not hold

in this situation, we call the map

$$K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}, \quad (x, y) \mapsto K(x, y) = \text{cov}\,(Z(x), Z(y))$$

the **non-stationary covariance** function of the random field

K is **positive definite**: given any finite system of real numbers $a_1, \ldots, a_m \in \mathbb{R}$ and locations $x_1, \ldots, x_m \in \mathbb{R}^d$,

$$\sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j \, K(x_i, x_j) = \text{var} \left(\sum_{j=1}^{m} a_j Z(x_j) \right) \geq 0$$

1.6 Non-stationary covariance functions

similar to the stationary and isotropic case, the **sample path properties** of non-stationary Gaussian random fields (Hausdorff dimension and smoothness of the realizations, ...) can be read off the **correlation function** (Adler 1981, 2009)

however, the probabilistic properties might be **localized** and thus might **vary spatially**

at least three basic **construction principles**:

- **space deformation** and **space embedding** approaches (Sampson and Guttorp 1992)
- spatial **moving average** or **process convolution** approaches (Higdon 1998; Paciorek and Schervish 2006)
- approaches based on **stochastic partial differential equations** (Lindgren, Rue and Lindström 2011)

1.6 Non-stationary covariance functions

space deformation and **space embedding** approaches

general idea: given a random field $\{ Z(x) : x \in \mathbb{R}^d \}$ find a generally nonlinear **transformation function**

$$\mathsf{T} : \mathbb{R}^d \times \mathbb{R}^e \to \mathbb{R}^n,$$

which might depend on both the **spatial coordinate**, $x \in \mathbb{R}^d$, and spatially varying **covariates**, $t(x) \in \mathbb{R}^e$, such that

$$K(x, y) = \sigma(x)\, \sigma(y)\, \varphi(|\mathsf{T}(x, t(x)) - \mathsf{T}(y, t(y))|),$$

where $\sigma : \mathbb{R}^d \to [0, \infty)$ is a spatially varying **standard deviation** and $\varphi \in \Phi_n$ is an **isotropic** correlation function

statistical challenges lie in

- the choice of a suitable family of **covariates**, $t(x) \in \mathbb{R}^e$, and of a suitable **embedding dimension**, n
- the parameterization and joint **estimation** of the spatially varying **standard deviation**, σ, the **transformation** function, T, and the **isotropic correlation** function, φ

1.6 Non-stationary covariance functions

moving average or **process convolution** approaches

a general **process convolution** is defined as the random field

$$\left\{ Z(x) = \int_{\mathbb{R}^d} u_x(y) \, \mathrm{d}W(y) \; : \; x \in \mathbb{R}^d \right\}$$

where W represents a latent, typically Gaussian **white noise** process or **Lévy basis**, and $u_x : \mathbb{R}^d \to \mathbb{R}$ is a square-integrable, locally varying moving average **kernel**

in particular, if

$$u_x(y) = v(x - y)$$

for some fixed function $v : \mathbb{R}^d \to \mathbb{R}$, the process convolution is **stationary**, with the **covariance** function

$$C : \mathbb{R}^d \to \mathbb{R}, \qquad h \mapsto C(h) = v * v(h) = \int_{\mathbb{R}^d} v(x) v(x + h) \, \mathrm{d}x$$

being the **self-convolution** of the **kernel** v

if v is the indicator function of a sphere and $d = 3$, we recover the **spherical** correlation function

1.6 Non-stationary covariance functions

in the general case, a **process convolution** has non-stationary **covariance**

$$K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d, \quad (x, y) \mapsto K(x, y) = \int_{\mathbb{R}^d} u_x(z)\, u_y(z)\, \mathrm{d}z$$

can **model** the **kernel**, which is arbitrary subject to sq. integrability, rather than the covariance, which needs to be positive definite

the **domain** of the process is easily **restricted** to irregular areas of interest, in both Gaussian and **non-Gaussian** settings

the approach allows for parametric families of **non-stationary covariance** functions in **closed form** (Paciorek and Schervish 2006; Schlather 2010; Anderes and Stein 2011)

e.g., if for each $x \in \mathbb{R}$ the matrix $\Sigma(x) \in \mathbb{R}^{d \times d}$ is positive definite, then

$$K(x, y) = \exp\left(-(x - y)' \left(\Sigma(x) + \Sigma(y)\right)^{-1} (x - y)\right)$$

is a non-stationary covariance function

Part 1: Further reading

Gelfand, A. E., Diggle, P. J., Fuentes, M. and Guttorp, P., eds. (2010). *Handbook of Spatial Statistics*. Chapman & Hall/CRC.

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. Wiley.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer.

Chilès, J. P. and Delfiner, P. (2012). *Geostatistics. Modeling Spatial Uncertainty*. Second edition. Wiley.

Schoenberg, I. J. (1938). Metric spaces and completely monotone functions. *Annals of Mathematics*, **39**, 811–841.

Guttorp, P. and Gneiting, T. (2006). Studies in the history of probability and statistics XLIX: On the Matérn correlation family. *Biometrika*, **93**, 989–995.

Gneiting, T. and Schlather, M. (2004). Stochastic models that separate fractal dimension and the Hurst effect. *SIAM Review*, **46**, 269–282.

Sampson, P. D. and Guttorp, P. (1992). Nonparametric estimation of non-stationary spatial covariance structure. *Journal of the American Statistical Association*, **87**, 108–119.

Paciorek, C. and Schervish, M. J. (2006). Spatial modelling using a new class of non-stationary covariance functions. *Environmetrics*, **17**, 483–506.

Part 2

Random fields on spheres

2.1 Global data

2.2 Isotropic correlation functions on spheres

2.3 Examples

2.4 Some challenges

2.1 Global data

in geophysical, meteorological and climatological applications, data and processes need to be modeled on the **surface** of the **Earth**, rather than a Euclidean space

this motivates the study of **random fields**

$$\left\{ Z(x) : x \in \mathbb{S}^d \right\},$$

where $\mathbb{S}^d = \{x \in \mathbb{R}^{d+1} : |x| = 1\}$ denotes the **unit sphere** in $\mathbb{R}^{d+1}$

the process is **isotropic** if there exists a function $\psi : [0, \pi] \to \mathbb{R}$ such that

$$\text{cov}(Z(x), Z(y)) = \psi(\theta(x, y)) \quad \text{for} \quad x, y \in \mathbb{S}^d,$$

where

$$\theta(x, y) = \text{arccos}(\langle x, y \rangle)$$

is the **great circle** or **geodesic distance** on $\mathbb{S}^d$ and $\langle \cdot, \cdot \rangle$ denotes the scalar product in $\mathbb{R}^{d+1}$

2.2 Isotropic correlation functions on spheres

class Ψ_d:

$$\Psi_d = \left\{ \psi : [0,\pi] \to \mathbb{R} \;\middle|\; \begin{array}{c} \psi \text{ is continuous, } \psi(0) = 1, \text{ and} \\ \theta \mapsto \psi(\theta) \text{ is a correlation function on } \mathbb{S}^d \end{array} \right\}$$

interpretation of the class **Ψ_d:**

- continuous and isotropic **correlation functions**: there exists a process $\{ Z(x) : x \in \mathbb{S}^d \}$ such that $\mathrm{corr}(Z(x), Z(y)) = \psi(\theta(x,y))$
- continuous and isotropic, standardized **positive definite functions**: given any finite system of real numbers $a_1, \ldots, a_m \in \mathbb{R}$ and locations $x_1, \ldots, x_m \in \mathbb{S}^d$,

$$\sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j \, \psi(\theta(x_i, x_j)) \geq 0$$

quite obviously,

$$\Psi_1 \supset \Psi_2 \supset \cdots \qquad \text{and} \qquad \Psi_d \downarrow \Psi_\infty = \bigcap_{d=1}^{\infty} \Psi_d$$

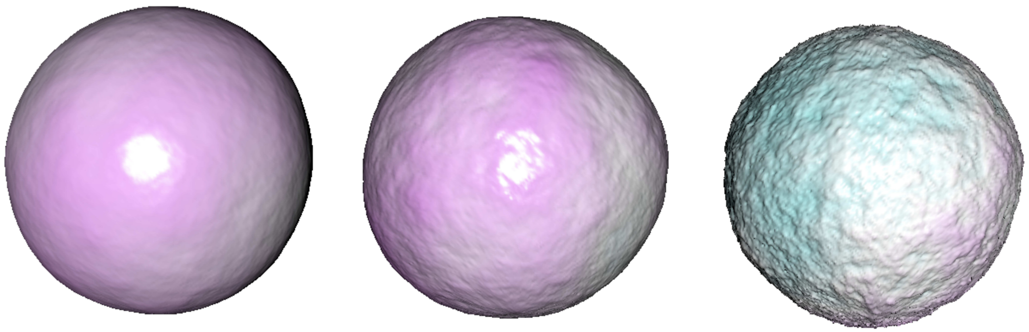
2.2 Isotropic correlation functions on spheres

natural construction: if $\varphi : [0, \infty) \to \mathbb{R}$ is a member of the class Φ_{d+1} for some $d \geq 1$, then the function

$$\psi : [0, \pi] \to \mathbb{R}, \qquad \theta \mapsto \varphi\left(2 \sin \frac{\theta}{2}\right)$$

corresponds to the **restriction** of an isotropic correlation function in $\mathbb{R}^{d+1}$ to the sphere $\mathbb{S}^d$ and thus it **belongs to** the class $\boldsymbol{\Psi_d}$

the construction **preserves** the **fractal index**: if φ has fractal index $\alpha \in (0, 2]$ then $\psi(0) - \psi(\theta) = \mathcal{O}(\theta^\alpha)$ as $\theta \downarrow 0$



$\alpha = 1.9$, $D = 2.05$ $\qquad$ $\alpha = 1.5$, $D = 2.25$ $\qquad$ $\alpha = 1.0$, $D = 2.50$

2.2 Isotropic correlation functions on spheres

however, this natural construction is restrictive

to **characterize** the class $\boldsymbol{\Psi_d}$, we consider **orthogonal polynomials** with argument $\cos\theta$

given $\boldsymbol{\lambda > 0}$ and an integer $\boldsymbol{k \geq 0}$, the function $\boldsymbol{C_k^\lambda(\cos\theta)}$ is defined by the expansion

$$\frac{1}{(1+r^2-2r\cos\theta)^\lambda} = \sum_{k=0}^{\infty} r^k \, C_k^\lambda(\cos\theta) \quad \text{for} \quad \theta \in [0,\pi],$$

where $r \in (-1,1)$ and $\boldsymbol{C_k^\lambda}$ is the **ultraspherical** or **Gegenbauer polynomial** of order λ and degree k

in particular, $\boldsymbol{C_0^\lambda(\cos\theta) \propto 1}$ and $\boldsymbol{C_1^\lambda(\cos\theta) \propto \cos\theta}$ for all $\lambda > 0$

the class $\boldsymbol{\Psi_d}$ is **convex**, and thus we expect **Choquet representations** in terms of **extreme members** in analogy to Φ_d

2.2 Isotropic correlation functions on spheres

Theorem (Schoenberg):

(a) The members of the class $\boldsymbol{\Psi_1}$ are of the form

$$\psi(\theta) = \sum_{k=0}^{\infty} b_{1,k} \cos(k\theta)$$

with coefficients $b_{1,k} \geq 0$ that sum to 1.

(b) If $\boldsymbol{d \geq 2}$, the members of the class $\boldsymbol{\Psi_d}$ are of the form

$$\psi(\theta) = \sum_{k=0}^{\infty} b_{d,k} \frac{C_k^{(d-1)/2}(\cos\theta)}{C_k^{(d-1)/2}(1)}$$

with coefficients $b_{d,k} \geq 0$ that sum to 1.

(c) The members of the class $\boldsymbol{\Psi_\infty}$ are of the form

$$\psi(\theta) = \sum_{k=0}^{\infty} b_{\infty,k} (\cos\theta)^k$$

with coefficients $b_{\infty,k} \geq 0$ that sum to 1.

2.2 Isotropic correlation functions on spheres

the members of the class $\boldsymbol{\Psi_1}$ admit the **Fourier cosine** representation

$$\psi(\theta) = \sum_{k=0}^{\infty} b_{1,k} \cos(k\theta),$$

where

$$b_{1,k} = \frac{2}{\pi} \int_0^{\pi} \cos(k\theta)\, \psi(\theta)\, \mathrm{d}\theta \quad \text{for} \quad k \geq 1$$

the members of the class $\boldsymbol{\Psi_2}$ admit the **Legendre** representation

$$\psi(\theta) = \sum_{k=0}^{\infty} \frac{b_{2,k}}{k+1} P_k(\cos\theta),$$

where $\boldsymbol{P_k}$ is the **Legendre polynomial** of degree $k \geq 0$ and

$$b_{2,k} = \frac{2k+1}{2} \int_0^{\pi} P_k(\cos\theta)\, \sin\theta\, \psi(\theta)\, \mathrm{d}\theta$$

2.2 Isotropic correlation functions on spheres

generally, if $d \geq 2$ the coefficient $b_{d,k}$ in the **Gegenbauer expansion** of a function $\psi \in \Psi_d$ equals

$$b_{d,k} = \frac{2k+d-1}{2^{3-d}\,\pi} \frac{(\Gamma(\frac{d-1}{2}))^2}{\Gamma(d-1)} \int_0^\pi C_k^{(d-1)/2}(\cos\theta)\,(\sin\theta)^{d-1}\,\psi(\theta)\,\mathrm{d}\theta$$

the **Fourier cosine** and **Gegenbauer** coefficients of a continuous function $\psi : [0,\pi] \to \mathbb{R}$ relate to each other in **dimension walks**, in that

$$b_{d+2,k} = \frac{(k+d-1)(k+d)}{d\,(2k+d-1)}\, b_{d,k} - \frac{(k+1)(k+2)}{d\,(2k+d+3)}\, b_{d,k+2}$$

for integers $d \geq 1$ and $k \geq 1$

in particular,

$$b_{3,k} = \frac{1}{2}(k+1)\left(b_{1,k} - b_{1,k+2}\right),$$

whence a function $\psi \in \Psi_1$ belongs to the class Ψ_3 if, and only if, the sequences of both the even and the odd **Fourier cosine** coefficients are **nonincreasing**

2.3 Examples

Theorem: If $\varphi \in \mathbf{\Phi}_\infty$ then the function

$$\psi : [0, \pi] \to \mathbb{R}, \qquad \theta \mapsto \varphi(\theta^{1/2})$$

belongs to the class $\mathbf{\Psi}_\infty$.

Family	Analytic Expression	Valid Parameter Range
Powered exponential	$\psi(\theta) = \exp\left(-\left(\frac{\theta}{c}\right)^\alpha\right)$	$c > 0;\ \alpha \in (0, 1]$
Cauchy	$\psi(\theta) = \left(1 + \left(\frac{\theta}{c}\right)^\alpha\right)^{-\tau/\alpha}$	$c > 0;\ \alpha \in (0, 1];\ \tau > 0$
Matérn	$\psi(\theta) = \frac{2^{\nu-1}}{\Gamma(\nu)} \left(\frac{\theta^{1/2}}{c}\right)^\nu K_\nu\left(\frac{\theta^{1/2}}{c}\right)$	$c > 0;\ \nu > 0$
Matérn	$\psi(\theta) = \frac{2^{\nu-1}}{\Gamma(\nu)} \left(\frac{\theta}{c}\right)^\nu K_\nu\left(\frac{\theta}{c}\right)$	$c > 0;\ \nu \in (0, \frac{1}{2}]$

2.3 Examples

Theorem: Suppose that $\boldsymbol{d = 1}$ or $\boldsymbol{d = 3}$. If the member $\boldsymbol{\varphi \in \Phi_d}$ satisfies $\varphi(t) = 0$ for $t \geq \pi$, the function defined by

$$\psi : [0, \pi] \to \mathbb{R}, \qquad \theta \mapsto \varphi(\theta)$$

belongs to the class $\boldsymbol{\Psi_d}$.

thus, if $d = 1$ or $d = 3$ **compactly supported** members of the class Φ_d induce **locally supported** function in the class Ψ_d

Family	Analytic Expression	Parameter Range in Ψ_3
Spherical	$\psi(\theta) = \left(1 + \frac{1}{2}\frac{\theta}{c}\right)\left(1 - \frac{\theta}{c}\right)_+^2$	$c \in (0, \pi]$
Askey	$\psi(\theta) = \left(1 - \frac{\theta}{c}\right)_+^\tau$	$c \in (0, \pi];\ \tau \geq 2$
Wendland	$\psi(\theta) = \left(1 + \tau\frac{\theta}{c}\right)\left(1 - \frac{\theta}{c}\right)_+^\tau$	$c \in (0, \pi];\ \tau \geq 4$

2.4 Some challenges

Problem 1. For $d \geq 1$ and $c \in (0, \pi)$, find the **infimum** of the **curvature** at the **origin** among the members ψ of the class Ψ_d with $\psi(\theta) = 0$ for $\theta \geq c$.

Problem 2. For $d \geq 1$ and $c \in (0, \pi)$, find the associated **Turán constant**, i.e., find the **supremum** of

$$\int_{\mathbb{S}^d} \psi(\theta(y_o, y))\, \mathrm{d}y$$

among the members ψ of the class Ψ_d with $\psi(\theta) = 0$ for $\theta \geq c$.

Problem 3. For $d \geq 1$, $c \in (0, \pi)$ and $\theta \in (0, c)$, find the associated **pointwise Turán constant**, that is, find the **supremum** of $\boldsymbol{\psi(\theta)}$ among the members ψ of the class Ψ_d with $\psi(\theta) = 0$ for $\theta \geq c$.

2.4 Some challenges

Conjecture 4. If $d = 2l + 1$ and $c \in (0, \pi)$ the **truncated power** function,

$$\psi(\theta) = \left(1 - \frac{\theta}{c}\right)_+^{\tau}$$

belongs to the class Ψ_d if, and only if, $\tau \geq l + 1$.

Problem 5. For what values of $\alpha \in (0, 2]$ and $c \in (0, \pi]$ does the **truncated sine power** function,

$$\psi(\theta) = \left(1 - \left(\sin \frac{\pi}{2} \frac{\theta}{c}\right)^{\alpha}\right) \mathbb{1}(\theta \leq c),$$

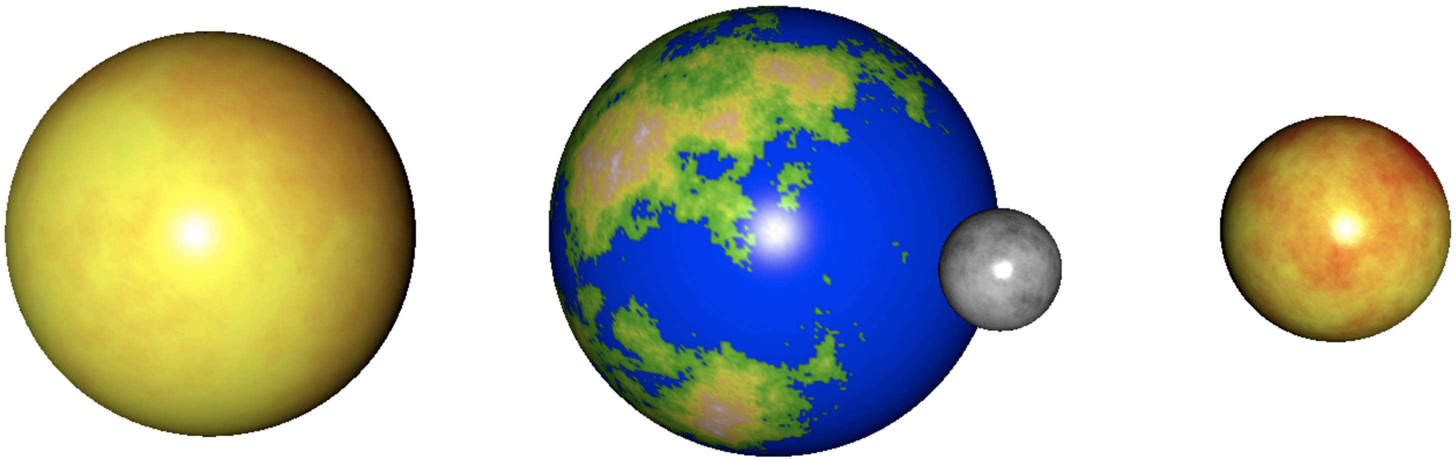
belong to the class Ψ_d?

Conjecture 6. If $\psi \in \Psi_d$ has **fractal index** $\alpha \in (0, 2]$ the associated regular Gaussian **particle** has fractal or **Hausdorff dimension** $d + 1 - \frac{\alpha}{2}$ almost surely.

2.4 Some challenges

Problem 7. Develop covariance models for **multivariate** stationary and isotropic random fields on $\mathbb{S}^d$.

Problem 8. Develop covariance models for **non-stationary** random fields on $\mathbb{S}^d$.



Problem 9. Develop covariance models for **spatio-temporal** random fields on the domain $\mathbb{S}^d \times \mathbb{R}$.

Part 2: Further reading

Gneiting, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli*, **19**, 1327–1349.

Ziegel, J. (2014). Convolution roots and differentiability of isotropic positive definite functions on spheres. *Proceedings of the American Mathematical Society*, in press.

Hansen, L. V., Thorarinsdottir, T. L. and Gneiting, T. (2011). Lévy particles: Modelling and simulating star-shaped random sets. Research Report 2011/04, Centre for Stochastic Geometry and Advanced Bioimaging, University of Aarhus, `http://data.imf.au.dk/publications/csgb/2011/imf-csgb-2011-04.pdf`.

Jun, M. and Stein, M. L. (2007). An approach to producing space-time covariance functions on spheres. *Technometrics*, **49**, 468–479.

Jun, M. (2011). Non-stationary cross-covariance models for multivariate processes on a globe. *Scandinavian Journal of Statistics*, **38**, 726–747.

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.

Part 3

Space-Time Covariance Functions

3.1 Geostatistical models for spatio-temporal data

3.2 Stationarity, separability and full symmetry

3.3 Stationary space-time covariance functions

3.4 Irish wind data

3.1 Geostatistical models for spatio-temporal data

spatio-temporal data in **environmental monitoring** or **meteorology** typically consist of multiple time series at scattered sites

Irish wind data (Haslett and Raftery 1989): daily averages of wind speed recorded at **11 meteorological stations** in Ireland in **1961–1978**

Station	Latitude	Longitude	Mean ($m \cdot s^{-1}$)
Valentia	51 56’	10 15’	5.48
Belmullet	54 14’	10 00’	6.75
Claremorris	53 43’	8 59’	4.32
Shannon	52 42’	8 55’	5.38
Roche’s Point	51 48’	8 15’	6.36
Birr	53 05’	7 53’	3.65
Mullingar	53 32’	7 22’	4.38
Malin Head	55 22’	7 20’	8.03
Kilkenny	52 40’	7 16’	3.25
Clones	54 11’	7 14’	4.48
Dublin	53 26’	6 15’	5.05

3.1 Geostatistical models for spatio-temporal data

a typical problem is **spatio-temporal prediction**:

given observations $Z(s_1, t_1), \ldots, Z(s_k, t_k)$, find a **point predictor** for the unknown value of Z the space-time coordinate (s_0, t_0) or, preferably, find a **conditional distribution** for $Z(s_0, t_0)$

geostatistical space-time models: spatio-temporal, typically Gaussian **random field**

$$\left\{ Z(s,t) : (s,t) \in \mathbb{R}^d \times \mathbb{R} \right\}$$

indexed in continuous **space, $s \in \mathbb{R}^d$**, and continuous **time, $t \in \mathbb{R}$**

allows for prediction and interpolation at any location and any time — typically, but not necessarily time-forward

alternative spatio-temporal **domains**: $\mathbb{R}^d \times \mathbb{Z}$ (discrete time), $\mathbb{S}^d \times \mathbb{R}$ or $\mathbb{S}^d \times \mathbb{Z}$ (**global** models with the **sphere** as spatial domain)

3.2 Stationarity, separability and full symmetry

spatio-temporal, typically Gaussian **random field**

$$\left\{ Z(s,t) : (s,t) \in \mathbb{R}^d \times \mathbb{R} \right\}$$

Gaussian processes are characterized by their **first** and **second moments**

after spatio-temporal **trend removal**, we may assume that $Z(s,t)$ has **mean zero**

simplifying assumptions on the **covariance structure**:

the spatio-temporal random field has **stationary** covariance structure if $\text{cov}\{Z(s_1,t_1), Z(s_2,t_2)\}$ depends only on the spatial separation vector, $s_1 - s_2 \in \mathbb{R}^d$, and the temporal lag, $t_1 - t_2 \in \mathbb{R}$

then there exists a **space-time covariance** function

$$C : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}, \quad (h,u) \mapsto C(h,u) = \text{cov}\{Z(s,t), Z(s+h,t+u)\}$$

3.2 Stationarity, separability and full symmetry

simplifying assumptions on the **covariance structure**:

the spatio-temporal random field has **fully symmetric** covariance structure if

$$\text{cov}\{Z(s_1, t_1), Z(s_2, t_2)\} = \text{cov}\{Z(s_1, t_2), Z(s_2, t_1)\}$$

physically interpretable in terms of **transport effects** and **prevailing winds** or **ocean currents**

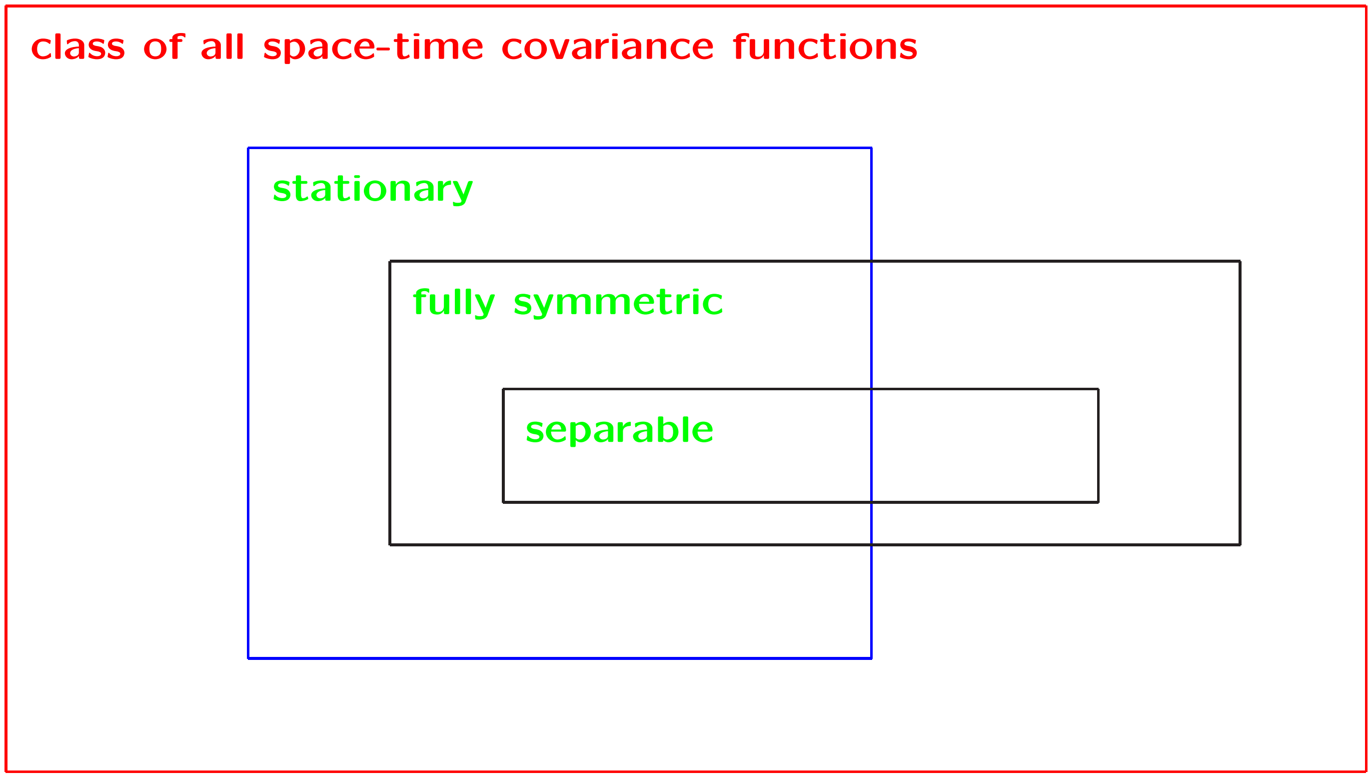
Westerly Station	Easterly Station	r_{WE}	r_{EW}
Valentia	Roche's Point	0.50	0.37
Belmullet	Clones	0.52	0.40
Claremorris	Mullingar	0.51	0.42
Claremorris	Dublin	0.51	0.38

the spatio-temporal random field has **separable** covariance structure if there exist purely spatial and purely temporal covariance functions cov_{S} and cov_{T} such that

$$\text{cov}\{Z(s_1, t_1), Z(s_2, t_2)\} = \text{cov}_{\text{S}}(s_1, s_2) \times \text{cov}_{\text{T}}(t_1, t_2)$$

3.2 Stationarity, separability and full symmetry

separability is a special case of **full symmetry** that may yield mathematical tractability and computational efficiency



3.3 Stationary space-time covariance functions

we consider a spatio-temporal, typically Gaussian **random field**

$$\left\{ Z(s,t) : (s,t) \in \mathbb{R}^d \times \mathbb{R} \right\}$$

with **stationary** covariance structure and **space-time covariance** function

$$C(h,u) = \text{cov}\{Z(s,t), Z(s+h,t+u)\}$$

the space-time covariance function is **separable** if there exist purely spatial and purely temporal covariance functions C_{S} and C_{T} such that

$$C(h,u) = C_{\mathsf{S}}(h) \times C_{\mathsf{T}}(u)$$

and it is **fully symmetric** if $C(h,u) = C(h,-u)$ for all spatio-temporal lags $(h,u) \in \mathbb{R}^d \times \mathbb{R}$

the space-time covariance function is a **positive definite** function in the Euclidean space $\mathbb{R}^{d+1} = \mathbb{R}^d \times \mathbb{R}$

in particular, **Bochner's theorem** applies

3.3 Stationary space-time covariance functions

Theorem (Cressie and Huang): A continuous, bounded, symmetric and integrable function

$$C : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$$

is a stationary **space-time covariance** function if, and only if, the function

$$C_\omega : \mathbb{R} \to \mathbb{R}, \qquad u \mapsto C_\omega(u) = \int_{\mathbb{R}^d} e^{-ih'\omega} C(h, u) \, \mathrm{d}h$$

is **positive definite** for almost all $\omega \in \mathbb{R}^d$.

Cressie and Huang (1999) used **Fourier inversion** of $C_\omega(u)$ with respect to $\omega \in \mathbb{R}^d$ to construct parametric classes of **non-separable space-time covariance** functions

3.3 Stationary space-time covariance functions

a function $\eta(t)$, defined for $t \geq 0$ or $t > 0$, is **completely monotone** if

$$(-1)^k \eta^{(k)}(t) \geq 0 \qquad (t > 0;\ k = 0, 1, 2, \ldots)$$

e.g., if $\varphi \in \Phi_\infty$ then $t \mapsto \eta(t) = \varphi(t^{1/2})$ is completely monotone

Theorem:

$\eta(t)$, $t \geq 0$ completely monotone
$\zeta(t)$, $t \geq 0$ strictly positive with a completely monotone derivative

Then

$$C : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}, \qquad (h, u) \mapsto C(h, u) = \frac{1}{\zeta(u^2)^{d/2}}\, \eta\left(\frac{|h|^2}{\zeta(u^2)}\right)$$

is a **space-time covariance** function

3.3 Stationary space-time covariance functions

the proof depends on the **Cressie-Huang** criterion and **Bernstein**'s theorem:

$$
\begin{aligned}
C_\omega(u) &= \int_{\mathbb{R}^d} e^{-ih'\omega}\, C(h;u)\, \mathrm{d}h \\
&= \frac{1}{\zeta(u^2)^{d/2}} \int_{\mathbb{R}^d} e^{-ih'\omega}\, \eta\!\left(\frac{|h|^2}{\zeta(u^2)}\right) \mathrm{d}h \\
&= \frac{1}{\zeta(u^2)^{d/2}} \int_{\mathbb{R}^d} \int_{(0,\infty)} e^{-ih'\omega}\, \exp\!\left(-r\frac{|h|^2}{\zeta(u^2)}\right) \mathrm{d}F(r)\, \mathrm{d}h \\
&= \pi^{d/2} \int_{(0,\infty)} \exp\!\left(-\frac{|\omega|^2}{4r}\,\zeta\!\left(u^2\right)\right) \frac{1}{r^{d/2}}\, \mathrm{d}F(r) \\
&= \varphi_\omega\!\left(u^2\right)
\end{aligned}
$$

where

$$
\varphi_\omega(t) = \pi^{d/2} \int_{(0,\infty)} \exp(-s\zeta(t))\, \mathrm{d}G_\omega(s)
$$

for a nondecreasing function G_ω

by Bernstein's theorem, φ_ω is **completely monotone**, whence C_ω is **positive definite**

3.3 Stationary space-time covariance functions

Family	Analytic Expression	Parameter Range
Matérn	$\eta(t) = \frac{2^{\nu-1}}{\Gamma(\nu)} \left(ct^{1/2}\right)^{\nu} K_{\nu}\left(ct^{1/2}\right)$	$c > 0;\ \nu > 0$
Power exp	$\eta(t) = \exp\left(-\, ct^{\gamma}\right)$	$c > 0;\ \gamma \in (0, 1]$
Cauchy	$\eta(t) = (1 + ct^{\gamma})^{-\tau}$	$c > 0;\ \gamma \in (0, 1];\ \tau > 0$
Logistic	$\eta(t) = 2^{\tau} \left(\exp(ct^{1/2}) + \exp(-ct^{1/2})\right)^{-\tau}$	$c > 0;\ \tau > 0$

Analytic Expression	Parameter Range
$\zeta(t) = (at^{\alpha} + 1)^{\beta}$	$a > 0;\ \alpha \in (0, 1];\ \beta \in [0, 1]$
$\zeta(t) = \frac{\ln(at^{\alpha} + b)}{\ln b}$	$a > 0;\ b > 1;\ \alpha \in (0, 1]$
$\zeta(t) = \frac{1}{b} \frac{at^{\alpha} + b}{at^{\alpha} + 1}$	$a > 0;\ b \in (0, 1];\ \alpha \in (0, 1]$

3.3 Stationary space-time covariance functions

example:

$\eta(t) = \exp(-ct^{\gamma})$

$\zeta(t) = (at^{\alpha} + 1)^{\beta}$

product with $C_{\mathsf{T}}(u) = \sigma^2(a|u|^{2\alpha} + 1)^{\beta d/2-1}$ where $d = 2$

$$C(h,u) = \frac{\sigma^2}{a|u|^{2\alpha} + 1} \exp\left(-\frac{c|h|^{2\gamma}}{(a|u|^{2\alpha} + 1)^{\beta\gamma}}\right)$$

parameters:

$c > 0$	space: scale
$\gamma \in (0, 1]$	space: smoothness
$a > 0$	time: scale
$\alpha \in (0, 1]$	time: smoothness
$\sigma^2 > 0$	space-time: variance
$\beta \in [0, 1]$	space-time: interaction

3.3 Stationary space-time covariance functions

full symmetry often violated as a result of prevailing winds or ocean currents and associated **transport effects**

useful idea: **Lagrangian** reference frame

specifically, if $C_{\mathsf{S}} : \mathbb{R}^d \to \mathbb{R}$ is a stationary **spatial covariance** function and V is an $\mathbb{R}^d$-valued **random vector** then

$$C(h,u) = \mathbb{E}\, C_{\mathsf{S}}(h - Vu)$$

is a stationary **space-time covariance** function

covariance function of a **frozen** spatial random field that moves with **random velocity** V in $\mathbb{R}^d$; e.g.,

- $V = v$ constant, representing the **prevailing wind**
- law of V is equal to or fitted from the **empirical distribution** of historical wind vectors
- law of V is **flow-dependent** and governed by the current state of the atmosphere

3.4 Irish wind data

Haslett and Raftery (1989)
`http://lib.stat.cmu.edu/datasets/`

11 meteorological stations in **Ireland**
1961 to 1978
daily average wind speed in $m \cdot s^{-1}$

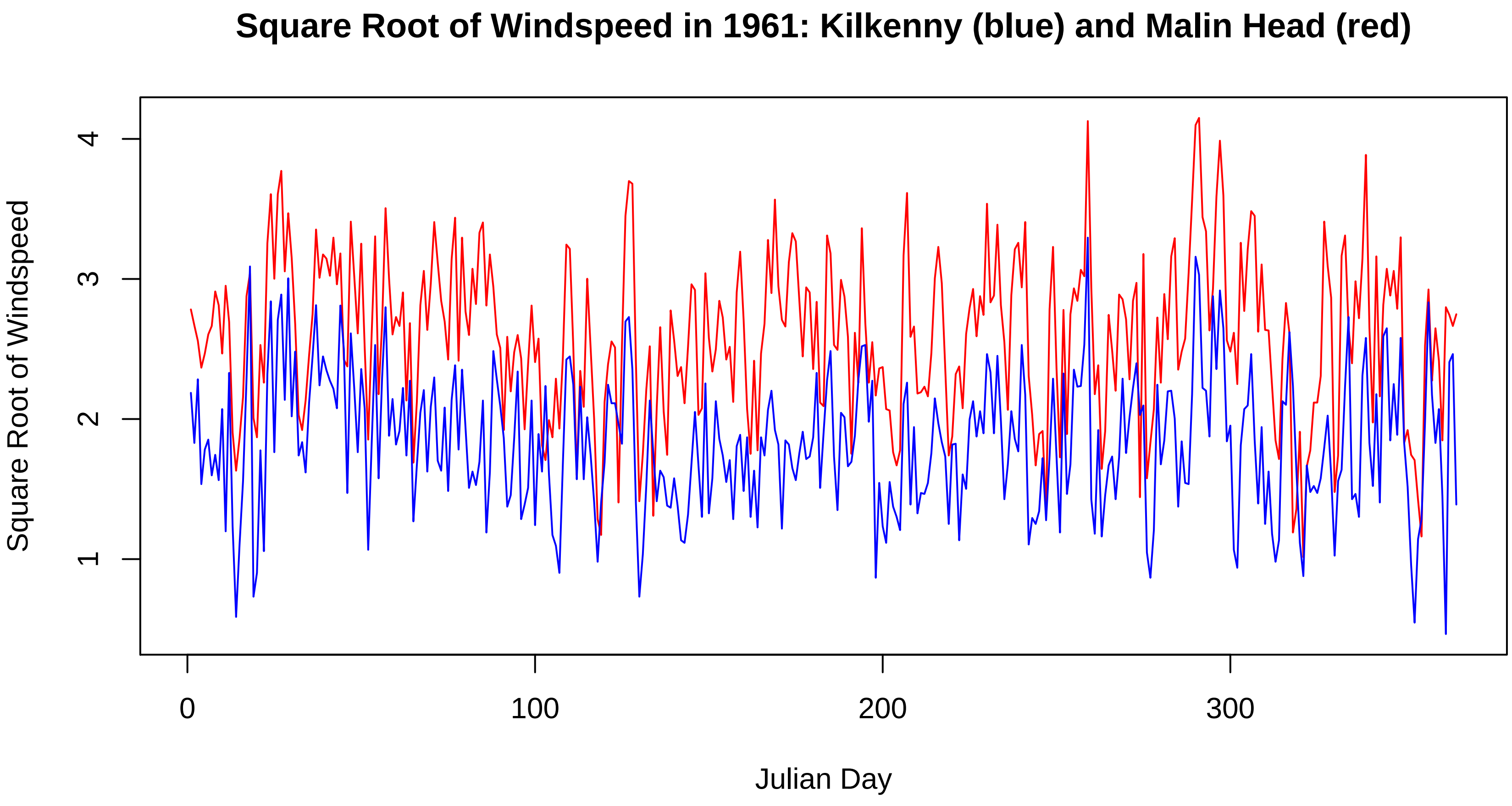
Station	Latitude	Longitude	Mean ($m \cdot s^{-1}$)
Valentia	51 56'	10 15'	5.48
Belmullet	54 14'	10 00'	6.75
Claremorris	53 43'	8 59'	4.32
Shannon	52 42'	8 55'	5.38
Roche's Point	51 48'	8 15'	6.36
Birr	53 05'	7 53'	3.65
Mullingar	53 32'	7 22'	4.38
Malin Head	55 22'	7 20'	8.03
Kilkenny	52 40'	7 16'	3.25
Clones	54 11'	7 14'	4.48
Dublin	53 26'	6 15'	5.05

3.4 Irish wind data

square root transform

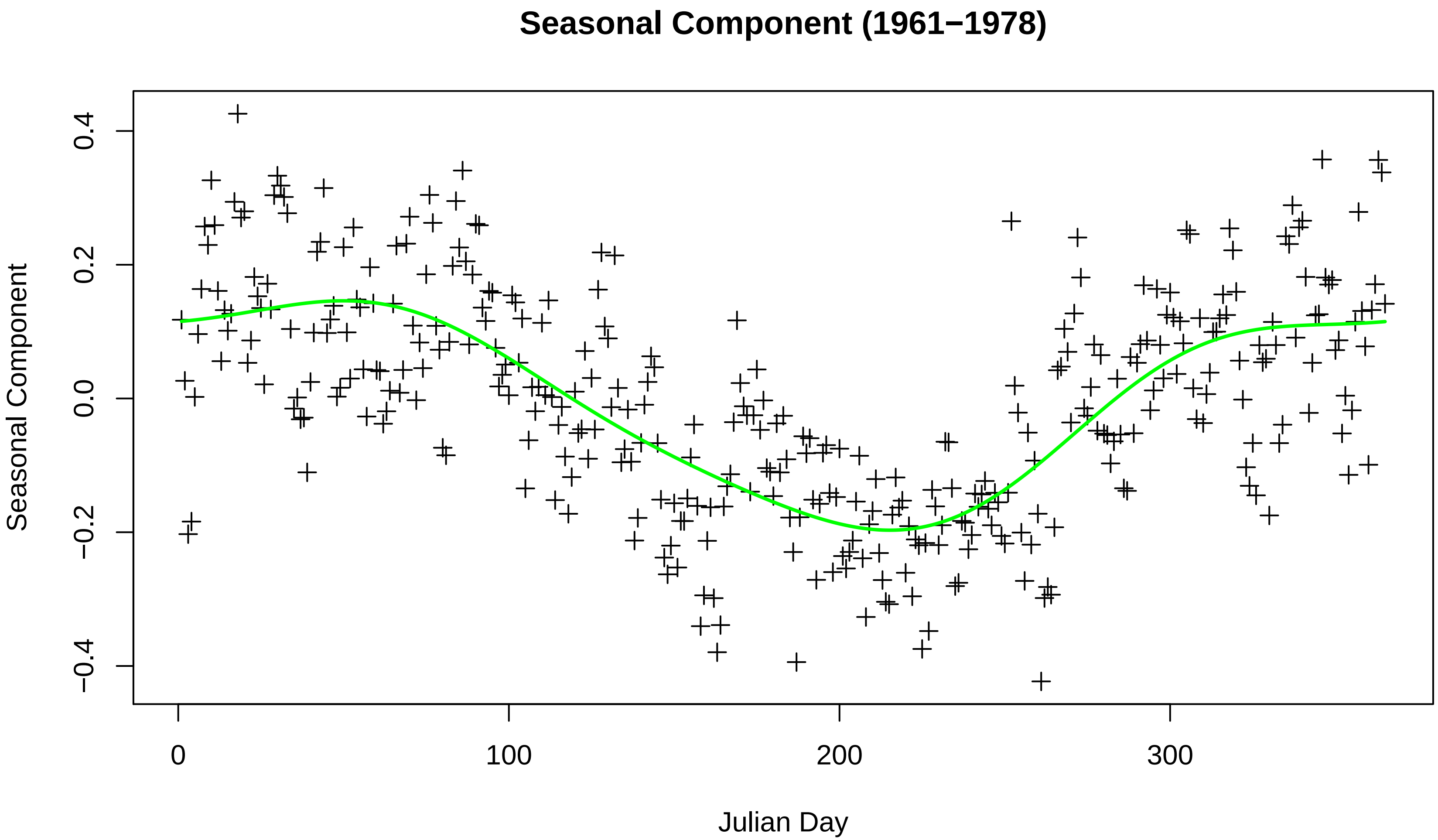
⇒ variance stabilized over stations and time

⇒ marginals approximately normal



3.4 Irish wind data

seasonal component estimated by least squares using a trigonometric regression model



3.4 Irish wind data

seasonal component and site specific mean removed
$\Rightarrow$ **velocity measures**

stationarity and **spatial isotropy** appropriate approximations

successively more general fits that we discuss in terms of **correlation** structures:

- **separable** space-time correlation function
- non-separable but **fully symmetric** correlation function
- general **stationary** correlation function

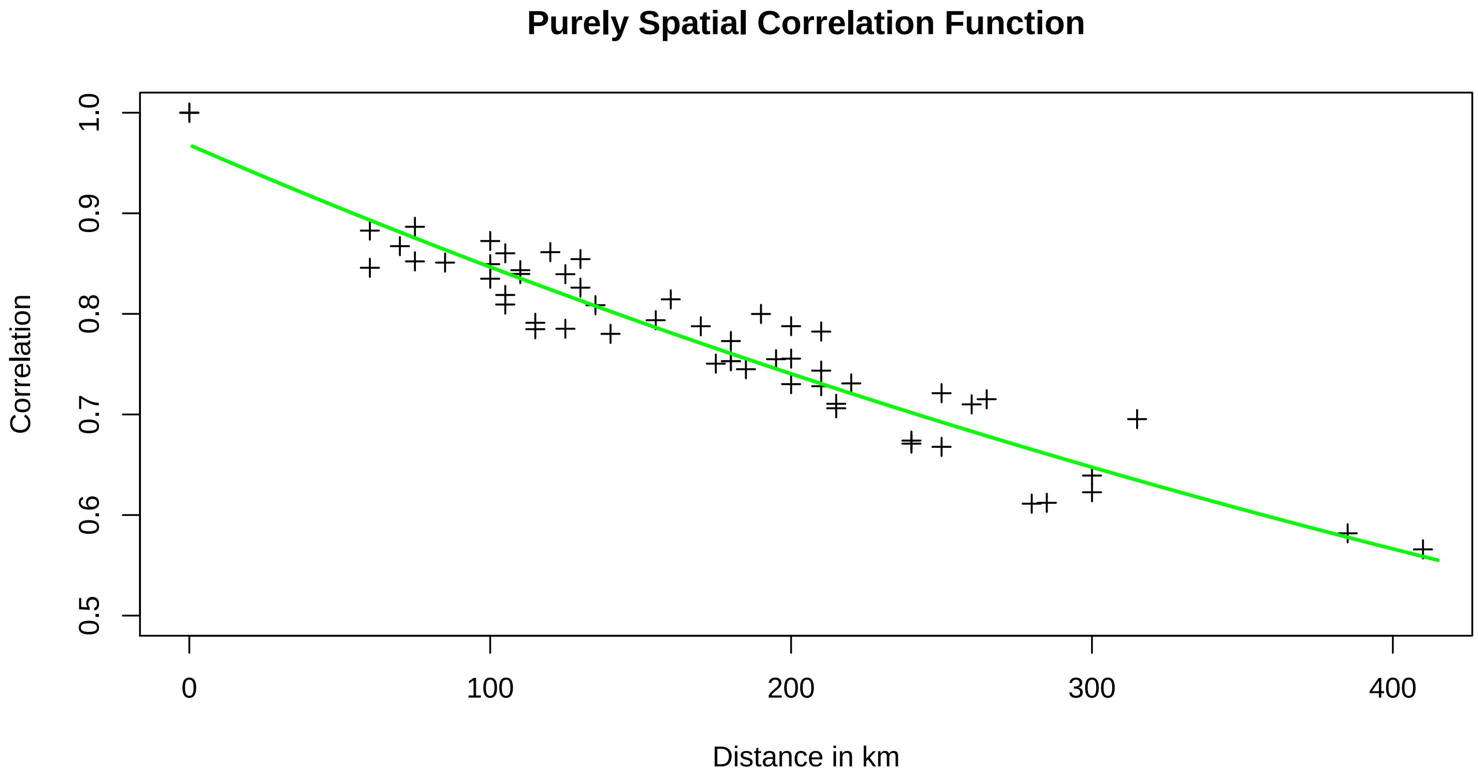
3.4 Irish wind data

purely spatial correlation function

convex combination of an **exponential** model and a **nugget effect**

$$C_S(h) = (1 - \nu) \exp(-c|h|) + \nu \delta_{h=0}.$$

with estimates $\hat{\nu} = 0.032$ and $\hat{c} = 0.00132$



3.4 Irish wind data

purely temporal correlation function

$$C_{\mathsf{T}}(u) = \frac{1}{1 + a|u|^{2\alpha}}$$

with estimates $\hat{a} = 0.901$ and $\hat{\alpha} = 0.772$

separable space-time correlation function

$$\begin{aligned} C_{\mathsf{SEP}}(h,u) &= C_{\mathsf{S}}(h) \times C_{\mathsf{T}}(u) \\ &= \frac{1-\nu}{1 + a|u|^{2\alpha}} \left(\exp(-c|h|) + \frac{\nu}{1-\nu}\, \delta_{h=0} \right) \end{aligned}$$

fitted to spatio-temporal lags (h, u) where $|h| \leq 400$ km and $|u| \leq 3$ days

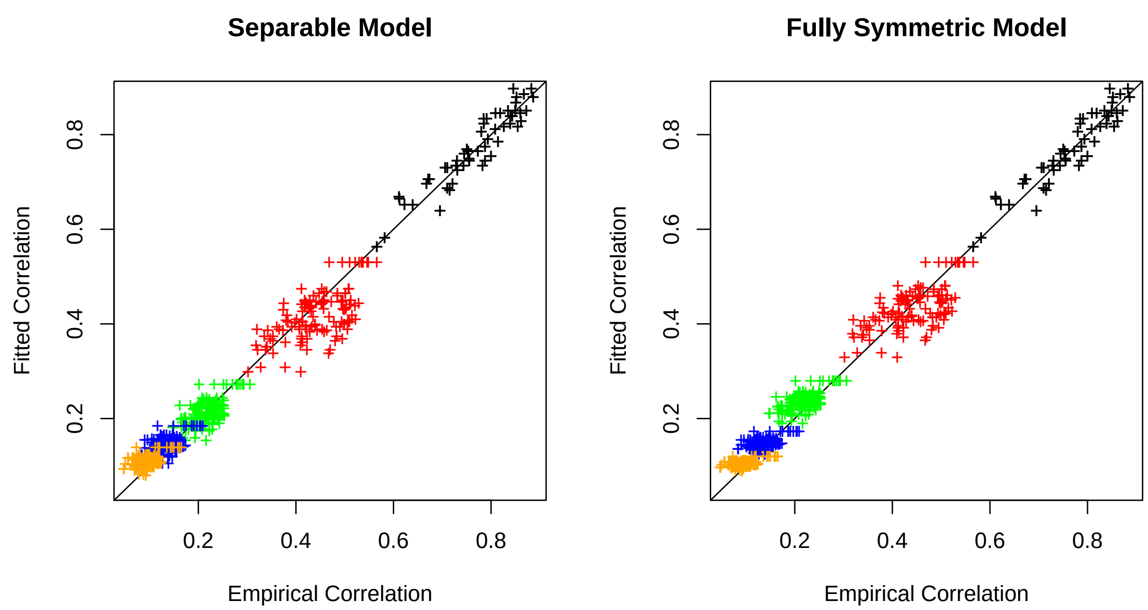
3.4 Irish wind data

we embed the separable model into a class of generally **non-separable** yet **fully symmetric space-time correlation** functions

$$C_{\mathrm{FS}}(h,u) = \frac{1-\nu}{1+a|u|^{2\alpha}} \left(\exp\left(-\frac{c|h|}{(1+a|u|^{2\alpha})^{\beta/2}} \right) + \frac{\nu}{1-\nu}\,\delta_{h=0} \right)$$

now including a space-time **interaction parameter** $\beta \in [0,1]$, where $\beta = 0$ corresponds to the separable model

weighted least squares estimate is $\widehat{\beta} = \mathbf{0.61}$

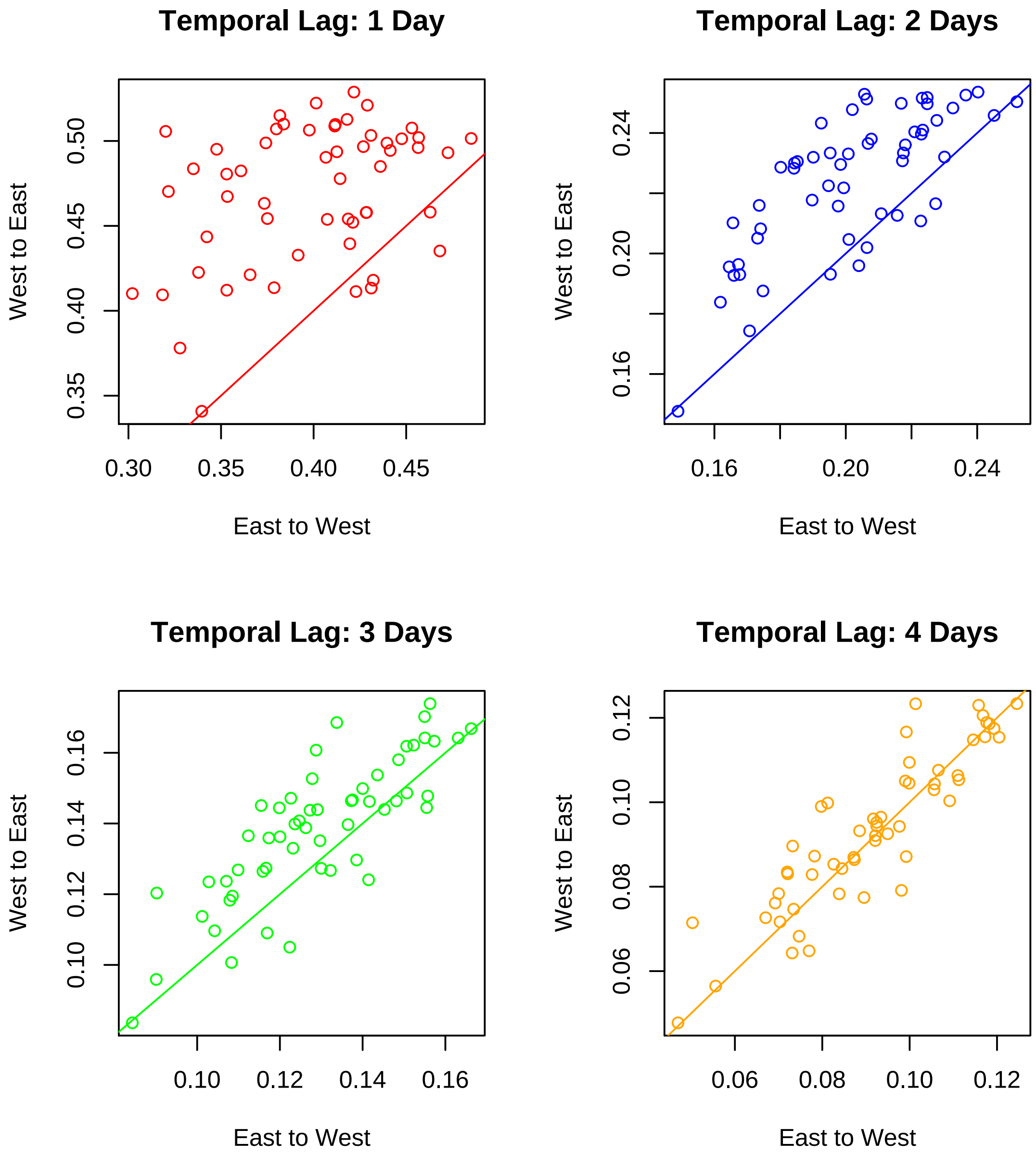


3.4 Irish wind data

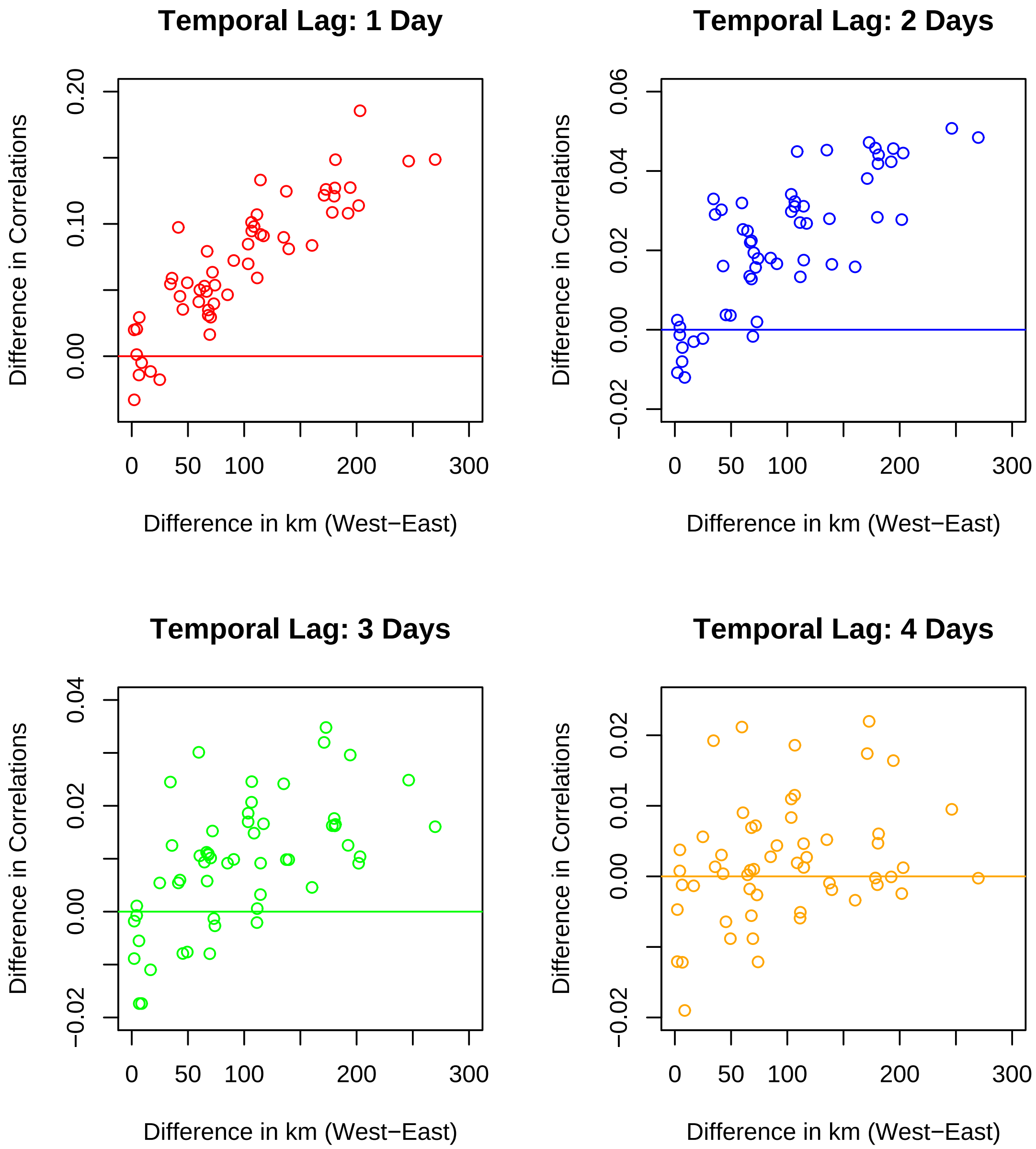
however, the assumption of **full symmetry** is clearly violated:

Westerly Station	Easterly Station	r_{WE}	r_{EW}
Valentia	Roche's Point	0.50	0.37
Belmullet	Clones	0.52	0.40
Claremorris	Mullingar	0.51	0.42
Claremorris	Dublin	0.51	0.38
Shannon	Kilkenny	0.53	0.42
Mullingar	Dublin	0.50	0.45

3.4 Irish wind data



3.4 Irish wind data



3.4 Irish wind data

general stationary space-time correlation function

convex combination

$$C_{\text{STN}}(h,u) = (1-\lambda)\,C_{\text{FS}}(h,u) + \lambda\,C_{\text{LGR}}(h,u)$$

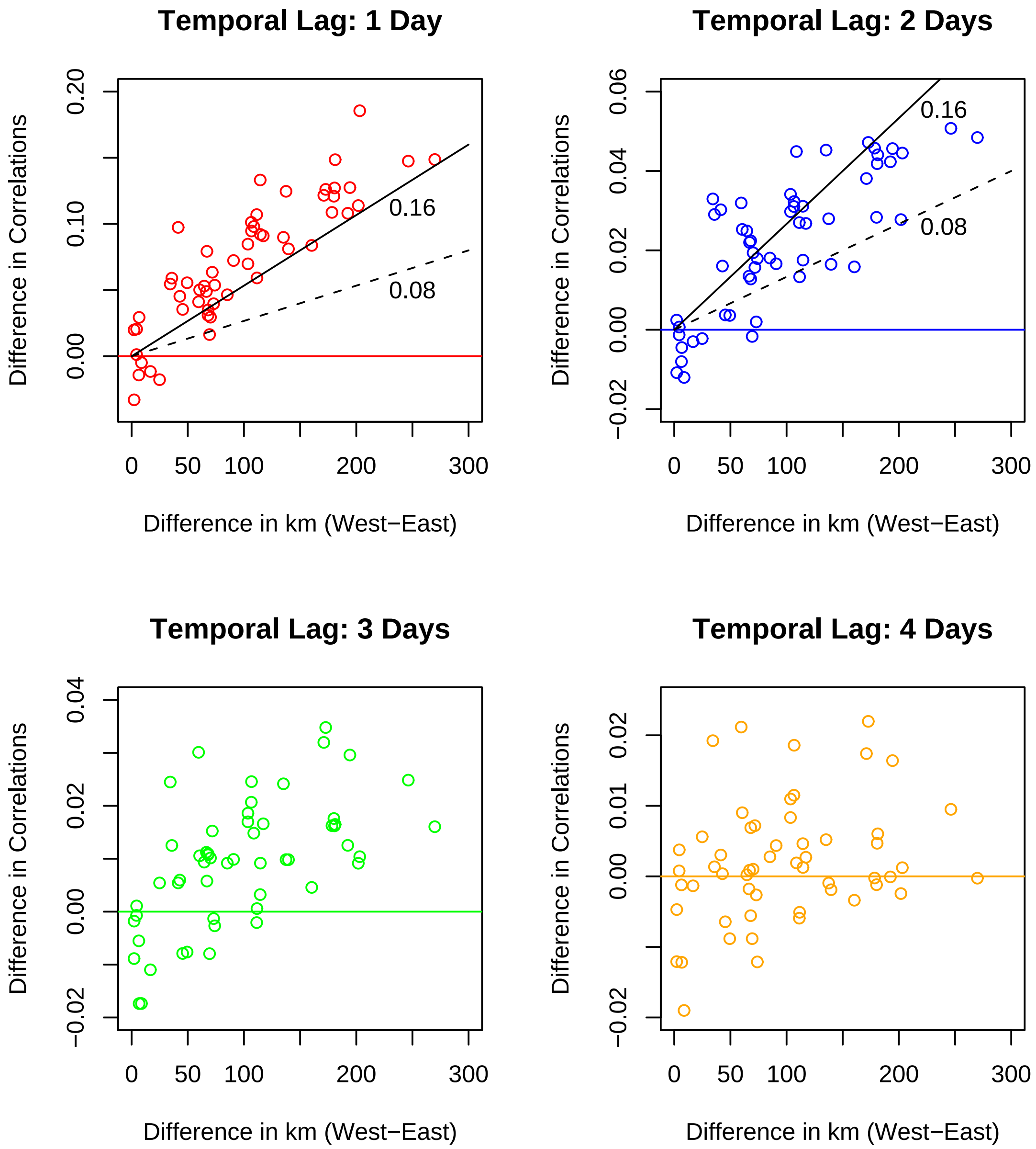
of the **fully symmetric** and a **Lagrangian** correlation model,

$$C_{\text{LGR}}(h,u) = \left(1 - \frac{1}{2v}\,|h_1 - vu|\right)_+$$

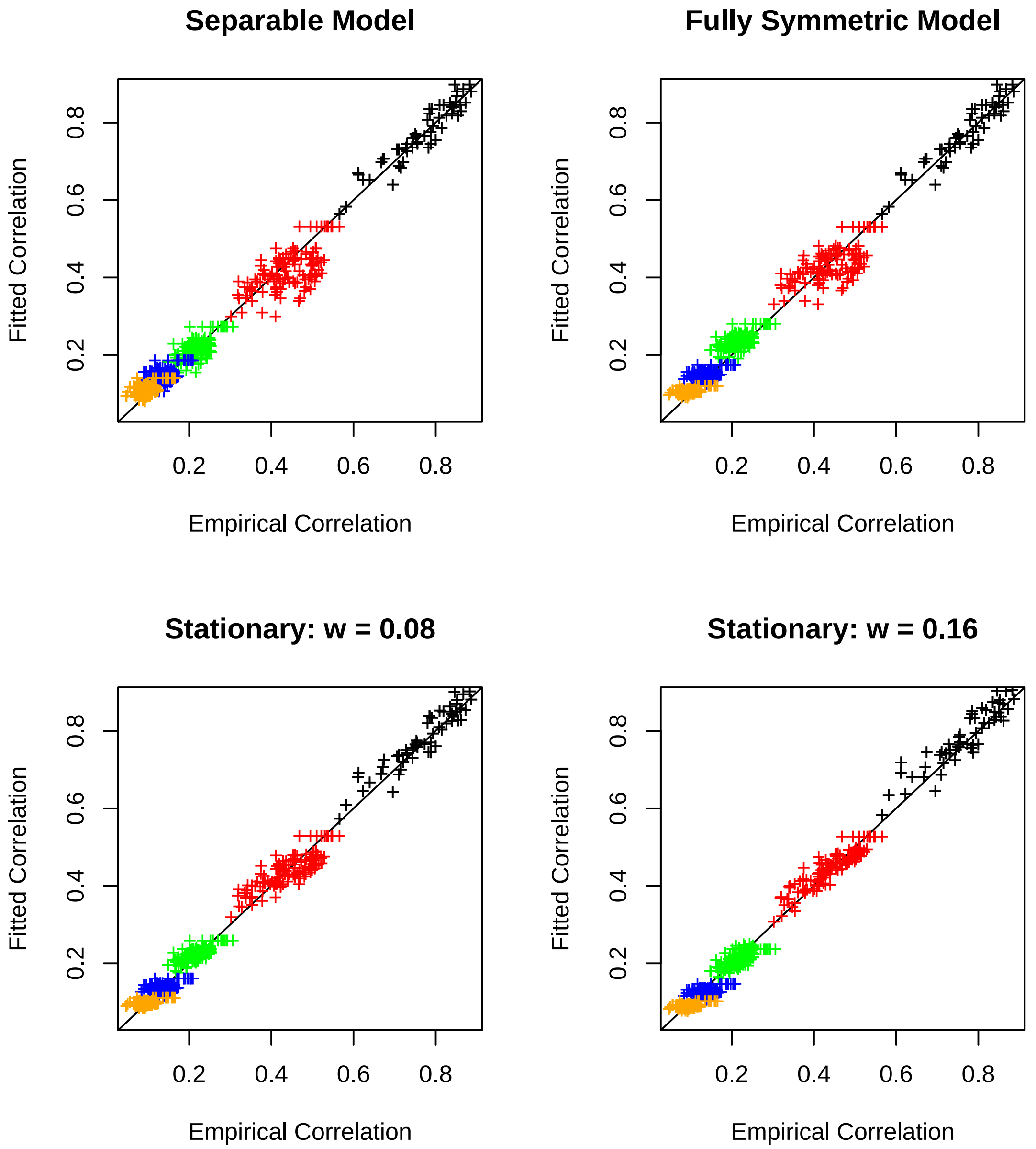
with **longitudinal velocity** $v = 300$ km$\cdot$d^{-1} or 3.5 m$\cdot$s^{-1}, where $h = (h_1, h_2)'$ has longitudinal (east-west) component h_1

a physically motivated and **physically justifiable** correlation model that is no longer fully symmetric

3.4 Irish wind data



3.4 Irish wind data



3.4 Irish wind data

for Gaussian processes, **conditional distributions** for unknown values are **Gaussian**

and we can take the **center** of the conditional distribution as a **point forecast**

mean absolute error (MAE) for 1-day ahead **point forecasts** of the velocity measures at the 11 meteorological stations during the 1971–1978 **test period**, based on fits in 1961–1970

	Val	Bel	Clo	Sha	Roc	Birr	Mul	Mal	Kil	Clo	Dub
SEP	.398	.395	.389	.372	.387	.375	.340	.399	.347	.385	.359
FS	.399	.396	.389	.372	.384	.373	.338	.396	.344	.382	.356
STAT	.397	.395	.387	.369	.379	.370	.334	.393	.339	.377	.351

the more complex and **more physically realistic** correlation models yield **improved** predictive performance

Part 3: Further reading

Haslett, J. and Raftery, A. E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland's wind-power resource. *Applied Statistics*, **38**, 1–50.

Cressie, N. and Huang, H.-C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, **94**, 1330–1340, **96**, 784.

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

Gneiting, T., Genton, M. G. and Guttorp, P. (2007). Geostatistical space-time models, stationarity, separability, and full symmetry. In *Statistical Methods for Spatio-Temporal Systems*, Finkenstädt, B., Held, L. and Isham, V., eds., Chapman & Hall/CRC, pp. 151–175.

Stein, M. L. (2005). Space-time covariance functions. *Journal of the American Statistical Association*, **100**, 310–321.

Schlather, M. (2010). Some covariance models based on normal scale mixtures. *Bernoulli*, **16**, 780–797.

Part 4

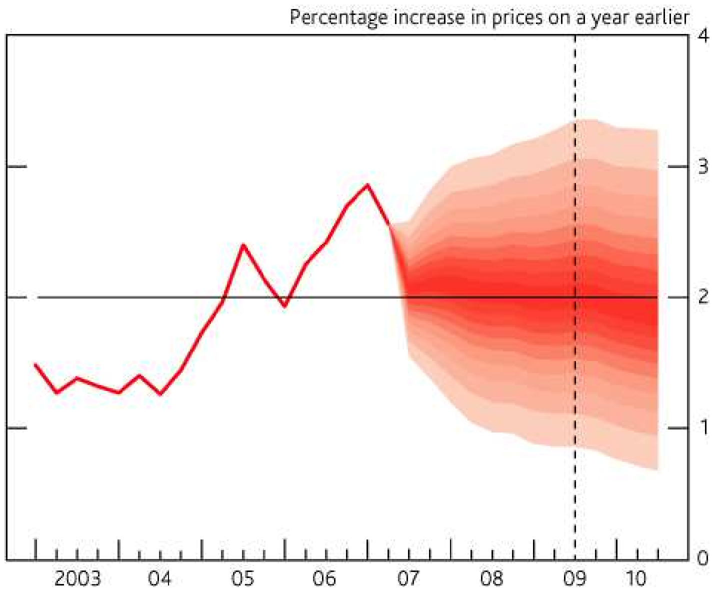
Calibrated probabilistic forecasting at the Stateline wind energy center

4.1 Probabilistic forecasts

consider a **real-valued** continuous **future quantity**, Y, such as an inflation rate or a wind speed

a **point forecast** is a single **real number**

a **probabilistic forecast** is a predictive **probability distribution** for Y, represented by a cumulative distribution function (**CDF**), F, or a probability density function (**PDF**), f



http://www.bankofengland.co.uk/inflationreport/irfanch.htm

4.2 The Stateline wind energy center

wind power: a clean, renewable energy source and the world's fastest growing energy source

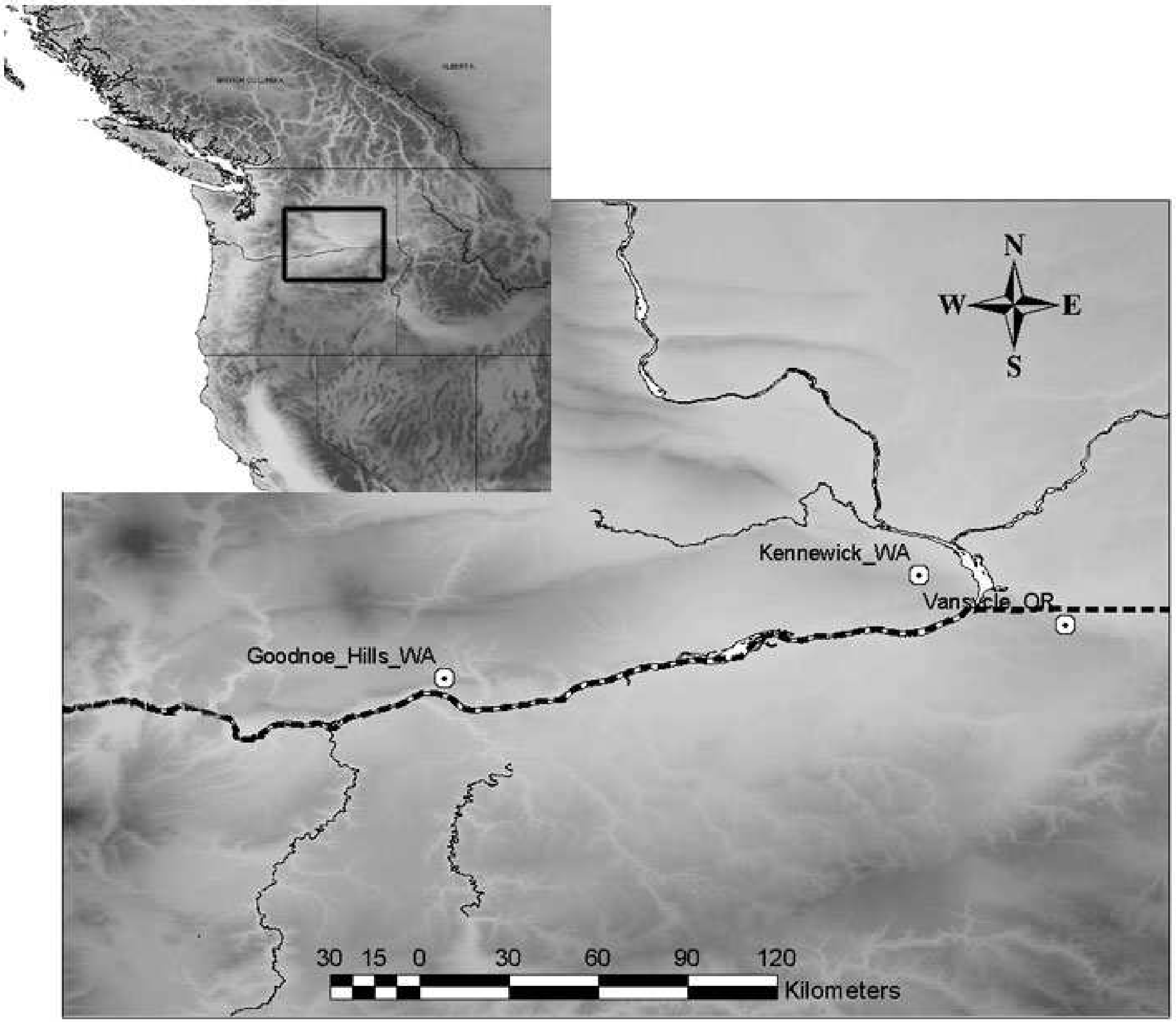
accurate wind energy forecasts critically important for power marketing, system reliability and scheduling

of central concern: **2-hour forecast horizon**

Stateline wind energy center

- a $300 million wind project located on the **Vansycle ridge** at the **Washington-Oregon** border in the US Pacific Northwest
- 454 wind turbines lined up over 50 square miles
- in the early 2000s the largest wind farm globally
- draws on westerly **Columbia Gorge** gap flows through the Cascade Mountains range

4.2 The Stateline wind energy center



4.2 The Stateline wind energy center

4.2 The Stateline wind energy center

2-hour forecasts of hourly average **wind speed** at **Vansycle**

joint project with **3TIER Environmental Forecast Group, Inc.**, Seattle (Gneiting, Larson, Westrick, Genton and Aldrich 2006)

data collected by Oregon State University for the Bonneville Power Administration

- **time series** of hourly average **wind speed** at **Vansycle, Kennewick** and **Goodnoe Hills**
- **time series** of **wind direction** at the three sites

notation: we denote the hourly average wind speed at time (hour) t at **V**ansycle, **K**ennewick and **G**oodnoe Hills by V_t, K_t and G_t, respectively

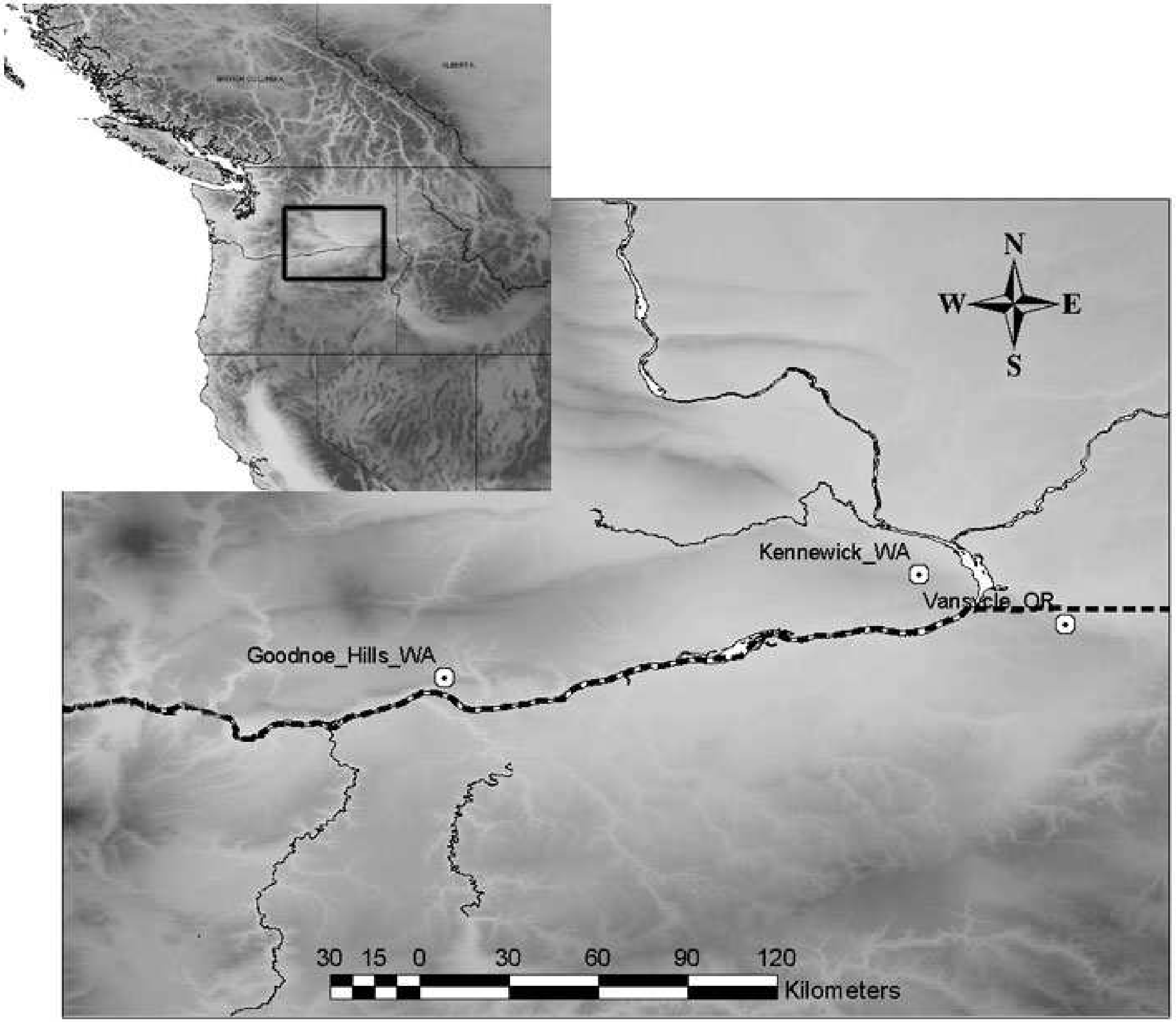
4.3 The regime-switching space-time (RST) method

merges meteorological and statistical expertize

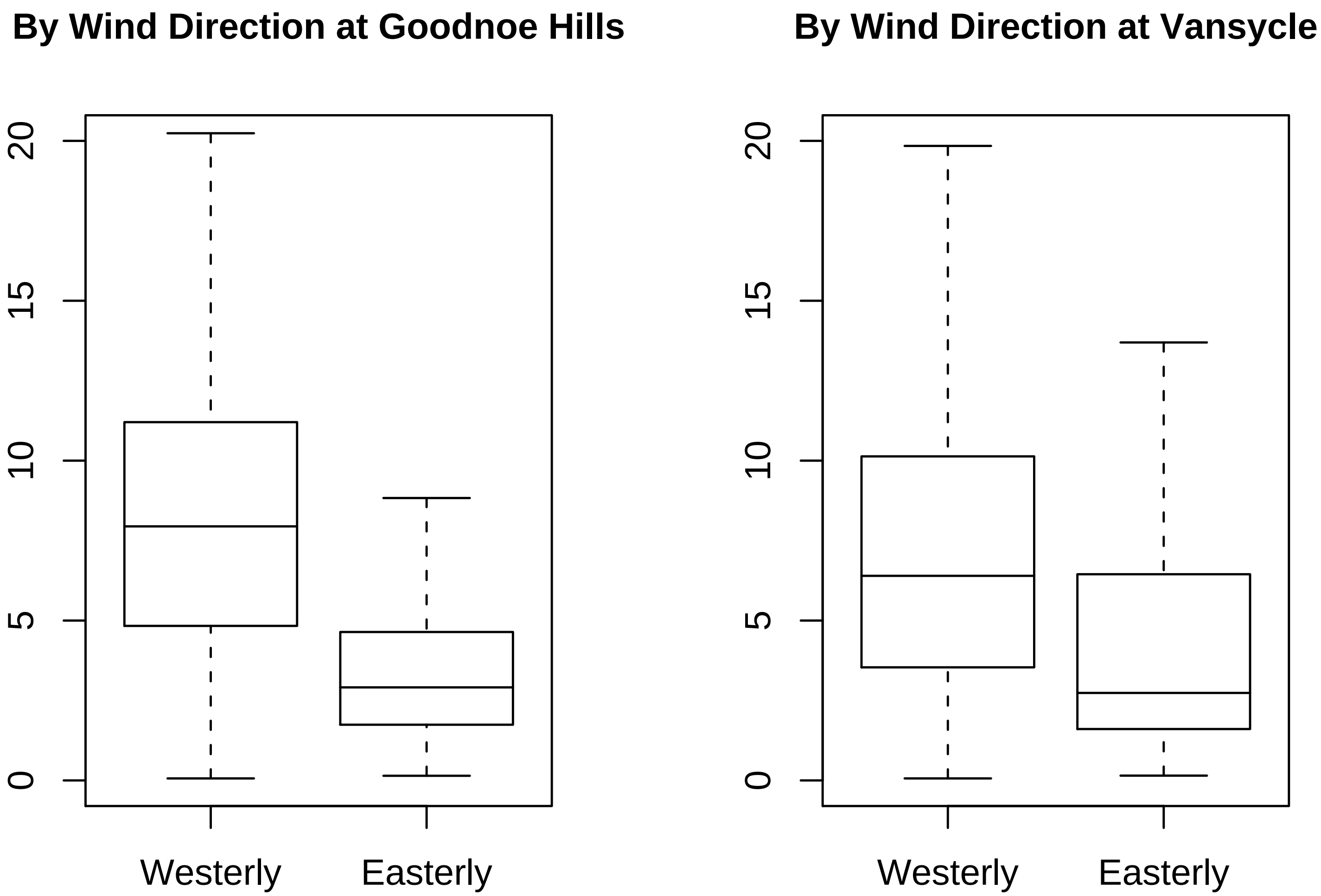
model formulation is parsimonious, yet **takes account** of the **salient features** of **wind speed**: alternating atmospheric regimes, temporal and spatial autocorrelation, diurnal and seasonal non-stationarity, conditional heteroskedasticity and non-Gaussianity

- **regime-switching**: identifies distinct **forecast regimes**
- **spatio-temporal**: utilizes geographically dispersed, recent meteorological observations in the vicinity of the wind farm
- **probabilistic**: provides probabilistic forecasts in the form of truncated normal **predictive PDFs**

4.3 The regime-switching space-time (RST) method



4.3 The regime-switching space-time (RST) method



hourly average wind speed two hours ahead at Vansycle

4.3 The regime-switching space-time (RST) method

regime-switching: 2-hour ahead forecasts of hourly average wind speed at Vansycle

- **westerly** regime:
  current wind direction at Goodnoe Hills westerly

  pronounced **diurnal component** with wind speeds that peak in the evening

- **easterly** regime:
  current wind direction at Goodnoe Hills easterly

  weak diurnal component

distinct **spatio-temporal dependence** structures

V_{t+2}	V_t	V_{t-1}	V_{t-2}	K_t	K_{t-1}	K_{t-2}	G_t	G_{t-1}	G_{t-2}
Westerly	.86	.78	.71	.76	.72	.67	.65	.66	.64
Easterly	.81	.72	.65	.56	.50	.45	.22	.21	.19

4.3 The regime-switching space-time (RST) method

forecasts in the **westerly regime**:

training set consists of the **westerly** cases within a **45-day rolling training period**

we fit and remove a **diurnal trend** component

$$\widehat{D}_t = b_0 + b_1 \sin\left(\frac{2\pi t}{24}\right) + \cdots + b_4 \cos\left(\frac{4\pi t}{24}\right)$$

to obtain **residuals** V_t^R, K_t^R and G_t^R

preliminary **point forecast** for the **residual** at **Vansycle**:

$$\widehat{V}_{t+2}^R = c_0 + c_1 V_t^R + c_2 V_{t-1}^R + c_3 K_t^R + c_4 K_{t-1}^R + c_5 G_t^R$$

preliminary **point forecast** for the **wind speed** at **Vansycle**:

$$\widehat{V}_{t+2} = \widehat{D}_{t+2} + \widehat{V}_{t+2}^R$$

4.3 The regime-switching space-time (RST) method

forecasts in the **westerly regime**:

probabilistic forecast takes the form of a **truncated normal** predictive **PDF**,

$$\mathcal{N}_{[0,\infty)}\left(\widehat{V}_{t+2}, \widehat{\sigma}^2_{t+2}\right) \quad \text{with} \quad \widehat{\sigma}_{t+2} = d_0 + d_1 v_t^R,$$

where v_t^R is a **volatility term** that averages recent increments in the residual components of the wind speed series

- truncated normal predictive PDF takes account of the **non-negativity** of wind speed: **tobit** model
- volatility term takes account of **conditional heteroskedasticity**

maximum likelihood plug-in estimators suboptimal for estimating predictive distributions

novel method of **minimum CRPS** estimation: M-estimator derived from a strictly proper scoring rule

4.3 The regime-switching space-time (RST) method

forecasts in the **easterly regime**:

training set consists of the **easterly** cases within a **45-day rolling training period**

we do not need to model the diurnal component

preliminary **point forecast** for the **wind speed** at **Vansycle**:

$$\widehat{V}_{t+2} = c_0 + c_1 V_t + c_2 V_{t-1} + c_3 K_t$$

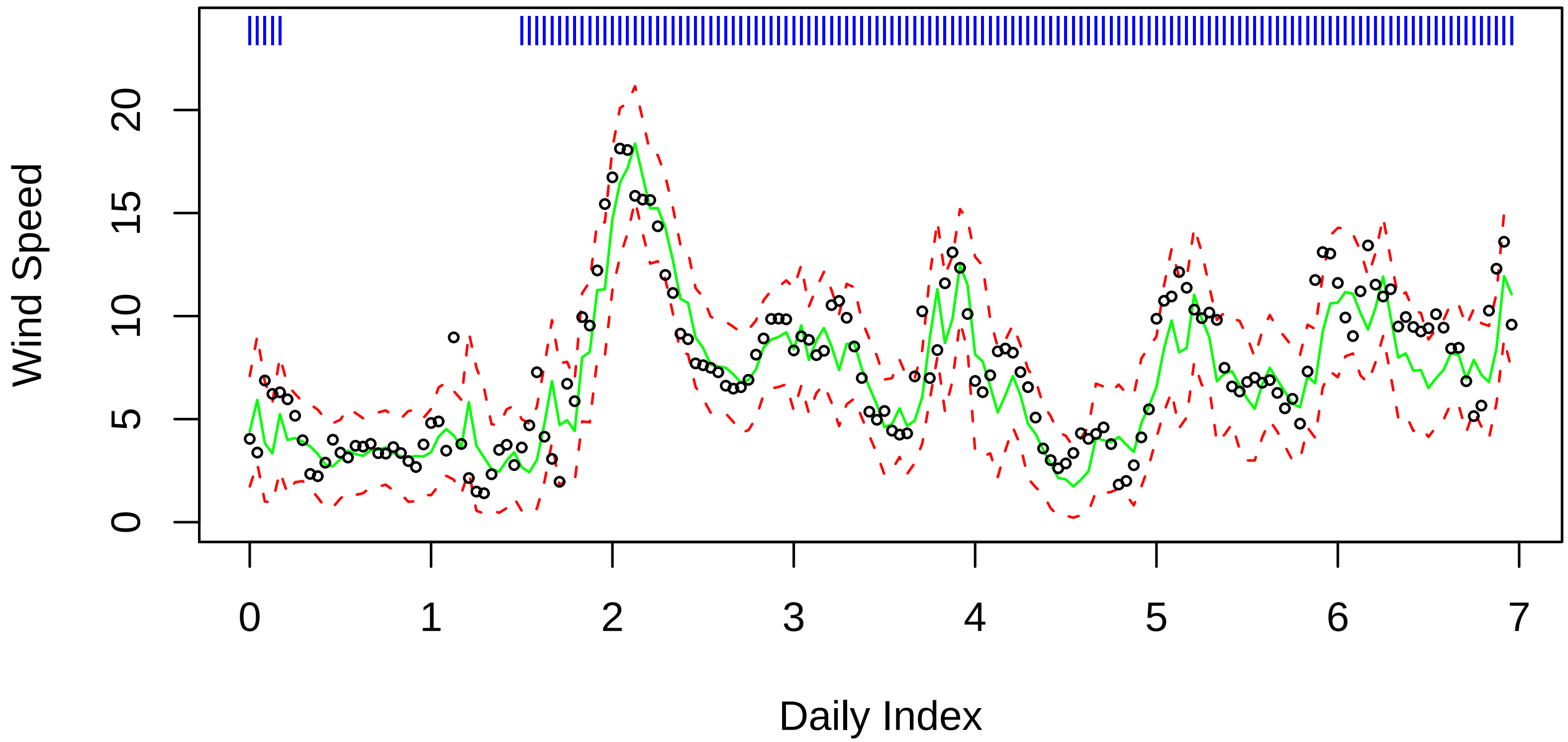
probabilistic forecast takes the form of a **truncated normal** predictive **PDF**,

$$\mathcal{N}_{[0,\infty)}\left(\widehat{V}_{t+2}, \widehat{\sigma}^2_{t+2}\right) \quad \text{with} \quad \widehat{\sigma}_{t+2} = d_0 + d_1 v_t,$$

where v_t is a **volatility term** that averages recent increments in the wind speed series

4.3 The regime-switching space-time (RST) method

2-hour ahead **RST** predictive **PDFs** of hourly average **wind speed** at **Vansycle** beginning June 28, 2003



4.4 Tools for evaluating probabilistic forecasts

Gneiting, Balabdaoui and Raftery (2007) contend that the goal of probabilistic forecasting is to **maximize** the **sharpness** of the predictive distributions **subject to calibration**

calibration

refers to the **statistical compatibility** between the **predictive PDFs** and the realizing **observations**

joint property of the forecasts and the observations

sharpness

refers to the **spread** of the **predictive PDFs**

property of the forecasts only

4.4 Tools for evaluating probabilistic forecasts

to assess **calibration**, we use the **probability integral transform** or **PIT** (Dawid 1984):

$$U = F(Y)$$

PIT histogram: histogram of the empirical PIT values

PIT histogram uniform $\iff$ prediction intervals at all levels have proper coverage

to assess **sharpness**, we consider the **average width** of the **90% central prediction interval**, say

4.4 Tools for evaluating probabilistic forecasts

a **scoring rule** is a function

$$\mathsf{s}(F, y)$$

that assigns a numerical score to each pair (F, y), where F is the **predictive CDF** and y is the realizing **observation**

we consider scores to be **negatively oriented** penalties that forecasters aim to **minimize**

a **proper** scoring rule s satisfies the expectation inequality

$$\mathbb{E}_G\, \mathsf{s}(G, Y) \leq \mathbb{E}_G\, \mathsf{s}(F, Y) \quad \text{for all} \quad F, G,$$

thereby encouraging **honest** and **careful** assessments (Gneiting and Raftery 2007)

the most popular example is the **logarithmic score**,

$$\mathsf{s}(f, y) = -\log f(y),$$

i.e., the negative of the **predictive PDF**, f, evaluated at the realizing **observation**, y

4.4 Tools for evaluating probabilistic forecasts

my favorite **proper scoring rule** score is the **continuous ranked probability score**,

$$\begin{aligned}\mathrm{crps}(F, y) &= \int_{-\infty}^{\infty} (F(x) - \mathbb{1}(x \geq y))^2 \, \mathrm{d}x \\ &= \mathbb{E}_F \, |X - y| - \frac{1}{2} \mathbb{E}_F \, |X - X'|\end{aligned}$$

where X and X' are independent random variables with cumulative distribution function F

the continuous ranked probability score is reported in the **same unit as the observations** and **generalizes** the **absolute error**, to which it reduces in the case of a point forecast

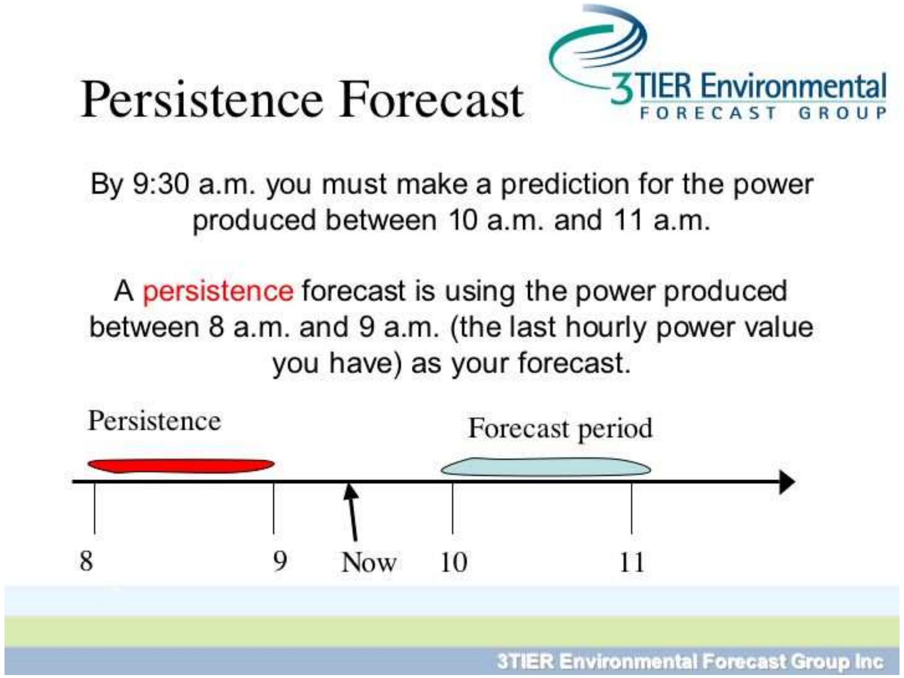
provides a **direct way** of **comparing point forecasts** and **probabilistic forecasts**

the **kernel score** representation allows for analogues on **Euclidean spaces** and **spheres** (Gneiting and Raftery 2007)

4.5 Predictive performance

competitors:

persistence forecast as reference standard: $\widehat{V}_{t+2} = V_t$



classical approach (Brown, Katz and Murphy 1984): **autoregressive (AR)** time series techniques

spatio-temporal approach: our new **regime-switching space-time (RST)** method

4.5 Predictive performance

evaluation period: May–November 2003

point forecasts: mean absolute error (MAE), using the **median** of the predictive distribution as point forecast (Gneiting 2011)

probabilistic forecasts: continuous ranked probability score (CRPS), PIT histogram, coverage and average width of the 90% central prediction interval

reporting scores **month by month** allows for **tests** of significance

4.5 Predictive performance

MAE ($m \cdot s^{-1}$)	May	Jun	Jul	Aug	Sep	Oct	Nov
Persistence	1.60	1.45	1.74	1.68	1.59	1.68	1.51
AR	1.54	1.38	1.50	1.54	1.53	1.67	1.53
RST	1.32	1.18	1.33	1.31	1.36	1.48	1.37

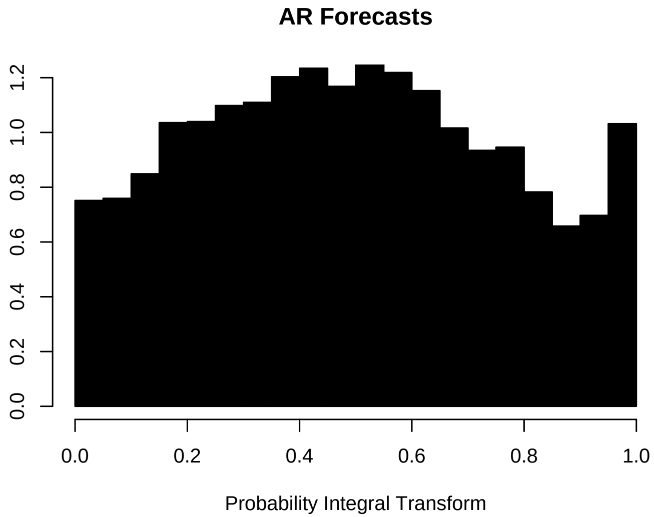
CRPS ($m \cdot s^{-1}$)	May	Jun	Jul	Aug	Sep	Oct	Nov
AR	1.11	1.01	1.10	1.11	1.10	1.22	1.10
RST	0.96	0.85	0.95	0.95	0.97	1.08	1.00

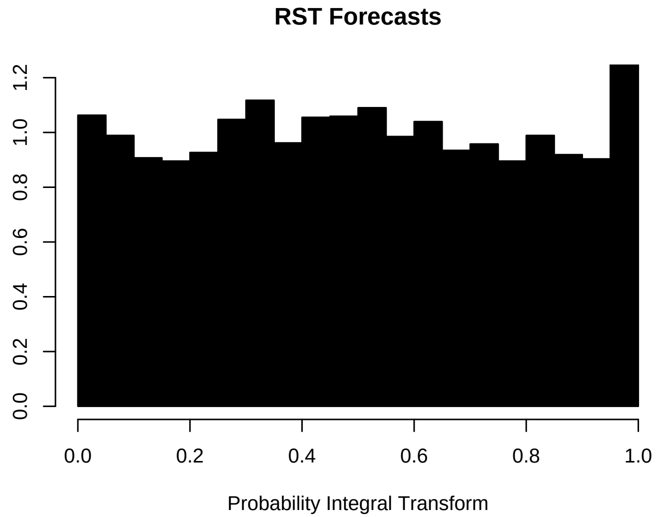
the **RST** forecasts had a **lower CRPS** than the **AR** forecasts in May, June, . . . , November

under the null hypothesis of equal forecast skill this only happens with probability

$$\left(\frac{1}{2}\right)^7 = \frac{1}{128}$$

4.5 Predictive performance





Coverage (%)	May	Jun	Jul	Aug	Sep	Oct	Nov
AR	91.1	91.7	89.2	91.5	90.6	87.4	91.4
RST	92.1	89.2	86.7	88.3	87.4	86.0	89.0

Ave Width ($m \cdot s^{-1}$)	May	Jun	Jul	Aug	Sep	Oct	Nov
AR	6.98	6.22	6.21	6.38	6.37	6.40	6.78
RST	5.93	4.83	5.14	5.22	5.15	5.45	5.46

Part 4: Further reading

Gneiting, T. (2008). Editorial: Probabilistic forecasting. *Journal of the Royal Statistical Society Series A: Statistics in Society*, **171**, 319–321.

Gneiting, T. and Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and its Application*, **1**, 125–151.

Gneiting, T., Larson, K., Westrick, K., Genton, M. G. and Aldrich, E. (2006). Calibrated probabilistic forecasting at the Stateline wind energy center: The regime-switching space-time method. *Journal of the American Statistical Association*, **101**, 968–979.

Zhu, X. and Genton, M. G. (2012). Short-term wind forecasting for power system operations. *International Statistical Review*, **80**, 2–23.

Pinson, P. (2013). Wind energy: Forecasting challenges for its operational management. *Statistical Science*, **28**, 564–585.

Gneiting, T., Balabdaoui, F. and Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **69**, 243–268.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102**, 359–378.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, **106**, 746–762.