Zeitliche und Räumliche Autokorrelation

Generelle Beobachtung: Eng benachbarte (zeitlich, räumlich) Messungen liefern tendenziell ähnliche Werte

Autokorrelation: 1-dimensional (Zeitreihenanalyse)
Variogramm: 2(n)-dimensional (Geostatistik)

=> Dieselben Prinzipien, aus historischen Gründen allerdings unterschiedliche Begriffe

Varianz, Korrelation

Varianz: $$\mathrm{var} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^{n}[(x_i - \bar{x}) \cdot (x_i - \bar{x})]}{n}$$

Kovarianz: $$\mathrm{cov} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{n}$$

Korrelation: $$r = \frac{1}{s_x \cdot s_y} \cdot \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{n}$$

Autokorrelation: $$r_t = \frac{1}{s_x \cdot s_x} \cdot \frac{\sum_{i=1}^{n}(f(x_i) - \overline{f(x)}) \cdot ([f(x_i + t)] - \overline{f(x)})}{n}$$

Räumliche Abhängigkeit: Variogramm

→ Erweiterung des Begriffs der Autokorrelation:

- ***n***-dimensionale Zusammenhänge (meist: ***n*** = 2)
- Äquidistanz der Datenpunkte nicht erforderlich
- inverse Darstellung: Zunahme der Varianz (Unabhängigkeit) als Funktion der Entfernung

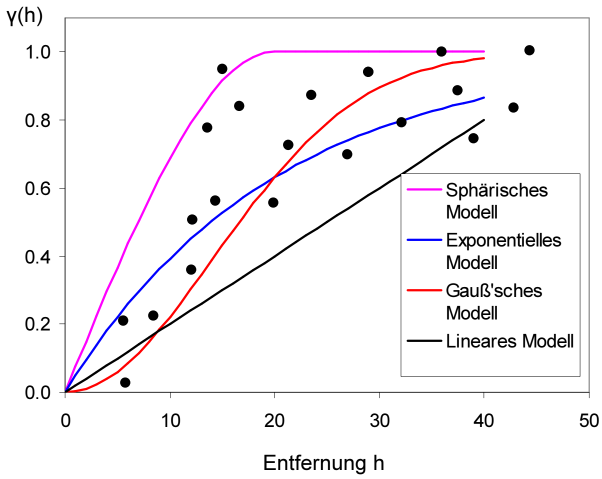
(Semi-)Variogramm-Funktion

Varianz: $\mathrm{var} = \frac{1}{n} \cdot \sum_{i=1}^{n} \left((f(x_i) - \overline{f(x)} \right)^2$

Semi-Varianz: $\gamma(h) = \frac{1}{2} \cdot \frac{1}{n} \cdot \sum_{i=1}^{n} \left[(z(x_i + h) - z(x_i) \right]^2$

Autokorrelation: $r_t = \frac{1}{s_x \cdot s_x} \cdot \frac{1}{n} \cdot \sum_{i=1}^{n} (f(x_i) - \overline{f(x)}) \cdot ([f(x_i + t)] - \overline{f(x)})$

Variogramm-Modelle



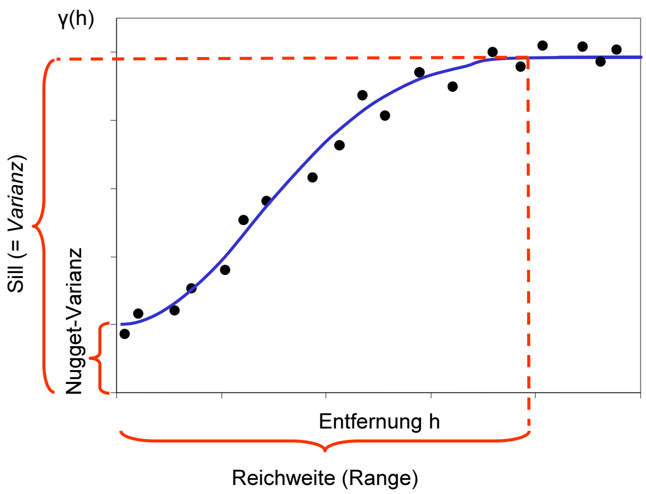
$$\gamma(h)=\begin{cases} C\cdot\left[\frac{3}{2}\cdot\frac{|h|}{a}-\frac{1}{2}\cdot\left(\frac{|h|}{a}\right)^3\right] & \text{für } |h|\le a \\ C & \text{für } |h|>a \end{cases}$$

$$\gamma(h)=C\cdot\left(1-e^{-\frac{|h|}{a}}\right)$$

$$\gamma(h)=C\cdot\left(1-e^{-\frac{|h|^2}{a^2}}\right)$$

$$\gamma(h)=\varpi\cdot|h|^p$$

Variogramm



Vorhersage-Modelle (Schätzer)

	zeitlich	räumlich
unabhängig von weiteren Variablen:		
- Modell:	Autoregression	Kriging
- Parametrisierung:	Autokorrelation	Semi-Variogramm
abhängig von weiteren Variablen:		
- Modell:	lin. Transferfunkt.	Co-Kriging
- Parametrisierung:	Kreuz-Korrelation	Kreuz-Variogramm

Kriging

Prinzip: Ausnutzung des räumlichen Zusammenhangs zur Schätzung der Werte zwischen den Stützpunkten

"**BLUE**": **B**est **L**inear **U**nbiased **E**stimator

- unbiased = erwartungsgetreu: $E\left[\hat{Z}(h) - Z(x)\right] = 0$
- linear: Wert an einem Punkt $\mathbf{x}_0$ wird geschätzt als Linearkombination (gewichtetes Mittel) von $\boldsymbol{n}$ beobachteten Werten
- exakter Schätzer: die geschätzten Werte sind an den Stützpunkten identisch mit den gemessenen
- wirkt stark glättend (*Low Pass Filter*)
- ermöglicht die Beurteilung der Zuverlässigkeit der Schätzung für jeden Schätzpunkt durch Angabe des Krigingfehlers (*Krigingvarianz*)

Kriging: Voraussetzungen

- **Normalverteilung** der Werte
  Grund: Schätzung der Kriging-Varianz (s.u.)
- **gleichmäßige Verteilung** der Stützpunkte im Raum (keine Cluster)
- **Stationarität 2. Ordnung**: Differenz von Werten der Messgrößen mit Abstand h [m] ist nur eine Funktion des Abstandsvektors, und nicht des Ortes
  $E[Z(x+h)] = E[Z(x)]$ und $\mathrm{cov}[Z(x+h), Z(x)] = \gamma(h)$ =>
  => keine Trends

(*Ordinary*) Kriging: Prinzip

- Der Wert z_0 für einen Punkt P_0 wird geschätzt als gewichteter Mittelwert (Linearkombination) der Werte der umgebenden Messstellen (Stützstellen) $P_1, P_2, \ldots, P_n$:
  $$z_0^* = \sum_{i=1}^{n} \lambda_i z(x_i)$$
- Die Gewichtung erfolgt gemäß des Semivariogramms anhand der Abstände zu den umgebenden Messstellen.
- Die Gewichte λ_i sind so zu bestimmen, dass gilt:
  - Erwartungstreue: $E[z_0^* - z_0] = 0$ und $\sum_{i=1}^{n} \lambda_i = 1$
  - Güte der Schätzung: $E[z_0^* - z_0]^2 \stackrel{!}{=} \min$

(*Ordinary*) Kriging: Vorgehensweise

- Bestimmung des quadrierten Fehlers anhand des Variogramms:

$$E\left[z_0^* - z_0\right]^2 = \mathrm{var}(z_0^* - z_0)$$
$$= 2 \cdot \sum_{i=1}^{n} \lambda_i \cdot \gamma(x_i - x_0) - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \gamma(x_i - x_j)$$

- Berücksichtigung der Nebenbedingung $\sum_{i=1}^{n} \lambda_i = 1$ mittels des Lagrange-Multiplikators μ:

$$\varphi = \mathrm{var}(z_0^* - z_0) - 2\mu\left(\sum_{i=1}^{n} \lambda_i - 1\right)$$

- Minimierung der Funktion φ führt zum Kriginggleichungssystem (KGS):

$$\begin{bmatrix} \gamma(x_1 - x_1) & \gamma(x_1 - x_2) & \cdots & \gamma(x_1 - x_n) & 1 \\ \gamma(x_2 - x_1) & \gamma(x_2 - x_2) & \cdots & \gamma(x_2 - x_n) & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma(x_n - x_1) & \gamma(x_n - x_2) & \cdots & \gamma(x_n - x_n) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma(x_1 - x_0) \\ \gamma(x_2 - x_0) \\ \vdots \\ \gamma(x_n - x_0) \\ 1 \end{bmatrix}$$

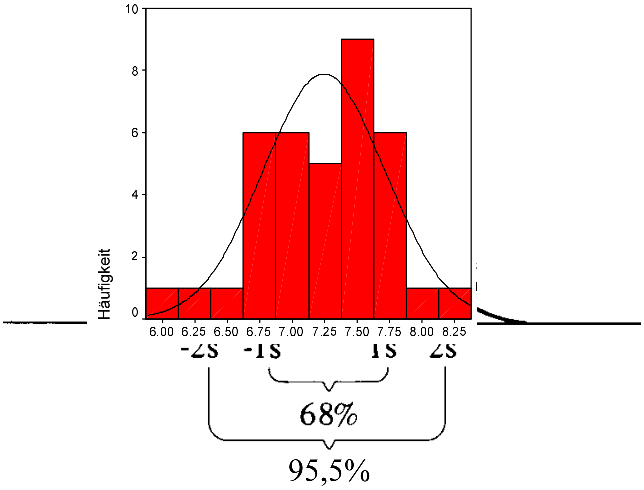
(*Ordinary*) Kriging: Schätzfehler

- Anhand der Krigingvarianz lässt sich der Schätzfehler $\boldsymbol{\sigma_k}$ bestimmen:

$$\mathrm{var}(z_0^* - z_0) = \mu + \sum_{i=1}^{n} \lambda_i \gamma(x_i - x_0) = \sigma_k^2$$

und daraus lassen sich die Konfidenzintervalle der Schätzung berechnen (unter Annahme einer Normalverteilung)

Dichtefunktion der Normalverteilung



Kriging-Verfahren: Übersicht

Punktkriging: Schätzung für einzelne Punkte

Blockkriging: Schätzung für einzelne Polygone (Blöcke)

Ordinary Kriging: gewöhnliches Kriging, berücksichtigt lokal unterschiedliche Mittelwerte

Simple Kriging: Kriging ohne Variogramm (multiple Regression), geht von einem globalen Mittelwert aus (*kriging with known mean*)
$\sum_{i=1}^{n} \lambda_i \cdot \operatorname{cov}\left(z(x_j) - z(x_i)\right) = \operatorname{cov}\left(z(x_j) - z(x_0)\right)$ für alle $j = 1, \ldots, n$

Indikator-Kriging: Schätzung eines Binärwertes (z.B. Zugehörigkeit zu best. Bodentyp, Überschreitung eines Grenzwertes, etc.)

Co-Kriging: Schätzung eines Wertes unter Nutzung der Kreuzkorrelation zu anderen Messgrößen

Kreuz-Variogramm

Variogramm: $$\gamma(h) = \frac{1}{2} \cdot \frac{1}{n} \cdot \sum_{i=1}^{n} \left[f(x_i + h) - f(x_i) \right]^2$$

$$= \frac{1}{2} \cdot \frac{1}{n} \cdot \sum_{i=1}^{n} \left[\{ (f(x_i + h) - f(x_i) \} \cdot \{ (f(x_i + h) - f(x_i) \} \right]$$

Kreuz-Variogramm: $$\gamma_{f,g}(h) = \frac{1}{2} \cdot \frac{1}{n} \cdot \sum_{i=1}^{n} \left[\{ f(x_i + h) - f(x_i) \} \cdot \{ g(x_i + h) - g(x_n) \} \right]$$

Zu grobe Messauflösung

=> Reichweite (Range) < Abstand der Stützpunkte

=> geschätzt wird der Mittelwert (Minimierung der Krigingvarianz), de facto aber ist jeder Wert möglich (Wahrscheinlichkeit: s. Verteilung der Werte)

=> Erstellung einer Karte nicht zulässig

Alternative:

- Angabe von Wahrscheinlichkeiten (s. Indikatorkriging)
- verschiedene zufällige Realisierungen

Zu grobe Messauflösung