Eliminating the practical boundary between Markov and other Gaussian random fields

Finn Lindgren[1]

Håvard Rue[2] Johan Lindström[1,3] David Bolin[1]

[1]Centre for Mathematical Sciences, Lund University
[2]Department of Mathematical Sciences, NTNU, Trondheim
[3]Department of Mathematical Statistics, University of Washington, Seattle

Trondheim
2009-05-15

LUND UNIVERSITY

Background

Large environmetric data sets require computationally efficient methods for

- Realistic modelling
- Estimation of parameters and influence of covariates
- Kriging (optimal prediction/reconstruction)
- Data assimilation (climate model + real data)
- Temporal forecasting

This talk will focus on a method for modelling the dependence structure of a random field, linking spatially discrete Markov specifications with continuous models.

Modelling spatial dependence

Spatial methods:

- Covariance function specifications
- Variogram specifications
- Spectral representations
- Convolution models
- Stochastic partial differential equations (SPDEs)
- Markov random field (MRF) specifications
- Non-stationary versions of most of the above

Spatio-temporal methods:

- All of the above...
- ...but many unsolved issues in both situations.

Gaussian random fields

- A discrete domain Gaussian field with expectation μ and *covariance matrix* Σ can be defined via the density

$$p(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\mu)^T\Sigma^{-1}(\boldsymbol{x}-\mu)\right)$$

- When the field is a discrete representation of a continuous-domain field,

$$\mu_i = \mathrm{E}(\boldsymbol{x}(\boldsymbol{u}))$$
$$\Sigma_{i,j} = \mathrm{C}(\boldsymbol{x}(\boldsymbol{u}), \boldsymbol{x}(\boldsymbol{v}))$$

- We want to construct a discrete representation of a continuous random field model, that is consistent between different spatial discretisations.

A Markov reformulation

- A Gaussian Markov random field with expectation μ and *precision matrix* $\boldsymbol{Q}$ can be defined via the density

$$p(\boldsymbol{x}) = \frac{|\boldsymbol{Q}|^{1/2}}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\mu)^T \boldsymbol{Q}(\boldsymbol{x}-\mu)\right)$$

- The (spatial) Markov property:

$$p(\boldsymbol{x}_i|\boldsymbol{x}_j, j \neq i) = p(\boldsymbol{x}_i|\boldsymbol{x}_j, j \in \mathcal{N}_i)$$

for some *neighbourhood* $\mathcal{N}_i$ of i.

- Equivalent: For any $i \neq j$,

$$\boldsymbol{Q}_{i,j} = 0 \iff (x_i \perp x_j \mid x_k, k \notin \{i,j\}) \iff j \notin \mathcal{N}_i$$

Kriging with GMRFs

- Prior distribution model: $\boldsymbol{x} \in \mathrm{N}(\mu, \boldsymbol{Q}^{-1})$
- Observation model: $\boldsymbol{y}|\boldsymbol{x} \in \mathrm{N}(\boldsymbol{Ax}, \Sigma_{\boldsymbol{e}})$
- Given the model parameters, the posterior density $p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})$ is also a GMRF, with precision

  $$\boldsymbol{Q}_{\boldsymbol{x}|\boldsymbol{y}} = \boldsymbol{Q} + \boldsymbol{A}^T \Sigma_{\boldsymbol{e}}^{-1} \boldsymbol{A}$$

  and expectation

  $$\mu_{\boldsymbol{x}|\boldsymbol{y}} = \boldsymbol{Q}_{\boldsymbol{x}|\boldsymbol{y}}^{-1} \left(\boldsymbol{Q}\mu + \boldsymbol{A}^T \Sigma_{\boldsymbol{e}}^{-1} \boldsymbol{y} \right)$$

- We only need the Cholesky factorisation, $\boldsymbol{Q}_{\boldsymbol{x}|\boldsymbol{y}} = \boldsymbol{R}^T \boldsymbol{R}$, which will be sparse if $\boldsymbol{A}^T \Sigma_{\boldsymbol{e}}^{-1} \boldsymbol{A}$ is sparse

Matérn covariances (Bertil Matérn, 1917–2007)

- The Matérn covariance family on $\boldsymbol{u} \in \mathbb{R}^d$:

$$r(\boldsymbol{u},\boldsymbol{v}) = \mathrm{C}(x(\boldsymbol{u}),x(\boldsymbol{v})) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)}(\varkappa\|\boldsymbol{v}-\boldsymbol{u}\|)^\nu K_\nu(\varkappa\|\boldsymbol{v}-\boldsymbol{u}\|)$$

  with scale (inverse range) $\varkappa > 0$ and shape/smoothness $\nu > 0$, and K_ν a modified Bessel function.
- Fields with Matérn covariances are solutions to an SPDE (Whittle, 1954) based on the Laplacian, $\Delta = \nabla^T \nabla$,

$$\left(\varkappa^2 - \Delta\right)^{\alpha/2} x(\boldsymbol{u}) = \tau^2 \mathcal{W}(\boldsymbol{u}),$$

  where $\mathcal{W}(\boldsymbol{u})$ is spatial white noise, and $\alpha = \nu + d/2$.
- Parameter link: $\sigma^2 = \tau^2 \frac{\Gamma(\nu)}{\Gamma(\alpha)\varkappa^{2\nu}(4\pi)^{d/2}}$

SPDE issues

- Non-uniqueness:
  If $x(\boldsymbol{u})$ is a solution to the SPDE for $\alpha = 2$, so is $x(\boldsymbol{u}) + c \cdot \exp(\varkappa \boldsymbol{e}^T \boldsymbol{u})$, for any unit length vector $\boldsymbol{e}$ and any constant c.
- Regularity:
  If $\alpha \leq d/2$, the fields are random measures, with no point-wise interpretation.
- Non-stationarity: On a bounded domain, the SPDE solutions are non-stationary, unless conditioned on suitable boundary distributions.
- Fractional differential operator:
  The SPDE is defined for any $\alpha \geq 0$, but we only consider integers $\alpha = 1, 2, 3, \ldots$.

Neumann boundaries

- Practical solution to the non-uniqueness and non-stationarity:
  Zero-normal-derivative (Neumann) boundaries reduce the impact of the null-space solutions:

$$\begin{cases} (\varkappa^2 - \Delta)^{\alpha/2} x(\boldsymbol{u}) = \mathcal{W}(\boldsymbol{u}), & \boldsymbol{u} \in \Omega \\ \partial_{\boldsymbol{n}} (\varkappa^2 - \Delta)^j x(\boldsymbol{u}) = 0, & \boldsymbol{u} \in \partial\Omega,\, j = 0, 1, \ldots, \lfloor (\alpha - 1)/2 \rfloor \end{cases}$$

- Resulting covariance, for $\Omega = [0, L] \subset \mathbb{R}$:

$$\begin{aligned} \mathrm{C}(x(u), x(v)) &\approx r_M(u, v) + r_M(u, -v) + r_M(u, 2L - v) \\ &= r_M(0, v - u) + r_M(0, v + u) + r_M(0, 2L - (v + u)). \end{aligned}$$

The finite element method

- A *stochastic weak formulation* of the SPDE states that

$$\left[\langle \varphi_k, (\Delta - \varkappa^2)^{\alpha/2} \boldsymbol{x} \rangle\right]_{k=1,\ldots,n} \stackrel{D}{=} \left[\langle \varphi_k, \mathcal{W} \rangle\right]_{k=1,\ldots,n}$$

  for each set of *test functions* $\{\varphi_k\}$.
- We use N simple piecewise linear test functions, equal to the piecewise linear basis functions, $\boldsymbol{x}(\boldsymbol{u}) = \sum_j \psi_k(\boldsymbol{u}) w_j$, and compute the resulting distributions by explicitly calculating the expectation vector and precision matrix for $\boldsymbol{w}$ $(= \boldsymbol{x})$.
- For $\alpha = 2$, the weak formulation can be written

$$\boldsymbol{K}\boldsymbol{w} = \left[\langle \varphi_i, (\Delta - \varkappa^2) \psi_j \rangle\right]_{i,j=1,\ldots,N} \boldsymbol{w} \stackrel{D}{=} \left[\langle \varphi_k, \mathcal{W} \rangle\right]_{k=1,\ldots,N}$$

Construction of $\boldsymbol{Q}$

- With the help of Green's first identity,

$$\begin{aligned} \boldsymbol{C}_{i,j} &= \langle \psi_i, \psi_j \rangle, \\ \boldsymbol{K}_{i,j} &= \left\langle \psi_i, (\varkappa^2 - \Delta)\psi_j \right\rangle = \varkappa^2 \boldsymbol{C}_{i,j} + \langle \nabla \psi_i, \nabla \psi_j \rangle, \end{aligned}$$

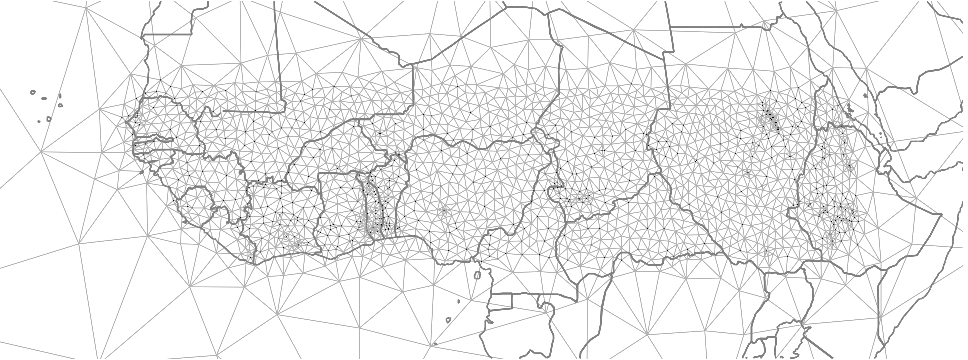
  for Neumann boundaries.
- Markovified Least Squares and Galerkin solutions:

$$\begin{aligned} \boldsymbol{C} &= \operatorname{diag}(\langle \psi_i, 1 \rangle), \quad (\text{“optimal” approximation}) \\ \boldsymbol{Q}_{1,\varkappa} &= \boldsymbol{K}, \quad (\text{Least Squares}) \\ \boldsymbol{Q}_{2,\varkappa} &= \boldsymbol{K}\boldsymbol{C}^{-1}\boldsymbol{K}, \quad (\text{Galerkin}) \\ \boldsymbol{Q}_{\alpha,\varkappa} &= \boldsymbol{K}\boldsymbol{C}^{-1}\boldsymbol{Q}_{\alpha-2,\varkappa}\boldsymbol{C}^{-1}\boldsymbol{K}, \quad \alpha = 3, 4, \ldots \ (\text{Galerkin recursion}) \end{aligned}$$

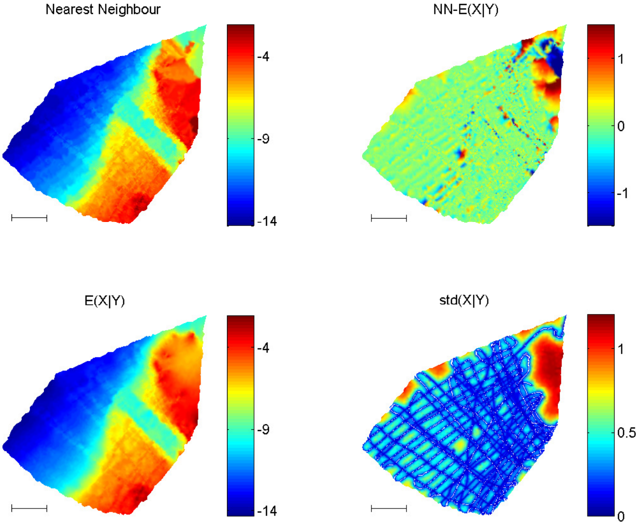
  Neighbourhood radius equals α.
- Simple closed-form expressions for the matrix elements.

Automatic triangulation

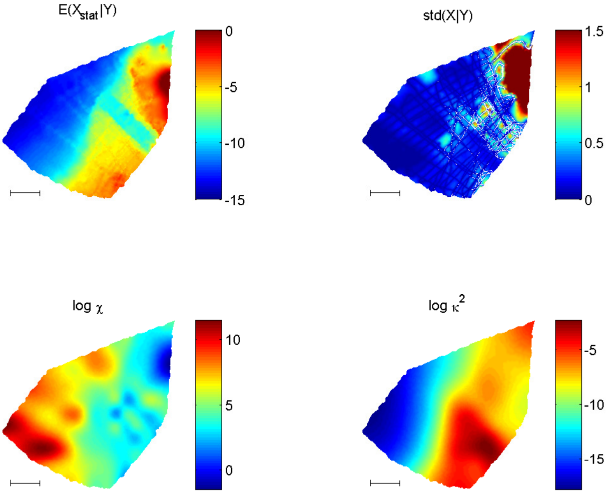
- We do not need a dense regular grid for the field.
- The GMRF construction allows for irregular triangulations, that can be constructed automatically.



Seabed depth reconstruction with INLA



Non-stationary model with B-splines



Extensions

- Well-defined SPDE on curved manifolds (e.g. a globe).
- Anisotropy: modify the Laplacian.
- Spatially oscillating fields can be introduced via a complex version of the SPDE:

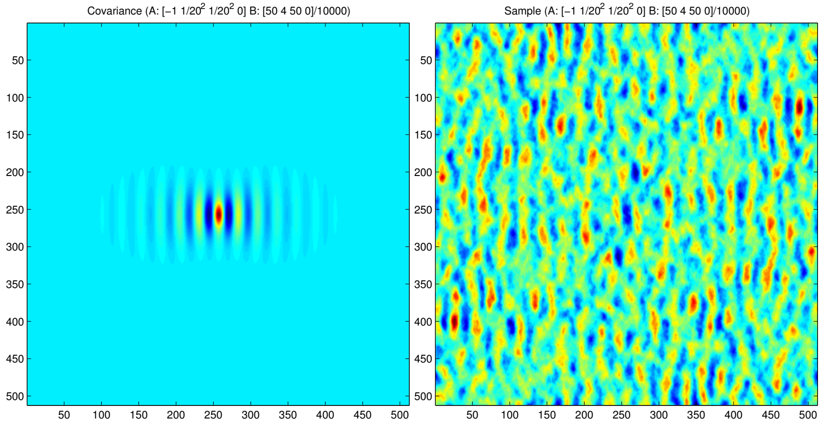
$$(h_1 + ih_2 - \nabla^T(H_1 + iH_2)\nabla)(x_1(\boldsymbol{u}) + ix_2(\boldsymbol{u})) = \mathcal{W}_1(\boldsymbol{u}) + iW_2(\boldsymbol{u})$$

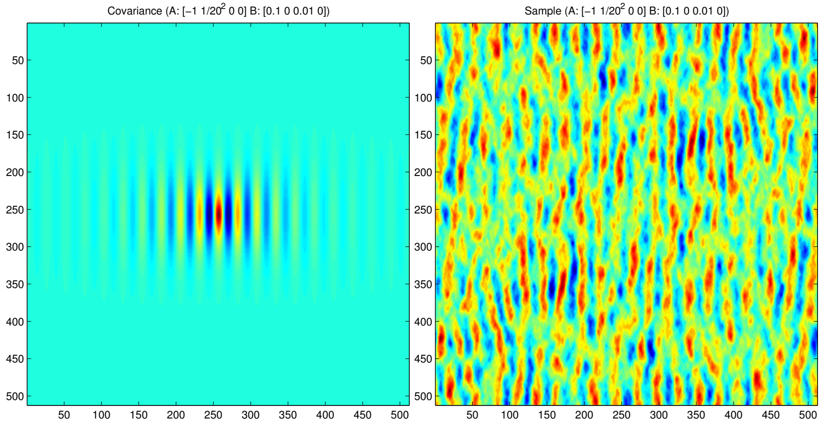
  Coupled system:

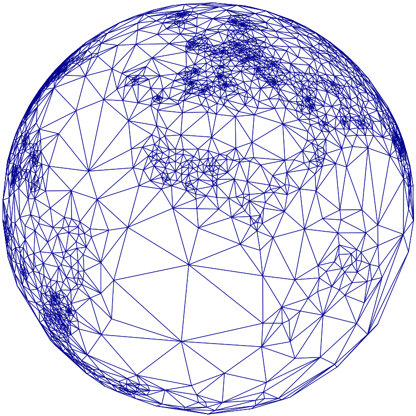
$$\begin{cases} (h_1 - \nabla^T \boldsymbol{H}_1 \nabla)x_1 - (h_2 - \nabla^T \boldsymbol{H}_2 \nabla)x_2 = \mathcal{E}_1, \\ (h_2 - \nabla^T \boldsymbol{H}_2 \nabla)x_1 + (h_1 - \nabla^T \boldsymbol{H}_1 \nabla)x_2 = \mathcal{E}_2, \end{cases}$$

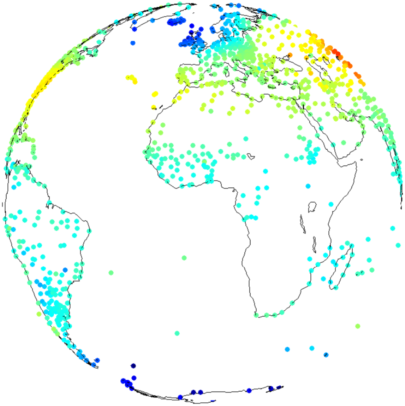
- x_1 and x_2 independent, with spectrum

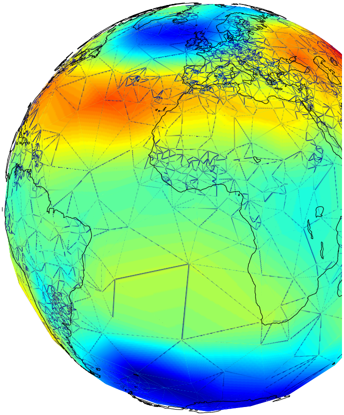
$$R_x(\boldsymbol{k}) = \frac{R_{\mathcal{E}}(\boldsymbol{k})}{(h_1 + \boldsymbol{k}^T \boldsymbol{H}_1 \boldsymbol{k})^2 + (h_2 + \boldsymbol{k}^T \boldsymbol{H}_2 \boldsymbol{k})^2},$$

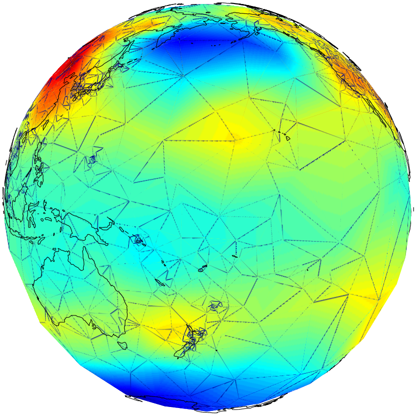

Covariance (A: [−1 1/20^2 1/20^2 0] B: [50 4 50 0]/10000)
Sample (A: [−1 1/20^2 1/20^2 0] B: [50 4 50 0]/10000)
50
100
150
200
250
300
350
400
450
500

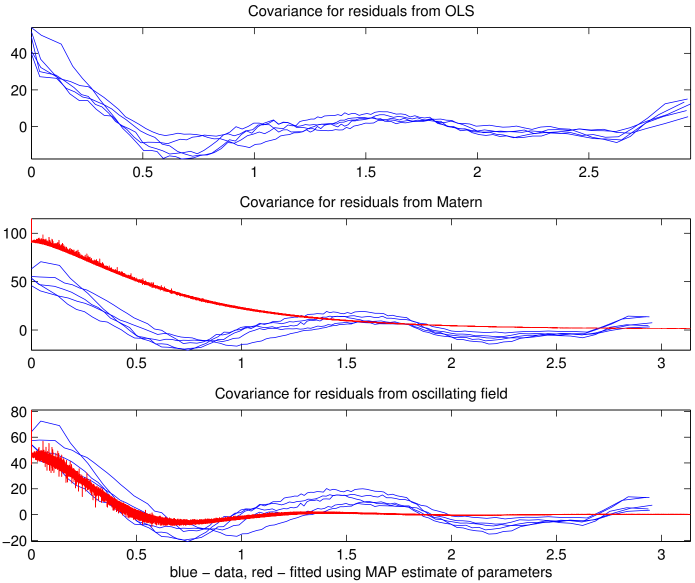

Covariance (A: [−1 1/20^2 0 0] B: [0.1 0 0.01 0])
Sample (A: [−1 1/20^2 0 0] B: [0.1 0 0.01 0])
50 100 150 200 250 300 350 400 450 500










Covariance for residuals from OLS
40
20
0
0
0.5
1
1.5
2
2.5
Covariance for residuals from Matern
100
50
0
0
0.5
1
1.5
2
2.5
3
Covariance for residuals from oscillating field
80
60
40
20
0
−20
0
0.5
1
1.5
2
2.5
3
blue – data, red – fitted using MAP estimate of parameters

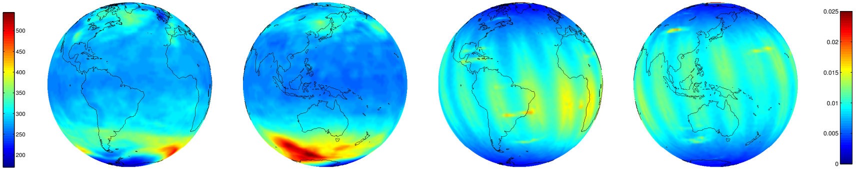
Spatial differential models

- Non-stationary continuous ARMA models:

$$(\varkappa(\boldsymbol{u})^2 - \Delta)^{\alpha/2} X_0(\boldsymbol{u}) = \mathcal{E}(\boldsymbol{u})$$

$$X(\boldsymbol{u}) = \mu(\boldsymbol{u}) + \varphi(\boldsymbol{u})(1 + \boldsymbol{b}(\boldsymbol{u})^T \nabla) X_0(\boldsymbol{u})$$

  Non-stationary parameters constructed from spherical harmonic basis functions.
- Ozone estimate (left) and Kriging standard errors (right) in Dobson units:



Flow

- The spatio-temporal particle transport/diffusion equation

$$\frac{\partial}{\partial t} x(\boldsymbol{u}, t) = -\left(\varkappa^2 + \nabla^T \mu - \nabla^T \boldsymbol{\Sigma} \nabla\right) x(\boldsymbol{u}, t) + \mathcal{E}(\boldsymbol{u}, t)$$

can be represented purely spatially by the similar equation

$$\left(\varkappa^2 + \nabla^T \mu - \nabla^T \boldsymbol{\Sigma} \nabla\right) x(\boldsymbol{u}) = \mathcal{E}(\boldsymbol{u})$$

where $\mu(\boldsymbol{u})$ is the vector flow field. This yields a physically motivated GMRF model for particle transport variation.

Comments

- Explicit link from a substantial class of SPDEs to GMRFs, allowing the use of INLA for inference, as long as the number of parameters is small.
- When the number of parameters is large, estimation is efficient with direct optimisation. What about the posterior densities? Hybrid INLA/MCMC approach?
- A need for categorising the different stationary and non-stationary SPDEs for user-friendly specification.
- Is there a single "best" parameterisation, e.g. one with direct control over the marginal variances, or are several alternative specifications needed?

F. Lindgren and H. Rue, *Explicit construction of GMRF approximations to generalised Matérn fields on irregular grids*, Preprints in Mathematical Sciences, 2007:12, MC, Lund, 2007.
(Full article is in preparation.)

Spectral interpretation using linear filter theory

In $\mathbb{R}^d$:

- Transfer function for $\varkappa^2 - \Delta$: $H(\boldsymbol{k}) = \varkappa^2 + \boldsymbol{k}^T\boldsymbol{k}$
- The covariance function is

$$r(\boldsymbol{u},\boldsymbol{v}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \frac{1}{|\varkappa^2 + \boldsymbol{k}^T\boldsymbol{k}|^\alpha} e^{i\boldsymbol{k}^T(\boldsymbol{v}-\boldsymbol{u})}\,\mathrm{d}\boldsymbol{k}$$

On the ordinary unit sphere, $\mathbb{S}^2$:

- Transfer function for $\varkappa^2 - \Delta$: $H(k,m) = \varkappa^2 + k(k+1)$
- The covariance function is

$$r(\boldsymbol{u},\boldsymbol{v}) = \sum_{k=0}^{\infty} \sum_{m=-k}^{k} \frac{1}{|\varkappa^2 + k(k+1)|^\alpha} Y_{k,m}(\boldsymbol{u}) Y_{k,m}(\boldsymbol{v})$$

$$= \frac{1}{4\pi} \sum_{k=0}^{\infty} \frac{2k+1}{|\varkappa^2 + k(k+1)|^\alpha} P_{k,0}(\boldsymbol{u}^T\boldsymbol{v})$$