

Todo list

■ BibLaTeX: Punkt nach Auflage einfügen: 2. Auflage	18
■ BibLaTeX: irgendein Strichpunkt stimmt nicht	18
■ BibLaTeX: bei Zitat mit Seitenzahl das S.tilde dazu	18

Inhaltsverzeichnis

1	Einleitung	5
2	Zielsetzung und Aufgabenstellung	6
3	Allgemeine Grundlagen	7
3.1	Stofftransport	7
3.2	Untersuchte Parameter	7
4	Material und Methoden	8
4.1	Datenaufbereitung (Preprocessing)	8
4.1.1	Standardisierung	8
4.1.2	Box-Cox Transformation	8
4.2	Verfahren zur numerischen Integration von Stromlinien	10
4.2.1	Integration nach Euler	11
4.2.2	Runge-Kutta Verfahren	12
4.3	Clustering	13
4.3.1	Partitionierendes Clustering	13
4.3.2	Hierarchisches Clustering	13
4.3.3	Unscharfes Clustering	13
4.4	Hauptkomponentenanalyse	13
4.5	Co-Kriging	13
4.6	Zeitreihenanalyse	13
5	Ergebnisse	14
6	Interpretation und Diskussion der Ergebnisse	15
7	Zusammenfassung	16
8	Ausblick	17

Abbildungsverzeichnis

4.1	Veränderung der Dichtefunktion durch eine Box-Cox Transformation	9
4.2	QQ-Plot. Annäherung an die Normalverteilung durch Box-Cox Transformation .	9
4.3	Vergleich numerischer Integrationsverfahren für Anfangswertprobleme	11

Tabellenverzeichnis

1 Einleitung

2 Zielsetzung und Aufgabenstellung

3 Allgemeine Grundlagen

3.1 Stofftransport

3.2 Untersuchte Parameter

4 Material und Methoden

Die Berechnungen in der vorliegenden Arbeit wurden zur Gänze mit der freien, statistischen Programmiersprache R (R Core Team, 2012) durchgeführt.

4.1 Datenaufbereitung (Preprocessing)

Vor allem multivariate Verfahren wie die Faktoren- oder die Hauptkomponentenanalyse setzen von den Zufallsvariablen voraus, dass der Erwartungswert Null und die Varianz Eins beträgt. Betrachtet man z.B. die Konzentration von gelösten Grundwasserinhaltsstoffen, so nimmt diese für die Hauptkomponenten (Na^+ , Mg_2^+ , HCO_3^- , O_2 , ...) häufig Werte $> 10 \text{ mg l}^{-1}$ an (Hütter, 1994)), währenddessen Spurenstoffe wie Cu_2^+ , Zn_2^+ oder Pb_2^+ in der Regel zwei Größenordnungen darunter liegen ($< 0.1 \text{ mg l}^{-1}$). Um trotz dieser Größenunterschiede die unterschiedlichen Variablen vergleichen zu können, werde diese — wie in Unterkapitel 4.1.1 beschrieben — standardisiert.

Verfahren wie beispielsweise die lineare Regression setzen eine Normalverteilung der Residuen voraus. Ist diese Annahme verletzt, so hat das zwar keine Auswirkungen auf die geschätzten Regressionskoeffizienten, jedoch auf Konfidenzintervalle und die Ergebnisse von Signifikanztests (Albers, 2009). Mithilfe einer geeigneten Transformation (z.B. Box-Cox Transformation) kann eine Approximation der Daten an eine Normalverteilung erzielt werden. Vergleiche Unterkapitel 4.1.2.

4.1.1 Standardisierung

Beim Standardisieren (auch z-Transformation genannt) wird von der Zufallsvariable X ihr Erwartungswert $E(X)$ abgezogen und diese Differenz durch die Standardabweichung $\sqrt{\text{Var}(X)}$ dividiert. Die Transformation erfolgt durch

$$Z = \frac{X - E(X)}{\sqrt{\text{Var}(X)}}, \quad (4.1)$$

sodass nun $E(X) = 0$ und $\text{Var}(X) = \sqrt{\text{Var}(X)} = 1$ sind. Eine Rücktransformation kann mit

$$X = (Z + E(X)) \cdot \sqrt{\text{Var}(X)}$$

durchgeführt werden.

Da in den meisten Fällen der Erwartungswert als auch die Varianz einer Zufallsvariable unbekannt sind, wird nicht eine Transformation der Zufallsvariablen selbst durchgeführt, sondern nur deren Realisationen in der vorliegenden Stichprobe. Erwartungswert und Varianz können stattdessen durch Stichprobenmittelwert und Stichprobenvarianz ersetzt werden.

4.1.2 Box-Cox Transformation

Die Box-Cox Transformation ist eine Potenztransformation mit reellen Exponenten. Sie ermöglicht bei geeigneter Wahl des Exponenten λ eine Annäherung der transformierten Variable $y^{(\lambda)}$ an die Normalverteilung und ist wie folgt definiert:

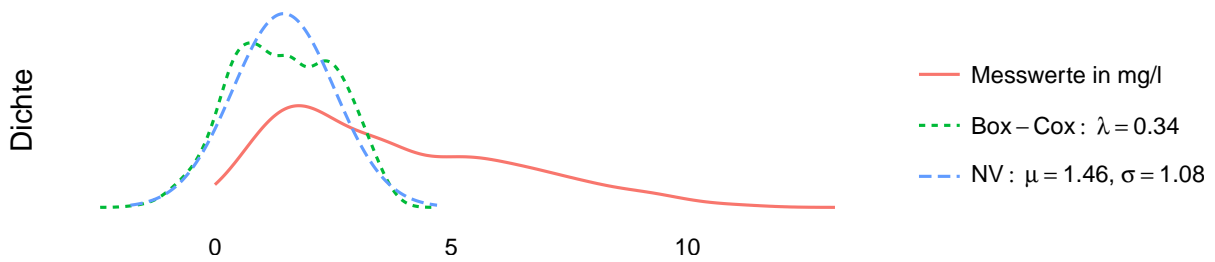


Abbildung 4.1: Führt man für die im Zeitraum 2005 bis 2012 an den Grundwassermonitoringstellen beobachteten O_2 -Konzentration (rot) eine Box-Cox Transformation durch, so nähert sich die Verteilung der transformierten Werte (grün) einer Normalverteilung (blau) an.

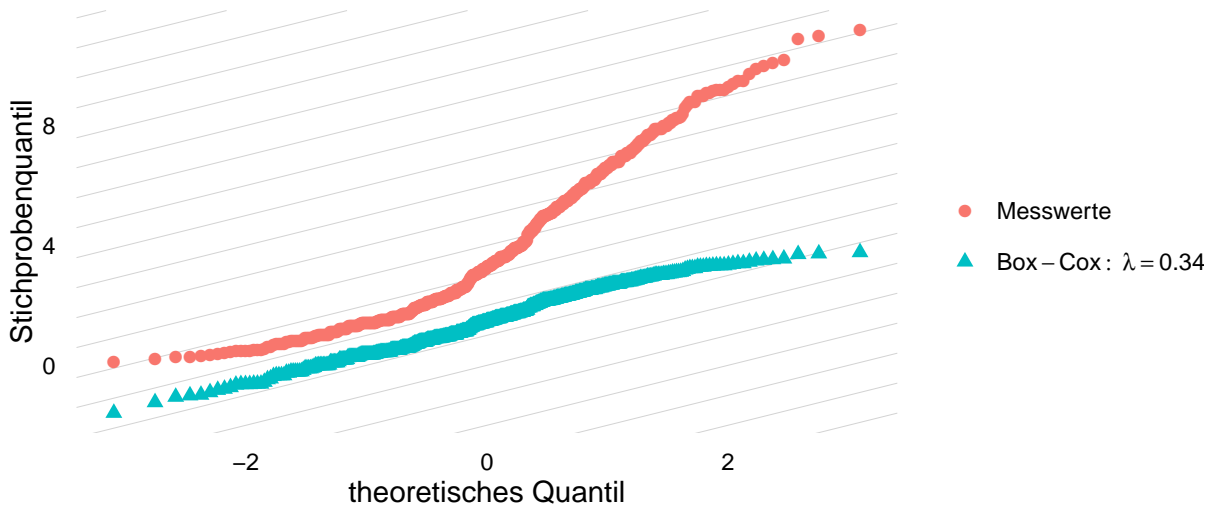


Abbildung 4.2: Mithilfe eines QQ-Diagramms kann die Ähnlichkeit einer Verteilung zur Standardnormalverteilung (45°-Gerade, bei gleicher Achsenskalierung) überprüft werden.

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \text{für } \lambda \neq 0, \\ \log(y), & \text{für } \lambda = 0. \end{cases}$$

Um für alle $\lambda < 0$ eine Änderungen des Vorzeichens von $y^{(\lambda)}$ und die damit einhergehende Änderung der Elementreihenfolge zu verhindern, wird der Ausdruck $y^\lambda - 1$ nochmals durch λ dividiert.

Da eine Funktion der Form $f = y^{\lambda-1}/\lambda$ an der Stelle $\lambda = 0$ aufgrund der Division durch Null eine Singularität (Definitionslücke) aufweist, ist die Box-Cox Transformation abschnittsweise, mit einer Log-Transformation an dieser Stelle, definiert.

Zur Schätzung von λ wird in dieser Arbeit auf die Funktion `powerTransform` aus dem R-Paket `car` (Fox und Weisberg, 2011) zurückgegriffen, welches eine Maximum-Likelihood-Schätzung des Koeffizienten durchführt.

Mithilfe statistischer Tests kann getestet werden, ob ein geschätzter Exponent λ signifikant von bestimmten, leicht interpretierbaren Werten unterschieden werden kann. Häufig getestete

Exponenten sind $\lambda = 0$ (Log-Transformation), $\lambda = 1/3$ (Transformation mit Wurzelexponenten 3, Kubikwurzel), $\lambda = 1/2$ (Transformation durch Ziehen der Quadratwurzel) und $\lambda = 1$ (keine Transformation). Aus Gründen der Anschaulichkeit werden selbst Exponenten gerundet, welche sich signifikant von den zuvor Genannten unterscheiden.

Als Tests eignen sich einerseits der Wald-Test welcher die Nullhypothese $H_0: \lambda = \lambda_0$ gegen die Alternativhypothese $H_1: \lambda \neq \lambda_0$ testet. Oftmals wird kontrolliert, ob λ_0 innerhalb des Konfidenzintervalls $[\lambda_u, \lambda_o]$ mit

$$\lambda_{u,o} = \hat{\lambda} \pm 1.96 \cdot \sigma(\hat{\lambda}),$$

liegt. Zum anderen können ebenfalls Likelihood-Ratio-Tests angewandt werden, welche gegenüber dem Wald-Test ein besseres Verhalten bei kleinem Stichprobenumfang zeigen.

Für eine eventuell erforderliche Rücktransformation kann folgende Umkehrfunktion hergeleitet werden:

$$y = \begin{cases} (y^{(\lambda)} \lambda + 1)^{1/\lambda}, & \text{für } \lambda \neq 0, \\ e^{y^{(\lambda)}}, & \text{für } \lambda = 0. \end{cases}$$

4.2 Verfahren zur numerischen Integration von Stromlinien

Folgt man in einem (zeitlich invarianten) Geschwindigkeitsfeld stets den Richtungsvektoren der Wasserbewegung, so entspricht die zurückgelegte Bahn einer Stromlinie. Hat man nun Kenntnis über das dreidimensionale Geschwindigkeitsfeld eines Grundwasserkörpers das jedem beliebigen Punkt $\vec{p} = (p_x, p_y, p_z)$ innerhalb des stationären Feldes einen Geschwindigkeitsvektor $\vec{v} = (v_x, v_y, v_z)$ zuordnet, kann für jeden dieser Punkte \vec{p} genau *eine einzige* Stromlinie bestimmt werden die durch \vec{p} geht. Mathematisch betrachtet muss hierfür eine Differentialgleichung erster Ordnung integriert werden, deren Lösung lediglich vom Anfangspunkt \vec{p} der Stromlinie abhängt, weshalb man auch von einem Anfangswertproblem spricht.

Wird Wasser aus einem Grundwasserkörper gefördert, welcher seitlich durch Fließgewässer begrenzt ist und sonst keine weiteren Quellen und Senken aufweist, so werden die meisten Stromlinien am Gewässerrand beginnen und im Brunnen enden. Ziel ist es nun diese Stromlinien aus einem gegebenen Geschwindigkeitsfeld und einem gewählten Punkt zu berechnen und somit eine Abschätzung der Fließzeit zum Brunnen zu erhalten.

Meist liegt das Vektorfeld der Grundwasserfließgeschwindigkeiten nicht als stetige Funktion vor sondern ist nur an bestimmten, diskreten Punkten auswertbar, beispielsweise an den Knoten eines Finite-Elemente-Modells. Das hat zur Folge, dass auch das Integral der Stromlinie nicht geschlossen lösbar ist und man auf Näherungsverfahren der numerischen Integration zurückgreifen muss Papula (2011).

Das einfachste Verfahren hierfür ist das Eulersche Polygonzugverfahren das verglichen mit z.B. Runge-Kutta Verfahren ein deutlich schlechteres Konsistenzverhalten zeigt. Zur Veranschaulichung soll die einfache Differentialgleichung erster Ordnung

$$y' = \sin(x)^2 \cdot y$$

mit der analytischen Lösung

$$y = 2 \cdot \exp\left(\frac{x - \sin(x)\cos(x)}{2}\right)$$

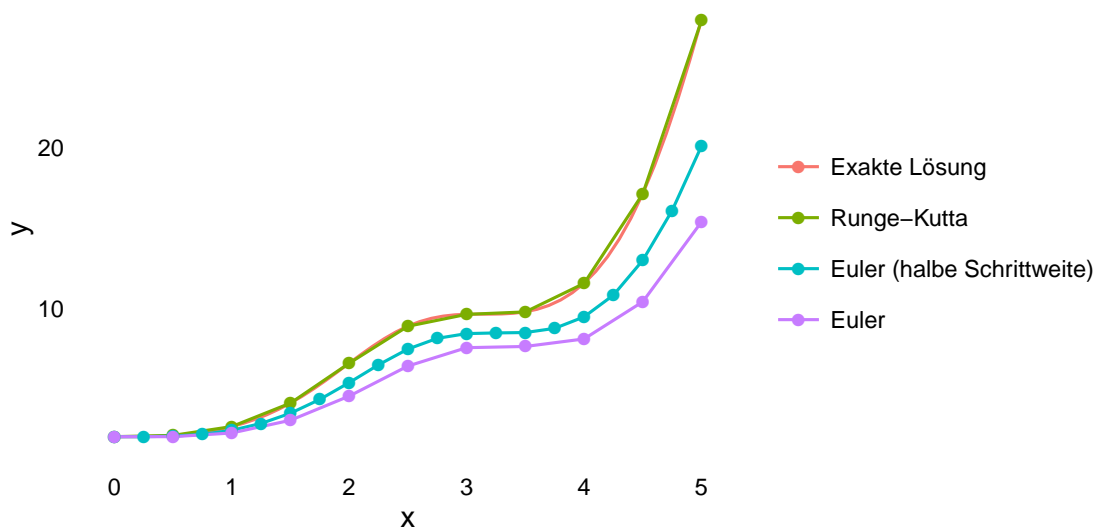


Abbildung 4.3: Vergleich numerischer Integrationsverfahren für Anfangswertprobleme. Man beachte die großen, lokalen Fehler des Euler-Verfahrens selbst bei Halbierung der Schrittweite.

in dem Intervall $x \in [0, 5]$ mit der festen Schrittweite $h = 0,5$ mit beiden Methoden gelöst werden. Als Anfangswert wird $x_0 = 0$ gewählt. Wie in Abbildung 4.3 zu sehen, weist das Euler-Verfahren bei gleicher Schrittweite deutlich größere lokale Fehler auf als das Runge-Kutta-Verfahren. Eine dem Runge-Kutta-Verfahren 4. Ordnung vergleichbare Genauigkeit erhält man für das explizite Euler-Verfahren erst bei ca. einem Tausendstel der Schrittweite ($h = 5 \cdot 10^{-4}$). Dies macht sich vor allem in Bereichen plötzlicher Richtungsänderung (z.B. große Grundwasserspiegelabsenkung in Brunnennähe bzw. Infiltration aus einem Fließgewässer) durch höhere, lokale Diskretisierungsfehler nachteilig bemerkbar.

In weiterer Folge wird auf beide Verfahren der numerischen Integration von Anfangswertproblemen genauer eingegangen und deren Implementierung in R vorgestellt. Aufgrund der zuvor aufgezeigten Nachteile der Integration nach Euler, wird diese jedoch nicht angewendet.

4.2.1 Integration nach Euler

Es sei das Anfangswertproblem

$$y' = f(x, y)$$

gegeben, welches mit der Schrittweite $h > 0$ und dem Startwert $y(x_0) = y_0$ gelöst werden soll. Somit ergibt sich für die ersten beiden Schritte

$$\begin{aligned} y(x_0) &= y_0 \\ y(x_1) &\approx y_1 = y_0 + h \cdot f(x_0, y_0) \end{aligned}$$

und in allgemeiner Form

$$y(x_{k+1}) \approx y_{k+1} = y_k + h \cdot f(x_k, y_k) \quad (k = 1, 2, \dots, n).$$

Ausgehend von einem gewählten Anfangspunkt $P_0 = (x_0, y_0)$ geht man h Einheiten in Richtung der Tangentensteigung $f(x_0, y_0)$ und erhält einen Ordinatenwert y_1 welcher als Anfangswert für den darauffolgenden Schritt gewählt wird.

Dieses Verfahren liefert nur eine grobe Näherung der exakten Lösungskurve da zur Schätzung nur die Tangentensteigung im Ausgangspunkt berücksichtigt wird. Ausreichende Genauigkeiten sind nur mit hohem Rechenaufwand durch kleine Schrittweiten h zu erhalten Papula (2011).

In R lässt sich dieses Verfahren vergleichsweise einfach als Einzeiler implementieren,

```
1 euler <- function(x, y, h, fun) c(x + h, y + h * fun(x, y))
```

wobei die Funktion die Argumente

x Vektor, Anfangswert der unabhängigen Variable x

y Vektor, Anfangswert der abhängigen Variable y

h Vektor, Schrittweite in Richtung x

fun Funktion, welche zu integrieren ist und als Argumente x und y akzeptiert

erfordert und als Rückgabewert die Approximation des nächsten Punktes $(x + h, y_{k+1})$ liefert.

4.2.2 Runge-Kutta Verfahren

Explizite Runge-Kutta Verfahren werten die zu integrierende Funktion in s Zwischenstufen der Schrittweite h aus. Es wird eine mittlere Steigung durch Gewichtung der s Zwischenstufen gebildet. Das klassische Runge-Kutta Verfahren besitzt die Ordnung $s = 4$ und ist definiert mit

$$y_{k+1} = y_k + \frac{1}{6} \cdot (k_1 + 2k_2 + 2k_3 + k_4)$$

mit den vier Zwischenstufen

$$\begin{aligned} k_1 &= h \cdot f(x_k, y_k) \\ k_2 &= h \cdot f(x_k + h/2, y_k + k_1/2) \\ k_3 &= h \cdot f(x_k + h/2, y_k + k_2/2) \\ k_4 &= h \cdot f(x_k + h, y_k + k_3). \end{aligned}$$

Für einen Schritt sind somit vier Auswertungen der zu integrierenden Funktion notwendig Papula (2011).

```
2 runge <- function(x, y, h, fun) {
  k <- numeric(4)
4  k[1] <- h * fun(x, y)
  k[2] <- h * fun(x + h / 2, y + k[1] / 2)
6  k[3] <- h * fun(x + h / 2, y + k[2] / 2)
  k[4] <- h * fun(x + h, y + k[3])
8
  return(c(x + h, y + 1 / 6 * sum(k * c(1, 2, 2, 1))))
10 }
```

4.3 Clustering

4.3.1 Partitionierendes Clustering

4.3.2 Hierarchisches Clustering

4.3.3 Unscharfes Clustering

4.4 Hauptkomponentenanalyse

4.5 Co-Kriging

4.6 Zeitreihenanalyse

5 Ergebnisse

Die in reduzierten Wässern herrschende Redoxspannung wirkt häufig bereits bakterizid, daher sind natürliche reduzierte Wässer meist keimarm. (Hütter, 1994, S. 40)

6 Interpretation und Diskussion der Ergebnisse

7 Zusammenfassung

8 Ausblick

BibLaTeX: Punkt nach Auflage einfügen: 2. Auflage

BibLaTeX: irgendein Strichpunkt stimmt nicht

BibLaTeX: bei Zitat mit Seitenzahl das S.tilde dazu

Literaturverzeichnis

ALBERS, S. (2009). Methodik der empirischen Forschung. Gabler, 3. Ausgabe.

FOX, J., WEISBERG, S. (2011). An R Companion to Applied Regression. Sage, Thousand Oaks CA, 2. Ausgabe.

HÜTTER, L. A. (1994). Wasser und Wasseruntersuchung: Methodik, Theorie und Praxis chemischer, chemisch-physikalischer, biologischer und bakteriologischer Untersuchungsverfahren. Salle ; Sauerländer, 6. Ausgabe.

PAPULA, L. (2011). Mathematik für Ingenieure und Naturwissenschaftler Band 2 : Ein Lehr- und Arbeitsbuch für das Grundstudium.

R CORE TEAM (2012). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.