

A structured random effects introduction, the GMRFs

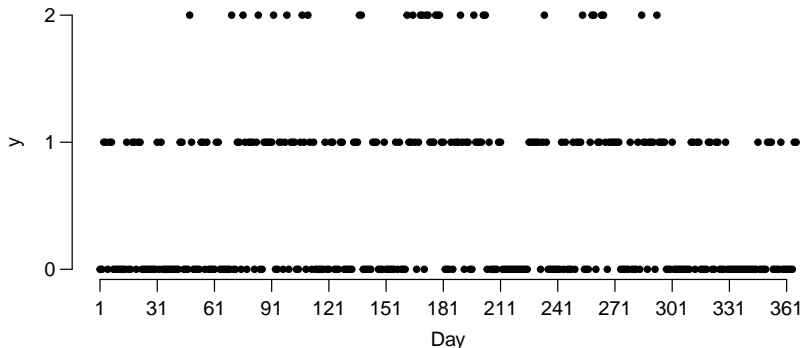
Elias T. Krainski

April 2022

Motivating examples

Tokyo example

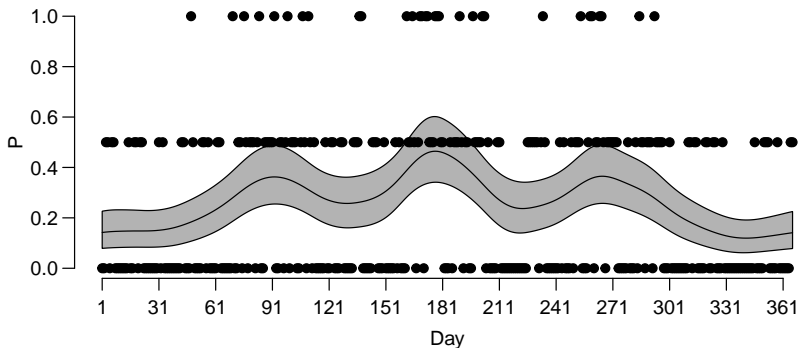
Rainfall over 1 mm in the Tokyo area for each calendar day during two years (1983-84) are registered.



Problem: model the probability of rain each day of the year

Tokyo example (fit)

Rainfall over 1 mm in the Tokyo area for each calendar day during two years (1983-84) are registered.



The probability of rain each day of the year

Scotland example (from WinBUGS)

The rates of lip cancer in 56 counties in Scotland have been analysed by Clayton and Kaldor (1987) and Breslow and Clayton (1993). The form of the data includes the observed and expected cases (expected numbers based on the population and its age and sex distribution in the county), a covariate measuring the percentage of the population engaged in agriculture, fishing, or forestry, and the "position" of each county expressed as a list of adjacent counties.

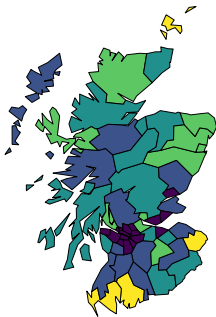
County	Observed cases O_i	Expected cases E_i	Percentage in agric. x_i	SMR	Adjacent counties
1	9	1.4	16	652.2	5,9,11,19
2	39	8.7	16	450.3	7,10
...
56	0	1.8	10	0.0	18,24,30,33,45,55

$$O_i \sim \text{Poisson}(\mu_i)$$

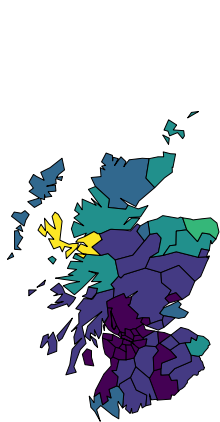
$$\log \mu_i = \log E_i + \alpha_0 + \alpha_1 x_i / 10 + b_i$$

Scotland maps

% in Agriculture



SMR



Scotland data: GLM

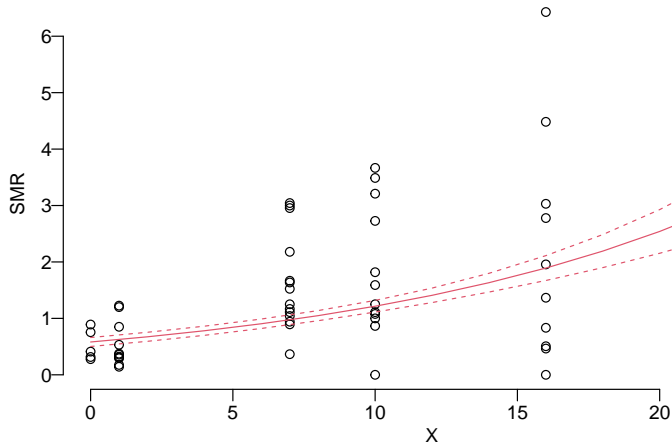
Number of **O**bserved cases as a function of the percentage working in agriculture (**X**)

```
m1 <- glm(O ~ X, poisson, offset=log(E), data=map@data)
summary(m1)
##
## Call:
## glm(formula = O ~ X, family = poisson, data = map@data, offset = log(E))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.763  -1.216   0.097   1.336   4.713
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.54227     0.06952   -7.8 6.2e-15 ***
## X            0.07373     0.00596   12.4 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 380.73  on 55  degrees of freedom
## Residual deviance: 238.62  on 54  degrees of freedom
## AIC: 450.6
##
## Number of Fisher Scoring iterations: 5
```

Scotland data: model 1

```
r1 <- inla(0 ~ X, family='poisson', offset=log(E), data=map@data,
          control.compute=list(cpo=TRUE))
summary(r1)
##
## Call:
## c("inla.core(formula = formula, family = family, contrasts = contrasts,
## ", " data = data, quantiles = quantiles, E = E, offset = offset, ", "
## scale = scale, weights = weights, Ntrials = Ntrials, strata = strata,
## ", " lp.scale = lp.scale, link.covariates = link.covariates, verbose =
## verbose, ", " lincomb = lincomb, selection = selection, control.compute
## = control.compute, ", " control.predictor = control.predictor,
## control.family = control.family, ", " control.inla = control.inla,
## control.fixed = control.fixed, ", " control.mode = control.mode,
## control.expert = control.expert, ", " control.hazard = control.hazard,
## control.lincomb = control.lincomb, ", " control.update =
## control.update, control.lp.scale = control.lp.scale, ", "
## control.pardiso = control.pardiso, only.hyperparam = only.hyperparam,
## ", " inla.call = inla.call, inla.arg = inla.arg, num.threads =
## num.threads, ", " blas.num.threads = blas.num.threads, keep = keep,
## working.directory = working.directory, ", " silent = silent, inla.mode
## = inla.mode, safe = FALSE, debug = debug, ", " .parent.frame =
## .parent.frame)")
## Time used:
## Pre = 0.934, Running = 0.188, Post = 0.0147, Total = 1.14
## Fixed effects:
##      mean      sd 0.025quant 0.5quant 0.975quant mode kld
## (Intercept) -0.542 0.070      -0.680   -0.541      -0.408  NA   0
## X              0.074 0.006        0.062    0.074        0.085  NA   0
##
## Marginal log-Likelihood: -234.10
## CPO, PIT is computed
## Posterior summaries for the linear predictor and the fitted values are computed
## (Posterior marginals needs also 'control.compute=list(return.marginals.predictor=TRUE)')
```


The fitted covariate effect



After the covariate effect, is there something left?

Covariate

- 0 to 5
- 5 to 10
- 10 to 15
- 15 to 20
- 20 to 25

SMR

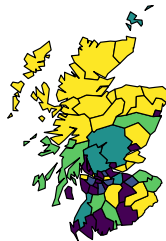
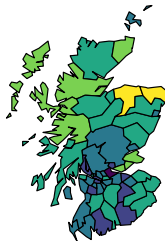
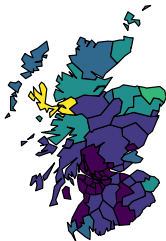
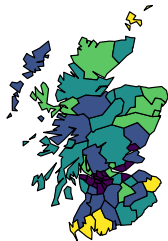
- 0 to 1
- 1 to 2
- 2 to 3
- 3 to 4
- 4 to 5
- 5 to 6
- 6 to 7

Residuals

- 6 to -4
- 4 to -2
- 2 to 0
- 0 to 2
- 2 to 4
- 4 to 6

PIT

- 0 to 0.2
- 0.2 to 0.4
- 0.4 to 0.6
- 0.6 to 0.8
- 0.8 to 1



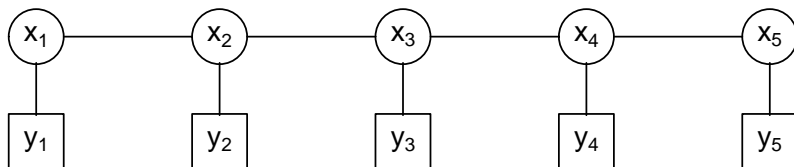
Structured random effects

Smoothed probability over time

- ▶ Temporally smooth probability of rain
 - ▶ is different for each day but similar for nearby days
 - ▶ p_i is similar to p_{i+1}
 - ▶ assume $\text{logit}(p_i) = x_i$

Smoothed probability over time

- ▶ Temporally smooth probability of rain
 - ▶ is different for each day but similar for nearby days
 - ▶ p_i is similar to p_{i+1}
 - ▶ assume $\text{logit}(p_i) = x_i$



- ▶ dependence on x
- ▶ y conditionally independent given x
 - ▶ y_i conditional on x_i is independent of y_{i-1} and of y_{i+1}

The RW1 prior

- ▶ It seems natural to borrow strength over time.
 - ▶ **x**: smoothing over time
 - ▶ ***R**andom **W**alk* - RW of first order: `rw1`
 - ▶ Gaussian distribution for the successive differences (**R** sparse)

$$\Delta_i = x_i - x_{i-1} \sim N(0, \tau^{-1})$$

The RW1 prior

- ▶ It seems natural to borrow strength over time.
 - ▶ \mathbf{x} : smoothing over time
 - ▶ *Random Walk* - RW of first order: `rw1`
 - ▶ Gaussian distribution for the successive differences (\mathbf{R} sparse)

$$\Delta_i = x_i - x_{i-1} \sim N(0, \tau^{-1})$$

- ▶ The log of the (joint) distribution for \mathbf{x} is

$$\log(\pi(\mathbf{x}|\tau)) \propto -\frac{\tau}{2} \sum_{i=2}^n (x_i - x_{i-1})^2 = -\frac{\tau}{2} \mathbf{x}' \mathbf{R} \mathbf{x},$$

where

$$\mathbf{R} = \begin{bmatrix} 1 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & & \ddots & & & \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 2 & -1 \\ & & & & & & -1 & 1 \end{bmatrix}$$

The cyclic RW1

- 1st of January is similar to December, 31: **cyclic random walk**

$$\begin{aligned}\pi(\mathbf{x}|\theta) &\propto \exp \left\{ -\frac{\theta}{2} \left[(x_1 - x_n)^2 + \sum_{i=2}^n (x_i - x_{i-1})^2 \right] \right\} \\ &= \exp \left\{ -\frac{\theta}{2} \mathbf{x}^T \mathbf{R} \mathbf{x} \right\}\end{aligned}$$

where, now,

$$\mathbf{R} = \begin{bmatrix} 2 & -1 & & & & & -1 \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & & \ddots & & & \\ & & & & -1 & 2 & -1 \\ & & & & -1 & 2 & -1 \\ -1 & & & & & -1 & 2 \end{bmatrix}$$

- Cyclic second order is analogous.

Tokyo example: the model

- ▶ y_i assume values 0, 1 or 2, for $i = 1, \dots, n$
 - ▶ assuming conditional independence, thus

$$y_i | p_i \sim \text{Binomial}(n_i, p_i)$$

- ▶ link function (logit)

$$p_i = 1 / (1 + \exp(-x_i))$$

Tokyo example: the model

- ▶ y_i assume values 0, 1 or 2, for $i = 1, \dots, n$
 - ▶ assuming conditional independence, thus

$$y_i | p_i \sim \text{Binomial}(n_i, p_i)$$

- ▶ link function (logit)

$$p_i = 1 / (1 + \exp(-x_i))$$

- ▶ \mathbf{x} is a *Gaussian Markov Random Field* - GMRF, Rue and Held (2005)

Tokyo example: the model

- ▶ y_i assume values 0, 1 or 2, for $i = 1, \dots, n$
 - ▶ assuming conditional independence, thus

$$y_i | p_i \sim \text{Binomial}(n_i, p_i)$$

- ▶ link function (logit)

$$p_i = 1 / (1 + \exp(-x_i))$$

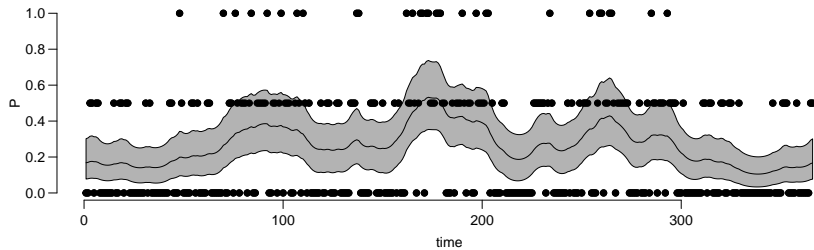
- ▶ \mathbf{x} is a *Gaussian Markov Random Field* - GMRF, Rue and Held (2005)
- ▶ τ : local precision parameter

Model fit in INLA

$$\begin{aligned}y_i|x_i &\sim \text{Binomial}(2, p_i) && \rightarrow \text{likelihood} \\ \mathbf{x}|\tau &\sim N(\mathbf{0}, (\tau \mathbf{R})^{-}) && \rightarrow \text{latent field, GMRF} \\ \tau &\sim p(\tau) && \rightarrow \text{prior distribution}\end{aligned}$$

```
head(Tokyo, 5)
##    y n time    P
##  1 0 2    1 0.0
##  2 0 2    2 0.0
##  3 1 2    3 0.5
##  4 1 2    4 0.5
##  5 0 2    5 0.0
model <- y ~ f(time, model='rw1', cyclic=TRUE)
result <- inla(model, family='binomial',
               data=Tokyo, Ntrials=n,
               control.compute=list(cpo=TRUE))
```

Result for the time series



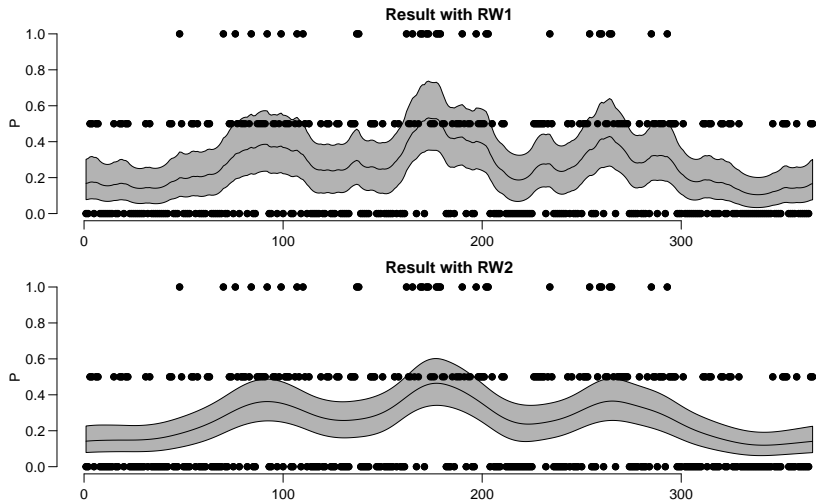
Smoothing more

Gaussian distribution for the second order differences (rw2)

$$\Delta_i^2 = x_i - 2x_{i-1} + x_{i-2} \sim N(0, \tau^{-1})$$

```
model2 <- y ~ f(time, model='rw2', cyclic=TRUE)
result2 <- inla(model2, family='binomial',
               data=Tokyo, Ntrials=n,
               control.compute=list(cpo=TRUE))
```

Both results for the time series



Smooth areal dependent risk

$$y_i \sim \text{Poisson}(E_i r_i)$$

$$\log(r_i) = \alpha + \beta X_i + s_i$$

where s_i may be spatial smooth

$s_i | s_j$, j the index for the neighbours of i

Besag: random walk over areas, Besag (1974)

$$\pi(x_i | \mathbf{x}_{-i}, \tau) \sim N\left(\frac{1}{n_i} \sum_{j \sim i} x_j, \frac{1}{n_i \tau}\right)$$

where $j \sim i$ means j *neighbour of* i . This gives:

Besag: random walk over areas, Besag (1974)

$$\pi(x_i | \mathbf{x}_{-i}, \tau) \sim N\left(\frac{1}{n_i} \sum_{j \sim i} x_j, \frac{1}{n_i \tau}\right)$$

where $j \sim i$ means j neighbour of i . This gives:

$$\pi(\mathbf{x} | \tau) \propto \tau^{\frac{n-1}{2}} \exp\left(-\frac{\tau}{2} \sum_i^n \left(x_i - \frac{1}{n_i} \sum_{j \sim i} x_j\right)^2\right)$$

Besag: random walk over areas, Besag (1974)

$$\pi(x_i | \mathbf{x}_{-i}, \tau) \sim N\left(\frac{1}{n_i} \sum_{j \sim i} x_j, \frac{1}{n_i \tau}\right)$$

where $j \sim i$ means j neighbour of i . This gives:

$$\begin{aligned}\pi(\mathbf{x} | \tau) &\propto \tau^{\frac{n-1}{2}} \exp\left(-\frac{\tau}{2} \sum_i^n \left(x_i - \frac{1}{n_i} \sum_{j \sim i} x_j\right)^2\right) \\ &= \tau^{\frac{n-1}{2}} \exp\left(-\frac{\tau}{2} \sum_{j \sim i}^n (x_i - x_j)^2\right) = \tau^{\frac{n-1}{2}} \exp\left(-\frac{\tau}{2} \mathbf{x}^T \mathbf{R} \mathbf{x}\right)\end{aligned}$$

$$\mathbf{R}_{ij} = \begin{cases} n_i & \text{if } i = j \\ -1 & \text{if } j \sim i \\ 0 & \text{otherwise} \end{cases}.$$

The Scotland graph

Scotland map



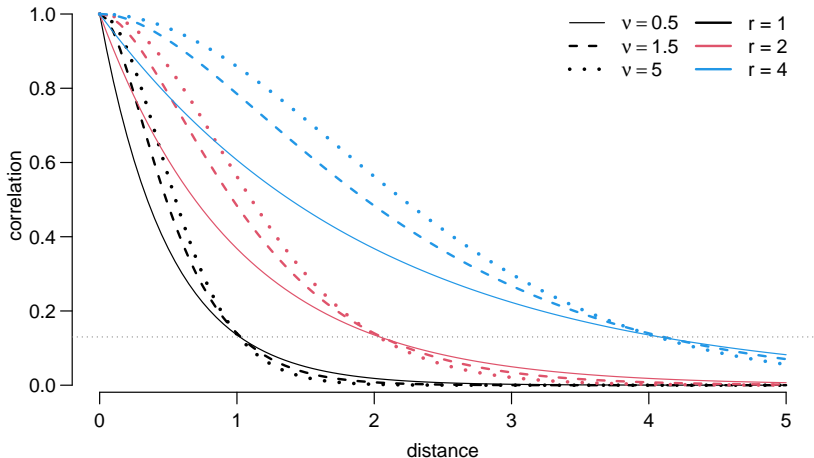
Neighborhood graph



The SPDE modeling approach

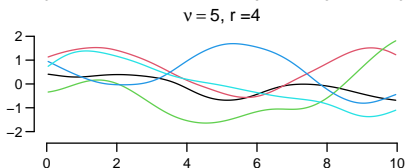
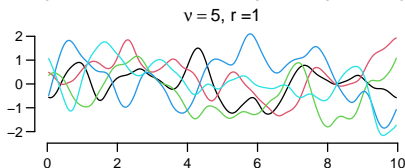
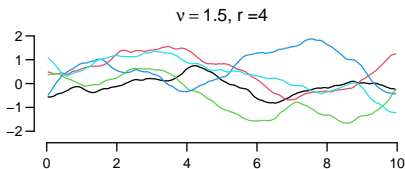
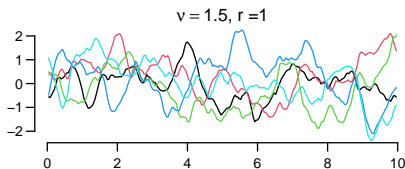
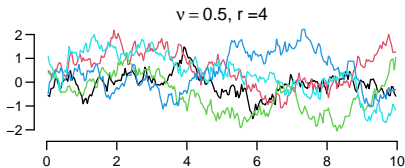
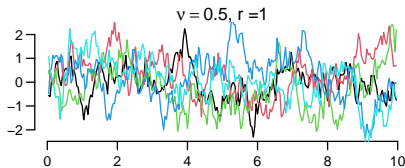
The Matérn covariance

$$\Sigma_{ij} = \sigma_x^2 \frac{2^{1-\nu} K_\nu(\kappa \|s_i - s_j\|)}{\Gamma(\nu) (\kappa \|s_i - s_j\|)^{-\nu}}, \text{ practical range} = r = \sqrt{8\nu}/\kappa$$



$$\text{corr}(\sqrt{8\nu}/\kappa) \approx 0.13$$

Simulations, 1D, $\sigma_x^2 = 1$



The SPDE approach

- Fields with Matérn covariance are solutions to the following Stochastic Partial Differential Equation (SPDE)

$$(\kappa^2 - \Delta)^{\alpha/2} \xi(\mathbf{s}) = \tau \mathcal{W}(\mathbf{s})$$

- $\kappa > 0$: scale parameter
- $\alpha = \nu + d/2$: smoothness
- Δ is the Laplacian

$$\Delta = \sum_{i=1}^d \frac{\partial^2}{\partial s_i^2}$$

- See Whittle (1954) and Lindgren, Rue, and Lindström (2011)

Regular grid, $d = 2$

- ▶ $\alpha = 1$: $\mathbf{Q}_{1,\kappa} = \mathbf{K}_\kappa = \kappa^2 \mathbf{C} + \mathbf{G}$
- ▶ $\mathbf{C} = \mathbf{I}$, $\mathbf{G} = \text{Laplacian (4 neighbours)}$
 - ▶ Laplacian-local pattern:

$$\begin{bmatrix} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{bmatrix}$$

- ▶ $\mathbf{Q}_{1,\kappa}$ -local pattern

$$\begin{bmatrix} & -1 & \\ -1 & 4 + \kappa^2 & -1 \\ & -1 & \end{bmatrix}$$

Regular grid, $d = 2$

- ▶ $\alpha = 1$: $\mathbf{Q}_{1,\kappa} = \mathbf{K}_\kappa = \kappa^2 \mathbf{C} + \mathbf{G}$
- ▶ $\mathbf{C} = \mathbf{I}$, \mathbf{G} = Laplacian (4 neighbours)
- ▶ Laplacian-local pattern:

$$\begin{bmatrix} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{bmatrix}$$

- ▶ $\mathbf{Q}_{1,\kappa}$ -local pattern

$$\begin{bmatrix} & -1 & \\ -1 & 4 + \kappa^2 & -1 \\ & -1 & \end{bmatrix}$$

- ▶ κ is a scale parameter
 - ▶ \rightarrow Sparse precision \mathbf{Q} {!!!}
 - ▶ remember: $(\kappa^2 - \Delta)^{\alpha/2} \xi(\mathbf{s}) = \tau \mathcal{W}(\mathbf{s})$
 - ▶ $\rightarrow (\mathbf{Q}_{1,\kappa})^{1/2} \xi = \text{independent noise}$
 - ▶ 'effective' range $(0.139) \approx \sqrt{8\nu/\kappa}$

Important fact: role of α

- ▶ Bigger $\alpha \rightarrow \mathbf{Q}$ less sparse \rightarrow smoother
 - ▶ $\alpha = 1$: $\mathbf{Q}_{1,\kappa} = \mathbf{K}_\kappa = \kappa^2 \mathbf{C} + \mathbf{G}$
 - ▶ $\alpha = 2$: $\mathbf{Q}_{2,\kappa} = \mathbf{K}_\kappa \mathbf{C}^{-1} \mathbf{K}_\kappa$
 - ▶ $\alpha = 3, 4, \dots$: $\mathbf{Q}_{\alpha,\kappa} = \mathbf{K}_\kappa \mathbf{C}^{-1} \mathbf{Q}_{\alpha-2,\kappa} \mathbf{C}^{-1} \mathbf{K}_\kappa$

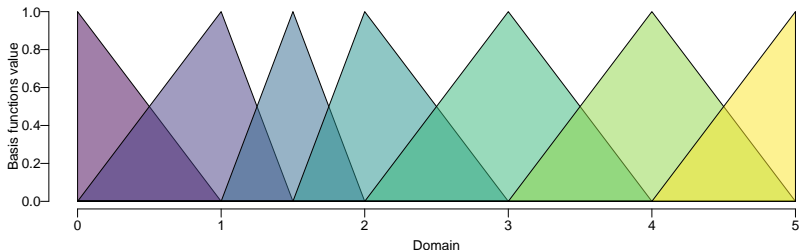
Important fact: role of α

- ▶ Bigger $\alpha \rightarrow \mathbf{Q}$ less sparse \rightarrow smoother
 - ▶ $\alpha = 1$: $\mathbf{Q}_{1,\kappa} = \mathbf{K}_\kappa = \kappa^2 \mathbf{C} + \mathbf{G}$
 - ▶ $\alpha = 2$: $\mathbf{Q}_{2,\kappa} = \mathbf{K}_\kappa \mathbf{C}^{-1} \mathbf{K}_\kappa$
 - ▶ $\alpha = 3, 4, \dots$: $\mathbf{Q}_{\alpha,\kappa} = \mathbf{K}_\kappa \mathbf{C}^{-1} \mathbf{Q}_{\alpha-2,\kappa} \mathbf{C}^{-1} \mathbf{K}_\kappa$
- ▶ $d = 1$ and $\kappa = 0$, the model at the knots is
 - ▶ $\alpha = 1$: like RW1
 - ▶ $\alpha = 2$: like RW2

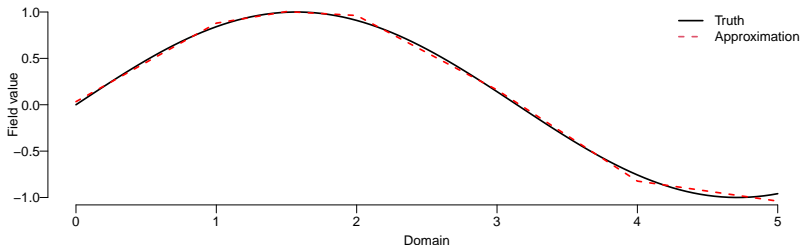
Important fact: role of α

- ▶ Bigger $\alpha \rightarrow \mathbf{Q}$ less sparse \rightarrow smoother
 - ▶ $\alpha = 1$: $\mathbf{Q}_{1,\kappa} = \mathbf{K}_\kappa = \kappa^2 \mathbf{C} + \mathbf{G}$
 - ▶ $\alpha = 2$: $\mathbf{Q}_{2,\kappa} = \mathbf{K}_\kappa \mathbf{C}^{-1} \mathbf{K}_\kappa$
 - ▶ $\alpha = 3, 4, \dots$: $\mathbf{Q}_{\alpha,\kappa} = \mathbf{K}_\kappa \mathbf{C}^{-1} \mathbf{Q}_{\alpha-2,\kappa} \mathbf{C}^{-1} \mathbf{K}_\kappa$
- ▶ $d = 1$ and $\kappa = 0$, the model at the knots is
 - ▶ $\alpha = 1$: like RW1
 - ▶ $\alpha = 2$: like RW2
- ▶ $d = 2$ (equivalent model at the mesh nodes)
 - ▶ $\alpha = 2$, Whittle (1954).
 - ▶ $\alpha = 1$, Besag (1974)
 - ▶ $\alpha = 1$ & $\kappa = 0$: intrinsic

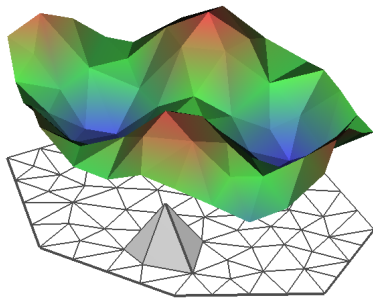
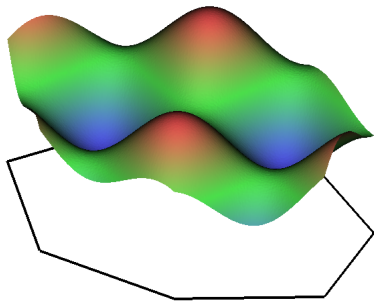
Continuous approximation: 1d case



- ▶ $\xi(\mathbf{s}) \approx \sum_{k=1}^m \psi_k(\mathbf{s}) \xi_k = \mathbf{A}(\mathbf{s}, \mathbf{s}_0) \xi(\mathbf{s}_0),$
 - ▶ ψ_k : basis functions evaluated at data locations \mathbf{s}
 - ▶ ξ_k : the process at the discretization points \mathbf{s}_0



Continuous approximation: 2d case



- ▶ $\xi(\mathbf{s}) \approx \sum_{k=1}^m \psi_k(\mathbf{s})\xi_k = \mathbf{A}(\mathbf{s}, \mathbf{s}_0)\xi(\mathbf{s}_0),$
 - ▶ ψ_k : basis functions evaluated at data locations \mathbf{s}
 - ▶ ξ_k : the process at the discretization points \mathbf{s}_0

SPDE as a random effect in INLA

- ▶ The (1d, 2d, ...) SPDE models are processes on a continuous domain
- ▶ Used as a random effect on a set of knots/nodes
 - ▶ Needs to be projected to the observation location

$$\eta = \mathbf{X}\beta + \mathbf{A}\xi$$

- ▶ \mathbf{X} is n (observation) times p (covariates)
 - ▶ \mathbf{A} is n (observation) times m (knots/nodes)
- ▶ see more on Krainski et al. (2018)

References

References

- Besag, J. 1974. "Spatial Interaction and the Statistical Analysis of Lattice Systems." *JRSS-B* 36 (2): 192–236.
- Krainski, E. T., V. Gómez-Rubio, H. Bakka, A. Lenzi, D. Castro-Camilio, D. Simpson, F. Lindgren, and H. Rue. 2018. *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*. New York: Chapman; Hall/CRC. <https://doi.org/10.1201/9780429031892>.
- Lindgren, F., H. Rue, and J. Lindström. 2011. "An Explicit Link Between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach (with Discussion)." *JRSS-B* 73 (4): 423–98.
- Rue, H., and L. Held. 2005. *Gaussian Markov Random Fields: Theory and Applications*. Monographs on Statistics & Applied Probability. Boca Raton: Chapman; Hall.
- Whittle, P. 1954. "On Stationary Processes in the Plane." *Biometrika* 41 (3/4): 434–49.