Regression Kriging (RK)
Kriging with an External Drift (KED)

D G Rossiter

Nanjing Normal University, Geographic Sciences Department
南京师范大学地理学学院

November 8, 2018

General objective: spatial prediction

- **Objective**: Given a set of **attribute values** at **known points**, **predict** the value of that attribute at other points.
  - Also with the **uncertainty** of the prediction.
- **Objective**: **Understand** why the attribute has its spatial distribution.
- This lecture: **Regression Kriging (RK)** for both objectives.

A taxonomy of prediction methods

Strata: divide area to be mapped into 'homogeneous' **strata**; predict **within each stratum** from all observations in that stratum.

Global: (or "regional") predictors: use **all observations** to build a model that allows to predict at **all points.**

Local: predictors: use only **'nearby'** observations to predict at each point.

- geostatistical: with an **model of local spatial dependence**, e.g., **Ordinary Kriging**
- empirical: directly adjusting to the data, e.g., **thin-plate splines**
- non-geostatistical: with an implicit model, not from the data, e.g., **inverse distance**

Mixed: predictors: some of structure is explained by strata or globally, the **residuals** from this are explained **locally**, e.g., **Regression Kriging**;

Which prediction method is "best"?

- There is **no theoretical answer.**
- It depends on how well the approach models the **'true' spatial structure**, and this is unknown (but we may have **prior evidence**).
- The method should correspond with what we know about the **process** that created the spatial structure.
- It should also be achievable with the available data.

(continued . . .)

Which prediction method is "best"? (continued)

- Check against an independent **validation** dataset
  - **Mean squared error** ("precision") of **prediction** vs. **actual** (residuals)
  - **Bias** ("accuracy") of predicted vs. actual mean
  - How well it reproduces the **spatial variability** of the calibration dataset.
- With **large** datasets, model with one part and hold out the rest for **validation**
- For **small** datasets use **cross-validation**

Mixed predictors

- When there is **both** long-range structure, (*trend*, strata, or covariables **and** local structure
  - Covariable values known at all prediction points
- One approach: **Regression Kriging** (RK): model trend/strata/covariables and their residuals **separately**, add for final result (see details below).
  - advantage: can use any kind of model for the "regression" part, not just a linear model
- Another approach: model everything **together**, e.g. **Kriging with External Drift** (KED)
  - must use a linear model for the "regression" part
  - problem: true spatial structure of model residuals not known

Universal model of spatial variation

$$Z(\mathbf{s}) = Z^*(\mathbf{s}) + \varepsilon(\mathbf{s}) + \varepsilon'(\mathbf{s}) \quad (1)$$

$(\mathbf{s})$ a location in space, designated by a **vector** of coördinates

$Z(\mathbf{s})$ **true** (unknown) value of some property at the location

- when modelled, expressed as **most likely** value and some **uncertainty**, or as a **probability distribution**

$Z^*(\mathbf{s})$ **deterministic** component, due to some known or modelled **non-stochastic** process

$\varepsilon(\mathbf{s})$ **spatially-autocorrelated stochastic** component

$\varepsilon'(\mathbf{s})$ pure (“white”) **noise**, no structure

Mixed vs. local geostatistical models

- In OK $Z^*(\mathbf{s}) \equiv \mu$, i.e., 1st order stationarity
  - same expected value everywhere
  - differences are due to spatially-correlated random variation $\varepsilon(\mathbf{s})$
- in RK or KED, relax that assumption: **non-stationary expected value** $Z^*(\mathbf{s})$
  - This must be predicted by some deterministic model.
- But we still assume 2nd order stationarity:
  - the covariance structure of $\varepsilon(\mathbf{s})$ is the same everywhere
  - it only depends on the separation between point-pairs

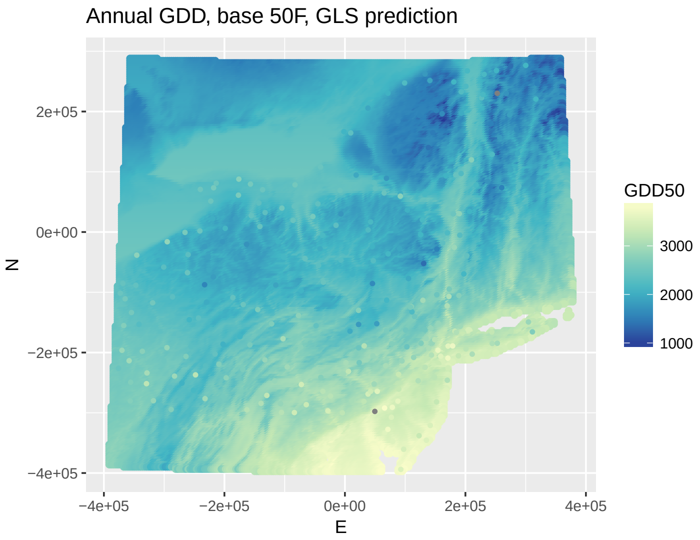
Regression kriging

1. Predict **trend** over the area
   - typically by multiple linear regression fit by GLS
   - Must know the predictor (indepenent) variables at all locations to be predicted
     - e.g., coördinates for a trend surface
     - e.g., covariables such as elevation, vegetation index, terrain parameters . . .
2. Model the spatial structure of the **residuals** from the trend
3. Predict **residuals** over the area with Ordinary Kriging (OK),
4. Add predictions from (1) + (2) $\longrightarrow$ **prediction**
5. Add prediction variances from (1) + (2) $\longrightarrow$ **prediction variance**

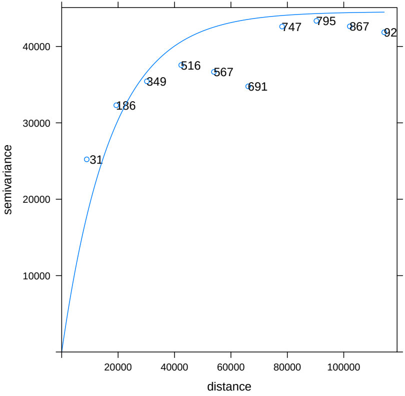
GLS trend surface model in R/nlme

```
> library(nlme)  ## this includes the gls method
> m.gls <- gls(model=ANN_GDD50 ~ sqrt(ELEVATION_) + N,
>             data=ne.df,
             correlation=corExp(value=c(50000, 0.1),
                          form=~E + N, nugget=FALSE))
> summary(m.gls)

Correlation Structure: Exponential spatial correlation
 Formula: ~E + N
 Parameter estimate: range 17460.45

Coefficients:
                    Value Std.Error   t-value p-value
(Intercept)      4090.972  60.15542  68.00671       0
sqrt(ELEVATION_)  -30.045   1.41095 -21.29447       0
N                  -0.002   0.00011 -16.63821       0

Residual standard error: 217.3552
Degrees of freedom: 305 total; 302 residual
```

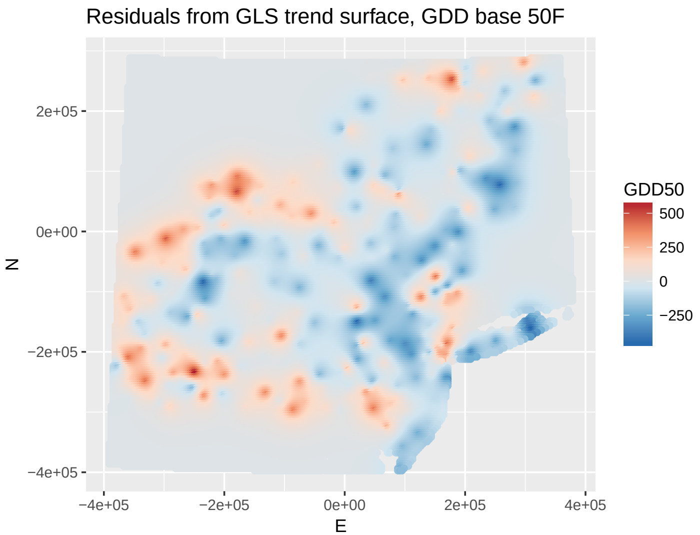
Example: (1) trend surface



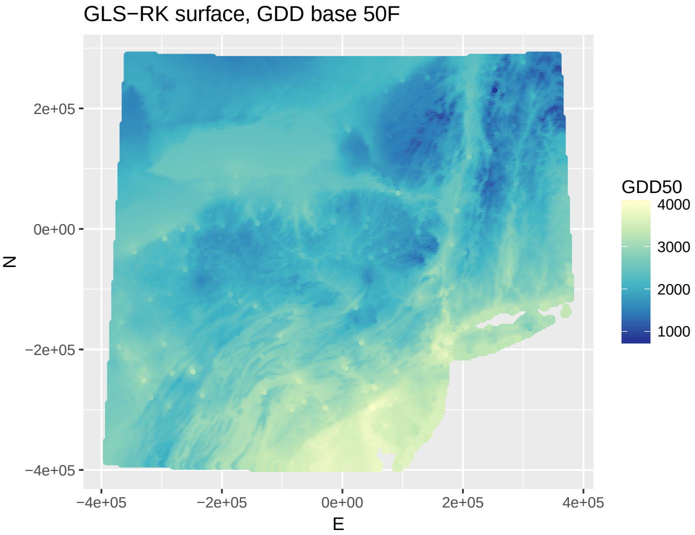
Example: (2) spatial correlation of residuals



Example: (3) OK of residuals



Example: (4) trend + OK residuals = RK



Kriging with External Drift (KED)

- Like RK, also predicts the trend plus local deviations
- One prediction step (KED), not two
  - RK: trend + kriged trend residuals
- Mathematically equivalent to Universal Kriging (UK)
  - UK uses only coördinates for the trend (some say)
- Two steps:
  1. Fit a variogram model of the OLS residuals, without explicitly computing an OLS model
     - This differs from GLS, which computes the covariance structure by REML, and from that the GLS trend.
  2. Krige with the same covariates; this does use the GLS estimate of the trend, assuming the residual variogram model from step (1)

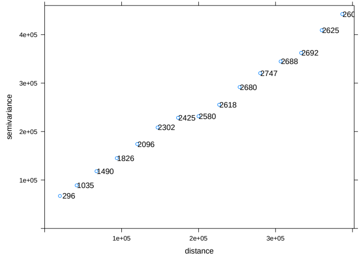
KED in R/gstat (1) - residual variogram model fitting

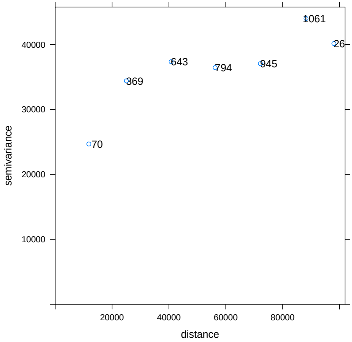
```
> v.ked <- variogram(ANN_GDD50 ~ sqrt(ELEVATION_) + N,
                     locations=ne.m,
                     cutoff=100000, width=16000)
> (vmf.ked <- fit.variogram(v.ked,
                    vgm(15000, "Exp", 20000, 20000)))
  model    psill    range
1   Nug  1435.79     0.00
2   Exp 37378.22 12104.31
```

Note the formula `ANN_GDD50 ~ sqrt(ELEVATION_) + N`, this first fits an OLS model to this formula, and then extracts the **residuals** to build the empirical variogram.

Any long-range structure is taken out by the global model, using the spatial distribution of the covariates and target.

Ordinary vs. residual variograms





Unbounded (no range) to $\geq$ 400 km

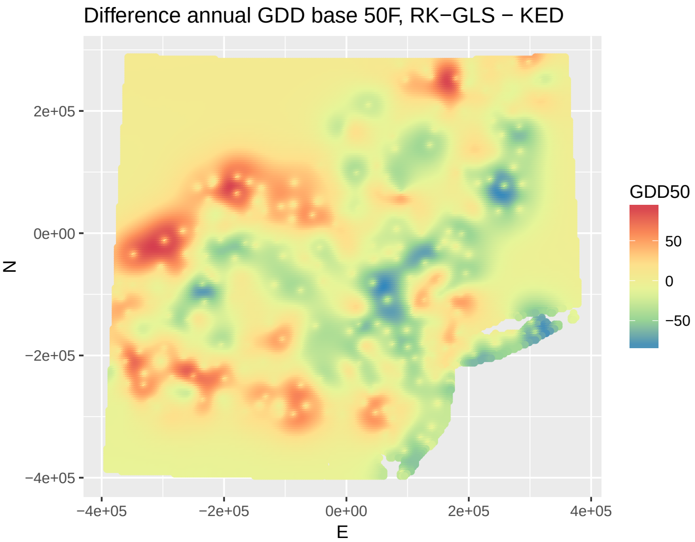
Range $\approx$ 40 km

KED in R/gstat (2) - prediction

```
> k.ked <- krige(ANN_GDD50 ~ sqrt(ELEVATION_)+ N,
               locations=ne.m,
               newdata=dem.ne.m.sp, model=vmf.ked)
[using universal kriging]

> summary(k.ked@data)

   var1.pred        var1.var
 Min.   : 912.6   Min.   : 2981
 1st Qu.:1943.9   1st Qu.:32459
 Median :2268.0   Median :36761
 Mean   :2321.0   Mean   :34751
 3rd Qu.:2608.3   3rd Qu.:39182
 Max.   :3943.4   Max.   :42476
```

Note the *same formula* is used in `krige` as in `variogram`.

Example: (5) difference RK-GLS - KED



Practical advantages of KED over RK-GLS

1. It is easier to implement;
2. The covariance structure is estimated beforehand, and there is no risk that the procedure might not converge on a solution, as in REML;
3. It gives a prediction variance in the same step.

But. . . its predictions are suboptimal.

End