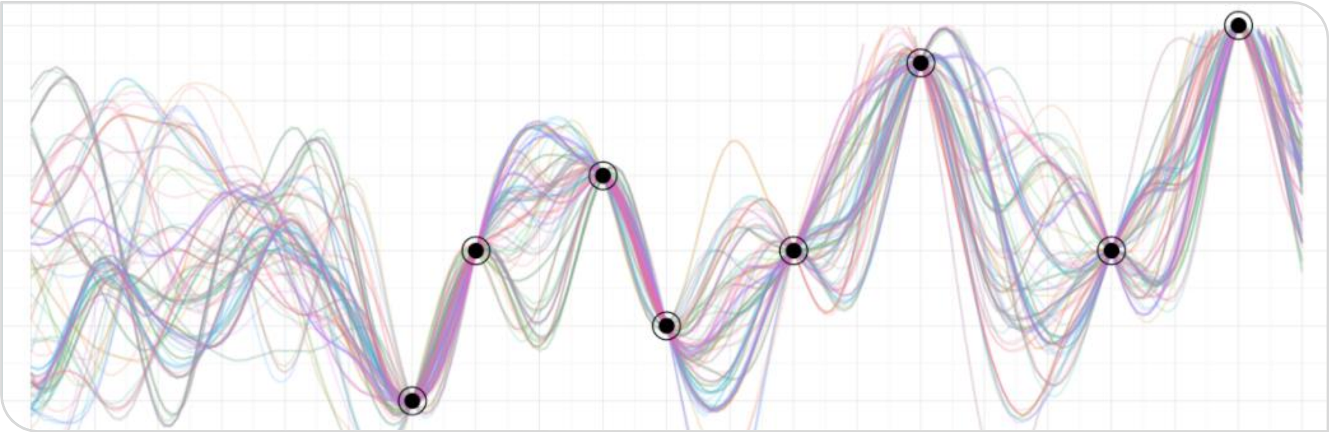
Geodatenanalyse I: Gauß-Prozesse

Kathrin Menberg



Stundenplan

	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 2.12

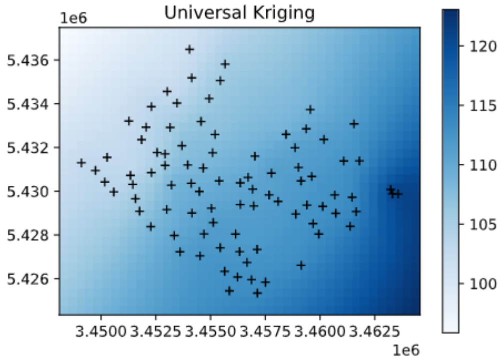
Am Ende der Stunde werden die Teilnehmer:

- … mit den mathematischen Grundlagen Gauß-Prozessen vertraut sein.
- … verschiedene Anwendungsgebiete von Gauß-Prozessen in den Geowissenschaften kennen.
- … Gauß-Prozess Regression in Python zur Interpolation von Geodaten anwenden können.

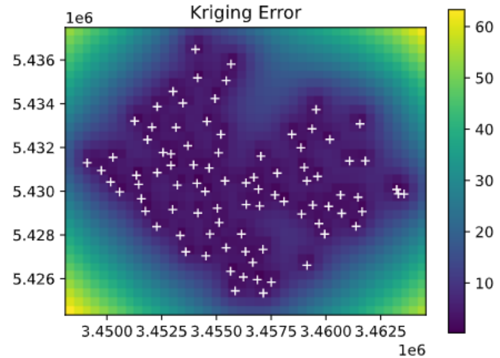
Fortführung der Kriging-Idee

- Interpolation, bzw. Vorhersage von Datenwerten

Mittelwert



Varianz

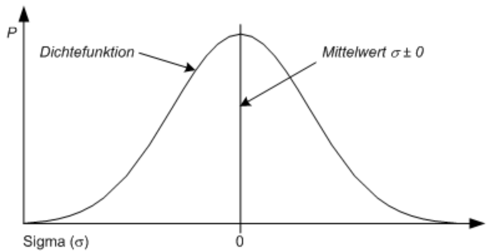


- Schließen auf Normalverteilungen von Werten
- Mittelwert & Varianz als Funktion von x- und y-Koordinaten
- Gauß-Prozesse schließen direkt auf diese Funktion

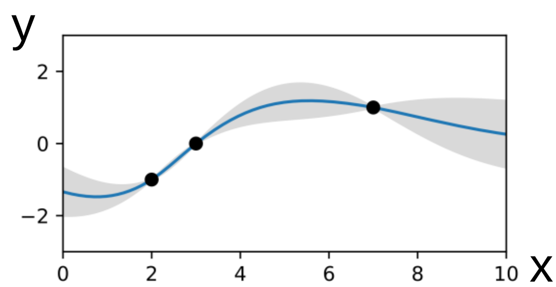
Gauß-Verteilung vs. Gauß-Prozesse

Normalverteilung

www.wikipedia.org



Gauß-Prozess



www.wikipedia.org

- Wahrscheinlichkeitsverteilung von <u>Zufallswerten</u>
- Definiert durch Mittelwert und Varianz
- 1-dimensional

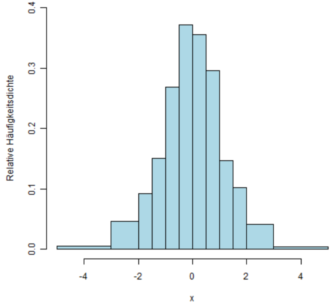
- Wahrscheinlichkeitsverteilung von <u>Zufallsfunktionen</u>
- Definiert durch Erwartungswertfunktion und Kovarianzfunktion
- multidimensional

Gauß-Verteilung vs. Gauß-Prozesse

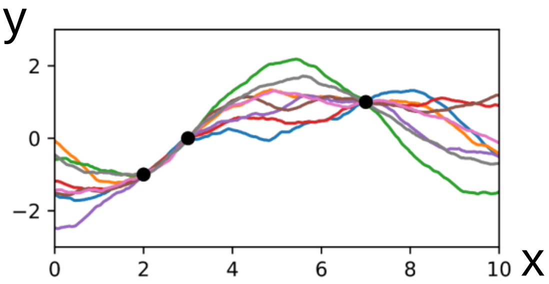
Karlsruher Institut für Technologie

www.wikipedia.org

Normalverteilung



Gauß-Prozess



www.wikipedia.org

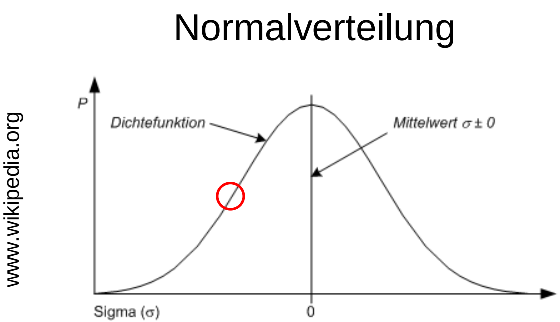
- Wahrscheinlichkeitsverteilung von <u>Zufallswerten</u>
- Definiert durch Mittelwert und Varianz
- 1-dimensional

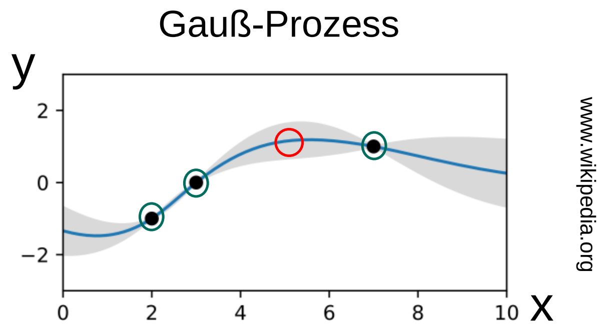
- Wahrscheinlichkeitsverteilung von <u>Zufallsfunktionen</u>
- Definiert durch Erwartungswertfunktion und Kovarianzfunktion
- multidimensional

Formulierung von Gauß-Prozessen

- Gauß-Prozess (GP): $p(x) \sim GP(m(x), k(x, x'))$
- m: Erwartungswertfunktion
  - oft als lineare Funktion: $m(x) = X\beta$
  - mit Erwartungswert $E = 0$, also $m(x) = 0$
- k: Kovarianzfunktion
  - oft als quadratische Exponentialfunktion $k(x, x') = \tau^2 \exp(-\frac{|x-x\prime|^2}{l^2})$
  - τ (magnitude) und l (lengthscale) als unbekannte Parameter
  - verschiedene Funktionen für periodische, nicht-stationäre, u.a. Datenstrukturen
  - beschreibt Korrelation zwischen benachbarten Punkten

Wahrscheinlichkeitsdichtefunktion





- für eindimensionale Normalverteilung analytisch bestimmbar
- Gauß Prozess: an Stützstellen analytisch lösbar
- An neuen, interpolierten Stellen Approximation nötig

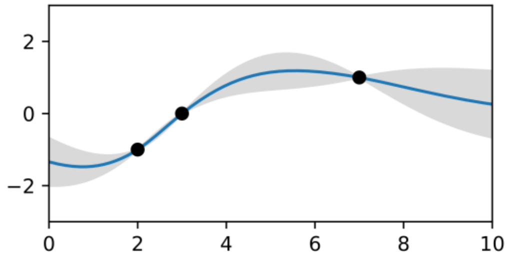
Schritte zur Interpolation mit GPs

1. Definition a-priori Erwartungswertfunktion
   - Berücksichtigung von Trend oder Drift
2. Definition a-priori Kovarianzfunktion
   - Beschreibung der Korrelation eines Punktes zu seiner Nachbarschaft
   - Unbekannte Parameter an Messwerte anpassen
3. Feinabstimmung der Parameter
   - Automatische Anpassung über die Randwahrscheinlichkeit
   - Maximierung Übereinstimmung von vermuteten Gaußprozess und vorhandene Messdaten
   - Abwägung zwischen Fehlerminimierung und Einfachheit der Theorie
   - Maximum-Likelihood-Methode

Schritte zur Interpolation mit GPs

4. Bedingter Gauß-Prozess an Stützstellen
   - Vorhersage, bzw. Interpolation an unbekannten Stellen
   - Basierend auf der Funktion an bekannten Stützstellen

5. Interpretation
   - A-posteriori Gauß-Prozess, der bekannte Information berücksichtigt
   - Gesamtheit aller möglichen Lösungen
   - Lösungsfunktionen mit unterschiedlichen Wahrscheinlichkeiten gewichtet
   - Quantifizierung der Unsicherheit

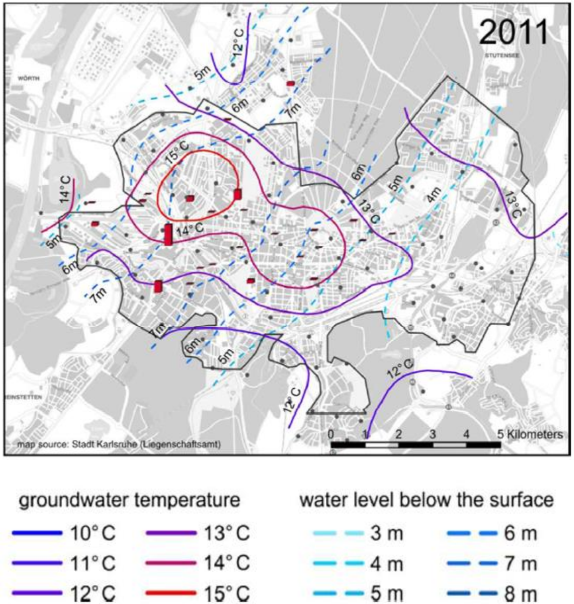


Anwendung von Gauß-Prozessen

- Interpolation, Extrapolation und Glättung von Messdaten (Gauß-Prozess Regression)
- Klassifikation in der multivariaten Statistik
- Überwachtes maschinelles Lernen zur abstrakten Modellierung (Gauß-Prozess Modelle)
- Emulatoren bzw. Surrogate-Modelle für komplexe numerische Modelle
  - Berücksichtigung von Unsicherheiten

Übung 2.12: Gauß-Prozesse

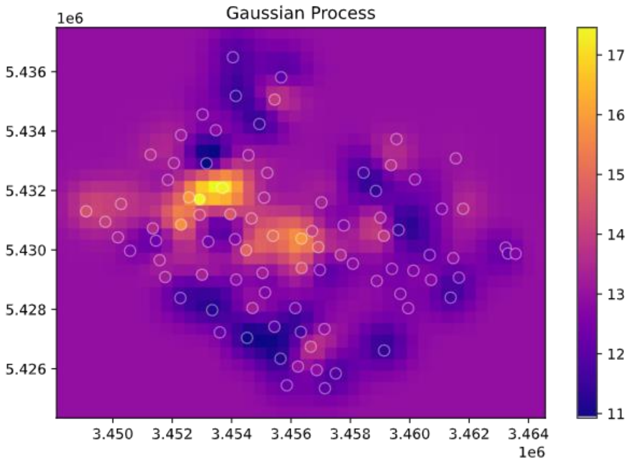
- ► Gauß-Prozess Regression
  - ► Interpolation von GW Daten
  - ► Definition der Kovarianzfunktion
  - ► Vorhersagen an neuen Stellen
  - ► Darstellung der Fehler und Unsicherheiten
- ► Aufgaben in Jupyter Notebook: geodatenanalyse_1-2-12

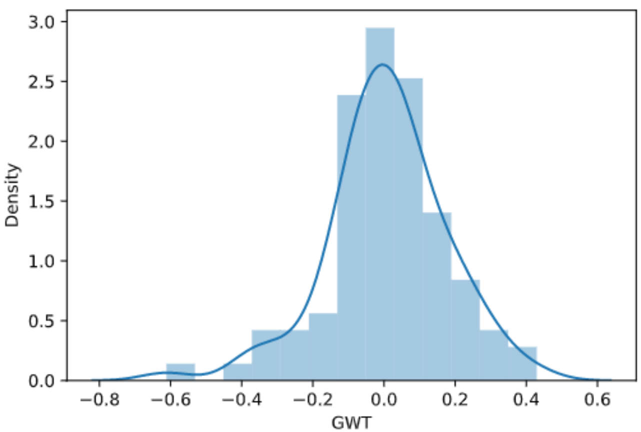


Menberg et al. (2013)

Aufgabenbesprechung

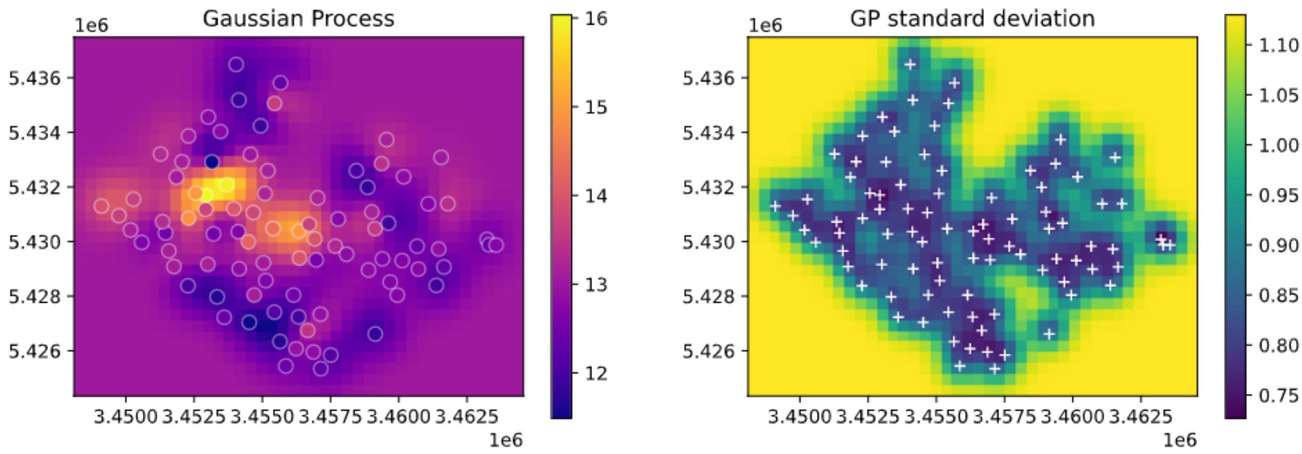
- Gauß-Prozess Regression mit einfacher Kovarianzfunktion
- RMSE = 0.17





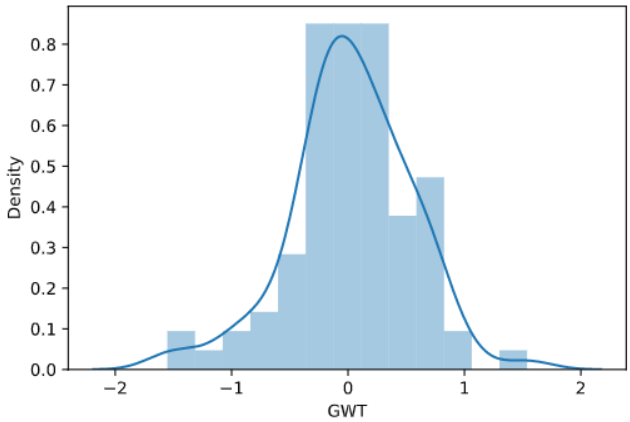
Aufgabenbesprechung

- Gauß-Prozess Regression mit Störgeräuschen



Aufgabenbesprechung

- Gauß-Prozess Regression mit Störgeräuschen
- RMSE = 0.27



Literatur

- Gelman et al. (2014): Bayesian Data Analysis, 2nd Ed., CRC Press
- C.E. Rasmussen, C.K. Williams (2006): Gaussian processes for machine learning, the MIT Press.


thank you
danke
謝謝
teşekkür ederim
gracias
спасибо
dank je
dziękuję
bedankt
obrigado
go raibh maith agat
mochchakkeram
sukriya
kop khun krap
arigatō
dakujem
grazie
terima kasih
merci
мерси
감사합니다
ありがとう
takk
hvala
sagolun
tapadh leat
Баярлалаа
рахмат
blagodaram
chnorakaloutioun
paldies
dhanyavadagalu
ευχαριστώ
хвала
asante
najis tuke
diolch
xiexie
shukriya