

Institut für Geoinformatik
Kurs: Modellierung dynamischer räumlicher Prozesse
Semester: WS 2008/2009
Dozent: Prof. Edzer Pebesma
Tutorin: Kristina Helle
Autoren: Stefan Arndt, Raimund Schnürer
Studiengang: Geoinformatik (Bachelor)

Analyse von Chloridwerten im südlichen Tullnerfeld (Österreich)



Abb.1: Blick über das Tullnerfeld ¹

Gliederung:

1. Einführung
 - 1.1. Arbeitshypothesen
 - 1.2. Beschreibung des Datensatzes
 - 1.3. Geografische Einordnung des Untersuchungsgebietes
 - 1.4. Charakterisierung des Indikators
2. Analyse
 - 2.1. Datenaufbereitung
 - 2.2. Analyse von Grenzwertüberschreitungen
 - 2.3. Analyse der Jahresmittelwerte
3. Fazit
4. Quellen
5. R-Code

¹ <http://de.wikipedia.org/wiki/Tullnerfeld>

1. Einführung

Der folgende Bericht befasst sich mit Chloridwerten, die im südlichen Tullnerfeld während eines Zeitraums von sechs Jahren gemessen wurden, und begutachtet diese unter dem Gesichtspunkt rechtlicher Rahmenbestimmungen. Grundlage der Untersuchung ist die in Österreich geltende Trinkwasserverordnung *BGBI. II Nr. 304/01*, die einen maximalen Schwellenwert von 200mg/l für die Chloridkonzentration im Wasser vorschreibt.

1.1. Arbeitshypothesen

In der vorliegenden Arbeit untersuchen wir folgende zwei Fragestellungen:

- 1.) Ist zu einem bestimmten Zeitpunkt der höchstzulässige Grenzwert für im Wasser gelöste Chloridionen überschritten worden, sodass eine Gefährdung der Gesundheit von Lebewesen in diesem Gebiet bestanden hat?
- 2.) Weisen die Jahresmittelwerte der Messstationen zeitlich-räumliche Trends auf?

1.2. Beschreibung des Datensatzes ¹

Die für die Analyse benötigten Daten entnehmen wir dem Datensatz *tull* aus dem Package *gstat*. Das Datenmaterial ist dabei in die drei Objekte *tull36*, *Chlorid92* und *TULLNREG* unterteilt. *tull36* enthält Namen sowie x- und y-Koordinaten von 36 Wasserqualitätsmessstationen, die unregelmäßig über das Untersuchungsgebiet verteilt sind. In *Chlorid92* findet man die Chloridkonzentrationen und 161 verschiedene Messzeitpunkte, die zwischen den Jahren 1992 und 1997 liegen. Es handelt sich bei den Werten um Tagesmittelwerte. *TULLNREG* beinhaltet 485 Koordinatenpaare, die das gesamte Testgebiet abdecken und als Vorlage für ein Gitternetz dienen.

1.3. Geografische Einordnung des Untersuchungsgebietes ²

Das Gebiet „Südliches Tullnerfeld“ befindet sich in Niederösterreich und erstreckt sich in einem Streifen von ca. 31 km Länge und 14 km Breite entlang der Donau. Es umfasst eine Fläche von rund 100 km² und ist ungefähr in 30 km nordwestlicher Richtung von der österreichischen Hauptstadt Wien entfernt. In der Nähe des Untersuchungsgebietes liegt die Stadt Tulln.



Abb.2: Bild-Overlay des Untersuchungsgebietes ^{1,2}

¹ http://www.ifas.jku.at/e5361/index_ger.html

² Google Earth

1.4. Charakterisierung des Indikators¹

Chlorid ist ein reaktives Ion mit negativer Ladung, das in fast allen Gewässern in gelöster Form enthalten ist. Als Feststoff kommt es gebunden in Natrium-, Kalium- oder Calciumchlorid vor. Je nach örtlichen Gegebenheiten treten in der Natur Konzentrationen von wenigen mg/l bis 100mg/l auf. Anthropogene Ursachen für einen erhöhten Chloridgehalt können Wasserenthärtungsanlagen oder Verunreinigungen durch Abwässer sein. Ebenso bewirken Straßenstreusalze oder Düngemittel einen Anstieg der Chloridkonzentration. Da eine Erhöhung nur durch Verdünnung beseitigt werden kann, ist Chlorid deshalb gut als permanenter Verschmutzungsindikator für die Wasserqualität geeignet.

2. Analyse

2.1. Datenaufbereitung

Am Anfang unserer Analyse luden wir zuerst den zu untersuchenden Datensatz „tull“ und die dazu erforderliche Bibliothek „gstat“ in den Arbeitsbereich des Statistikprogramms R. Nachdem wir uns einen Überblick über das Datenmaterial verschafft hatten, wurde uns klar, dass wir die gegebenen Daten entsprechend aufbereiten mussten, um sie später einfacher handzuhaben.

Hierfür erstellten wir einerseits die Tabelle „tulln“, welche die Messwerte an einzelnen Tagen, gespeichert in „Chlorid92“, mit den zugehörigen Koordinaten der Stationen, enthalten in „tull36“, verknüpft. In den Spalten der Tabelle stellten wir die x- und y-Werte der Koordinaten als auch die einzelnen Messzeitpunkte dar. Als Zeilen verwendeten wir die unterschiedlichen Messstationen. Zudem wurde uns bewusst, dass der Umgang mit den vielen nicht vorhandenen (NA) Werten sich schwierig gestalten würde.

Andererseits mussten wir das in „TULLNREG“ vorhandene Gitternetz überarbeiten, da nicht nur die Abstände zwischen den Koordinaten unregelmäßig waren, sondern auch zwei kleinere Randgebiete eine Lücke zum zentral gelegenen Untersuchungsgebiet aufwiesen. Dies machte eine Umwandlung des Data-Frames zu einem SpatialPixelsDataFrame unmöglich. Um dieses Problem zu beheben, rundeten wir die Gitternetzpunkte, sodass diese anschließend gleiche Abstände hatten, und entfernten die zwei äußeren Teilgebiete. Die erhaltenen Punkte bildeten nun unser neues Gitternetz „tulln.grid“ (siehe Abb. 3).

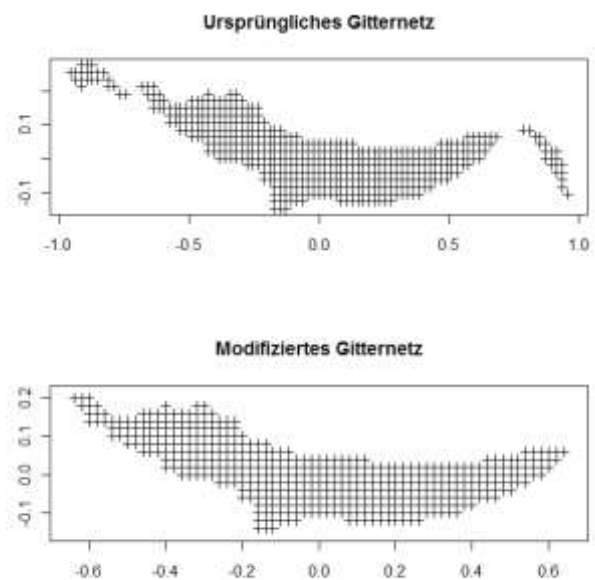


Abb. 3: Aufbereitung des Gitternetzes

2.2. Analyse von Grenzwertüberschreitungen

Nachdem wir die Daten aufbereitet hatten, konnten wir nun mit der Untersuchung der ersten Arbeitshypothese beginnen. Hierfür überprüften wir zunächst, ob und wenn ja, welche Messwerte

¹ <http://marktcheck.greenpeace.at/uploads/media/chlorid.pdf>

eine höhere Chloridkonzentration als der zulässige Grenzwert von 200mg/l aufweisen. Es stellte sich heraus, dass lediglich ein Messwert diesen Schwellenwert überschreitet. Nach einer näheren Betrachtung erfuhren wir, dass dieser 42mg/l über dem Grenzwert liegt und an der Station „S2066“ am 1. Februar 1994 gemessen wurde. Da nicht nur an dem betreffenden, sondern auch am vorangegangenen und nachfolgenden Tag je eine Messung an anderen Stationen durchgeführt wurde, konnten wir diese drei Spalten zu einer zusammenfassen, sodass wir 30 verschiedene Messpunkte für unsere Stichprobe während dieses kurzen Zeitraums zur Verfügung hatten. Die nicht verfügbaren Werte schlossen wir dabei aus.

Zu einer ersten Visualisierung stellten wir die Messwerte auf der Karte dar und erkannten, dass die Station mit dem herausgefundenen Höchstwert sich im östlichen Teil der Karte befindet. Insgesamt sind die Werte im westlichen Teil der Karte geringer als die im östlichen, mit Ausnahme des östlichsten Messpunktes. Da jedoch alle anderen Werte deutlich niedriger als der Maximalwert sind, wurde die voreingestellte Klassifikation stark davon beeinflusst, sodass feine Unterschiede innerhalb der niedrigeren Werte verloren gingen. Daher prüften wir die Möglichkeit, die Daten geeignet zu transformieren. Anhand des Box-Cox-Plots sahen wir, dass es sinnvoll wäre, den natürlichen Logarithmus auf die Daten anzuwenden, um eine annähernd normalverteilte Stichprobe zu erreichen.

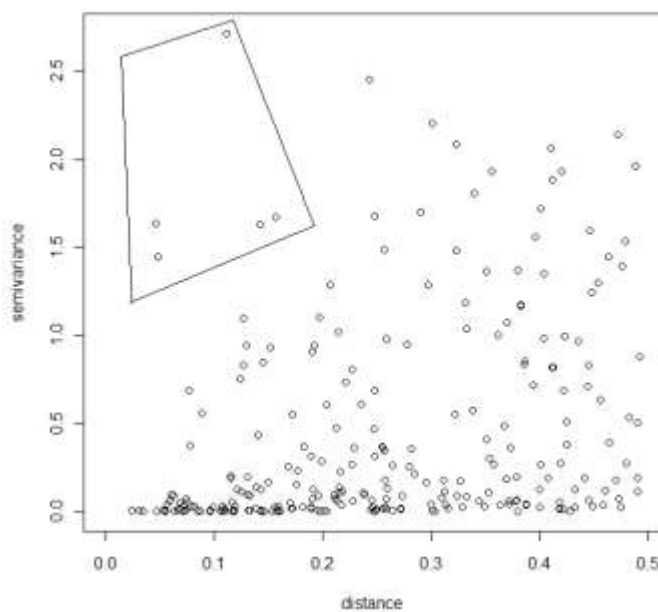


Abb. 4: Markieren von Punkten mit hoher Semivarianz in geringer Distanz in einer Variogramm-Cloud

Mithilfe der Variogramm-Cloud prüften wir nun die Beziehung der Distanz zu der Semivarianz einzelner Punktepaare. Uns interessierte hierbei, welche Punktepaare dicht beieinander liegen und gleichzeitig hohe Unterschiede aufweisen. Nach Auswahl von fünf Punkten in der Variogramm-Cloud, auf die diese Eigenschaften zutreffen (siehe Abb. 4), ließen wir uns diese auf der Karte anzeigen und auch die dazugehörigen Wertepaare ausgeben. Dabei ist zum Einen der Höchstwert im Osten der Karte beteiligt sowie ein sehr niedriger Wert in einer Umgebung von höheren Werten im Süden. Als nächstes erstellten wir das Variogramm aus den Stichprobenwerten. Als passendes Modell verwendeten wir eine exponentielle Funktion, deren

Parameter wir optimieren ließen. Dies ermöglichte uns nun das Ordinary Kriging durchzuführen, wodurch Werte der Rasterzellen berechnet wurden, an denen keine Messwerte vorlagen. Diese Schätzung visualisierten wir neben der Varianz auf einer Karte. Auch hier lässt sich eine Abnahme der Chloridkonzentrationen nach Westen hin beobachten. Auch den geringen Wert im Süden und den hohen Wert im Osten sieht man deutlich hervorstechen. Die Varianz ist nahezu an allen Stellen konstant. Anschließend führten wir eine Kreuzvalidierung durch, um zu ermitteln, wie gut unser aufgestelltes Modell die Messwerte beschreibt. Sowohl der Mittelwert der Abweichungen als auch der Mittelwert des Z-Scores lagen sehr dicht an Null. Der Mittelwert der quadrierten Abweichungen und die Varianz waren jedoch höher als im Optimalfall. Die Ursache hierfür ist auf der Karte der

hohen Residuen erkennbar. An den Punktepaaren, die wir mithilfe der Variogramm-Cloud herausgefunden haben, treten die größten Abweichungen auf.

Leider war es uns an dieser Stelle nicht möglich, vergleichend das Universal Kriging durchzuführen. Der zusätzliche Parameter - die Entfernung jedes Messpunktes zur Donau - ließ sich nicht ermitteln, da für das interne Koordinatensystem keine Referenz angegeben war bzw. das Gebiet sich nicht genau auf einer Satellitenkarte lokalisieren ließ, sodass dadurch eine Georeferenzierung ausgeschlossen war (siehe Abb. 2).

Zuletzt ließen wir das 95%-Konfidenzintervall zu dem gegebenen Grenzwert von 200mg/l berechnen und stellten es auf einer Karte dar. Von statistischer Seite ist es wahrscheinlich, dass der Schwellenwert größtenteils nicht überschritten wurde. Zu einem geringeren Teil auf der Karte - in der Umgebung vom Höchstwert - lässt sich jedoch nicht mit 95-prozentiger Sicherheit beantworten, ob dort eine Grenzwertüberschreitung vorliegt.

2.3. Analyse der Jahresmittelwerte

Nach Abschluss des ersten Teils unserer Analyse, widmeten wir uns nun der zweiten Fragestellung. Dafür untersuchten wir zuerst, ob zeitliche Korrelationen vorliegen, indem wir die Messwerte jeder Station in Abhängigkeit von der Zeit betrachteten. Hierbei lässt sich jedoch keine einheitliche Struktur erkennen, da an den Stationen sowohl steigende und fallende, als auch zyklische und variierende Trends zu beobachten sind. Dies lässt sich mit dem Mittelwert aller Korrelationskoeffizienten belegen, der mit einem Wert von 0,06 sehr nahe an Null liegt und somit auf eine insgesamt unkorrelierte Stichprobe hindeutet. Bei dieser Berechnung haben wir jedoch nicht die unteren fünf Stationen mit einbezogen, da hier nur wenige Messwerte vorlagen. Resultierend untersuchen wir deshalb nur den räumlichen-zeitlichen Trend, der sich bei der Mittelwertbildung der einzelnen Jahre ergibt.

Für die weiteren Schritte errechneten wir also die Jahresmittelwerte der jeweiligen Stationen und fügten diese der Tabelle „tulln“ hinzu. Außerdem selektierten wir alle nicht vorhandenen Werte, damit wir diese bei den späteren Interpolationsverfahren ausschließen können. Im Anschluss prüften wir, ob eine Transformation der Messwerte notwendig sei. Dazu erstellten wir Histogramme, Normalverteilungsdiagramme und Box-Cox-Plots der sechs Jahresmittelwerte. Wir entschieden uns, den Mittelwert von den Exponenten λ zu nehmen, der bei ungefähr $2/3$ liegt. Die nachfolgenden drei Diagramme der transformierten Mittelwerte weisen zwar keine optimale Normalverteilung auf, im Gegensatz aber zu keiner Umwandlung der Werte lässt sich eine Verbesserung feststellen (siehe Abb. 5).

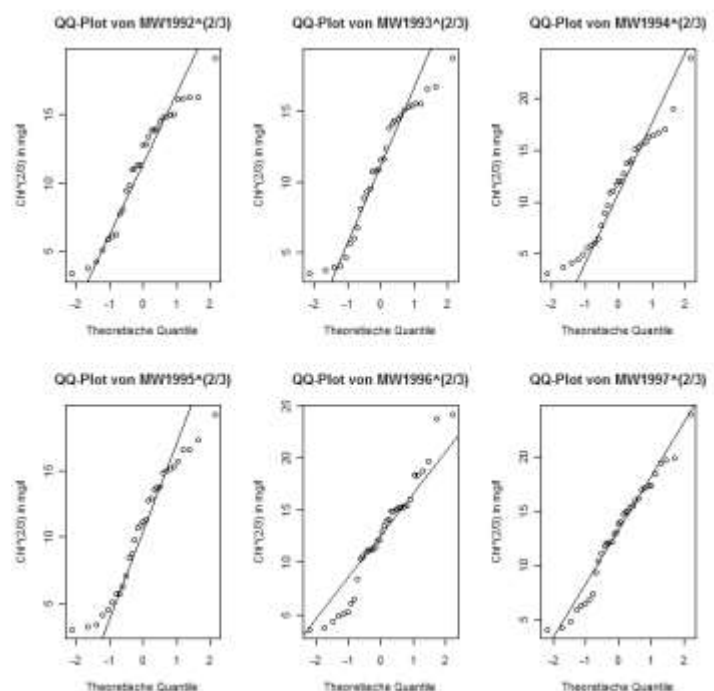


Abb. 5: Normalverteilungsdiagramme der transformierten Jahresmittelwerte

Als erstes Interpolationsverfahren probierten wir das Inverse-Distance-

Weighting aus. In der erstellten Grafik lässt sich grundsätzlich eine hohe Ähnlichkeit zwischen den Jahresmittelwerten erkennen. Im Zentrum ist jedoch ein leichter Anstieg im Laufe der Jahre zu beobachten. Zudem lassen sich einzelne Messpunkte mit erhöhten Chloridkonzentrationen feststellen, die allerdings nicht den zulässigen Grenzwert überschreiten.

Um die Oberfläche zu glätten und dadurch weitere räumliche Muster zu erfassen, führten wir nun Block-Kriging als zweite Interpolationsvariante durch. Dafür benutzten wir wiederum die Exponentialfunktion als geeignetes Modell, die wir optimieren ließen. In der resultierenden Karte erkennt man bei der Schätzung sehr deutlich, dass die Chloridkonzentrationen von Westen und von Osten ins Zentrum hin zunehmen. Eine Ausnahme bildet das Jahr 1994, bei der sich die höchste Konzentration nach Osten verlagert hat. Grund dafür ist der Maximalwert, den wir in Abschnitt 2.2 untersucht haben. Die Ränder der Karte weisen die höchsten Varianzen der Jahresmittelwerte auf, die zentralen Gebieten die niedrigsten, da dort mehr Werte in einem Block zur Mittelung vorliegen.

In Vorbereitung auf das Cokriging, welches die Vorteile der vorangegangenen Interpolationsmethoden bündeln soll, ließen wir uns die Korrelationen zwischen den Jahresmittelwerten untereinander ausgeben. Alle Werte sind sehr hoch und veranschaulichen somit eine starke Korrelation. Selbst der geringste Wert zwischen den Jahren 1994 und 1997 ist mit 0,86 noch vergleichsweise hoch. Daraus können wir schließen, dass sich die Werte gut für das Cokriging eignen werden.

Im Zuge dessen stellten wir zunächst die direkten Variogramme und Kreuzvariogramme auf. Wir erhöhten dabei geringfügig den Grenzwert der Distanz, um auch die Varianzen weiter entfernter Punktepaare mit einfließen zu lassen. Darüberhinaus mussten wir einen weiteren Parameter modifizieren, damit das Verfahren fehlerfrei abläuft.¹ Jetzt konnten wir das Ergebnis des Cokriging auf einer Karte visualisieren (siehe Abb. 6). Dabei fällt auf, dass der südliche Teil stark von den hohen Messwerten beeinflusst wird, die an zwei neueren Stationen in den Jahren 1996 und 1997 gemessen wurden. Diese scheinen ebenfalls die Ursache für die deutlich höheren Varianzen in den beiden letzten Jahren der Messungen zu sein.

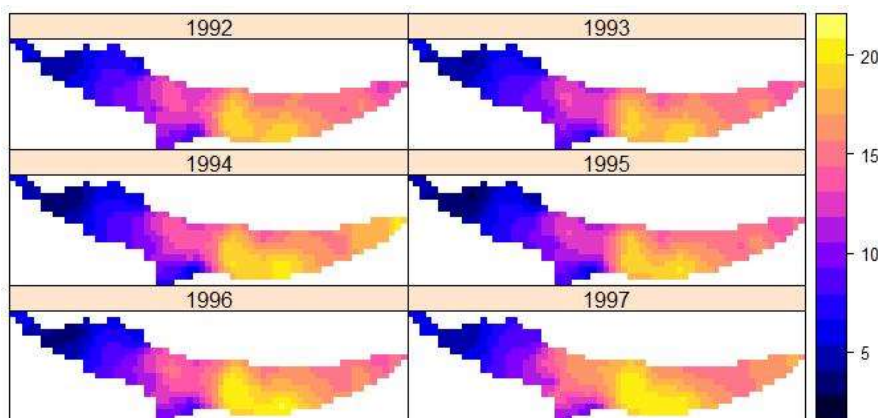


Abb. 6: Schätzungen der Jahresmittelwerte mittels Cokriging

3. Fazit

Zusammenfassend lässt sich sagen, dass eine einmalige Überschreitung des zulässigen Grenzwertes für die Chloridkonzentration im Wasser während des Untersuchungszeitraums vorkommt. Dass es sich hierbei um einen Messfehler handelt, ist weitestgehend auszuschließen, da die vorliegenden Werte aus den Mittelwerten von verschiedenen Messungen an einem Tag berechnet wurden. Aus

¹ <https://stat.ethz.ch/pipermail/r-sig-geo/2009-January/004850.html>

diesem Grund vermuten wir eine andere Ursache für dieses Ereignis. Bei unseren Recherchen haben wir herausgefunden, dass es nur wenige Tage vor dieser Messung einen schweren Eissturm über Deutschland und Österreich gegeben hat.¹ Möglicherweise wurde infolgedessen Streusalz nahe dieser Station eingesetzt, was eine hohe Konzentration an Chloridionen zur Konsequenz hatte. Unsere angewendeten Analysemethoden ergaben, dass die statistische Aussage, ob eine Übertretung des Schwellenwertes des Chloridgehalts an diesem Tag vorlag, nicht sicher zu beantworten sei. Dabei ist anzuführen, dass innerhalb der Stichprobe auf der einen Seite die Messwerte sich mit zunehmender Entfernung nur leicht veränderten, es aber auf der anderen Seite auch große Abweichungen in ziemlich geringen Abständen gab. Außer dieser einen Ausnahme lagen die Messwerte immer in einem zulässigen Bereich, sodass die Wasserqualität bezüglich der Chloridkonzentration in Ordnung war und keine Gefahr für Lebewesen in diesem Zeitraum darstellte.

Bei der Untersuchung der jährlichen Mittel von 1992 bis 1997 ergaben sich einige räumliche Muster und Trends, wobei der Anstieg der Werte im Süden besonderer Beachtung bedarf. Ob aus diesem Anlass die zusätzlichen Messstationen in Betrieb genommen wurden, ist uns nicht bekannt. Feststeht jedoch, dass dieses Gebiet auch in Zukunft genau beobachtet werden sollte, damit eine ausreichend gute Wasserqualität gewährleistet wird.

4. Quellen

- Bivand, Roger S. et al. (2008): Applied Spatial Data Analysis with R. New York
- Sachs, Lothar et al. (2006): Angewandte Statistik - Methodensammlung mit R (12. Auflage). Berlin, Heidelberg, New York
- R Packages und Help
- als Fußnoten vermerkte Internetquellen

5. R-Code

```
## 2.1. Datenaufbereitung

> library(gstat)
> data(tull)
> edit(apply(Chlorid92, 2, as.character))
> edit(tull36)

> tulln = cbind(tull36[,1:2], t(Chlorid92[1:161,2:37]))
> colnames(tulln) = c("x", "y", as.character(Chlorid92[,1]))

> edit(TULLNREG)
> tmp.grid = apply(TULLNREG, 2, function(x) round(x/21, 3)*20)
> tmp.grid = as.data.frame(tmp.grid)
> sel = which((tmp.grid$x > -0.7 & tmp.grid$x < 0.7) == FALSE)
> tulln.grid = tmp.grid[-sel,]
> coordinates(tulln.grid) = ~x + y
> gridded(tulln.grid) = TRUE
> par(mfrow=c(2,1))
> plot(TULLNREG, main = "Ursprüngliches Gitternetz", xlab="", ylab="", pch = 3)
> plot(tulln.grid, axes = TRUE); title("Modifiziertes Gitternetz")
```

¹<http://www.wetterzentrale.de/cgi-bin/wetterchronik/home.pl?read=820&jump1=topic&jump2=17>

```

## 2.2. Analyse von Grenzwertüberschreitungen

> exc = c() ## exc = exceeding
> for (i in 1:36) { for (j in 3:163) { if (!is.na(tulln[i,j]) && tulln[i,j] > 200)
exc = c(exc, paste("[" ,i ,",",j,""]", sep = "")) } }
> exc
[1] "[26,63]"
> tulln[26,63]; rownames(tulln[26,]); colnames(tulln[63])
[1] 242
[1] "S2066"
[1] "1994-02-01 00:00:00"
> edit(apply(Chlorid92[59:63,], 2, as.character))

> get.mean = function(matrix) { ##selbst geschriebene Funktion, um den Mittelwert
von Zeilen einer Matrix zu bilden
  get = c()
  for (i in 1:length(matrix[,1])) {
    m = mean(as.numeric(matrix[i, 1:length(matrix[1,])]), na.rm = TRUE)
    if (m == "NaN") { m = "NA" }
    get = c(get, m)
  }
  get = as.numeric(get)
  return(get)
}
> chl = get.mean(tulln[,62:64])
> tmp = cbind(tull36, chl)

> sel.na = function(list) { ##selbst geschriebene Funktion, um NA-Werte in einer
Liste auszuwählen
  sel = c()
  for (i in 1:length(list)) {
    if (is.na(list[i])) sel = c(sel, i)
  }
  if (length(sel) == 0) { sel = seq(-1,-length(list),by=-1) }
  return(sel)
}
> na = sel.na(chl)
> tulln.exc = tmp[-na,]

> coordinates(tulln.exc) = ~x + y
> sp.theme(TRUE)
> splot(tulln.exc, "chl", main = "Erste Visualisierung der Chloridkonzentrationen
in mg/l \n am 31.01. / 01.02. / 02.02.1994")
> library(MASS)
> boxcox(lm(chl ~ 1)); title("Box-Cox-Plot der Stichprobenwerte")
> hist(log(chl), main = "Histogramm der logarithmierten Stichprobenwerte", xlab =
"log(Chl) in mg/l", ylab = "Häufigkeit")

> par(mfrow=c(1,1))
> sel = plot(variogram(log(chl) ~ 1, tulln.exc, cloud = TRUE), digitize = TRUE,
main = "Cloud-Variogramm \n (um Punktepaare mit hoher Varianz \n in geringer
Entfernung zu markieren)", xlab = "Entfernung", ylab = "Semivarianz")
> cbind(tulln.exc$chl[sel$head], tulln.exc$chl[sel$tail])
  [,1] [,2]
[1,] 48.6 8.0
[2,] 43.9 8.0
[3,] 8.0 49.8
[4,] 242.0 39.7
[5,] 8.0 82.3
> plot(sel, tulln.exc, main = "Punktepaare mit hoher Varianz in geringer
Entfernung")

> var = variogram(log(chl) ~ 1, tulln.exc)
> var.est = vgm(0.8, "Exp", 0.3, 0)
> var.fit = fit.variogram(var, var.est);
> plot(var, var.fit, main = "Variogramm der Stichprobenwerte mit angepasstem
Modell", xlab = "Entfernung", ylab = "Semivarianz")
> ok = krige(log(chl) ~ 1, tulln.exc, tulln.grid, var.fit);
> splot(ok, c("var1.var", "var1.pred"), names.attr = c("Varianz","Schätzung"),
main = "Chloridkonzentrationen (log) in mg/l \n am 31.01. / 01.02. / 02.02.1994 \n
mittels Ordinary Kriging")

> cv.ok = krige.cv(log(chl)~1, tulln.exc, var.fit, nfold = 5)
> mean(cv.ok$residual)

```



```

[1] 0.002006220
> mean(cv.ok$residual^2)
[1] 0.3878332
> mean(cv.ok$zscore)
[1] 0.003983765
> var(cv.ok$zscore)
[1] 1.691224
> bubble(cv.ok, "residual", main = "Entstandene Abweichungen bei der
Kreuzvalidierung")

> ci = function(cutoff, mean, var, alpha = 0.05) { ##Funktion aus der Übung, um
Konfidenzintervalle für einen Grenzwert zu berechnen
  p = 1 - pnorm(cutoff, mean, sqrt(var))
  f = ifelse(p < alpha/2, "niedriger", ifelse(p < (1 - alpha/2), "nicht
eindeutig", "höher"))
  factor(f, levels = c("niedriger", "nicht eindeutig", "höher"))
}
> ok$ci200 = ci(log(200), ok$var1.pred, ok$var1.var)
> spplot(ok, c("ci200"), col.regions = c("green","yellow","red"))
> title("95%-Konfidenzintervall \n zur Überschreitung des Grenzwertes")

## 2.3. Analyse der Jahresmittelwerte

> par(mfrow=c(6,6))
> for(i in 2:37) { plot(Chlorid92[,i] ~ Chlorid92$Datum, main = paste("",
colnames(Chlorid92[i])), xlab = "", ylab = "Chl in mg/l" ) }
> b = c()
> for (i in 2:32) b = c(b, cor(Chlorid92[,i], Chlorid92$Datum, "complete"))
> mean(b)
[1] 0.05929298

> tulln$MW1992 = get.mean(tulln[,3:37]) ##Mittelwerte von 1992
> tulln$MW1993 = get.mean(tulln[,38:61]) ##Mittelwerte von 1993
> tulln$MW1994 = get.mean(tulln[,62:91]) ##Mittelwerte von 1994
> tulln$MW1995 = get.mean(tulln[,92:122]) ##Mittelwerte von 1995
> tulln$MW1996 = get.mean(tulln[,123:156]) ##Mittelwerte von 1996
> tulln$MW1997 = get.mean(tulln[,157:163]) ##Mittelwerte von 1997
> NA1992 = sel.na(tulln$MW1992); NA1993 = sel.na(tulln$MW1993); NA1994 =
sel.na(tulln$MW1994)
> NA1995 = sel.na(tulln$MW1995); NA1996 = sel.na(tulln$MW1996); NA1997 =
sel.na(tulln$MW1997)
> par(mfrow=c(2,3))
> for (i in 164:169) hist(tulln[,i], main = paste("Histogramm von",
colnames(tulln[i])), xlab = "Chl in mg/l", ylab = "Häufigkeit")
> for (i in 164:169) { qqnorm(tulln[,i], main = paste("QQ-Plot von",
colnames(tulln[i])), xlab = "Theoretische Quantile", ylab = "Chl in mg/l");
lines(qqline(tulln[,i])) }
> for (i in 164:169) { boxcox(lm(tulln[,i] ~ 1, tulln)); title(paste("Box-Cox-Plot
von", colnames(tulln[i]))) }
> for (i in 164:169) hist(tulln[,i]^(2/3), main = paste("Histogramm von ",
colnames(tulln[i]), "^(2/3)", sep = ""), xlab = "Chl^(2/3) in mg/l", ylab =
"Häufigkeit")
> for (i in 164:169) { qqnorm(tulln[,i]^(2/3), main = paste("QQ-Plot von ",
colnames(tulln[i]), "^(2/3)", sep = ""), xlab = "Theoretische Quantile", ylab =
"Chl^(2/3) in mg/l"); lines(qqline(tulln[,i]^(2/3))) }
> for (i in 164:169) { boxcox(lm(tulln[,i]^(2/3) ~ 1, tulln)); title(paste("Box-
Cox-Plot von ", colnames(tulln[i]), "^(2/3)", sep = "")) }

> coordinates(tulln) = ~x + y
> idw = idw(MW1992^(2/3) ~ 1, tulln[-NA1992,], tulln.grid); idw$mw1992 =
idw$var1.pred
> tmp = idw(MW1993^(2/3) ~ 1, tulln[-NA1993,], tulln.grid); idw$mw1993 =
tmp$var1.pred
> tmp = idw(MW1994^(2/3) ~ 1, tulln[-NA1994,], tulln.grid); idw$mw1994 =
tmp$var1.pred
> tmp = idw(MW1995^(2/3) ~ 1, tulln[-NA1995,], tulln.grid); idw$mw1995 =
tmp$var1.pred
> tmp = idw(MW1996^(2/3) ~ 1, tulln[-NA1996,], tulln.grid); idw$mw1996 =
tmp$var1.pred
> tmp = idw(MW1997^(2/3) ~ 1, tulln[-NA1997,], tulln.grid); idw$mw1997 =
tmp$var1.pred
> spplot(idw, c("mw1996", "mw1997", "mw1994", "mw1995", "mw1992", "mw1993"),
names.attr = c("1996", "1997", "1994", "1995", "1992", "1993"), main = "Schätzungen

```

der Jahresmittelwerte \n von Chloridkonzentrationen^(2/3) in mg/l \n mittels Inverse-Distance-Weighting")

```
> v = variogram(MW1992^(2/3) ~ 1, tulln[-NA1992,])
> v.fit = fit.variogram(v, vgm(20, "Exp", 0.35, 3))
> bk = krige(MW1992^(2/3) ~ 1, tulln[-NA1992,], tulln.grid, v.fit, block =
c(0.2,0.2))
> bk$mw1992.pred = bk$var1.pred; bk$mw1992.var = bk$var1.var
> v = variogram(MW1993^(2/3) ~ 1, tulln[-NA1993,])
> v.fit = fit.variogram(v, vgm(20, "Exp", 0.35, 3))
> tmp = krige(MW1993^(2/3) ~ 1, tulln[-NA1993,], tulln.grid, v.fit, block =
c(0.2,0.2))
> bk$mw1993.pred = tmp$var1.pred; bk$mw1993.var = tmp$var1.var
> v = variogram(MW1994^(2/3) ~ 1, tulln[-NA1994,])
> v.fit = fit.variogram(v, vgm(20, "Exp", 0.35, 3))
> tmp = krige(MW1994^(2/3) ~ 1, tulln[-NA1994,], tulln.grid, v.fit, block =
c(0.2,0.2))
> bk$mw1994.pred = tmp$var1.pred; bk$mw1994.var = tmp$var1.var
> v = variogram(MW1995^(2/3) ~ 1, tulln[-NA1995,])
> v.fit = fit.variogram(v, vgm(20, "Exp", 0.35, 3))
> tmp = krige(MW1995^(2/3) ~ 1, tulln[-NA1995,], tulln.grid, v.fit, block =
c(0.2,0.2))
> bk$mw1995.pred = tmp$var1.pred; bk$mw1995.var = tmp$var1.var
> v = variogram(MW1996^(2/3) ~ 1, tulln[-NA1996,])
> v.fit = fit.variogram(v, vgm(20, "Exp", 0.35, 3))
> tmp = krige(MW1996^(2/3) ~ 1, tulln[-NA1996,], tulln.grid, v.fit, block =
c(0.2,0.2))
> bk$mw1996.pred = tmp$var1.pred; bk$mw1996.var = tmp$var1.var
> v = variogram(MW1997^(2/3) ~ 1, tulln[-NA1997,])
> v.fit = fit.variogram(v, vgm(20, "Exp", 0.35, 3))
> tmp = krige(MW1997^(2/3) ~ 1, tulln[-NA1997,], tulln.grid, v.fit, block =
c(0.2,0.2))
> bk$mw1997.pred = tmp$var1.pred; bk$mw1997.var = tmp$var1.var
> spplot(bk, c("mw1996.pred", "mw1997.pred", "mw1994.pred", "mw1995.pred",
"mw1992.pred", "mw1993.pred"), names.attr = c("1996", "1997", "1994", "1995",
"1992", "1993"), main = "Schätzungen der Jahresmittelwerte \n von
Chloridkonzentrationen^(2/3) in mg/l \n mittels Block-Kriging")
> spplot(bk, c("mw1996.var", "mw1997.var", "mw1994.var", "mw1995.var",
"mw1992.var", "mw1993.var"), names.attr = c("1996", "1997", "1994", "1995", "1992",
"1993"), main = "Varianzen der Jahresmittelwerte \n von
Chloridkonzentrationen^(2/3) in mg/l \n mittels Block-Kriging")

> cor((as.data.frame(tulln)[c("MW1992", "MW1993", "MW1994", "MW1995", "MW1996",
"MW1997")])^(2/3), use = "complete.obs")
      MW1992  MW1993  MW1994  MW1995  MW1996  MW1997
MW1992 1.0000000 0.9698774 0.8726617 0.9556653 0.9331945 0.8968798
MW1993 0.9698774 1.0000000 0.9400590 0.9762435 0.9331259 0.8797276
MW1994 0.8726617 0.9400590 1.0000000 0.9275718 0.8906369 0.8645558
MW1995 0.9556653 0.9762435 0.9275718 1.0000000 0.9345175 0.8762184
MW1996 0.9331945 0.9331259 0.8906369 0.9345175 1.0000000 0.9616014
MW1997 0.8968798 0.8797276 0.8645558 0.8762184 0.9616014 1.0000000

> g = NULL
> g = gstat(g, "MW1992", MW1992^(2/3) ~ 1, tulln[-NA1992,])
> g = gstat(g, "MW1993", MW1993^(2/3) ~ 1, tulln[-NA1993,])
> g = gstat(g, "MW1994", MW1994^(2/3) ~ 1, tulln[-NA1994,])
> g = gstat(g, "MW1995", MW1995^(2/3) ~ 1, tulln[-NA1995,])
> g = gstat(g, "MW1996", MW1996^(2/3) ~ 1, tulln[-NA1996,])
> g = gstat(g, "MW1997", MW1997^(2/3) ~ 1, tulln[-NA1997,])
> vm = variogram(g, cutoff = 0.6)
> vm.fit = fit.lmc(vm, g, vgm(20, "Exp", 0.35, 3), correct.diagonal = 1.01)
> plot(vm, vm.fit, main = "Direkte Variogramme und Kreuzvariogramme mit angepassten
Modellen", xlab = "Entfernung", ylab = "Semivarianz")
> ck = predict(vm.fit, tulln.grid)
> spplot(ck, c("MW1996.pred", "MW1997.pred", "MW1994.pred", "MW1995.pred",
"MW1992.pred", "MW1993.pred"), names.attr = c("1996", "1997", "1994", "1995",
"1992", "1993"), main = "Schätzungen der Jahresmittelwerte \n von
Chloridkonzentrationen^(2/3) in mg/l \n mittels Cokriging")
> spplot(ck, c("MW1996.var", "MW1997.var", "MW1994.var", "MW1995.var",
"MW1992.var", "MW1993.var"), names.attr = c("1996", "1997", "1994", "1995", "1992",
"1993"), main = "Varianzen der Jahresmittelwerte \n von
Chloridkonzentrationen^(2/3) in mg/l \n mittels Cokriging")
> edit(apply(Chlorid92[,c(1, 33:37)], 2, as.character))
```