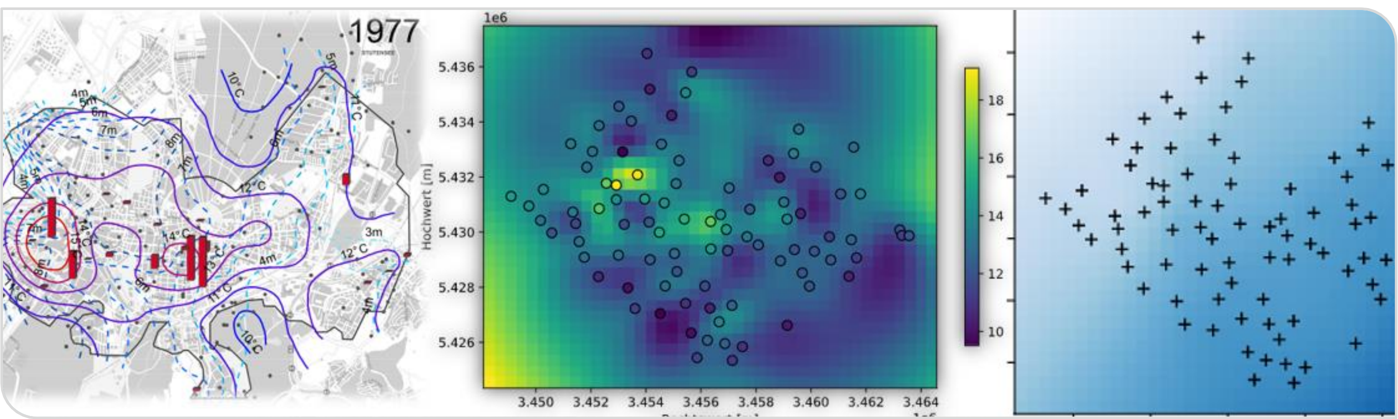
Karlsruher Institut für Technologie

Geodatenanalyse I: Interpolation – Kriging

Kathrin Menberg



Stundenplan

Karlsruher Institut für Technologie

	08:30 – 12:30 Uhr	13:30 – 17:30 Uhr
Montag	Tag 1 / Block 1	Tag 1 / Block 2
Dienstag	Tag 2 / Block 1	Tag 2 / Block 2
Mittwoch	Tag 3 / Block 1	Tag 3 / Block 2
Donnerstag	Tag 4 / Block 1	**Tag 4 / Block 2**
Freitag	Tag 5 / Block 1	Tag 5 / Block 2

- 2.10 Interpolation: Deterministische Verfahren
- **2.11 Interpolation: Kriging**
- 2.12 Gauß-Prozesse

Lernziele Block 2.11

Am Ende der Stunde werden die Teilnehmer:

- … mit den mathematischen Grundlagen der Semivariogramm Analyse vertraut sein.
- … verschiedenen Typen von Variogrammen kennen und diese auf Datensätze anpassen können.
- … in Python Kriging für Datensätze mit und ohne Trend durchführen können.

Stochastische Interpolation

Karlsruher Institut für Technologie

- Geostatistische Verfahren
- Ergebnis ist eine von vielen möglichen Realisationen
- Interpolation basierend auf den statistischen Eigenschaften der Daten
- Statistische Auswertung der Wahrscheinlichkeit und Unsicherheit der Interpolationsergebnisse
- Möglichkeit der Qualitätsbestimmung der Interpolation
- z.B. Kriging

Kriging

- Motivation: Bestimmung der Wahrscheinlichkeit an einem Ort Gold zu finden, basierend auf Proben von ein paar wenigen Bohrlöchern
- Verfahren basierend auf Inverse Distance Weighting…
- … unter Berücksichtigung der **räumlichen Varianz** der Daten
- Kriging gliedert sich in zwei Schritte:
  - Analyse der räumlichen Korrelation, bzw. Varianz der Daten
  - eigentliche Kriging-Interpolation

Kriging Grundgleichung

Karlsruher Institut für Technologie

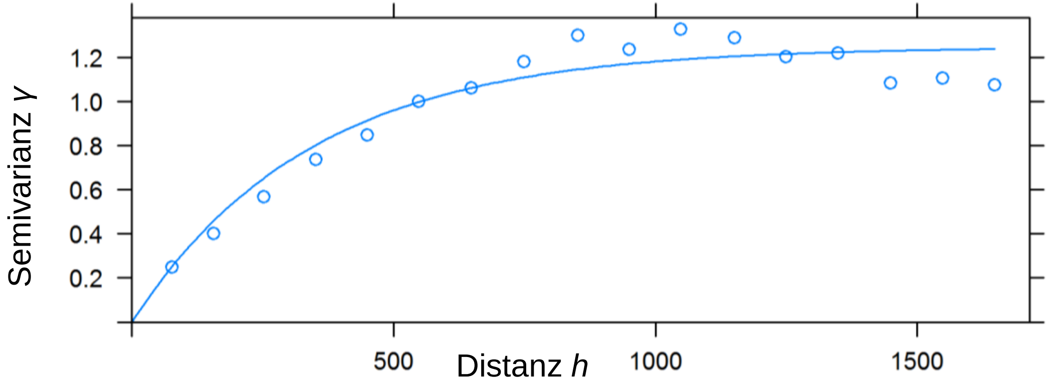
$$\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i \cdot Z(x_i), \qquad mit \sum_{i=1}^{n} \lambda_i = 1$$

- $\hat{Z}(x_0)$ = Schätzwert am Punkt x_0
- *n* = Anzahl Datenwerte, die zur Schätzung herangezogen werden
- $Z(x_i)$ = Messwert am Punkt x_i
- λ_i = Gewichte, mit der der jeweilige Datenwert bei der Interpolation gewichtet wird

- Unterschiede zu IDW:
  - Berechnung der Gewichte
  - Annahme dass $\hat{Z}(x_0)$ eine Zufallsgröße ist

Kriging Gewichtung

- Berücksichtigung von Heterogenität der Messpunkte
  - Gewichte von Punkten innerhalb von Clustern werden gesenkt
- Bestimmung der Gewichte so, dass Varianz des Schätzfehlers möglichst gering ist
- Ermittlung der Gewichte mittels (Semi-) Variogrammanalyse



Variogrammanalyse

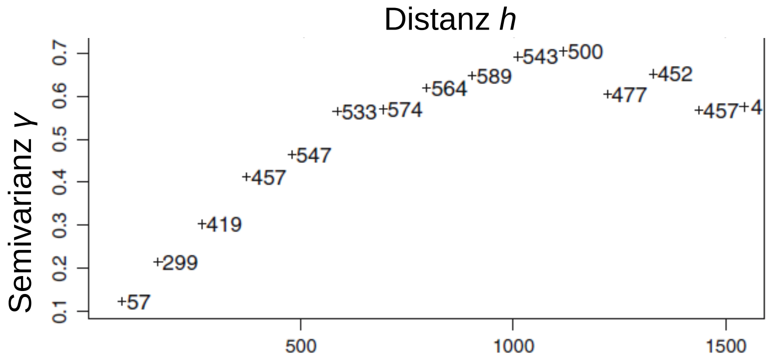
- Semivarianz $\gamma(h)$: Maß für den Grad der räumlichen Abhängigkeit von Messwerten

$$\gamma(h) = \frac{1}{2n}\sum_{i=1}^{n}(Z(x_i) - Z(x_i + h))^2$$

- n = Anzahl Datenwerte, die zur Schätzung herangezogen werden
- $Z(x_i)$ = Messwert am Punkt xi
- $Z(x_i + h)$ = Messwert an einem Punkt im Abstand h

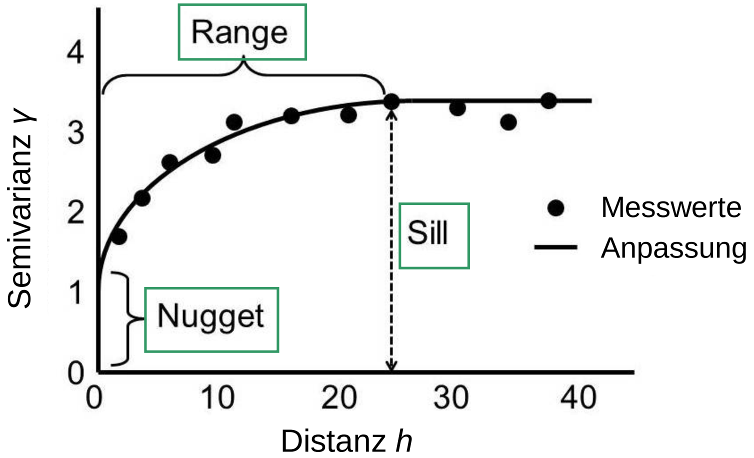
Bivand et al. (2008)

- halbe, mittlere, quadrierte euklidische Distanz zwischen zwei Messwerten



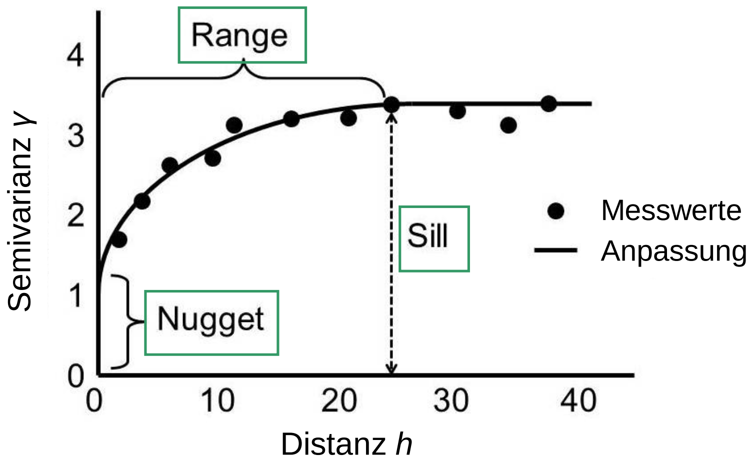
Experimentelles Variogramm

- Auftragen der (Semi-)Varianz $\gamma(h)$ über dem Abstand h
- Experimentelles Variogramm → Messwerte



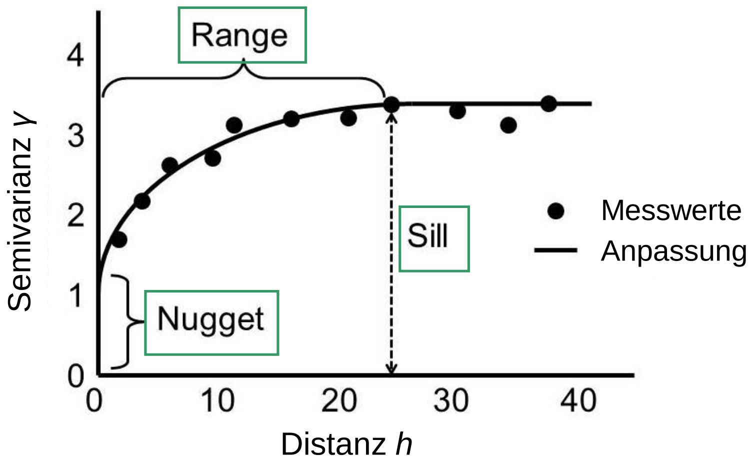
Schwellenwert und Reichweite

- Schwellenwert (Sill): verglichene Werte haben keinen Bezug mehr zueinander, ihre quadrierten Differenzen entsprechen der Varianz um den Mittelwert
- Reichweite (Range) = Abstand, unterhalb dem die Werte als räumlich in Beziehung stehend gelten können, darüber keine räumliche Korrelation mehr



Nugget-Effekt

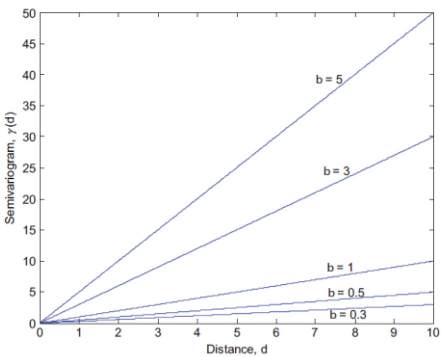
- Nugget (y-Achsen-Abstand): charakterisiert Variablen, deren Variabilität kleinräumiger als die geringsten Probenabstände ist.
- kann aber auch durch Fehler im Datensatz, z.B. durch aufgrund der Probenahme oder der Analyse, verursacht werden.



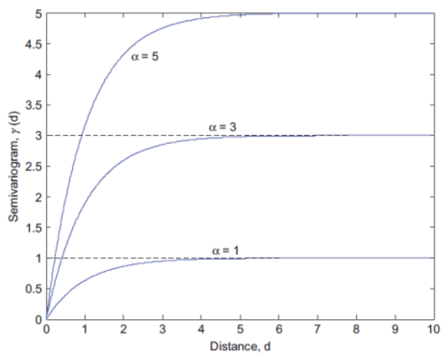
Theoretisches Variogramm

- Anpassung des experimentellen Variogramms mit einer Funktion
- z.B. sphärische Funktion, lineare Funktion, usw.

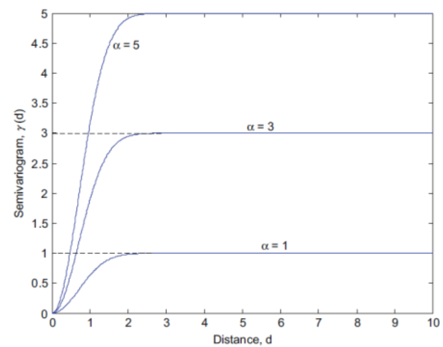
Lineares Variogramm



Exponentielles Variogramm



Gauß-Variogramm



Bivand et al. (2008)

Ordinary Kriging

- Es liegt kein Trend (und keine Drift) in den Daten vor, d.h. die Differenz zwischen Schätzung und den wahren Werten soll im Mittel gleich 0 sein
- Die Varianz (Kriging-/Schätzvarianz) des Schätzfehlers soll minimal sein.
- Die Summe aller Gewichte muss 1 ergeben

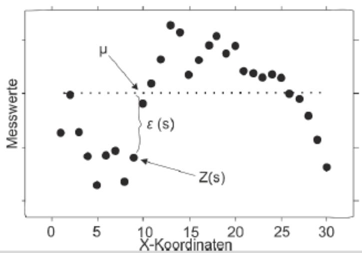
$$\hat{\sigma}^2 = \sum_{i=1}^{n} \lambda_i \cdot \gamma_{i0} + \mu \quad \rightarrow \text{min.}$$

und $$\sum_{i=1}^{n} \lambda_i = 1$$

$\hat{\sigma}^2$ = Varianz
λ_i = Gewicht für Messpunkt i
γ_{i0}= Semivarianz für Messpunkt i und Punkt 0
μ = Lagrange-Faktor

$$Z(s) = \mu(s) + \varepsilon(s)$$

Z(s): Vorhersagewert (an Lokalität *s*)
μ(s): deterministischer Trend (Erwartungswert)
ε(s): autokorrelierte Zufallsfehler

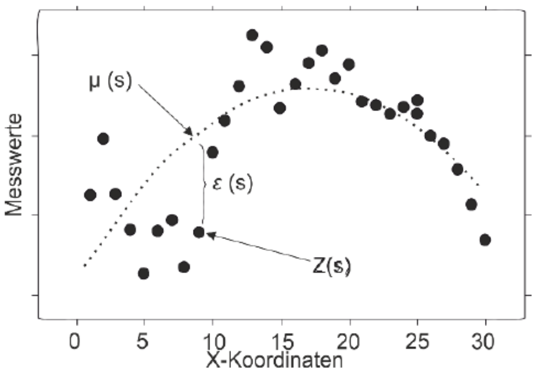


Universal Kriging

- Finden Anwendung, wenn Daten mit einem Trend (oder Drift) behaftet sind
- Trend: über das gesamte Gebiet (globaler Trend)
- Drift: nur lokal (lokale Drift)

$$Z(s) = \mu(s) + \varepsilon(s)$$

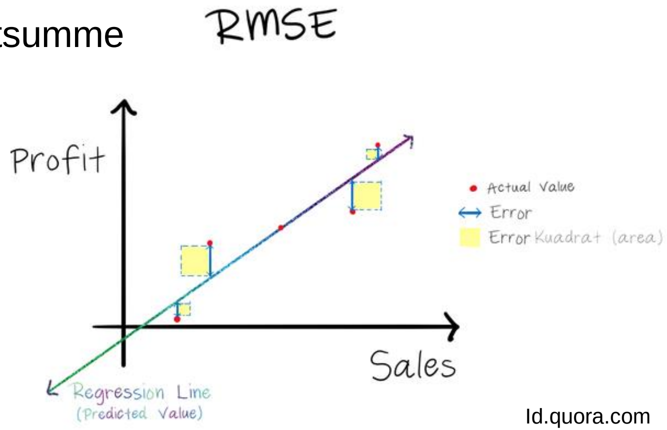
mit: $Z(s)$: Vorhersagewert (an Lokalität s)
$\mu(s)$: deterministischer Trend (Erwartungswert)
$\varepsilon(s)$: autokorrelierte Zufallsfehler



Root Mean Square Error (RMSE)

- dt. Wurzel der mittleren Fehlerquadratsumme
- y: Beobachtungen, $\hat{y}_i$ Vorhersagen

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

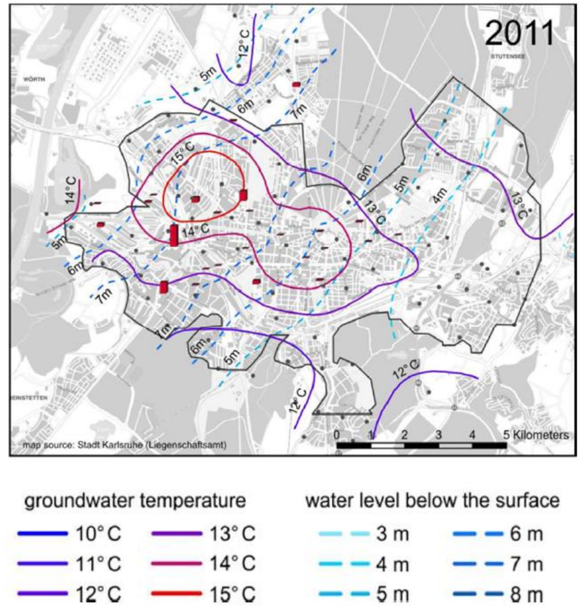


Id.quora.com

- Maß für die Güte der Anpassung, bzw. Genauigkeit der Vorhersagen
- Perfekte Anpassung, bzw. Übereinstimmung $RMSE = 0$
- Je größer, desto schlechter ist die Anpassung
- Magnitude abhängig vom Maßstab der Datenwerte
  - Relatives Fehlermaß!

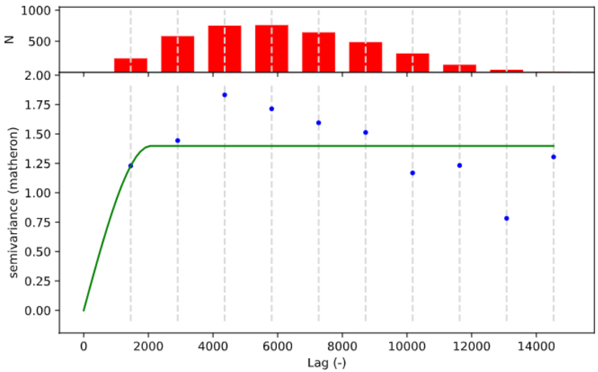
Übung 2.11: Interpolation II

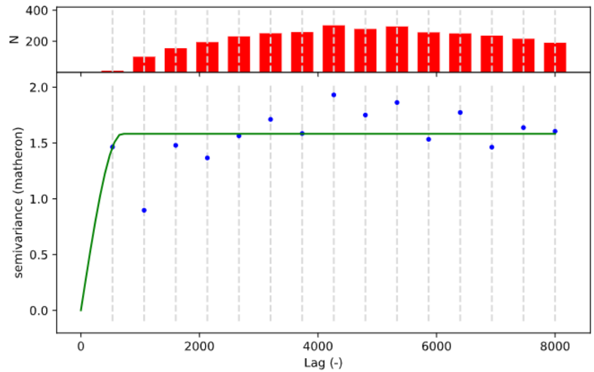
- Kriging mit Grundwasserdaten aus Karlsruhe
  - Ordinary Kriging der Grundwassertemperaturen
  - Universal Kriging der Grundwasserstände
  - Analyse der Vorhersagegenauigkeit
- Aufgaben in Jupyter Notebook: geodatenanalyse_1-2-11

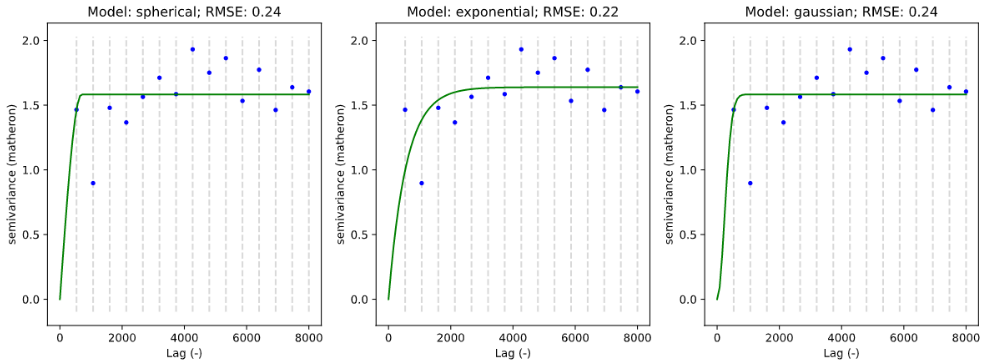


Menberg et al. (2013)

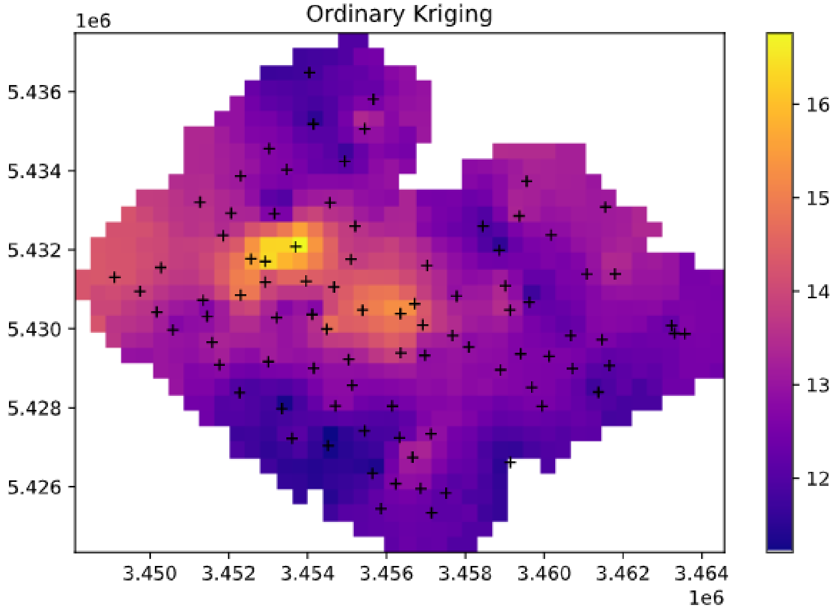
Aufgabenbesprechung





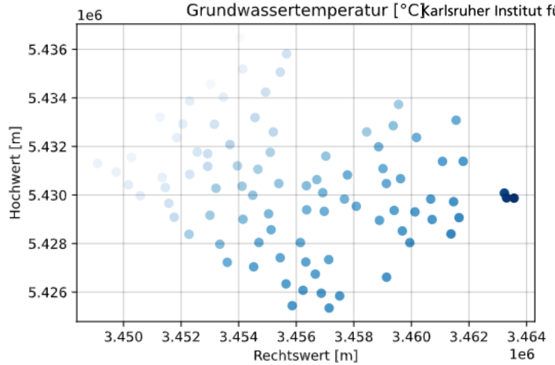


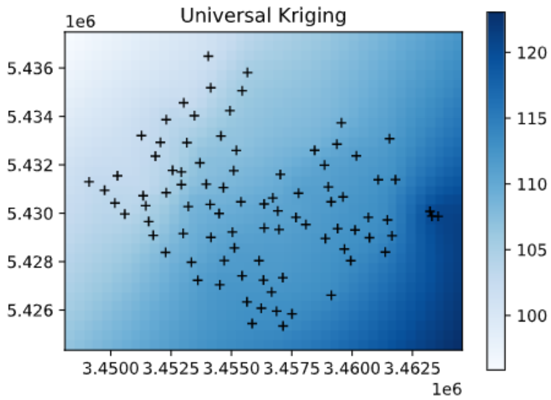
Aufgabenbesprechung

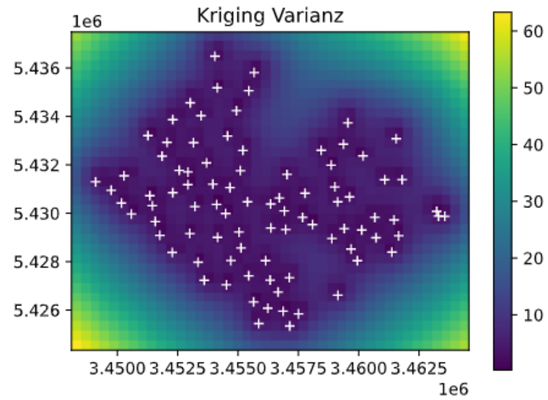


Aufgabenbesprechung

Karlsruher Institut für Technologie







Literatur

- Bivand, Pebesma & Gomez-Rubio (2008): Applied Spatial Data Analysis with R, Springer
- Oliver & Webster (2015): Basic Steps in Geostatistics: The Variogram and Kriging, Springer
- Menberg et al. (2013): Long-term evolution of anthropogenic heat fluxes into a subsurface urban heat island, Environ. Sci. Technol. 47(17) (2013) 9747-9755

Nützliche Weblinks:

- https://towardsdatascience.com/what-does-rmse-really-mean-806b65f2e48e

Karlsruher Institut für Technologie