



Erstellen von Diagrammen

Die Darstellung von erhobenen Daten ist ein integraler Bestandteil wissenschaftlicher Arbeit. Durch eine anschauliche Darstellungsweise können Zusammenhänge erkannt und Ergebnisse der Arbeit effektiv kommuniziert werden. Diagramme müssen in allen Bereichen der Geowissenschaften erstellt werden, im wissenschaftlichen Bereich für akademische Arbeiten, Vorträge und Publikationen sowie im angewandten Ingenieurbereich für Berichte und Gutachten. Meist gibt es für jeden Verwendungszweck eigene, verbindliche Vorgaben, z.B. DIN und Richtlinien einzelner wissenschaftlicher Fachzeitschriften. Es gibt jedoch einige Grundregeln zum Erstellen wissenschaftlicher Diagramme, die im Folgenden dargelegt werden sollen.

1. Streudiagramme

Wissenschaftliche Diagramme zur Darstellung von Messwerten sind immer Streudiagramme (scatter plots), d.h. es wird die statistische Abhängigkeit zweier Variablen dargestellt. Balkendiagramme, Kuchendiagramme und ähnliche Darstellungsweisen sind nicht dazu geeignet, wissenschaftliche Messwerte darzustellen, da sie keine direkte, lineare Abhängigkeit der Variablen zeigen. Diese Art von Diagrammen wird vor allem für wirtschaftliche Daten benutzt (Einnahmen, GDP etc.). Bei Streudiagrammen wird dargestellt, wie sich der Wert einer Messgröße (abhängige Variable) ändert, wenn der Wert der Stellgröße (unabhängige Variable) verändert wird. Besteht ein kausaler Zusammenhang zwischen den beiden Variablen, so wird die unabhängige Variable als Ursache und die abhängige Variable als Wirkung bezeichnet. Aus einer Korrelation kann allerdings noch keine Kausalität abgeleitet werden (cum hoc ergo propter hoc). In einem kartesischen Koordinatensystem wird nach wissenschaftlicher Konvention die unabhängige Variable auf der Abszisse (x-Achse) und die abhängige Variable auf der Ordinate (y-Achse) dargestellt. In Einzelfällen wird das Diagramm auch anders gestaltet, um es besser lesbar zu machen. So wird in geowissenschaftlichen Diagrammen oft die Tiefe nach unten dargestellt, obwohl sie eigentlich die unabhängige Variable darstellt.

2. Achsen, Skalierung und Einheiten

In wissenschaftlichen Diagrammen wird normalerweise ein Rahmen (box) statt zwei Achsen gezeichnet. Absolute Werte müssen auf beiden Achsen in gleichmäßigen Abständen mit Teilstrichen dargestellt werden. Dabei sollen die Teilstriche in einem sinnvollen Abstand gewählt werden. Zu dichte Teilstriche machen das Diagramm unübersichtlich, zu weite Abstände erschweren das Ablesen von Werten. Als Inkremente sollten 1, 2, 5 oder Zehnerpotenzen davon verwendet werden. Sehr große oder kleine Werte können abgekürzt werden, indem SI-Präfixe für Maßeinheiten verwendet werden oder eine Zehnerpotenz auf der jeweiligen Achse angegeben wird, mit der jeder Wert auf der Achse multipliziert werden soll. Die Achsen sollten so skaliert werden, dass die Datenpunkte möglichst das gesamte Diagramm ausfüllen. Verwenden Sie keine Gitterlinien, da diese das Diagramm überfüllen. Alle Achsen müssen sinnvoll beschriftet sein. Dabei muss die jeweilige Variable sowie deren physikalische Einheit in runden oder eckigen Klammern angegeben werden (z.B. „Tiefe d [km]“).

3. Datenpunkte, Fehler und Legende

Es empfiehlt sich, für unterschiedliche Datensätze, die zusammen in einem Diagramm dargestellt werden, verschiedene Symbole (Kreuze, Sterne, Kreise, Quadrate, Rauten etc.) bzw. Linienarten (fett, gestrichelt, gepunktet) zu verwenden. Symbole müssen groß genug und Linien breit genug sein, um gut erkennbar zu sein. Wenn zwei Messreihen nahe aneinander liegen, sollten gut unterscheidbare Symbole verwendet werden (z.B. ungefüllte Kreise und schwarz gefüllte Dreiecke). Wird eine große Zahl unterschiedlicher Datensätze im selben Diagramm dargestellt, sind Farben hilfreich. Mit Farben sollte allerdings sparsam umgegangen werden. Es gibt nur eine geringe Zahl verschiedener Farben, die vom menschlichen Auge gut und eindeutig unterschieden werden können. Es ist zudem auf Farbsehschwächen zu achten, vor allem Rot-Grün-Sehschwäche, die sehr weit verbreitet ist (ca. 9% bei Männern, 1% bei Frauen). Zudem verfügen die meisten nur über Schwarz-Weiß-Drucker und -Photokopierer, was die Reproduktion von Dokumenten in Farbe erschwert. Daher sind Helligkeitsunterschiede (Graustufen) und Symbole bzw. Linienarten Farben vorzuziehen. Alle verwendeten Symbole und Linien müssen in einer Legende erklärt werden, die meist innerhalb des Rahmens gezeichnet wird. Bei Messdaten muss immer ein Fehler mit angegeben werden. Dieser wird über Fehlerbalken ausgehend vom Mittelpunkt des Symbols nach oben und unten (Fehler in y) oder links und rechts (Fehler in x) dargestellt. Die Länge der Balken muss den Fehlerwerten entsprechen.

4. Regression

Einzelne Datenpunkte dürfen nicht durch Linien miteinander verbunden werden. Es gibt Ausnahmen, z.B. bei geochemischen Datensätzen, bei denen die Punkte verbunden werden, um Trends leichter erkennbar zu machen. Bei Modellen mit definierten, diskreten Werten können die Einzelpunkte ebenfalls verbunden werden. Dies ist jedoch auf bestimmte Einzelfälle beschränkt, bei denen diese Konvention angewandt wird. Um eine Abhängigkeit der Variablen zu untersuchen und die absoluten Werte der Parameter der Abhängigkeitsbeziehung sowie deren Fehler zu berechnen, wird eine Regression der Daten durchgeführt. Dabei wird eine Funktion an die Datenpunkte angenähert, deren Abweichung der Messwerte von der Funktion (Residuen) möglichst gering ist, z.B. nach der Methode der kleinsten Quadrate (least squares method). Am häufigsten tritt ein linearer Zusammenhang der Form $y = mx + c$ auf, wobei m die Steigung und c der y-Achsenabschnitt ist. Die Werte von m und c können über eine lineare Regression ermittelt werden, wobei auch das Bestimmtheitsmaß (R^2) berechnet wird, das den Anteil der Varianz der abhängigen Variablen angibt. Die Regressionsgerade wird als Funktion im Diagramm aufgetragen. Die Gleichung der Geraden mit den Werten der ermittelten Parameter sowie das Bestimmtheitsmaß sollte in der Legende oder als Beschriftung der Geraden angegeben werden.

5. Verwendung im Text

Wenn ein Diagramm in einem Text verwendet wird muss es als Abbildung (figure) bezeichnet werden. Zu den Abbildungen zählen außer Diagrammen noch Photographien, Karten und andere graphische Darstellungen (außer Tabellen). Abbildungen müssen immer nach der ersten Erwähnung im Text eingebunden werden und der Reihenfolge nach nummeriert werden. Unterschriften ist das Wort „Abbildung“ sowie die jeweilige Nummer voranzustellen, gefolgt von einem Doppelpunkt. Dann folgt eine kurze Beschreibung des Diagramms und der enthaltenen Daten.

Beispiel: „Abbildung 5: Der Zusammenhang zwischen Tiefe und Temperatur, gemessen in Bohrloch X mit Wärmesonde Y. Jeder Punkt ist der gemittelte Wert von 20 Einzelmessungen mit 2σ -Fehler.“

Ein alleinstehendes Diagramm, das nicht in einen Text eingebunden wird – z.B. als Teil einer Folie für einen Vortrag – benötigt einen kurzen Titel, z.B. „Die Abhängigkeit von y und x“, der oberhalb des Diagramms platziert wird.

6. Beispiele

Die folgenden Beispiele für gute und schlechte Diagramme verwenden alle denselben Datensatz (Tab. 1). Hierbei handelt es sich um Messwerte der Temperatur mit der Tiefe, die in einem Bohrloch gemessen wurden.

Tabelle 1: Beispieldatensatz für alle Diagramme.
In bestimmten Bohrlochtiefen gemessene Temperatur mit Fehlerangaben.

Tiefe d [m]	Temperatur T [°C]	2σ -Messfehler Temperatur [°C]
500	10	2
650	15	2
1300	28	3
2300	39	2
2500	45	1
3000	54	3

a) Schlechte Beispiele

Das erste Beispiel (Abb. 1) zeigt eine komplett falsche Darstellungsweise. Es handelt sich hierbei um ein Liniendiagramm aus Excel. Die Ordinatenachse ist nicht linear, es werden die einzelnen Werte in gleichem Abstand dargestellt, obwohl sie nicht gleichmäßig verteilt sind. Dies zeigt nicht die Abhängigkeit der beiden Variablen auf.

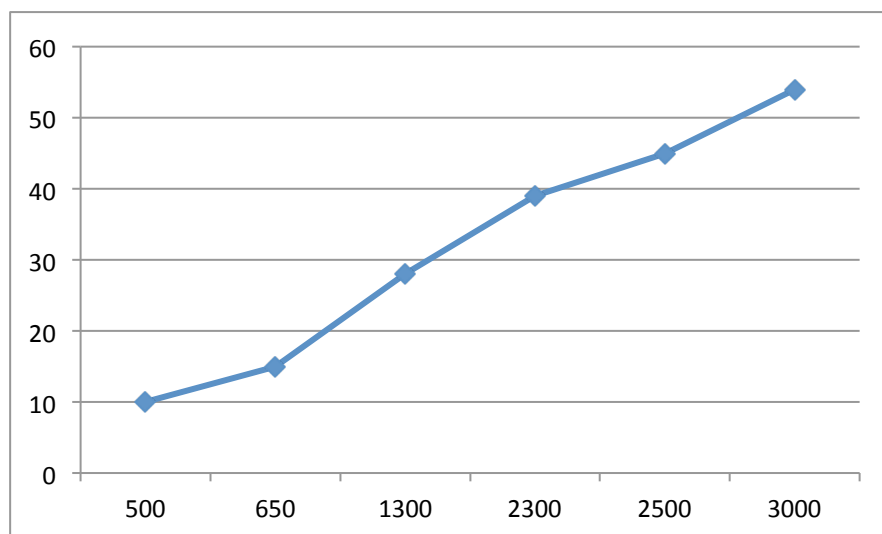


Abbildung 1: Beispiel für falsche Darstellungsweise.
Es handelt sich um ein Liniendiagramm, das mit Excel erstellt wurde.

Das Diagramm in Abbildung 2 zeigt ein Beispiel für ein schlechtes Diagramm. Die Diagrammart (Streudiagramm) ist zwar schon eine korrekte Darstellungsweise, allerdings sind die Achsen nicht beschriftet, die Messwerte sind durch Linien verbunden, die Gitterlinien machen das Diagramm unübersichtlich und die Beschreibung der Symbole („Datenreihe 1“) ist nicht sehr aufschlussreich.

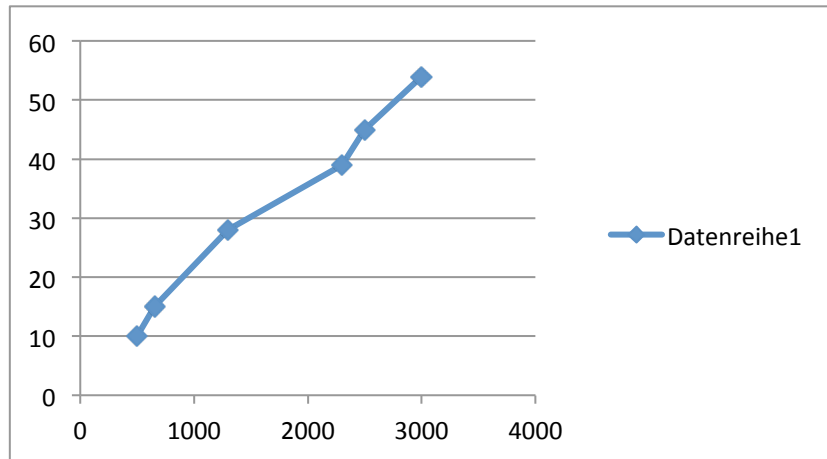


Abbildung 2: Beispiel für ein schlechtes Diagramm. Streudiagramm erstellt mit Excel.

Abbildung 3 zeigt einige weitere Fehler auf. Zwar ist der Rahmen und die Beschriftung der Achsen korrekt, die Teilstriche auf beiden Achsen sind jedoch viel zu eng gesetzt. Zudem erschwert der unglücklich gewählte Abstand der Teilstriche auf der Ordinate von 150 m das Ablesen. Die Symbole sind viel zu klein dargestellt und die Legende ist nicht deutlich abgegrenzt, so dass das Beispielsymbol der Legende mit einem Messpunkt verwechselt werden kann. Die Verwendung von Farben zur Darstellung einer einzelnen Messreihe ist nicht sinnvoll.

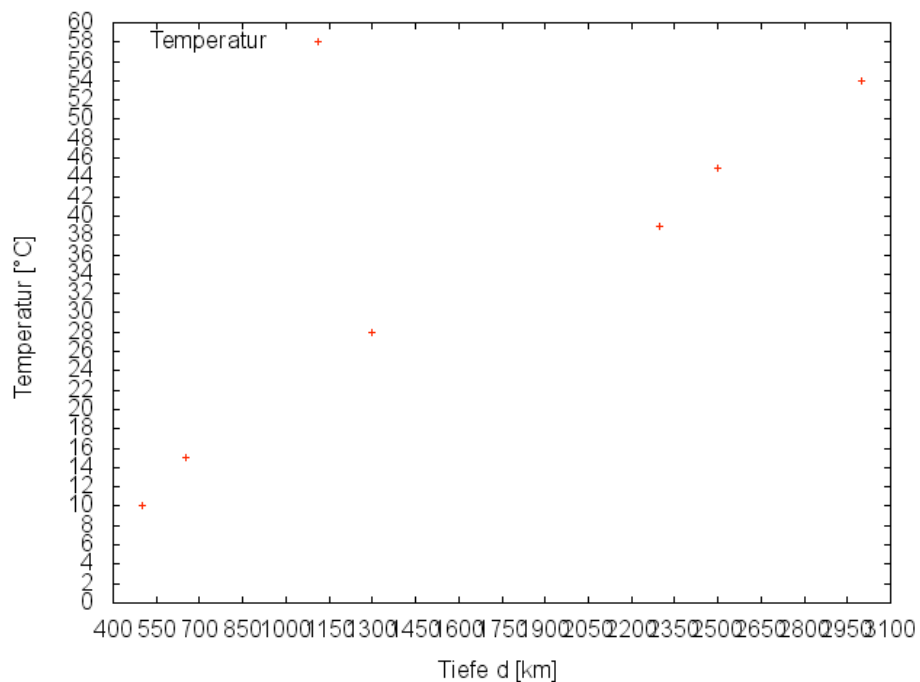


Abbildung 3: Beispiel für Diagramm mit offensichtlichen Fehlern. Erstellt mit gnuplot.

b) Gutes Beispiel

Abbildung 4 ist ein Beispiel für ein gutes Diagramm. Die Vorgaben wurden eingehalten, alle Elemente sind groß genug und leserlich, der Teilstrichabstand beider Achsen ist sinnvoll gewählt, beide Achsen sind beschriftet und die Legende ist klar von den Daten getrennt. Es wurde eine lineare Regression der Messwerte durchgeführt und die Gleichung sowie das Bestimmtheitsmaß angegeben. Die Regressionsgerade verläuft nicht genau durch die Punkte, ist aber meistens noch innerhalb der Fehlergrenzen. Es ergibt sich eine relativ gute Korrelation der beiden Variablen ($R^2 = 0.83$).

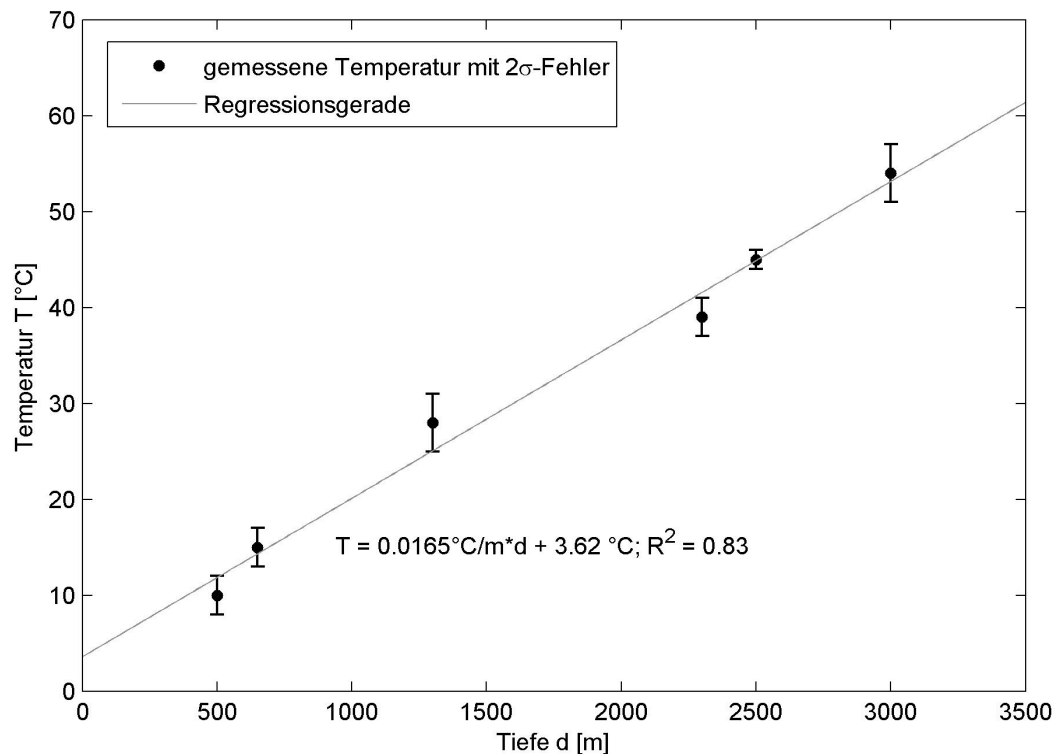


Abbildung 4: Beispiel für ein gutes Diagramm. Erstellt mit MATLAB.

7. Programme

Es gibt eine Vielzahl verschiedener Computerprogramme, mit denen man wissenschaftliche Diagramme erstellen kann. Je weiter Ihr Studium fortschreitet, desto mehr Diagramme müssen Sie für Hausaufgaben und Berichte und später für die Bachelor-/Master-/Doktorarbeit und Publikationen erstellen. Übung im Erstellen von Diagrammen und Kenntnis eines geeigneten Programms erleichtern diese Arbeit und lassen mehr Zeit für den wissenschaftlichen Teil Ihrer Arbeit (Zeit ist vor allem bei Ihrer Bachelorarbeit ein entscheidender Faktor!). Neben reinen Plot-Programmen verfügen auch Computeralgebrasysteme über gute Fähigkeiten zur Erstellung von wissenschaftlichen Diagrammen. Diese haben zusätzlich noch die Möglichkeit, Berechnungen durchzuführen und die Ergebnisse graphisch darzustellen und diese Vorgänge durch Skripte zu automatisieren. Im Folgenden werden mehrere dieser Programme vorgestellt und die Vor- und Nachteile aufgezählt. Die Wahl des Programms bleibt Ihnen überlassen. Es lohnt sich jedoch, einige dieser Programme zu testen und das Programm zu finden, das einem am meisten zusagt.

- **MATLAB**
MATLAB ist das in den Geowissenschaften am weitesten verbreitete Rechenprogramm (siehe Trauth, 2010). Es zeichnet sich vor allem durch seine effiziente Rechenleistung, die starke Erweiterungsfähigkeit durch Module und eine ausgedehnte, vielseitige Skriptsprache aus. Das Programm ist matrixbasiert und vor allem für numerische Berechnung gedacht. Mit der Symbol Math Toolbox ist allerdings auch symbolische Algebra möglich. MATLAB verfügt über sehr komfortable Plot-Möglichkeiten für Funktionen und diskrete Daten. Die Bedienung ist einfach zu erlernen und kann zum Großteil über die graphische Benutzeroberfläche erfolgen. Besonders praktisch ist auch der Export als *.ai oder *.eps mit dem das fertige Diagramm in Vektorgraphikprogrammen weiter bearbeitet werden kann. Für Studenten der TUM gibt es kostenlose Lizenzen (<https://matlab.rbg.tum.de>). MATLAB läuft auf Windows, MacOS und Linux (über XServer).
Dokumentation: <http://www.mathworks.de/help/techdoc/>
Anleitungen: <http://www.ma.tum.de/foswiki/pub/Ferienkurse/SoSe09/MatLab2706/vorlesungsskript.pdf>
<http://people.inf.ethz.ch/arbenz/MatlabKurs/matlabintro.pdf>
- **Veusz**
Kostenlose OpenSource-Software zum Erstellen von wissenschaftlichen Plots mit Publikations-Qualität (<http://home.gna.org/veusz>). Einfach zu bedienen und gut dokumentiert. Läuft auf Windows, Mac OS, Linux und Unix.
Dokumentation: <http://home.gna.org/veusz/docs/manual.pdf>
Einführungsvideo: <http://vimeo.com/14577084>
- **gnuplot**
Dieses OpenSource-Programm ist auf <http://www.gnuplot.info> verfügbar. Es ist befehlzeilenbasiert, allerdings existiert eine graphische Benutzeroberfläche, mit der einige Aktionen durchgeführt werden können.
Dokumentation: http://www.gnuplot.info/docs_4.6/gnuplot.pdf
Anleitungen: http://www.rz.uni-osnabrueck.de/Zum_Nachlesen/Skripte_Tutorials/Gnuplot_Einfuehrung/pdf/gnuplot.pdf
<http://userpage.fu-berlin.de/~voelker/gnuplotkurs/gnuplotkurs.html>
- **Graph**
Sehr einfach zu benutzendes, kostenloses Programm zum Plotten von Funktionen und Daten (<http://www.padowan.dk>). Die Standard-Einstellungen sind auf mathematische Anwendungen ausgelegt und muss für unsere Zwecke etwas angepasst werden (Achsen boxed etc.).
Dokumentation: <http://www.padowan.dk/bin/Graph-German.pdf>
- **PyXPlot**
An gnuplot angelehnte OpenSource-Software für Linux und Mac OS (<http://pyxplot.org.uk>). Die Steuerung erfolgt mehr auf der graphischen Benutzeroberfläche als bei gnuplot.
Dokumentation: <http://pyxplot.org.uk/0.8/doc/pyxplot.pdf>

- LabPlot
OpenSource-Programm basierend auf graphischer Benutzeroberfläche. Nur für Linux erhältlich (<http://labplot.sourceforge.net>).
Dokumentation: <http://prdownloads.sourceforge.net/labplot/LabPlot-1.6.0-de.pdf>
- Mathcad
Proprietäre Software, die komplett auf einer graphischen Oberfläche basiert (<http://de.ptc.com/product/mathcad>). Die Berechnungen finden auf Arbeitsblättern statt und alle Gleichungen werden in Standardnotation dargestellt. Mit diesem Programm lassen sich nicht nur Berechnungen und graphische Darstellungen gestalten, sondern alle Formeln und Abbildungen können in einen Text eingebettet werden. Dieses Programm ist vor allem bei Ingenieuren beliebt, da sich sehr einfach Dokumentationen erstellen lassen. Mathcad ist ein guter Kompromiss zwischen einfacher Bedienung und Rechenleistung. Allerdings kostet eine Studentenversion derzeit 100 Euro.

Die hier genannten Programme stellen nur eine kleine Auswahl dar. Es gibt viele andere kommerzielle und frei verfügbare Computeralgebrasysteme (Maple, Mathematica, Maxima, Sage etc.) sowie Plot-Programme (PLplot, HippoDraw, matplotlib, SciPy etc.).

8. Literaturhinweise

- Gustavii B (2008): How to Write and Illustrate a Scientific Paper. 2. Auflage, Cambridge University Press, 168 S.
- Marsal D (1979): Statistische Methoden für Erdwissenschaftler. 2. Auflage, E. Schweitzerbart'sche Verlagsbuchhandlung, 191 S.
- Schönwiese C-D (2006): Praktische Statistik für Meteorologen und Geowissenschaftler. 4. Auflage, Gebrüder Borntraeger, 302 S.
- Trauth MH (2010): MATLAB Recipes for Earth Sciences. 3. Auflage, Springer-Verlag, 336 S.

9. Anwendungsbeispiele

Die folgenden Beispiele mit MATLAB, gnuplot und Veusz sollen das Prinzip des Erstellens von Diagramm mit diesen Programmen veranschaulichen. Weitere Informationen finden Sie in den angegebenen Handbüchern und Anleitungen. Pfeile beziehen sich auf Schaltflächen (anklicken), der Rest sind Eingaben in der Befehlszeile.

- MATLAB
Öffnen des Ordners, in dem die Textdatei gespeichert ist:
Current Folder (linkes Fenster) → Ordner auswählen
Laden einer Textdatei:
load ,Beispiel.txt'
Plotten der Daten (x in erster Spalte, y in zweiter Spalte):
plot(Beispiel(:,1), Beispiel(:,2))
Plotten der Daten mit Fehlerbalken (x in erster Spalte, y in zweiter Spalte, y-Fehler in dritter Spalte):
errorbar(Beispiel(:,1), Beispiel(:,2), Beispiel(:,3))
Erstellen von Achsenbeschriftungen, Titel und Legende im Plot-Fenster:

Insert → *X Label/Y Label/ Title/Legend*

Verändern der Form und Farbe der Punkte und Entfernen der Linie:

Edit → *Current Object Properties...*, dann anklicken der Punkte

Line → *no line*

Color → Farbe auswählen

Marker → Symbol auswählen

Lineare Regression einfügen:

Tools → *Basic Fitting* → *linear* und → *show equations*

Speichern des Diagramms als MATLAB-Plot:

File → *Save as...* → **.fig*

Speichern des Diagramms als JPG:

File → *Save as...* → **.jpg*

Mit Exportoptionen (Größe, Auflösung):

File → *Export Setup...* → *Size (Width, Height)* und → *Rendering (Resolution)*

- gnuplot

Öffnen des Ordners, in dem die Textdatei gespeichert ist:

ChDir (Schaltfläche) → Ordner auswählen

Plotten der Daten aus Textdatei (x in erster Spalte, y in zweiter Spalte, y-Fehler in dritter Spalte) in schwarz mit einem Datensatz, der in der Legende als „Messdaten“ angezeigt werden soll:

plot 'Beispiel.txt' using 1:2:3 with errorbars title „Messdaten“ lt rgb „black“

Erstellen von Achsenbeschriftungen, Titel und Legende:

Axes → *X Label/Y Label*

Nach jeder Veränderung muss der Graph neu geplottet werden:

replot

Verändern der Abschnitte in denen x und y dargestellt werden:

Axes → *X Range/Y Range*

Verändern der Form und Größe der Punkte:

Style → *Data Style* → *Points*

set pointsize 3

Definieren der Gleichung für die lineare Regression:

$f(x) = m \cdot x + c$

Lineare Regression:

fit f(x) 'Beispiel.txt' using 1:2 via m,c

Plotten der Ausgleichsgeraden und der Datenpunkte zusammen:

plot 'Beispiel.txt' using 1:2:3 with errorbars title „Messdaten“ lt rgb „black“, f(x) title „Ausgleichsgerade“ lt rgb „grey“

Einstellen des Dateityps (png, leider kein JPG möglich):

set terminal png

Umleiten des Outputs an die Bilddatei:

set output „Beispiel.png“

Erneutes Plotten (diesmal nicht an den Bildschirm, sondern in die Bilddatei):

replot

- **Veusz**
 Erstellen eines neuen Datensatzes:
Insert → *Add xy*
 Importieren der Textdatei:
Data → *Import*
Browse → Ordner auswählen → Datei auswählen → *Open*
 Bei *Dataset names* eingeben: *x,y,+*
Import
Close
 Einlesen der x- und y-Daten:
Editing → *xy* anklicken
Properties → *x-Data* → *x*
Properties → *y-Data* → *y*
 Verändern der Form und Größe der Punkte:
Formatting → *Main Formatting* → *Marker/Marker size/Error style*
 Erstellen von Achsenbeschriftungen, Verändern der Abschnitte in denen x und y dargestellt werden:
Editing → *x*
Properties → *Label* → Name eingeben
Properties → *Min* → minimaler Wert
Properties → *Max* → maximaler Wert
Editing → *y*
Properties → *Label* → Name eingeben
Properties → *Min* → minimaler Wert
Properties → *Max* → maximaler Wert
 Entfernen der Linie:
Formatting → *Plot Line* → *Hide*
 lineare Regression:
Insert → *Add fit*
Properties → *X dataset* → *x*
Properties → *Y dataset* → *y*
Properties → *Fit function*
 Speichern des Diagramms:
File → *Export* → Ordner auswählen → Dateinamen eingeben → *Save*