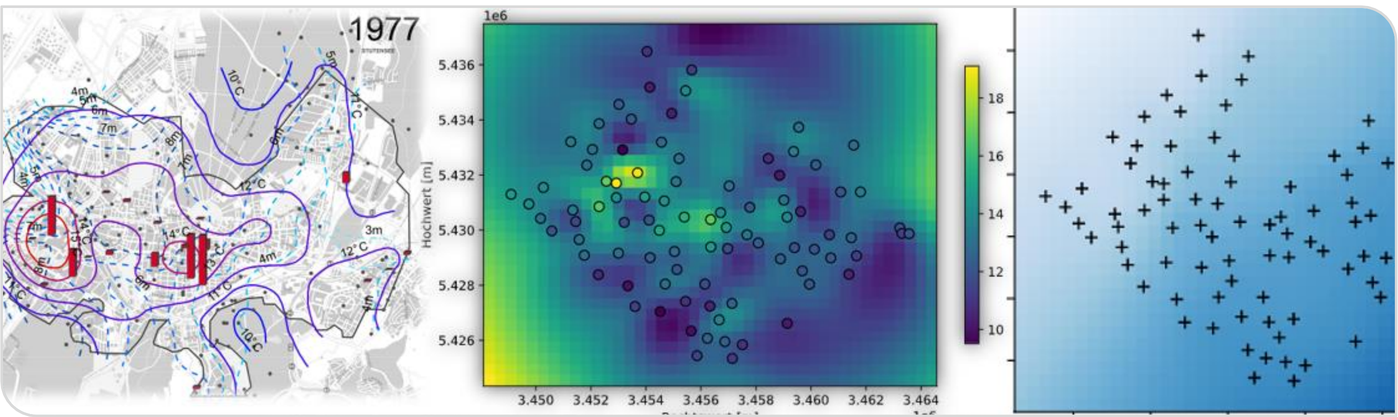
Karlsruher Institut für Technologie

Geodatenanalyse I: Räumliche Interpolation

Kathrin Menberg



Stundenplan

Vorläufiger Stundenplan		
Datum	**Thema**	**Dozent**
20.10.2021	Einführung in die Programmierung mit *Python*	Gabriel Rau
25.10.2021	Univariate Statistik und statistisches Testen	Kathrin Menberg
01.11.2021	*Feiertag*	
08.11.2021	Umgang und Berechnung von Datensätzen	Gabriel Rau
15.11.2021	Bivariate und schließende Statistik	Kathrin Menberg
22.11.2021	Datenvisualisierung mit *matplotlib*	Gabriel Rau
29.11.2021	Multivariate Statistik	Kathrin Menberg
06.12.2021	Datenformate, Datenspeicherung und Datenbanken	Gabriel Rau
13.12.2021	Monte-Carlo Methoden	Kathrin Menberg
20.12.2021	Analyse und Visualisierung von Geodaten	Gabriel Rau
27.12.2021	*Weihnachtsferien*	
03.01.2022	*Weihnachtsferien*	
10.01.2022	Sensitivitätsanalyse	Kathrin Menberg
17.01.2022	Datenethik, Lizensierung und Entwicklungstools	Gabriel Rau
24.01.2022	Räumliche Interpolation	Kathrin Menberg
31.01.2022	Fragen zur Programmierung	Gabriel Rau
07.02.2022	Regressionsanalyse	Kathrin Menberg

Vorlesungsplan

Uhrzeit	Inhalt
10:00 – 10:20	**Deterministische Interpolation**
10:20 – 11:00	Übung
11:00 – 11:10	Diskussion und Reflexion
11:10 – 11:25	Pause
11:25 – 11:45	Kriging
11:45 – 12:20	Übung
12:20 – 12:30	Diskussion und Reflexion

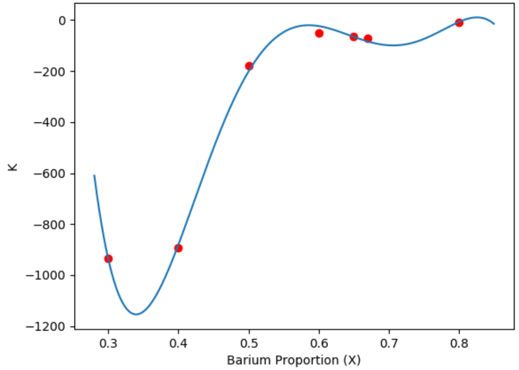
Lernziele

Am Ende der Stunde werden die Teilnehmer:

- … mit den mathematischen Grundlagen der Interpolation vertraut sein.
- … verschiedene Methoden zur Interpolation voneinander abgrenzen können.
- … in Python deterministische Methoden zur Interpolation von Datensätzen anwenden und die Ergebnisse graphisch darstellen können.

Interpolation

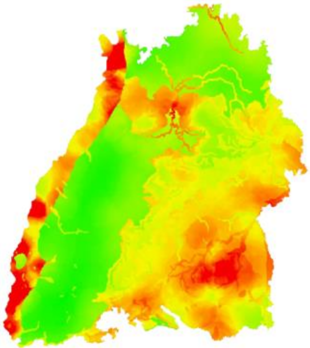
- Motivation
  - $X_i \rightarrow f(x)$
  - räumliche Interpolation



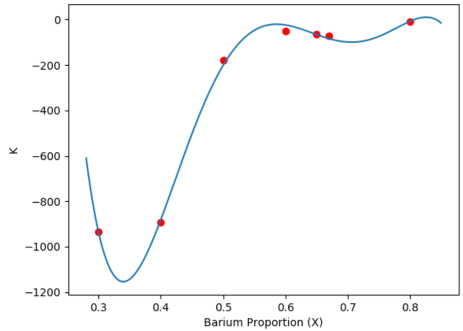
https://datascience.com/

Punkte (Linien)

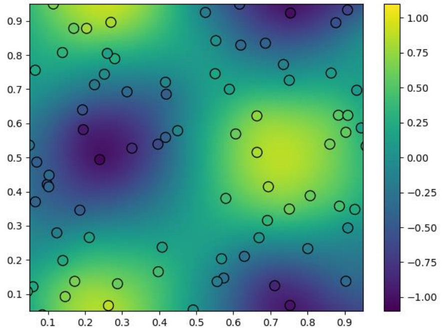
Isolinien, Rasterdaten



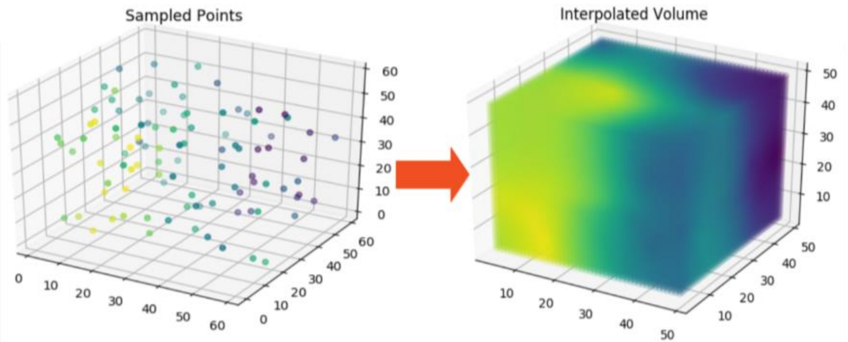
1D, 2D, n-dimensionale Interpolation



https://datascience.stackexchange.com/



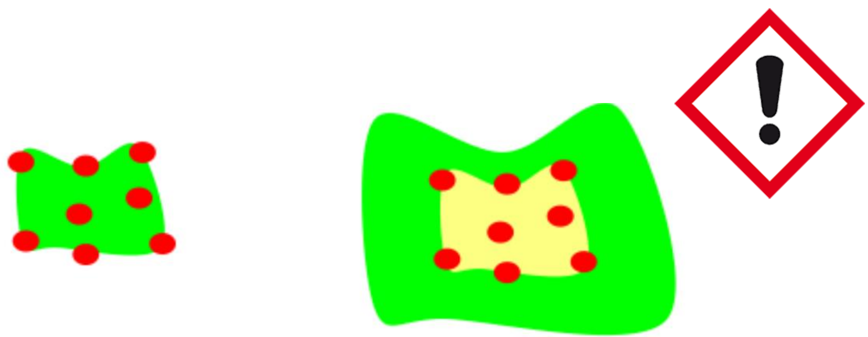
https://rbf.readthedocs.io/



https://innolitics.com/

Interpolation vs. Extrapolation

- Interpolation: Schätzung von Werten zwischen bekannten Datenpunkten
- Extrapolation: Schätzung von Werten außerhalb bekannter Datenpunkte



Interpolation **Extrapolation**

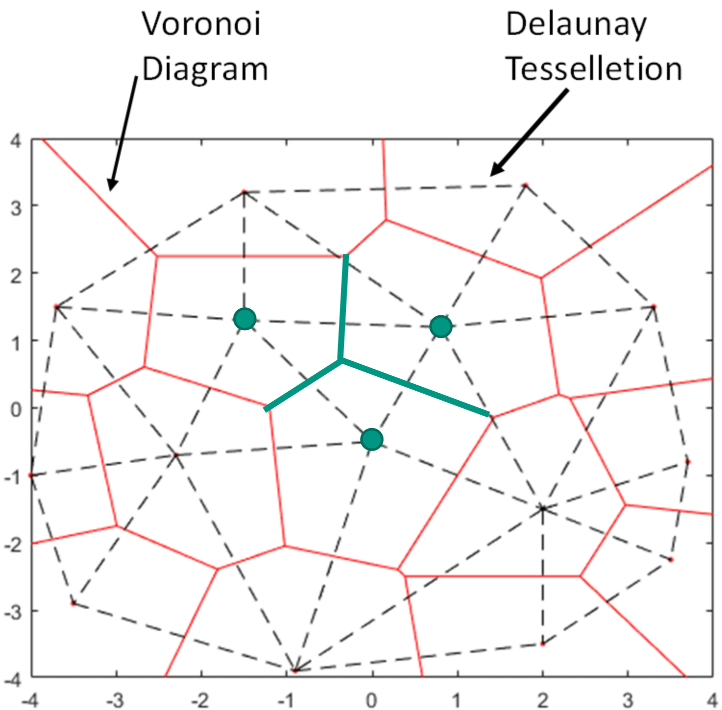
https://www.informatik.uni-augsburg.de/

Deterministische Interpolation

- Interpolation mittels festgelegter (z.B. linearer) Funktion
- Eindeutiges, immer gleiches Ergebnis
- Einfache Berechnung, aber keine Aussage zur Qualität der Interpolation möglich
- Beispiele:
  - Triangulation
  - Nearest Neighbour
  - Natural Neighbour
  - Spline Interpolation
  - Polynomische Interpolation
  - Inverse Distance Weighting
  - usw.

Nearest Neighbour

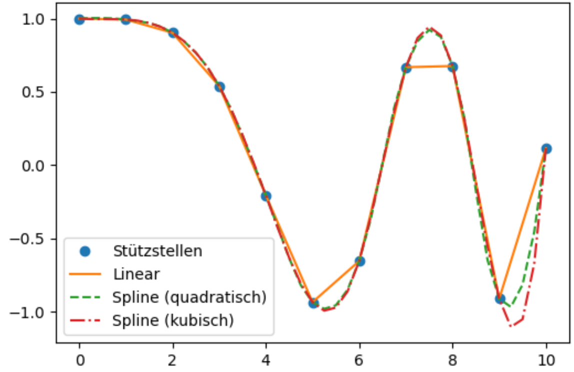
- Verbindung jeweils dreier benachbarter Punkte mittels Delaunay-Triangulation
- Mittelsenkrechten der Dreiecksmaschen ergeben sog. Thiessen-Polygone oder auch Voronoi-Polygone
- Zuordnung des jeweils nächstgelegenen Messwertes („nearest neighbour“) für das ganze Polygon



De Carvalho & Prati (2018)

Radiale Basisfunktionen (RBF)

- Reelle Funktion φ, deren Wert nur vom Abstand zum Ursprung abhängt
- Abstand: Euklidische Distanz
- Verschiedene Funktionstypen:
  - Linear
  - Spline $\varphi(r) = r^k$, $\varphi(r) = r^k \ln(r)$
  - Multi-quadratisch $\varphi(r) = \sqrt{1 + (ar)^2}$
  - Gauss $\varphi(r) = e^{-(ar)^2}$
  - …
- Annahme Formparameter a, bzw. k



- Approximation von Werten zwischen Stützpunkten
  - Interpolation
  - Maschinelles Lernen (z.B. Neuronale Netze)

Inverse Distance Weighting

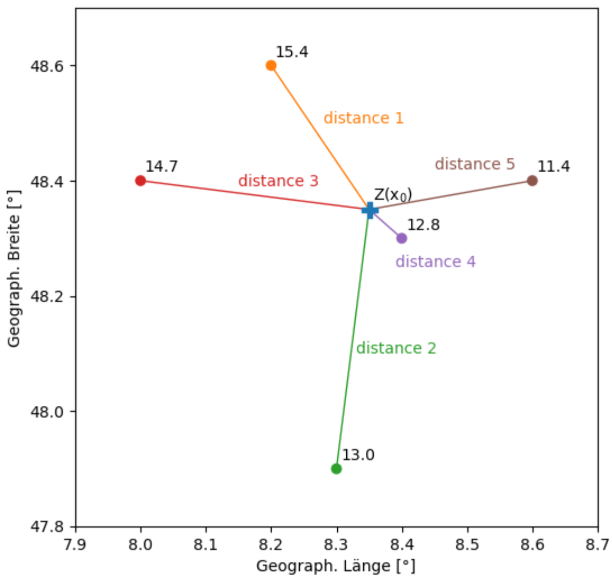
- Interpolationswert $\hat{Z}$ an Stelle x_0 wird berechnet aus den Messwerten $Z(x_i)$ der benachbarten Punkte $x_1 \ldots x_i$

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

- λ_i ist die Gewichtung, mit dem der Wert x_i an Punkt i in die Berechnung einfließt

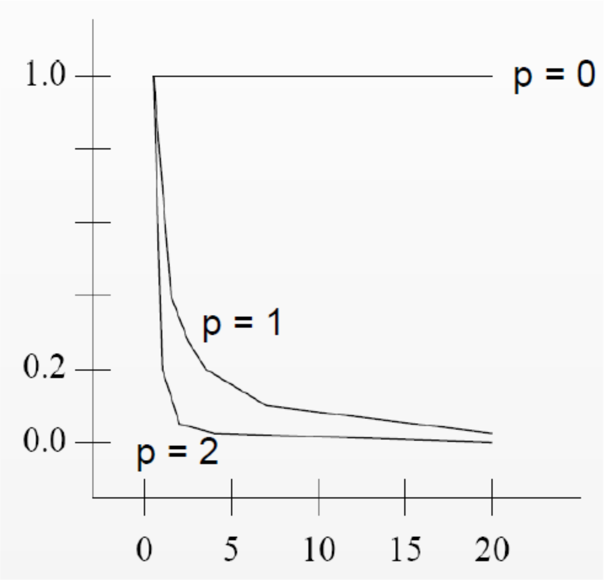
$$\lambda_i = \frac{d_{i0}^{-p}}{\sum_{i=1}^{n} d_{i0}^{-p}}$$

- Mit zunehmender Entfernung d nimmt das Gewicht ab.



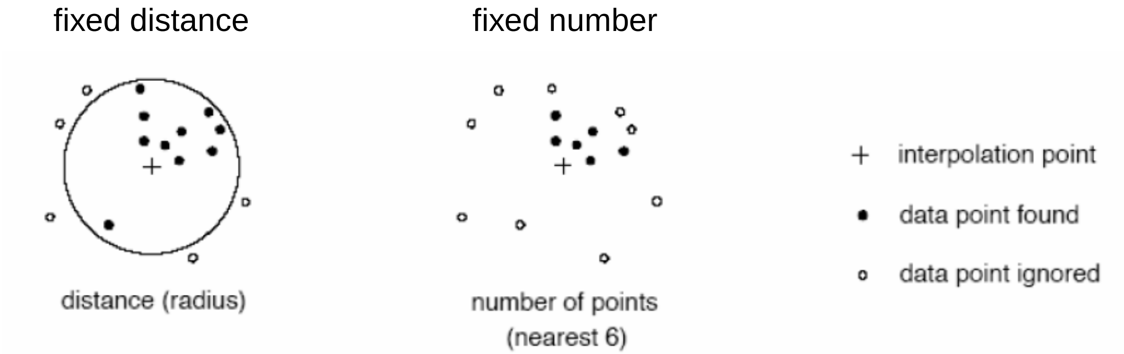
Inverse Distance Weighting

- p „Power“ ist ein Maß für Abnahme (i.d.R. >1, z.B. 2)
- Einfluss des Wertes p
- Welches der optimale p-Wert ist, kann man über Validierungsverfahren feststellen.



Inverse Distance Weighting

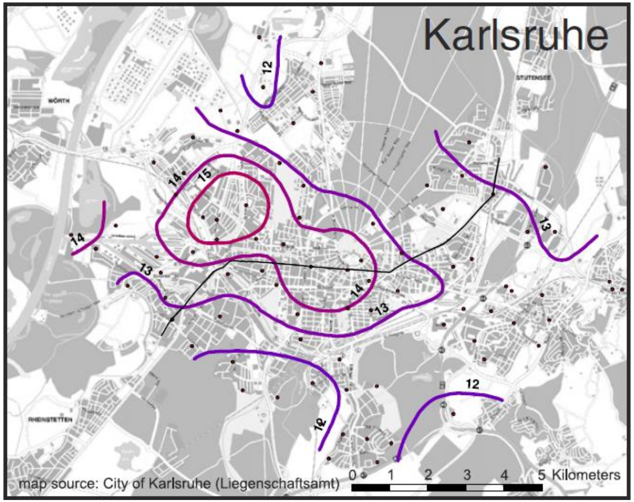
- Anzahl der bei der Interpolation berücksichtigten Punkte



- Welche Methode man wählt, hängt sehr stark von der Verteilung der Messpunkte ab
- Bei homogener Verteilung liefern beide Methoden das gleiche Ergebnis

Übung 9: Interpolation 1

- Interpolation von Grundwasserdaten in Karlsruhe
  - Delauney Triangulation
  - Nearest Neighbour
  - Radiale Basisfunktionen
  - Visualisierung



Menberg et al. (2013)

- Aufgaben in Jupyter Notebook: 09_Räumliche Interpolation_1

Pause

… bis 11:25 Uhr

Vorlesungsplan

Uhrzeit	Inhalt
10:00 – 10:20	**Deterministische Interpolation**
10:20 – 11:00	Übung
11:00 – 11:10	Diskussion und Reflexion
11:10 – 11:25	Pause
11:25 – 11:45	Kriging
11:45 – 12:20	Übung
12:20 – 12:30	Diskussion und Reflexion

Lernziele

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

- 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

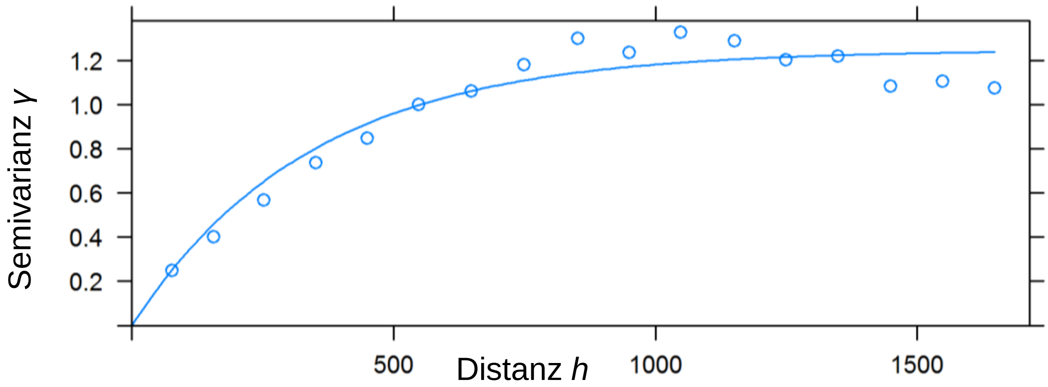
$$\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i \cdot Z(x_i), \qquad mit \sum_{i=1}^{n} \lambda_i = 1$$

- $\hat{Z}(x_0)$ = 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

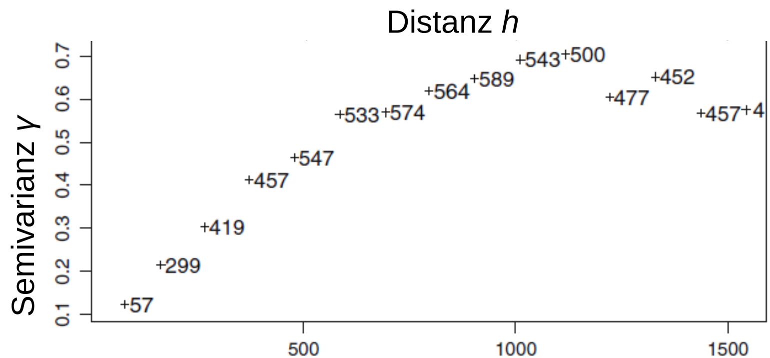
- Semivarianz $\gamma(h)$: Maß für den Grad der räumlichen Abhängigkeit von Messwerten

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

- 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

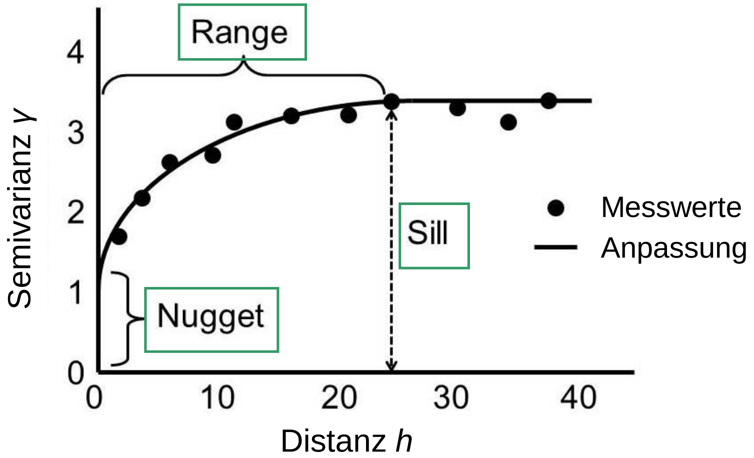
Bivand et al. (2008)

- halbe, mittlere, quadrierte euklidische Distanz zwischen zwei Messwerten



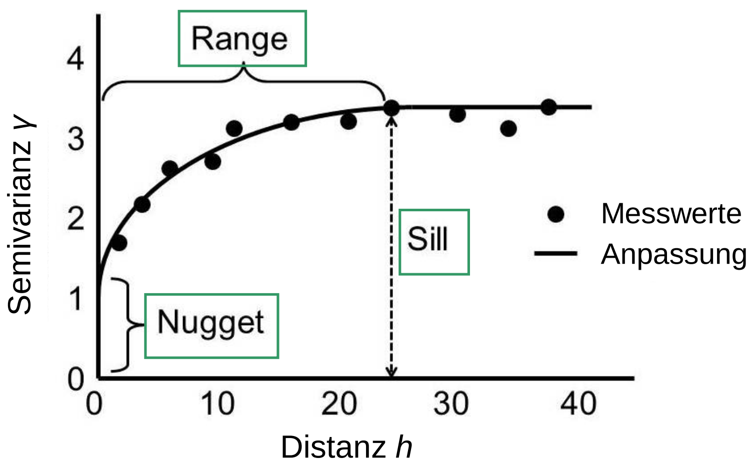
Experimentelles Variogramm

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



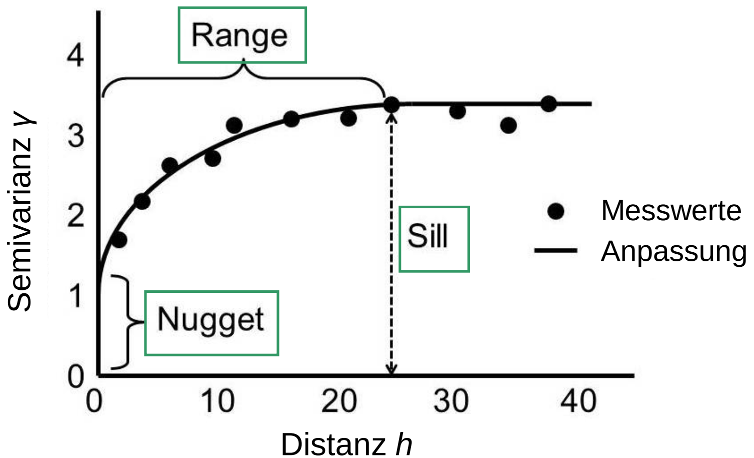
Schwellenwert und Reichweite

- 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

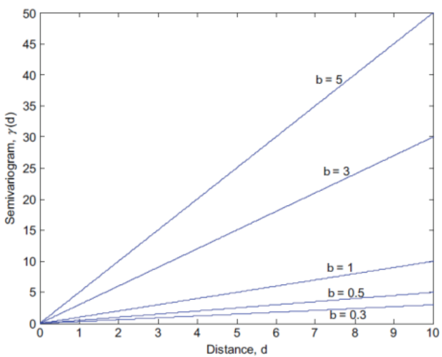
- Nugget (y-Achsen-Abstand): charakterisiert Variablen, deren Variabilität kleinräumiger als die geringsten Probenabstände ist.
- kann aber auch durch Fehler im Datensatz, z.B. durch aufgrund der Probenahme oder der Analyse, verursacht werden.



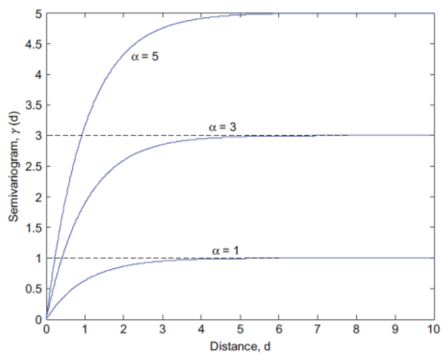
Theoretisches Variogramm

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

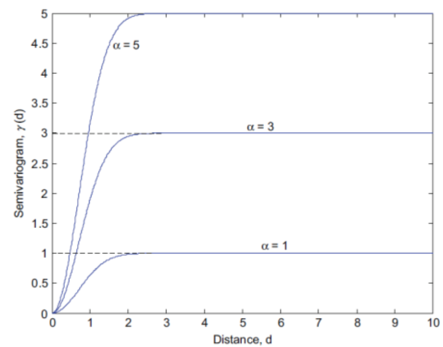
Lineares Variogramm



Exponentielles Variogramm



Gauß-Variogramm



Bivand et al. (2008)

Ordinary Kriging

Karlsruher Institut für Technologie

- 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

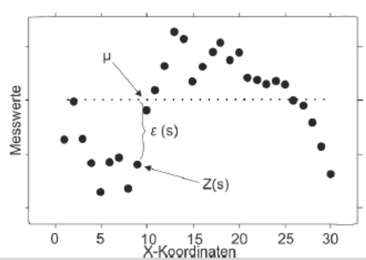
$$\hat{\sigma}^2 = \sum_{i=1}^{n} \lambda_i \cdot \gamma_{i0} + \mu \;\rightarrow \text{min.}$$

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)
$\mu(s)$: deterministischer Trend (Erwartungswert)
$\varepsilon(s)$: autokorrelierte Zufallsfehler

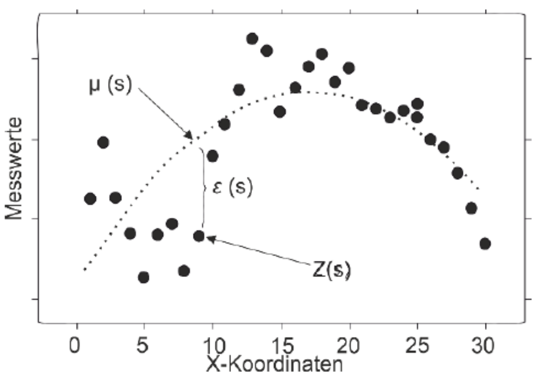


Universal Kriging

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

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

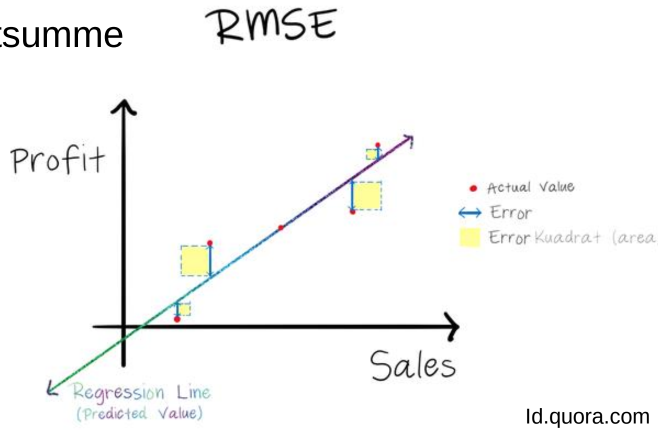
mit: $Z(s)$: Vorhersagewert (an Lokalität s)
$\mu(s)$: deterministischer Trend (Erwartungswert)
$\varepsilon(s)$: autokorrelierte Zufallsfehler



Root Mean Square Error (RMSE)

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

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



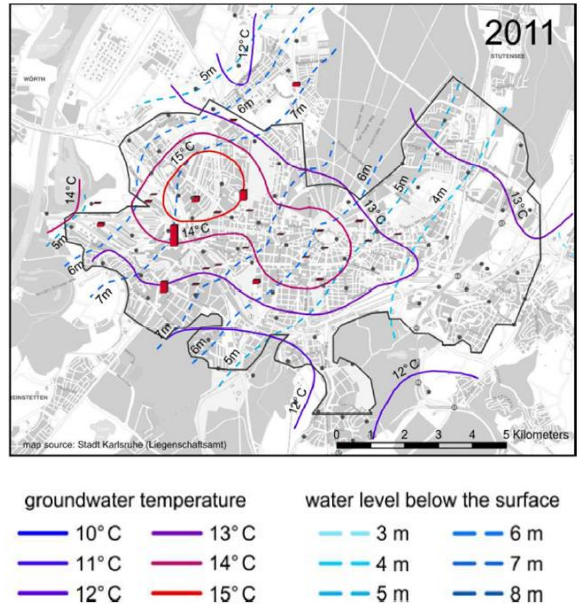
Id.quora.com

- 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 10: Interpolation 2

- Kriging mit Grundwasserdaten aus Karlsruhe
  - Ordinary Kriging der Grundwassertemperaturen
  - Universal Kriging der Grundwasserstände
  - Analyse der Vorhersagegenauigkeit

- Aufgaben in Jupyter Notebook: 10_Räumliche Interpolation_2



Menberg et al. (2013)

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): Subsurface urban heat islands in German cities, Sci. Tot. Environ. 442 (2013) 123-133.
- 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


danke
謝謝
teşekkür ederim
спасибо
thank you
gracias
dank je
tapadh leat
bedankt
dziękuję
mochchakkeram
go raibh maith agat
obrigado
sukriya
kop khun krap
arigatō
dakujem
grazie
terima kasih
merci
мерси
ありがとう
감사합니다
hvala
sagolun
takk
blagodaram