

„Grundwassersysteme und Numerik“

Veranstaltung im Modul Hydrosystemanalyse

- Vorlesung: Modellkalibrierung und Optimierung

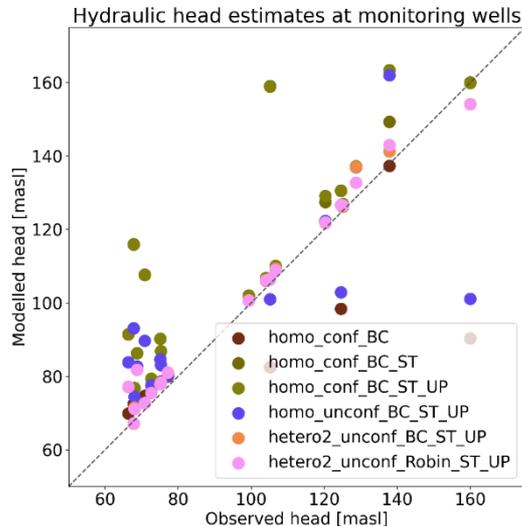
Prof. Dr. Olaf Kolditz

Dr. Erik Nixdorf

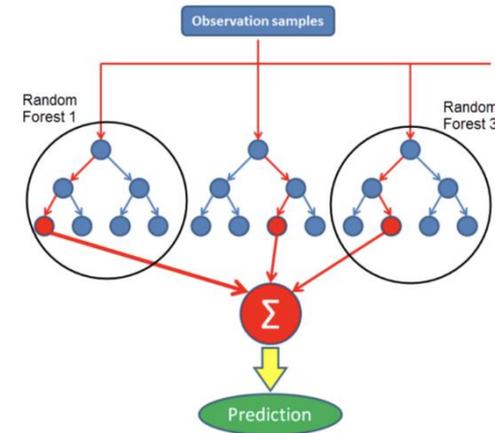
16.07.2021

Einleitung

- Simulation des Grundwasserfließverhaltens im Selke Einzugsgebiet (V7 und U9-U11)
- Kalibrierung** der Prozessparameter und **Optimierung** des Modellaufbaus führen zu einer besseren Performance der numerischen Simulation



- Datenbasierte Methoden (V12)
- Parameter und Hyperparameter **Optimierung** führen zu einer besseren Performance des generischen Algorithmus



<https://tinyurl.com/3wwvrt4b>

Einleitung

- Die heutige Veranstaltung ist eine gemische Übung/Vorlesung:
- Die Daten der zwei Übungsteile finden sie wie üblich auf der Cloud
- <https://nc.ufz.de/s/9WyZYSokq8Y2q3m> (pw: **grundwasser**)
- Für den Übungsteil „Modellkalibrierung mit PEST“ nutzen wir das erstellte Selkemodelle aus den Übungen 9-11

Einleitung

- Allgemeine Beschreibung eines Regressionsproblems:

$$\hat{y} = f(x, \vec{\theta})$$

- Optimierung der Parameter durch die Minimierung einer Verlustfunktion

$$\vec{\theta} = \operatorname{argmin}[L(x, y, \vec{\theta})]$$

- Beispiel lineare Regression:

- Zielfunktion: $\hat{y} = f(x, \vec{\theta}) = \theta_0 + \theta_1 x$

- Verlustfunktion: $L(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (\hat{y}(x^{(i)}) - y^{(i)})^2$

Mean Squared Error

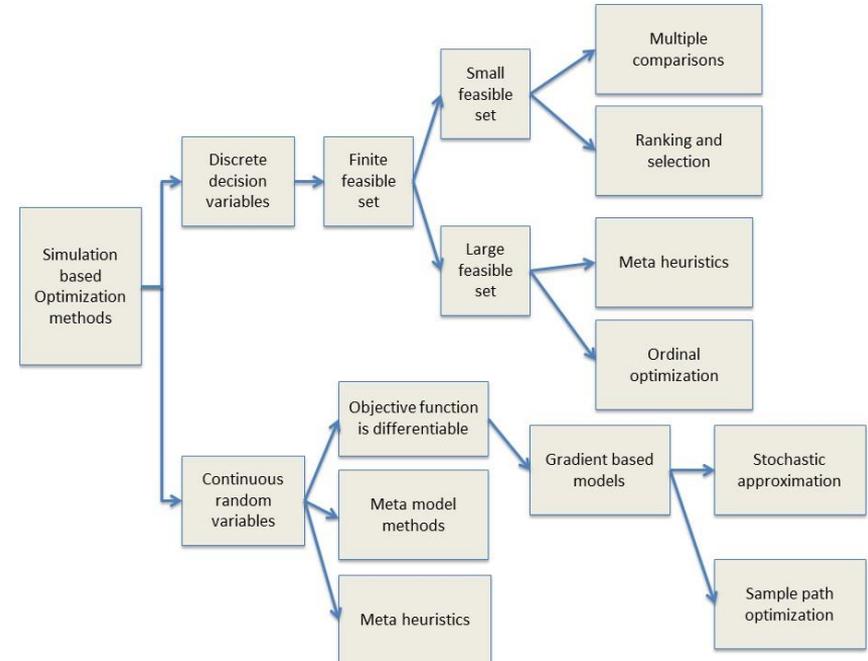
Table 2.1 Classical objective functions.

| ID | Categories | Objective functions | Formula ¹ | Characteristics | References |
|----|------------------------------------|--|---|--|---|
| 1 | Weak form-based objective function | Coefficient of Determination (R^2) | $\frac{(\sum_1^n (obs_i - \overline{obs})(sim_i - \overline{sim}))^2}{\sum_1^n (obs_i - \overline{obs})^2 \sum_1^n (sim_i - \overline{sim})^2}$ | Inappropriate for model-performance assessment; emphasize on high flows | Legates and McCabe (1999) |
| 2 | | Volume/cumulative error (CE) | $\sum_1^n (obs_i - sim_i)$ | Monotony; cannot be used alone | Guinot et al. (2011) |
| 3 | | Mean squared error (MSE) | $\frac{1}{n} \sum_1^n (obs_i - sim_i)^2$ | Most common; emphasize on high flows; neglect the low flows | McCuen et al. (2006); Krause et al. (2005) |
| 4 | | Mean absolute error (MAE) | $\frac{1}{n} \sum_1^n obs_i - sim_i $ | Balance consideration of the high- and low- flows | Legates and McCabe (1999); Krause et al. (2005) |
| 5 | Distance-based objective function | Mean quadrupled error (MS4E) | $\frac{1}{n} \sum_1^n (obs_i - sim_i)^4$ | Put greater emphasis on high flows | Baratti et al. (2003); Bennett et al. (2013) |
| 6 | | Square-root transformed MSE (RTMSE) | $\frac{1}{n} \sum_1^n (\sqrt{obs_i} - \sqrt{sim_i})^2$ | Put equal emphasis on high- and low- flows; focus on mean flows | Oudin et al. (2006); Pushpalatha et al. (2012) |
| 7 | | Log transformed MSE (LTMSE) | $\frac{1}{n} \sum_1^n (\log(obs_i) - \log(sim_i))^2$ | Put greater emphasis on low flows but still show sensitivity to high flows | Oudin et al. (2006); Pushpalatha et al. (2012) |
| 8 | | Inverse transformed MSE (ITMSE) | $\frac{1}{n} \sum_1^n \left(\frac{1}{obs_i} - \frac{1}{sim_i}\right)^2$ | Focus on low flows totally | Pushpalatha et al. (2012) |
| 9 | | Generalized efficiency (E_j) | $1 - \frac{\sum_1^n obs_i - sim_i ^j}{\sum_1^n obs_i - \overline{obs} ^j}$ | Modelers define the power (j) | Legates and McCabe (1999) |

Auswahl alternativer Verlustfunktionen (Cheng, 2014)

Einleitung Optimierung

- Optimierungsverfahren oder Algorithmen werden verwendet, um eine Verlustfunktion zu minimieren.
- Die Verfahren testen verschiedene Kombinationen von Variablen/parametern um die bestmöglichen Parametersatz für die Zielfunktion zu finden, während alle Gleichheits-, Ungleichheits- und Nebenbedingungen erfüllt werden.
- Es gibt viele Möglichkeiten zur Klassifizierung der verfügbaren Optimierungsalgorithmen (Art der Eingangsvariablen, Lokale und globale Optimierung, etc)



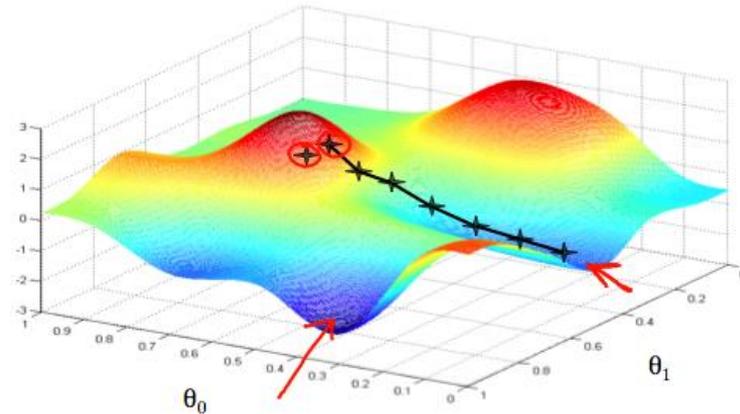
Klassifizierung der simulationsbasierten Optimierung nach Variablentypen [wiki]

Gradientenverfahren

- **Gradientenverfahren** eine Optimierungsmethode, bei der die Abstiegsrichtung durch Gradienteninformation gewonnen wird
- Der Algorithmus bewegt die Parameter **schrittweise** in die Richtung des steilsten Anstiegs
- Gradientenverfahren benötigt kontinuierliche Parameterverteilungen und eine differenzierbare Zielfunktion
- Die Größe der einzelnen Schritte wird durch den Parameter α bestimmt, der als **Lernrate** bezeichnet wird.
- Für eine Verlustfunktion ist das Gradientenverfahren:

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} L(\theta_0, \theta_1)$$

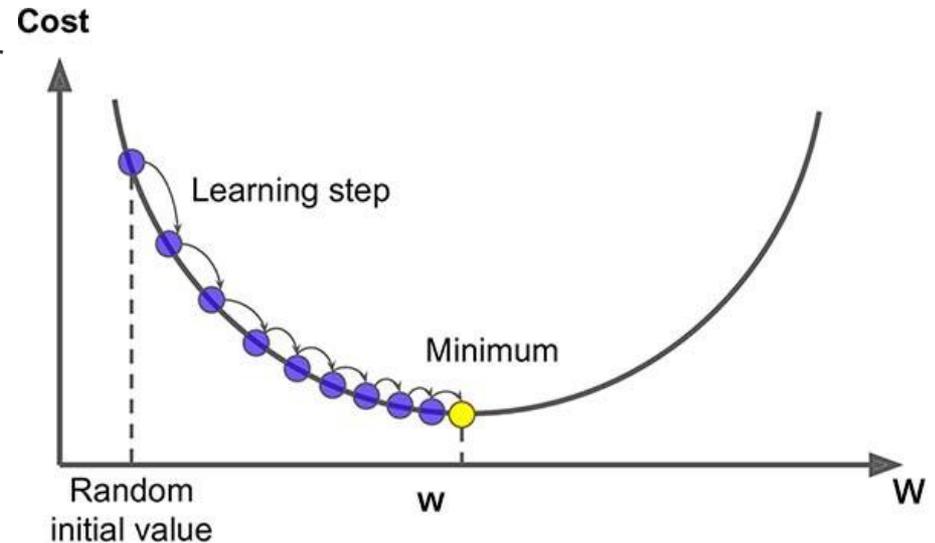
- In jedem Iterationsschritt werden alle Parameter neu berechnet



<https://tinyurl.com/c5pex8u9>

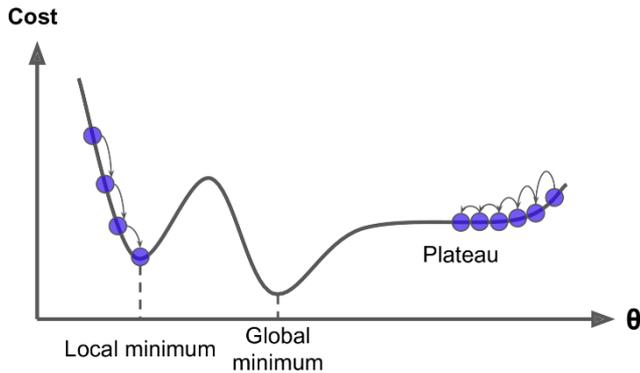
Gradientenverfahren: Anleitung

1. Leiten sie den Gradienten der Verlustfunktion in die Richtung jedes Parameters ab
2. Wählen Sie zufällige Werte für die Parameter
3. Einsetzen der Parameterwerte in die Ableitungen (den Gradienten)
4. Berechnen der Schrittweiten
5. Berechnen der neuen Parameter



Probleme Gradientenverfahren

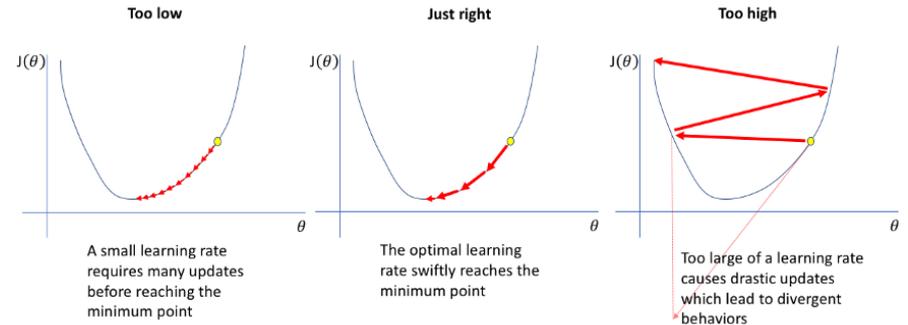
Lokale Extrempunkte Sattelpunkte



Lösungsansätze:

- Stochastisches /Mini-batch Gradientenverfahren
- Auswahl einer konvexen Verlustfunktion
- Auswahl anderer Globaler Optimierungsverfahren (https://en.wikipedia.org/wiki/Global_optimization)
- Skalierung_der_Eingangsdaten
- Anpassung der Lernrate
- Verfahren höherer Ordnung (z.B. Newton Verfahren)
- Stochastisches /Mini-batch Gradientenverfahren

Divergenz/Konvergenzprobleme

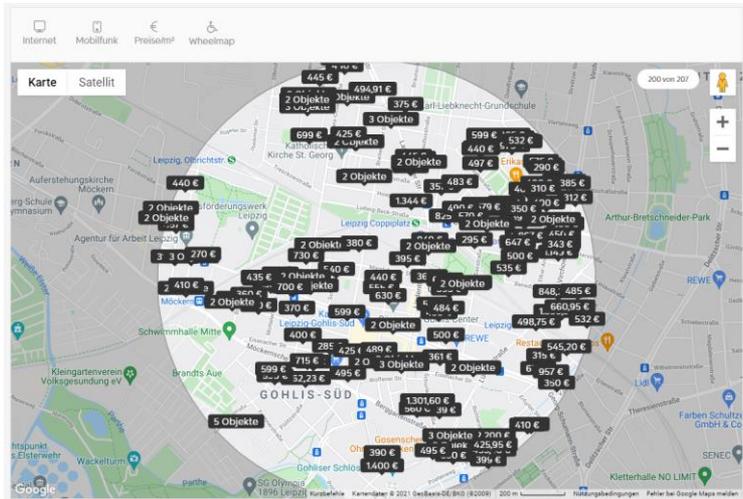


<https://tinyurl.com/23nhct7j>

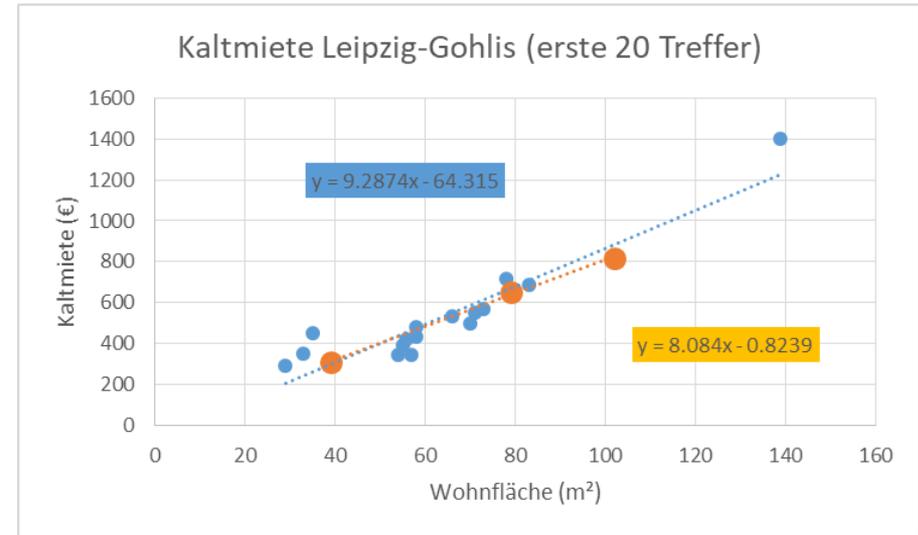
Lösungsansätze:

Gradientenverfahren: Ein Beispiel

- Aktuelle Kaltmieten (Neuvermietung) in Leipzig Gohlis (1km Radius)
- Treffer 1. Seite sortiert nach neu (13.07.2021)



Immobilienscout.com (Sortiert nach Neu)



- Zwischen Wohnfläche und Kaltmiete besteht ein Zusammenhang, der durch ein lineares Regressionsmodell hinreichend beschrieben werden kann

Gradientenverfahren Beispiel

$$\hat{y} = f(x) = \theta_0 + \theta_1 x$$

- Aufstellen der Verlustfunktion

$$L(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m \left((\theta_0 + \theta_1 x^{(i)}) - y^{(i)} \right)^2$$

- Ableiten der Gradienten

$$\frac{\partial L}{\partial \theta_0} = \frac{1}{m} \sum_{i=1}^m \left((\theta_0 + \theta_1 x^{(i)}) - y^{(i)} \right)$$

$$\frac{\partial L}{\partial \theta_1} = \frac{1}{m} \sum_{i=1}^m \left((\theta_0 + \theta_1 x^{(i)}) - y^{(i)} \right) x^{(i)}$$

Gradientenverfahren Beispiel

- Startwerte:

$$\theta_0=0, \theta_1=8.24, \alpha=0.0001$$

- Berechnung des Gradienten mit den Daten

| Step | 0 | 1 | n | 10 |
|----------------|----------|----------|------|----------|
| θ_0 | 0 | -0.00123 | ... | -0.0021 |
| θ_1 | 8.24 | 8.139472 | ... | 8.074025 |
| α | 0.0001 | 0.0001 | ... | 0.0001 |
| $dL/d\theta_0$ | 12.26667 | 4.893387 | ... | 0.093046 |
| dL/θ_1 | 1005.28 | 396.4595 | | 0.090385 |

| Wohnfläche | Kaltmiete |
|------------|-----------|
| 39 | 310 |
| 79 | 650 |
| 102 | 816 |

★★★★★ (554)

Mietspiegel Leipzig 2021

Entwicklung des Wohnungsmarktes und aktuelle Mietpreise in Leipzig

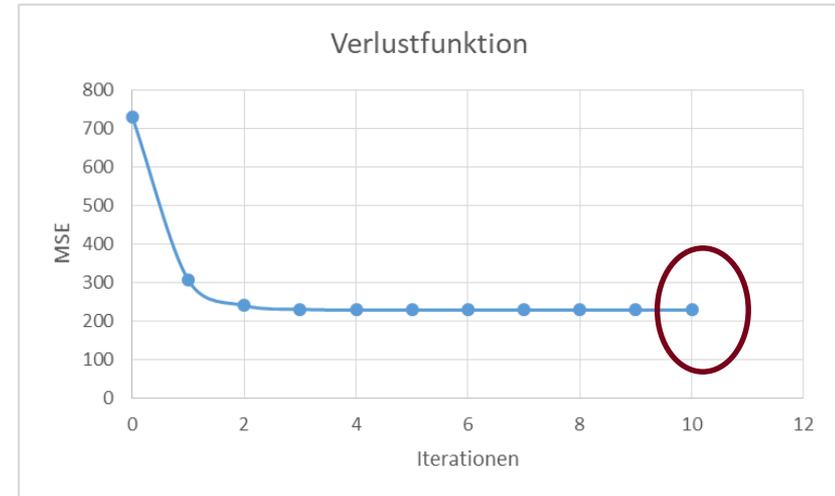
Hier finden Sie unseren **Mietspiegel von Leipzig - 7/2021**.

Diese Mietspiegel dienen zur Bestimmung der ortsüblichen Vergleichsmiete für Mietwohnungen und Häuser in Leipzig. Der Mietspiegel wird fortlaufend aktualisiert.

Wenn Sie sich für die Immobilienpreise von Leipzig interessieren: hier finden Sie aktuelle **Immobilienpreise in Leipzig** mit fortlaufend aktualisierten Immobilienpreisen und die Preisentwicklung der vergangenen Jahre.!

Der durchschnittliche Mietpreis in Leipzig liegt bei 8,24€/m².

<https://tinyurl.com/vhcxw7wc>

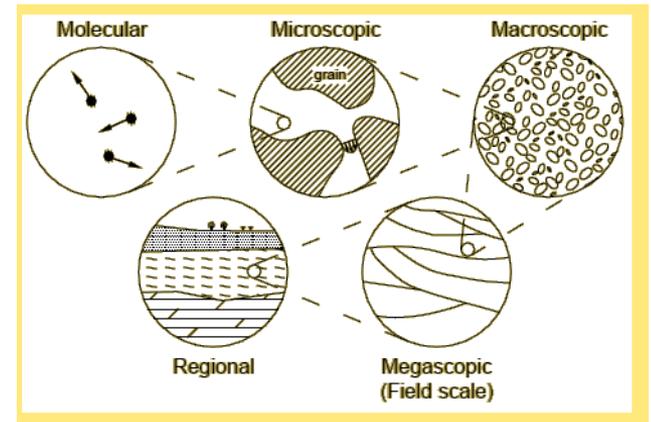
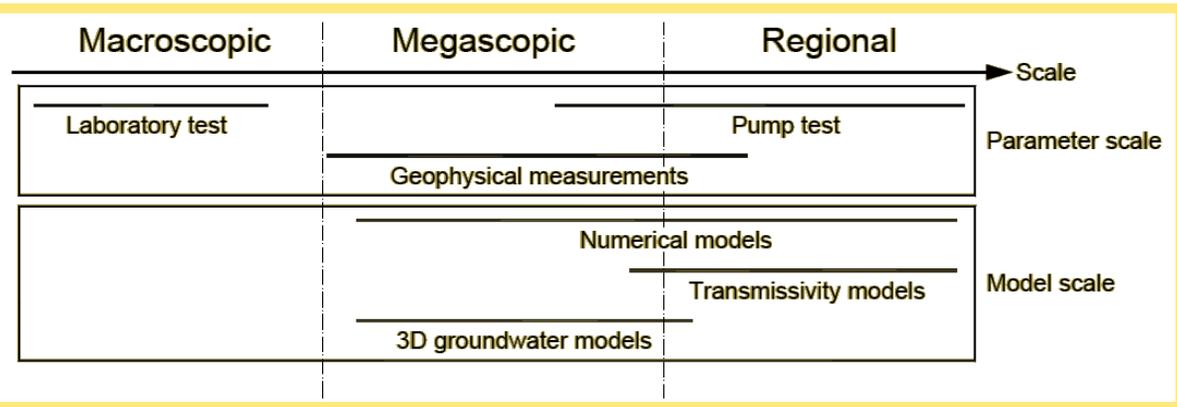


Kalibrierung in Hydrologischen und Hydrogeologischen Modellen

- Aufgrund der hydrologischen Komplexität/Heterogenität können auch physikalisch-basierte Modelle die natürlichen hydrologischen Prozesse nicht vollständig korrekt beschreiben
- Ihre Prozessparameter können nur als „**effektive Parameter**“ interpretiert werden, die das integrierte Verhalten auf der Modellelementskala darstellen.
- Kalibrierung ist der Prozess der Bewertung verschiedener alternativer Sätze von konzeptionellen Modellen, Parameter und Variablen nach dem Grad der Übereinstimmung zwischen Simulation und Beobachtungen.
- Klassische Kalibrierungsverfahren zielen darauf ab, einen einzigen eindeutigen Parametersatz zu finden, der dem globalen Optimum in Bezug auf den Anpassungsgrad entspricht

Messskala und Nutzungsskala von Parameter

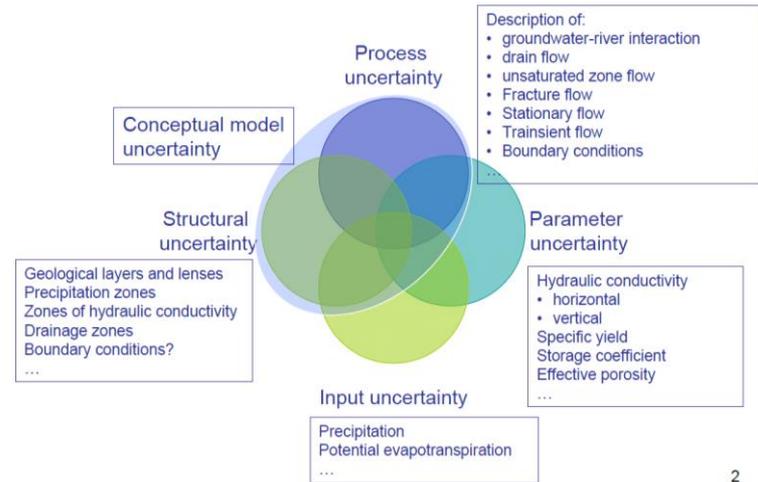
- Der Begriff "effektiver Parameter" wird oft verwendet, um anzudeuten dass die Parameter keine direkte physikalische Bedeutung haben. Die Parameterwerte sind oft abhängig vom Messvolumen (Repräsentatives Elementatvolumen REV)
- Eigentlich sollten gemessene Parameter nur in den herrschenden Gleichungen verwendet werden auf der Skala, auf der sie gefunden werden. Diese Bedingung kann aber oftmals nicht erfüllt werden
→ **Skalenunsicherheit**



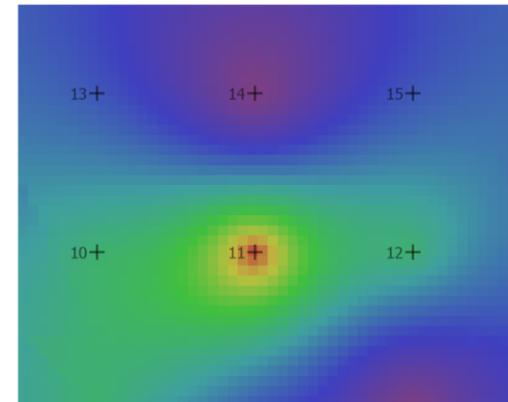
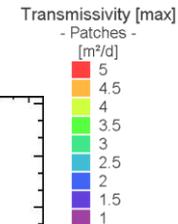
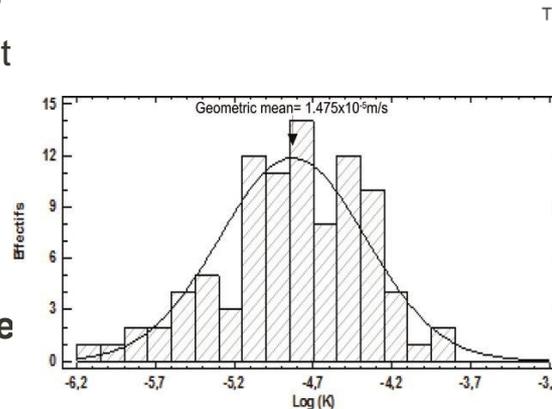
Messskala und Nutzungsskala der hydrogeologischer Parameter (Jensen, 2003)

Modellunsicherheiten

- Prozessparameterunsicherheiten sind nur ein Teil der mit der Modellierung verbundenen Unsicherheiten
- Grundsätzlich können alle Quellen von Modellunsicherheiten in die Kalibrierung einbezogen werden
- In der Praxis werden die Parameter kalibriert die die größte Unsicherheit enthalten und die auch über die Beobachtungen (invers) bestimmt werden können
- Klassische Beispiele sind die **hydraulische Leitfähigkeit und die Grundwasserneubildung**

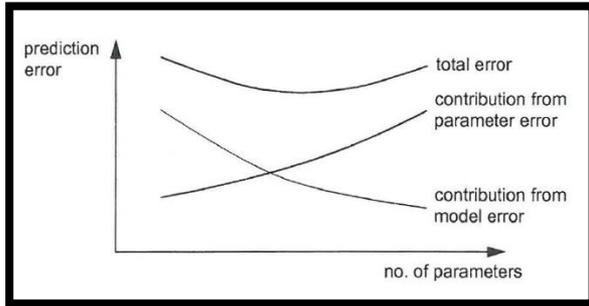
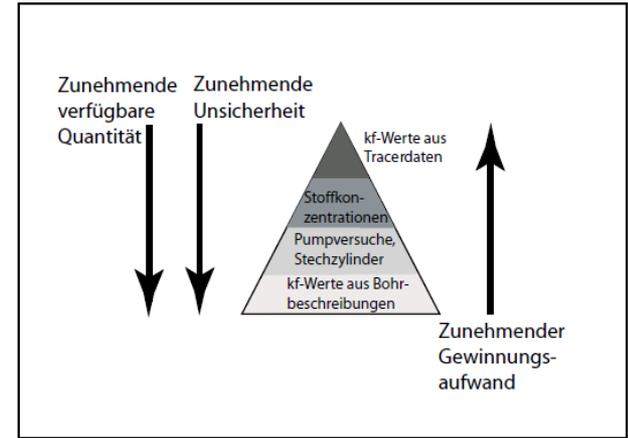


Unsicherheit in Grundwassermodellen (Jensen, 2003)

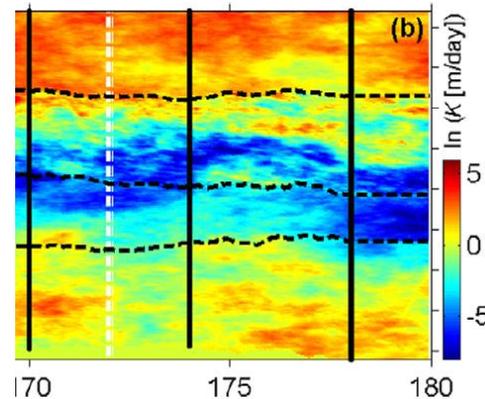


Parameterisierung: Warum messen und beschreiben wir nicht alles so genau und hochaufgelöst wie möglich

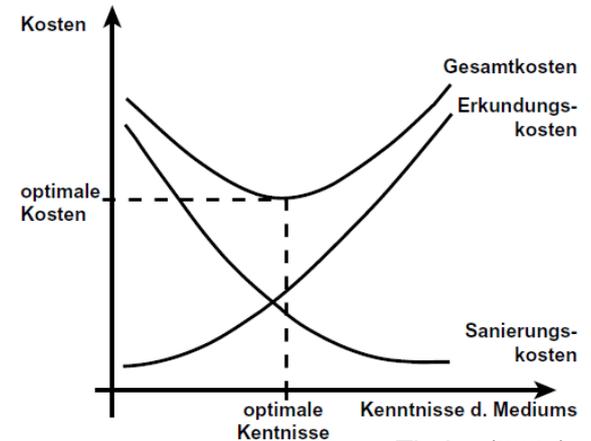
- Selbst wenn präzise Messtechniken auf der entsprechenden Skale zur Verfügung stünden, sollte nur soviel gemessen werden wie in der Kosten-Nutzen Rechnung vertretbar ist
- Des Weiteren führen zuviele Parameter (Theoretisch ein Satz je Zelle) zu einer höheren Modellunsicherheit → Beschreibung mit stochastischer oder diskreter Heterogenitätsverteilung



Jensen 2003



MEERSCHAERT et al 2013



Theiss (2005)

Unsicherheit in den Beobachtungsdaten

- Typische Beobachtungsdaten die für die Kalibrierung von Grundwassermodellen genutzt werden sind u.a :

1. **Hydraulische Potentiale** in Grundwassermessstellen
2. **Abflüsse** und **Abflussdifferenzen** in Vorflutern
3. **Grundwasserfließgeschwindigkeiten** und **Grundwasseralter**
4. **Konzentrationen**
5. ...

- Die Werte dieser Größen sind mit Unsicherheiten behaftet, z.B für das hydraulische Potential:

| | Error contribution | Standard deviation [m] | Reference |
|-------------------|---------------------------|---|-----------------|
| Observation error | Measuring equipment | 0.03 | |
| | Reading and bookkeeping | 0.05 | |
| | Reference level | 0.05 - 2.0 | |
| | Atmospheric pressure | 0.0 - 0.15 | |
| Scale error | Horizontal discretisation | $0.5 \Delta x J$ ¹⁾ | Sonnenborg 2001 |
| | Vertical discretisation | $0.5 J_v dz - 2.0$ ²⁾ | |
| | Topographical variations | σ_{topo}/d ³⁾ | Sonnenborg 2001 |
| | Heterogeneity | $\sqrt{\frac{1}{3} J^2 \sigma_{ln K}^2 \alpha_l^2}$ ⁴⁾ | Gelhar 1986 |
| Time effects | Non-stationarity | $\Delta H/2$ ⁵⁾ | Sonnenborg 2001 |
| Total error | | $\sqrt{\sum \sigma^2}$ | |

Äquifinalität von Parametersätzen

- Äquifinalität bedeutet, dass verschiedene Parameterkombinationen zu einer ähnlichen Güte der Kalibrierung führen. Dies bedeutet allerdings nicht, dass jeder dieser Parametersätze gleichwertige Ergebnisse für neue neue Anwendungen (wie Prognosen) liefert.

$$h^2(x) = \frac{h_D^2 - h_0^2}{D} x + h_0^2 + \frac{N}{K} x(D - x)$$

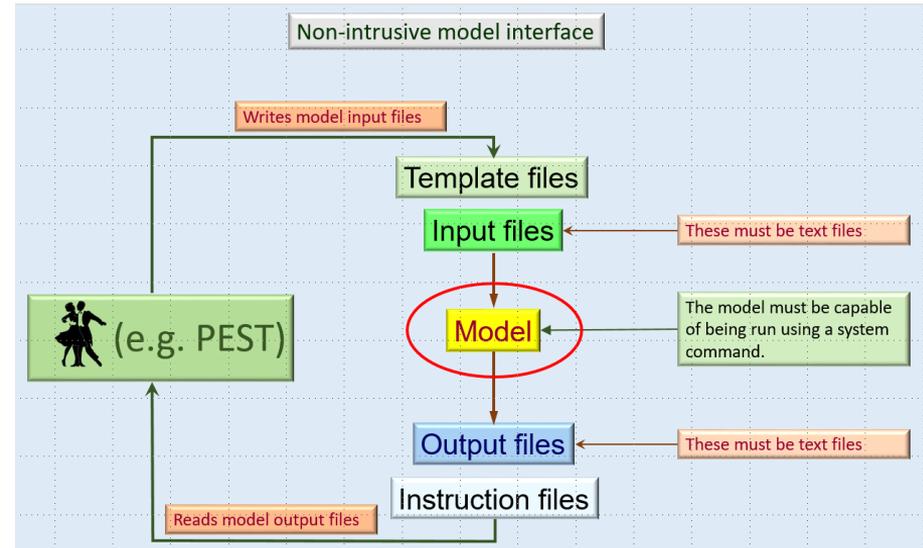
- Mögliche Lösungsansätze:

1. **Parsimonie:** Das gewählte Modell ist zu komplex. Die Parameterunsicherheit kann u.U. durch die Verringerung der Modellkomplexität reduziert werden (Wagener et al., 2002, 2003b).
2. **Effiziente Informationsnutzung:** Einerseits kann durch Multikriterielle Optimierung die Informationsmenge erhöht und der Äquifinalität entgegengewirkt werden (**Pareto Optimierung**). Andererseits stehen Algorithmen (PIMLI, BARE, DYNIA) zur Verfügung die durch Unterteilung und gradueller Assimilierung der Beobachtungsdaten besser unterteilen.
3. **Akzeptanz der Äquifinalität:** Äquifinalität ist eine inhärente Eigenschaft hydro(geo)logischer Modelle. Die logische Konsequenz besteht darin alle potentiellen Lösungen des Kalibrierungsproblems beizubehalten und in die Modellierung einzubeziehen , z.B. **Generalised Likelihood Uncertainty Estimation (GLUE)** [Beven & Binley 1992]

- Nach der Kalibrierung ist es erforderlich das Modell an Daten zu validieren, die bisher nicht in den Kalibrierungsprozess eingeflossen sind.
- Die Modellvalidierung ist ein wichtiges Werkzeug der Qualitätssicherung. Sie weist die Eignung eines Modells für einen intendierten Zweck nach.
- Durch die Validierung wird u.a. die **Generalisierungsfähigkeit** des Modells überprüft. Sie bezeichnet die Fähigkeit des Modells seine Funktion auch bei Änderungen des Systems und der Modellanwendung aufrecht zu erhalten
- Eine klassischer Ansatz zur Generierung von Beobachtungsdaten für die Validierung wäre die Aufteilung der verfügbaren Zeitreihen in Grundwassermessstellen in einen Kalibrierungs und einen Validierungszeitraum

Kalibrierung mit Pest

- PEST ist eines der am häufigsten verwendeten Softwarepakete für die Kalibrierung von Grundwassermodellen.
- PEST ist modellunabhängig, parallelisierbar und kann mit jeder Modellierungssoftware verwendet werden, die Parameter akzeptiert und Beobachtungen als numerische Werte liefert, wenn die entsprechenden Modelldateien zugänglich sind und ihr Format dokumentiert ist
- Die Beschaffenheit von PEST - ein hauptsächlich kommandozeilengesteuertes Tool (wie OGS!) - stellte für die Benutzer oft ein weiteres Hindernis dar, mit der Software zu beginnen → Es existieren GUIs für z.B. FEFLOW und MODFLOW



PEST Integration <https://tinyurl.com/7fcark4y>

Kalibrierung mit Pest: Dateien

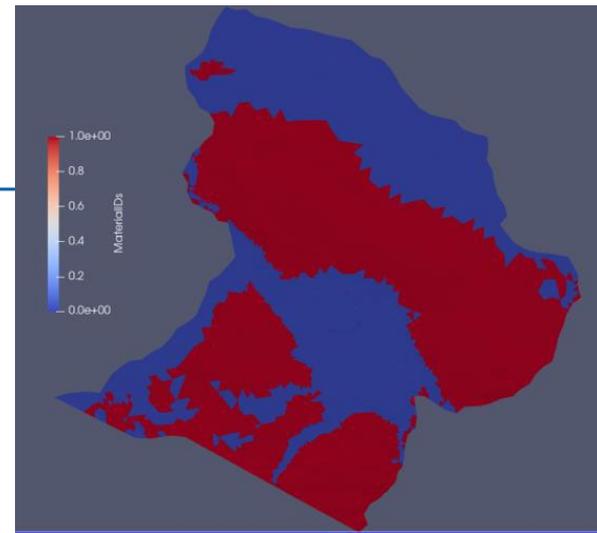
1. **Template file:** Sagt PEST, wie die Prozessparameter in die Modelleingabe geschrieben werden
2. **Instruction file:** Sagt PEST wo die Modellergebnisse ausgelesen werden können
3. **PEST control file:** Beinhaltet die Initialbedingungen, die oberen und unteren Parametergrenzen, die Beobachtungen, maximale Zahl der Iterationen, etc

```
pcf
* control data
RSTFLE PESTMODE
NPAR NOBS NPARGP NPRIOR NOBSGP
NTPLFLE NINSFLE PRECIS DPOINT NUMCOM JACFILE MESSFILE
RLAMBDA1 RLAMFAC PHIRATSUF PHIREDLAM NUMLAM
RELPARMAX FACPARMAX FACORIG
PHIREDSWH
NOPTMAX PHIRE DSTP NPHISTP NPHINORED RELPARSTP NRELPAR
ICOV ICOR IEIG
* parameter groups
PARGPNME INCTYP DERINC DERINCLB FORCEN DERINCMUL DERMTHD
(one such line for each of the NPARGP parameter groups)
* parameter data
PARNME PARTRANS PARCHGLIM PARVAL1 PARLBNB PARUBND PARGP SCALE OFFSET DERCOM
(one such line for each of the NPAR parameters)
PARNME PARTIED
(one such line for each tied parameter)
* observation groups
OBSNME
(one such line for each observation group)
* observation data
OBSNME OBSVAL WEIGHT OBGNME
(one such line for each of the NOBS observations)
* model command line
write the command which PEST must use to run the model
* model input/output
TEMPFLE INFLE
(one such line for each model input file containing parameters)
INSFLE OUTFLE
(one such line for each model output file containing observations)
* prior information
PILBL PIFAC * PARNME + PIFAC * log(PARNME) ... = PIVAL WEIGHT OBGNME
(one such line for each of the NPRIOR articles of prior information)
```

Example 4.1 Construction details of the PEST control file.

Übung: Kalibrierung mit PEST

- Wir nehmen das in Übung 10 erstellte Grundwassermodell und kalibrieren die hydraulischen Leitfähigkeiten
- 3 Kalibrierparameter : kf_1 , kf_2 und kf_3 , wobei kf_3 die hydraulische Leitfähigkeit im Flussbett beschreibt
- Arbeitsschritte (Daten auf der Cloud):
 1. Kopieren Sie die für die OGS Simulation notwendigen Dateien (+OGS) in einen neuen Ordner *pest_calibration*
 2. Aufbau des **Template, Instruction und Control Files** von PEST
 3. Kalibrierung und Parameteranpassung



| Parameter | Initialwert | Minimum | Maximum |
|--------------------------|-------------|---------|---------|
| Kf_1 [m ²] | 5.0E-9 | 1E-11 | 1E-8 |
| Kf_2 [m ²] | 9.5E-11 | 1E-13 | 1E-10 |
| Kf_3 [m ²] | 5.0E-9 | 1E-11 | 1E-8 |