

GChaos

November 6

2012

Gchaos ist eine Software die Verfahren der nichtlinearen
Zeitreihenanalyse zur Verfügung stellt. Die Software wird
seit 1992 von Guido Strunk entwickelt.

Einführung

1. Vorbemerkung

GChaos ist von Guido Strunk seit 1992 entwickelt worden, um neuere Verfahren der nichtlinearen Zeitreihenanalyse zu erproben. Die Software ist das tägliche Arbeitswerkzeug seines Autors und entspricht dabei vor allen seinen Vorstellungen und Anforderungen. Ob das Bedienkonzept auch für andere User sinnvoll erscheint ist weder getestet noch von vornherein mitbedacht worden.

Von GChaos existieren zwei Programmversionen, eine Arbeits- und eine Onlineausgabe. Die aktuelle „Arbeitsversion“ enthält zahlreiche zusätzliche Features, die jedoch noch nicht genügend getestet oder wahrscheinlich nicht von allgemeinem Interesse sind. Die „Onlineversion“ enthält derzeit z.B. keine Skriptsprachen-Funktionen, ist also nicht mit Hilfe einer Skriptsprache steuerbar. Diese Funktion soll jedoch in Zukunft von der „Arbeitsversion“ auf die „Onlineversion“ portiert werden.

GChaos beherrscht derzeit ca. 20 Verfahrensgruppen der statistischen Datenanalyse und der nichtlinearen Zeitreihenanalyse.

GChaos ist ein eigenständiges Analyseprogramm mit eigener Benutzeroberfläche

GChaos ist eine eigenständige Statistiksoftware mit eigener Benutzeroberfläche (Excel-like Tabelle mit derzeit maximal 300.000 Zeilen und 1.000 Spalten). Es verfügt über Funktionen zum Daten In- und Export, zur Generierung klassischer nichtlinearer Simulationsmodelle (Verhulst-System, Lorenz-System etc.), zur Datenmanipulation und -transformation, der Datenvisualisierung (Zeitreihen- und Phasenraumdiagramme sowie Recurrence Plots), Verfahren der Deskriptiv-, Regressions- und Inferenzstatistik, sowie Verfahren der nichtlinearen Zeitreihenanalyse. GChaos ist über Menü und Dialoge unter Windows steuerbar. Grafiken und Daten können auch über die Zwischenablage von Windows ausgetauscht werden.

Methoden der nichtlinearen Datenanalyse verarbeiten große Datenmengen und sind sehr rechenzeitintensiv. In der Regel werden mit Dialogen und *Progressbars* Rückmeldungen über den Status einer gestarteten Berechnung gegeben. Es kann aber durchaus sein, dass das Programm den Eindruck macht, nicht mehr zu reagieren, obwohl es tatsächlich rechnet. Auch lassen sich nur wenige Berechnungen nach einem Start wieder abbrechen. Berechnungszeiten von mehreren Stunden sind keine Seltenheit und man sollte zunächst mit kleinen Datensätzen probieren, bevor man aufs Ganze geht.

Die Funktionen von GChaos beruhen auf sorgfältigen Implementierungen von in der Literatur diskutierten Verfahren und auf verschiedenen eigenen Entwicklungen. Insbesondere nutzen einige Algorithmen selbst entwickelte Methoden zur Bestimmung gut passender Berechnungsparameter, die zwar in den meisten Fällen korrekte Parameter liefern, aber im Einzelfall auch daneben liegen können. Alle Parameter können von Hand gewählt werden, aber das erfordert mitunter einige Erfahrung.

Da die Methoden der nichtlinearen Zeitreihenanalyse erst in den letzten Jahren vorgeschlagen wurden, sind für viele Verfahren noch keine verbindlichen Standards über die Wahl von Berechnungsparametern ausgehandelt worden. Tatsäch-

lich finden sich sogar zahlreiche widersprüchliche Hinweise in der Literatur. GChaos entspricht an vielen Stellen den Erfahrungen des Autors, so dass manche Empfehlung aus der Literatur umgesetzt, andere aber abgelehnt wurden. Eine ausführliche Darstellung der verwendeten Algorithmen und der Verwendung von GChaos findet sich in: Strunk, Guido (2012) Die Messung von Komplexität in der Wirtschaftswissenschaft – Grundlagen, Methoden, Software und Beispiele. TU Dortmund.

Im Folgenden werden grundlegende Merkmale von GChaos beschrieben. Eine ausführliche Dokumentation ist derzeit in Arbeit.

2. Programmoberfläche

GChaos besitzt drei wesentliche Programmteile. Die Oberfläche des Hauptfensters stellt davon zwei bereit. Das ist zum einen der Dateneditor und zum anderen der Outputeditor. Der Dateneditor verwaltet die Daten. Berechnungsergebnisse werden im Outputeditor angezeigt.

Das Hauptfenster besitzt ein Menü und eine Symbolleiste. Menüpunkte und Symbolleistenbutton sind nur dann verfügbar, wenn der dazu passende Editor den Fokus besitzt, also dort der Cursor blinkt.

Um den Fokus zuzuweisen, ist also entweder in den Daten- oder den Outputeditor zu klicken.

Das dritte Fenster zeigt Grafiken von Zeitreihen und wird nur auf Anforderung geöffnet. Dieses Fenster verfügt über eine eigene Menüleiste und eine eigene Symbolleiste.

3. Datenimport

Im Dateneditor stehen frühe Daten oben und späte unten. Die Zeit verläuft von oben nach unten.

Der Dateneditor verarbeitet nur Zahlen. Texte lassen sich weder eingeben noch speichern. Auch Datumsangaben lassen sich nicht direkt in einer Datenspalte ablegen. Sollen Datumsangaben verarbeitet werden, müssen Spalten für Tag, Monat, Jahr, Stunde etc. getrennt definiert werden.

Daten können per Menü (Fokus im Dateneditor) geöffnet werden oder per Zwischenablage eingefügt werden.

Zwischenablage: Daten aus Excel, SPSS etc. können direkt per Zwischenablage in den Dateneditor eingefügt werden. Hier können nur Zahlen verarbeitet werden, aber keine Texte oder Beschriftungen der Daten-Spalten (Variablen). Unabhängig davon, wo der Cursor im Dateneditor steht, werden die Daten immer oben links startend eingefügt. Für das Einfügen der Daten gilt die Tastenkombination STRG-V oder Shift-Einfüg. Anders arbeitet das Kontextmenü des kleinen Editfeldes im Dateneditor. Dieses fügt die Daten in das Editfeld ein, verteilt sie aber nicht auf das Arbeitsblatt.

Datendateien können beliebige Endungen tragen, vorgesehen ist aber die Endung *.dat. Die Installationsroutine schlägt die Verknüpfung dieser Datenendung mit GChaos vor. Wurde diese Option gewählt, lassen sich Datendateien auch direkt durch einen Doppelklick im Explorer öffnen.

Datendateien sind einfache Text-Dateien, in denen die Daten spaltenweise durch Tabulatoren getrennt sind. Diese Dateien müssen in der ersten Zeile den Namen der Variable / Spalte enthalten.

Werden Daten im SPSS mit „Speichern unter“ als „Tabulator getrennt“ abgespeichert, so ist diese Datei direkt benutzbar.

Eine Beispieldatei findet sich im Programmverzeichnis.

Spaltenüberschriften können geändert werden. Rechter Mausklick auf die Spaltenüberschrift.

Beim Datenimport spielt es keine Rolle, ob die Daten deutsch (Dezimal-Trennzeichen ,) oder englisch (Dezimaltrennzeichen .) kodiert sind.

Durch Speichern des Datenblattes werden die Daten im passenden Format mit Spaltenüberschriften gespeichert.

4. Arbeit mit dem Dateneditor

4.1 Navigation

Im Dateneditor kann mit den Pfeiltasen, Bild hoch, Bild runter, Pos1, Ende und STRG+Pos1, STRG+Ende navigiert werden.

4.2 Dateneingabe und -änderung

Eine Dateneingabe und eine Veränderung von Daten ist im Dateneditor leicht möglich.

4.3 Datenorganisation

Daten sollten immer links oben starten. Zwischen benutzen Spalten sollten keine Leerspalten stehen.

4.4 Missing Data

Zellen in denen Daten fehlen werden bei der Eingabe leergelassen.

Bei einigen Berechnungsmethoden können zudem bestimmte Zahlenwerte als Missing definiert werden. Das ist aber nicht bei allen Berechnungsmethoden vorgesehen.

Einige Berechnungsmethoden verlangen vollständige Daten, so z.B. das D2/PD2.

4.5 Datenmanipulation direkt im Dateneditor

Klickt man mit der rechten Maustaste auf einen Spaltennamen wird ein Kontextmenü geöffnet, welches auch einige Berechnungsoptionen bietet. Die Manipulation der Daten erfolgt direkt in der jeweiligen Spalte.

Soll die Spalte aber erhalten bleiben kann man mit „Am Ende neu einfügen“ eine Kopie der Spalte ganz rechts im Datenblatt neu einfügen.

Es ist auch möglich in den „Settings“ eine Option zu setzen, die dazu führt, dass die per Kontextmenü durchgeführten Datenmanipulationen immer ans Ende des Datenblattes kopiert werden. Die Originaldaten bleiben dann unverändert.

4.6 Datenauswahl

Ein Klick mit der linken Maustaste auf einen Spaltennamen wählt die entsprechende Spalte aus. Wird gleichzeitig STRG betätigt, können auch mehrere Spalten ausgewählt werden.

Einige vom Menü angebotenen Datenanalysen werden direkt für die Spalten ausgeführt, die ausgewählt worden sind. Sind keine Spalten ausgewählt wird entweder keine Aktion durchgeführt oder es öffnet sich ein Dialog, der detaillierte Einstellungen erlaubt. Das bedeutet, dass sich einige Menüpunkte unterschiedlich verhalten, je nachdem ob Spalten ausgewählt wurden.

4.7 Export von Daten über die Zwischenablage

Über das Menü „Bearbeiten“ können ausgewählte Spalten oder das gesamte Datenblatt in die Zwischenablage gelegt werden. Die Spaltennamen werden mitkopiert. Das Zahlenformat ist deutsch. Die Daten lassen sich direkt ins Excel einfügen.

4.8 Speichern von Daten im Excel-Format

Excel kann die Daten von GChaos direkt lesen. Wenn man die Endung der Datei in xls ändert, können die Daten direkt im Excel geöffnet werden. Sie bleiben aber einfache Textdateien, solange man Excel nicht explizit um eine Datenkonvertierung durchführt.

5. Berechnungen/Analysen/Auswertungen

5.1 Allgemeines

Einige Analysen im Menü beziehen sich auf zuvor ausgewählte Spalten. Es können dazu gleichzeitig mehrere Spalten markiert sein.

Andere Analysen im Menü erfordern eine Auswahl der zu analysierenden Variablen in einem Dialog.

Der Befehl „Deskriptive Statistik“ funktioniert für ausgewählte Spalten automatisch und erlaubt weitere Optionen, wenn keine Spalte ausgewählt ist. Nicht alle Optionen sind derzeit implementiert.

Berechnungsergebnisse werden im Outputeditor angezeigt. Das Zahlenformat ist hier in der Regel englisch.

Ergebnisse im Output-Editor lassen sich nur über die Zwischenablage sichern.

Dialoggesteuerte Analysen erlauben eine Auswahl von Variablen im Dialog. Je nach Verfahren können hier einzelne Variablen in einer Liste oder gleichzeitig mehrere Variablen ausgewählt werden. In der Regel sind zudem einige Berechnungsparameter zu wählen, bevor eine Analyse möglich wird.

5.2 Statistics – Inference Statistics

Das Menü „Inference Statistics“ enthält einige einfache inferenzstatistische Verfahren. Der sich jeweils öffnende Dialog ist in jedem Fall gleich. Zunächst muss entschieden werden, ob die zu vergleichenden Daten in verschiedenen Variablen abgelegt sind. Ist das der Fall, kann man eine Variable in der linken Listbox und eine in der rechten auswählen. Die Mittelwerte oder Häufigkeiten dieser beiden Variablen werden dann inferenzstatistisch verglichen. Oder man arbeitet mit einer Gruppenvariable, die zeilenweise markiert zu welcher Gruppe die Daten der Datenvariable gehören. Die Kodierung der Gruppen muss dann zusätzlich angegeben werden.

5.3 Statistics – Descriptive Statistics

Einfache deskriptive Kennwerte können durch den Menüpunkt „Descriptive Statistics“ angefordert werden. Nicht alle Optionen sind derzeit implementiert. Beide Untermenüs arbeiten mit Standardeinstellungen direkt für im Datenblatt ausgewählte Variablen. Sind keine Variablen vorgewählt, öffnet sich ein Dialog, der weitere Einstellungen erlaubt.

5.4 Statistics – Regression/Correlation

Unter dem Menüpunkt „Regression/Correlation“ stehen zahlreiche Regressions- und Korrelationsverfahren zur Verfügung. Diese sind zum Teil anspruchsvoll in ihrer Bedienung und Interpretation. Eine ausführliche Dokumentation ist in Arbeit.

5.5 Statistics – D2/PD2

Beim D2/PD2 handelt es sich um ein Verfahren der nichtlinearen Zeitreihenanalyse. Ausgehend von Überlegungen von Mandelbrot (1977) über fraktale Strukturen wurden von Grassberger und Procaccia (Grassberger & Procaccia 1983a, 1983b, 1983c) Verfahren vorgeschlagen, die die fraktale Dimensionalität einer Dynamik ermitteln können. Diese beruhen auf dem Korrelationsintegral und ergeben das D2 und das K2. Erweiterungen der Grundidee erlauben zudem eine zeitpunktbezogene Analyse PD2 bzw. PK2. Zahlreiche Optionen erlauben die Anwendung der genannten Verfahren und einige spezielle Tests zur Absicherung der Berechnungsergebnisse. Ein Kernstück der Algorithmen bildet die von Strunk vorgeschlagene Methode zur automatischen Identifikation von Skalierungs- und Sättigungsbereich.

Eine ausführliche Dokumentation ist enthalten in: Strunk, Guido (2012) Die Messung von Komplexität in der Wirtschaftswissenschaft – Grundlagen, Methoden, Software und Beispiele. TU Dortmund.

5.6 Statistics – Lyapunov Exponent

Der Lyapunov Exponent beziffert die Stärke des so genannten Schmetterlingseffektes. Es handelt sich hierbei also um ein Verfahren der nichtlinearen Zeitrei-

henanalyse. Drei Methoden zur Vermessung des Chaos sind implementiert (Wolf et al. 1985, Rosenstein et al. 1993, Kantz 1994).

Eine ausführliche Dokumentation ist enthalten in: Strunk, Guido (2012) Die Messung von Komplexität in der Wirtschaftswissenschaft – Grundlagen, Methoden, Software und Beispiele. TU Dortmund.

5.7 Statistics – Symbolic Dynamics

Hinter dem Menüpunkt „Symbolic Dynamics“ finden sich zahlreiche Methoden zur Bezifferung der Komplexität von Symbolfolgen. Wichtig ist hier z.B. die so genannte Permutationsentropie (Bandt & Pompe 2002), die klassische Informationsdefinition nach Shannon (1948) sowie die Gruppe der GEntropien. Es handelt sich hierbei um Verfahren der nichtlinearen Zeitreihenanalyse.

Eine ausführliche Dokumentation ist enthalten in: Strunk, Guido (2012) Die Messung von Komplexität in der Wirtschaftswissenschaft – Grundlagen, Methoden, Software und Beispiele. TU Dortmund.

5.8 Statistics – Grammar Complexity

Die Grammar Complexity geht davon aus, dass Datensätze, die sich komprimieren lassen weniger komplex sind als solche für die das nicht gilt. Es handelt sich hierbei um Verfahren der nichtlinearen Zeitreihenanalyse (Darstellungen finden sich in Ebeling & Jiménez-Montano 1980, Jiménez-Montano 1984).

Eine ausführliche Dokumentation ist enthalten in: Strunk, Guido (2012) Die Messung von Komplexität in der Wirtschaftswissenschaft – Grundlagen, Methoden, Software und Beispiele. TU Dortmund.

5.9 Statistics – Forward DFT

Klassische diskrete Fourier Transformation.

Eine ausführliche Dokumentation ist enthalten in: Strunk, Guido (2012) Die Messung von Komplexität in der Wirtschaftswissenschaft – Grundlagen, Methoden, Software und Beispiele. TU Dortmund.

5.10 Statistics – Time Lag

Drei Methoden zur Schätzung eines geeigneten Time Lag finden sich hinter diesem Menüpunkt. Man sollte darauf achten, beim generalisierten Korrelationsintegral zunächst nur wenige Zeitverschiebungen zu erproben, da die Rechenzeit sonst sehr lang werden kann (vgl. Grassberger & Procaccia 1983a, 1983b, Fraser & Swinney 1986, Liebert & Schuster 1989, Schiepek & Strunk 1994). Es handelt sich hierbei um Verfahren der nichtlinearen Zeitreihenanalyse.

Eine ausführliche Dokumentation ist enthalten in: Strunk, Guido (2012) Die Messung von Komplexität in der Wirtschaftswissenschaft – Grundlagen, Methoden, Software und Beispiele. TU Dortmund.

5.11 Statistics – Recurrence Plots

Recurrence Plots sind als graphische Darstellungen seit einiger Zeit bekannt. Relativ neu sind Quantifizierungen über die Eigenschaften dieser Darstellungen (Marwan 2003, 2006). Diese erlauben es die Komplexität einer Zeitreihe zu bestimmen. Es handelt sich hier um Methoden der nichtlinearen Zeitreihenanalyse.

Eine ausführliche Dokumentation ist enthalten in: Strunk, Guido (2012) Die Messung von Komplexität in der Wirtschaftswissenschaft – Grundlagen, Methoden, Software und Beispiele. TU Dortmund.

5.12 Statistics – BDS

Der BDS-Test beruht auf dem Korrelationsintegral (Grassberger & Procaccia 1983a, 1983b, 1983c) und testet eine Zeitreihe auf iid (Brock et al. 1987, Brock et al. 1996). Es handelt sich hier um Methoden der nichtlinearen Zeitreihenanalyse.

Eine ausführliche Dokumentation ist enthalten in: Strunk, Guido (2012) Die Messung von Komplexität in der Wirtschaftswissenschaft – Grundlagen, Methoden, Software und Beispiele. TU Dortmund.

5.13 Statistics – Polynom Fitting

Kleinster Quadrate Schätzer für ein Polynom mit maximal 8ter Ordnung.

6. Literatur

- Bandt C. & Pompe B. (2002) Permutation Entropy: A Natural Complexity Measure for Time Series. *Physical Review Letters*, 88 (17), 174102-1 - 174102-4
- Brock W. A., Dechert W. D. & Scheinkman J. (1987) *A Test for Independence Based on the Correlation Dimension, Working Paper No. 8702 (Revised Version, 1996: Brock, W.A., W.D. Dechert, J. Scheinkman & B. LeBaron)*. University of Wisconsin, University of Houston and University of Chicago, Wisconsin, Houston, Chicago
- Brock W. A., Dechert D., Scheinkman J. & LeBaron B. (1996) A test for independence based on the correlation dimension. *Econometric Reviews*, 15, 197–235
- Ebeling W. & Jiménez-Montano M. A. (1980) On Grammars, Complexity, and Information Measures of Biological Macromolecules. *Mathematical Biosciences*, 52, 53-71
- Fraser A. M. & Swinney H. (1986) Independent Coordinates from Strange Attractors from Mutual Information. *Physical Review A*, 33, 1134-1140
- Grassberger P. & Procaccia I. (1983a) On the Characterization of strange Attractors. *Physical Review Letters*, 50, 346-356
- Grassberger P. & Procaccia I. (1983b) Measuring the Strangeness of strange Attractors. *Physica D*, 9, 189-208
- Grassberger P. & Procaccia I. (1983c) Estimation of the Kolmogorov entropy from a chaotic signal. *Physical Review A*, 28, 2591-2593
- Jiménez-Montano M. A. (1984) On the Syntactic Structure of Protein Sequences and the Concept of Grammar Complexity. *Bulletin of Mathematical Biology*, 46, 641-659
- Kantz H. (1994) A Robust Method to Estimate Maximal Lyapunov Exponents of a Time Series. *Physical Letters A*, 185, 77-87
- Liebert W. & Schuster H. G. (1989) Proper Choice of Time Delay for the Analysis of Chaotic Time Series. *Physical Letters A*, 142, 107-111
- Mandelbrot B. B. (1977) *The Fractal Geometry of Nature*. Freeman, New York
- Marwan N. (2003) *Encounters With Neighbours. Current Developments of Concepts Based on Recurrence Plots and their Applications*. Institut für Physik. Fakultät Mathematik und Naturwissenschaften. Universität Potsdam, Potsdam
- Marwan N. (2006) *Quantification of Recurrence Plots (Recurrence Quantification Analysis)*. www.recurrence-plot.tk, abgefragt am: 02.02.2007.
- Rosenstein M. T., Collins J. J. & De Luca C. J. (1993) A Practical Method for Calculating Largest Lyapunov Exponents from Small Data Sets. *Physica D*, 65, 117
- Schiepek G. & Strunk G. (1994) *Dynamische Systeme. Grundlagen und Analysemethoden für Psychologen und Psychiater*. Asanger, Heidelberg
- Shannon C. E. (1948) A Mathematical Theory of Communication. *Bell System Technical Journal*, 27, 379-423 and 623-656
- Wolf A., Swift J. B., Swinney H. L. & Vastano J. A. (1985) Determining Lyapunov Exponents from a Time Series. *Physica D*, 16, 285-317