

Webbasierte, interaktive Methoden zur Visualisierung von Umwelt-Massendaten

Mandy Koch und Ingrid Rohloff

Universität Potsdam · Mandy.Koch@uni-potsdam.de

Short paper

Zusammenfassung

Effektive Visualisierungen sorgen dafür, dass sich massenhaft vorrätige Zeitreihendaten, wie sie oftmals in Umweltprojekten generiert werden, dem Nutzer verständlich erschließen. Besonders dann, wenn solche Daten der Öffentlichkeit zugänglich gemacht werden sollen, ist es unabdingbar, geeignete Präsentationsformen zu wählen. Zu diesem Zweck werden interaktive, webbasierte Visualisierungstechniken von Massendaten mittels der Webtools D3.js und Highcharts untersucht, adaptiert und evaluiert. Die Erkenntnisse dieser Arbeit können Umweltprojekten Grundlage und Anregung für grafikgestützte Informationsvermittlung im Internet bieten.

1 Einleitung

Die Aufnahme von Informationen mithilfe von Bildern und Grafiken fällt Menschen leichter als die Extraktion von Wissen aus reinem Text oder Datentabellen. Im Zeitalter des Internets spielt auch dort die grafische Präsentation von Daten, Sachverhalten und Zusammenhängen eine wichtige Rolle. Im Internet werden unzählige Visualisierungswerkzeuge mit einem breiten Spektrum an Funktionalitäten und verschiedenen Interaktivitätsniveaus angeboten. Viele dieser Tools sind sehr gut dokumentiert, werden durch andauernde Aktivitäten in Foren sowie Blogs unterstützt und stetig weiterentwickelt. Oftmals mangelt es jedoch an fachwissenschaftlicher Expertise.

Im Rahmen eines Studienprojektes im Masterprogramm Geoinformation und Visualisierung an der Universität Potsdam werden Visualisierungstechniken nutzerspezifisch untersucht, adaptiert und evaluiert. Im Fokus steht dabei die Optimierung der webbasierten Darstellung von massenhaft vorrätigen Zeitreihendaten in einer hohen temporalen Auflösung sowohl für Nutzer mit wissenschaftlichem Hintergrund als auch für die interessierte Öffentlichkeit. Ein Testdatensatz, generiert aus frei verfügbaren TERENO-Messdaten, wird hierfür verwendet. Es handelt sich dabei um Messwerte verschiedener Umweltparameter, wie Lufttemperatur, Bodenfeuchte, Niederschlag etc., die im Zeitraum von einem Jahr mit einer Wertausprägung je 15 Minuten zur Verfügung stehen. TERENO ist ein Forschungsprojekt, an dem sich unter anderem das Deutsche GeoForschungsZentrum Potsdam beteiligt.

2 Visualisierungsmethoden

Die Methoden zur Visualisierung müssen zu einem expressiven, effektiven und angemessenen Ergebnis (SCHUMANN & MÜLLER 2000) führen. Mit zunehmender Datendimensionalität wird die Erfüllung dieser Kriterien komplizierter. Durch eine steigende Parameteranzahl verlieren z. B. Liniendiagramme an Effektivität. Sie finden daher Anwendung in Form eines Zeitreihendiagramms (vgl. Abb. 2) für die Darstellung einer oder zweier Messgrößen. Diese Visualisierungsmethode verfügt über zwei Diagrammsegmente, wobei das untere eine Überblicksfunktion bietet. Dort kann der Nutzer einen Datenbereich auswählen, der im oberen Diagrammteil detailliert dargestellt wird. Der selektierte Ausschnitt kann beliebig vergrößert, verkleinert und verschoben werden. Über eine Mouse-Over-Funktion werden konkrete Wertausprägungen angezeigt.

Weiterhin wird ein Liniendiagramm als mehrachsige Variante für die Darstellung von drei bis vier Parametern bzw. Parametergruppen aufbereitet. Sofern mehrere Messgrößen über die gleiche Einheit und Achsenskalierung verfügen, können sie im Multi-Achsen-Diagramm (vgl. Abb. 3) als Parametergruppe zusammengefasst werden. Neben der linienhaften Darstellung werden die Messgrößen je nach semantischem Bezug in Einzelfällen auf Balken oder Symbole referenziert. Jeder Parameter verweist über seine Linien-/Balkenfarbe auf die zugehörige Achse. Eine Mouse-Over-Funktion sowie die Möglichkeit, einzelne Parameter über die Legende in der Darstellung ab- bzw. anzuwählen, sind ebenfalls integriert.

Bei einer simultanen Darstellung von mehr als fünf Parametern gleicher Einheit ist keine ausreichende Effektivität gegeben. Dem Nutzer wird daher über einen Parallel Coordinates Plot (PCP) nach INSELBERG (1985) eine Datenvisualisierung angeboten. Die zwingende Orthogonalität eines kartesischen Koordinatensystems wird dabei aufgegeben, sodass eine parallele Achsenanordnung möglich ist. Für einen d -dimensionalen Datenvektor werden d Achsen erzeugt. Die Messwerte werden an den Achsen abgetragen und mit Linien verbunden (vgl. Abb. 1). Der PCP bietet durch Einschränkungsoptionen der Wertebereiche und Veränderung der Achsenanordnung Interaktionsmöglichkeiten.

Durch die Zuweisung verschiedener Diagrammtypen in Abhängigkeit der darzustellenden Messgrößenanzahl existiert für jede Parametermenge ein effektives Visualisierungsergebnis. Auch führen die Visualisierungsmethoden zu einem angemessenen Ergebnis. Nach einmaligem Programmieraufwand werden bei jeder Nutzeranfrage vollautomatisch Visualisierungen erzeugt. Weiterhin ist bei den Darstellungen auch das Expressivitätskriterium erfüllt. Abgebildet werden lediglich in den Ausgangsdaten enthaltene Informationen.

3 Visualisierungskonzept

Um für die jeweilige Nutzeranfrage die sinnvollste Darstellungsmethode auszuwählen, ist die Erstellung eines widerspruchsfreien Visualisierungskonzepts unabdingbar. Ziel dabei ist es, durch möglichst einfach verständliche, stufenweise aufgebaute Entscheidungsabfragen des Nutzers zur passenden Visualisierung zu gelangen.

Im ersten Schritt wird die Entscheidung gefällt, ob alle Parameter oder eine Auswahl dargestellt werden. Ist Ersteres der Fall, folgt die Abfrage, ob eine Aggregation der Daten durchzuführen ist. Möchte man sich einen groben Überblick über die Daten verschaffen, kann

eine Datenbündelung zur schnelleren Erfassung und höheren Übersichtlichkeit beitragen. Ist dies gewünscht, wird die Art der Aggregation ausgewählt (z. B. Monat, Tag oder Stunde). Da die Aggregationsart die Auswahl des Zeitraums bestimmt, sind im nächsten Menüpunkt einzig dem Aggregationstyp entsprechende Zeitschritte wählbar. Als Ergebnis wird ein PCP mit jeweils einer Linie je Aggregationsschritt erzeugt. Entscheidet sich der Anwender gegen eine Zusammenfassung der Daten, ist eine Beschränkung des auswählbaren Zeitraums sinnvoll, um lange Ladezeiten zu vermeiden und die visuelle Erfassbarkeit einzelner Datenpunkte innerhalb des PCP zu gewährleisten.

Fällt die Nutzerentscheidung auf die Darstellung einer Parameterauswahl, folgen die Abfragen der Datenaggregation und ggf. ihrer Art sowie die Selektion der relevanten Messgrößen. Bei mehr als vier gewählten Parametern ergibt sich nach der Angabe des darzustellenden Zeitraums ein PCP mit je einer Linie je Aggregationsabschnitt. Beschränkt sich die Anzahl der gewählten Parameter bzw. Parametergruppen auf drei bis vier, wird nach Selektion des Darstellungszeitraums ein Multi-Achsen-Diagramm ausgegeben. Bei ein bis zwei Parametern erfolgt eine aggregierte Darstellung im gewünschten Zeitraum als Zeitreihendiagramm. Ohne Aggregation kann der Nutzer ein bis zwei Parameter auswählen und den interessierenden Zeitraum angeben. Als Ergebnis erhält er ein Zeitreihendiagramm mit einem Datenpunkt je 15 Minuten.

4 Visualisierungswerkzeuge

Für eine effiziente, webbasierte Umsetzung des vorgestellten Visualisierungskonzepts wird ein möglichst optimal geeignetes Werkzeug benötigt. Hierfür werden folgende Anforderungen an einen generischen Ansatz definiert:

1. Die Datenvisualisierung soll für viele Projekte nutzbar sein. Deshalb werden frei verfügbare und quelloffene Webwerkzeuge verwendet. Der Einsatz proprietärer Software schließt eventuell Projekte ohne ausreichende finanzielle Mittel aus.
2. Die Unterstützung von ASCII-Formaten ist relevant, da Messdaten i. d. R. in dieser Form vorliegen.
3. Zusätzlich sollen auch über einen Sensor Observation Service (SOS) abgerufene Daten verarbeitet werden können.
4. Die Visualisierungsergebnisse sollen den Kriterien einer effektiven Darstellung genügen. Mithilfe des Tools werden dem Nutzer Interaktionsmöglichkeiten und Exportfunktionen (von Daten und Visualisierungen) angeboten.
5. Das Visualisierungslayout soll möglichst an verschiedene Projekte anpassbar sein.

Anhand dieses Kriterienkatalogs werden aus zahlreichen im Internet verfügbaren Visualisierungstools D3.js und Highcharts für beispielhafte Umsetzungen ausgewählt.

4.1 D3.js

Bei diesem Werkzeug handelt es sich um eine JavaScript-Bibliothek für dynamische grafische Darstellungen digitaler Daten. Durch eine Integrierung in ein HTML-Dokument entstehen interaktive, browserbasierte Datenvisualisierungen. D3.js kann mit Stildateien (.css) arbeiten, die eine leichte und schnelle Änderung des Visualisierungsstils ermöglichen. Die

freie Bibliothek wird online von BOSTOCK (2013) zur Verfügung gestellt. Vorteilhaft ist, dass dieses Webtool sehr gut mit Beispielcodes und Videotutorials dokumentiert ist.

Der PCP (vgl. Abb. 1) ist in D3.js umgesetzt mithilfe der Skripte „brushing“, „Linking with Data Table“ und „Progressive Rendering and SlickGrid“ von SYNTAGMATIC (2015). Ausschnitte der Code-Beispiele sind neu kombiniert und sinnvoll ergänzt, um ein ansprechendes Ergebnis zu erzielen.

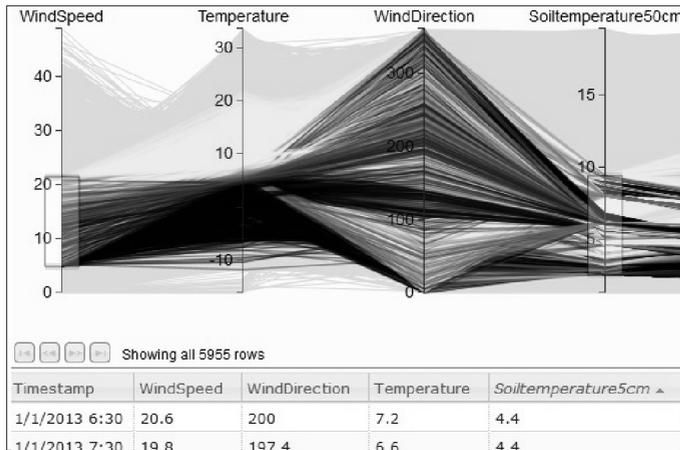


Abb. 1: Parallel Coordinates Plot (Quelle: eigener Entwurf)

Die Testdaten verfügen über ca. 28.000 Datensätze. Eine solche Liniendichte ist unübersichtlich und nicht optisch auswertbar. Es wird daher der Empfehlung von HEINRICH & WEISKOPF (2013) gefolgt und ein Farbverlauf implementiert. Der PCP bietet umfangreiche Interaktionsmöglichkeiten. Dazu gehören die Umkehrbarkeit und Verschiebung der Achsen sowie die Beschränkung der Wertebereiche. Unter dem Diagrammbereich werden in einer Tabelle alle im PCP selektierten Datensätze angezeigt. Eine weitere Interaktionsmöglichkeit wird durch Brushing geboten. Auch dies stellt eine Form von Datenselektion dar (HEINRICH & WEISKOPF 2013) und eignet sich, um in hyperdimensionalen Daten eine visuelle Verbindung zwischen der Grafik und den Ausgangsdaten zu schaffen (TURKAY et al. 2011). Grundsätzlich ist eine direkte Datenabfrage eines SOS umsetzbar. Ein entsprechendes Skript wird von 52°NORTH (2014) zur Verfügung gestellt.

4.2 Highcharts

Ähnlich wie D3.js ist Highcharts eine JavaScript-Bibliothek zur Erstellung von interaktiven Diagrammen für Webseiten sowie Webanwendungen. Für den persönlichen Gebrauch und die Nutzung innerhalb von Non-Profit-Projekten gilt die Creative Commons Attribution-NonCommercial 3.0 Lizenz. Beim Download von Highcharts (HIGHCHARTS 2014) werden Beispiele für alle unterstützten Diagrammtypen mitgeliefert, die an die jeweilige Fragestellung angepasst oder als Grundlage verwendet werden können.

Highcharts kann unter anderem Dateien im CSV-Format, in denen NoData-Values mit „null“ gekennzeichnet sind, verarbeiten. Ausgehend von bestehenden Code-Beispielen wer-

den auf Grundlage der vorhandenen Testdaten exemplarische Umsetzungen des Zeitreihendiagramms (vgl. Abb. 2) sowie des Multi-Achsen-Diagramms (vgl. Abb. 3) realisiert.

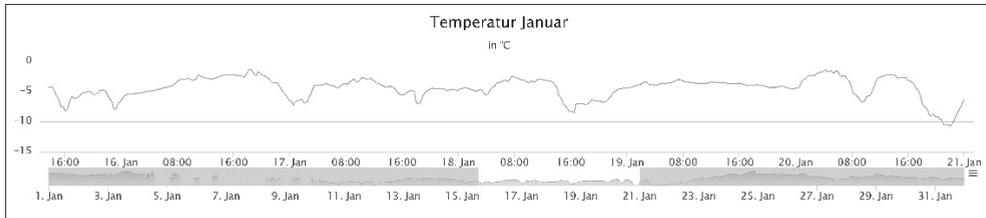


Abb. 2: Zeitreihendiagramm (Quelle: eigener Entwurf)

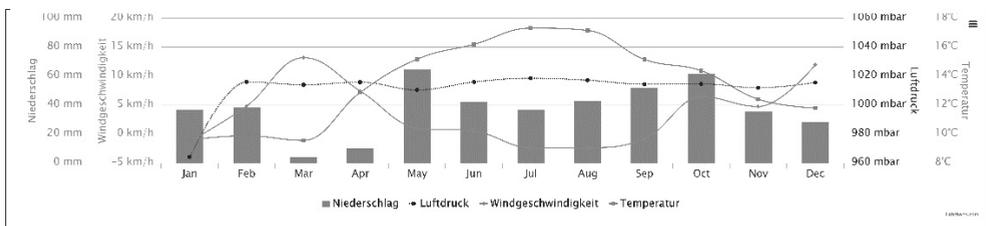


Abb. 3: Multi-Achsen-Diagramm (Quelle: eigener Entwurf)

Ein PCP sowie die Einbindung eines SOS werden von Highcharts per se bisher nicht angeboten. Jedoch können XML-Dateien verarbeitet werden, wodurch eine indirekte Kommunikation mit einem SOS möglich ist.

Besonders nutzerfreundlich sind die integrierbaren Exportfunktionen. Die Visualisierungen können über einen Downloadbutton in vier verschiedenen Formaten heruntergeladen werden (.png, .jpeg, .pdf, .svg). Diese Funktion ist in den Code-Beispielen von Highcharts bereits enthalten. Zusätzlich kann auch der Export der Daten (.xls oder .csv) implementiert werden. Des Weiteren wird das Arbeiten mit CSS-Dateien unterstützt. Die Möglichkeit zur unkomplizierten Anpassung an andere Projektlayouts ist somit gegeben.

4.3 Auswertung

Sowohl bei D3.js als auch bei Highcharts handelt es sich um frei verfügbare JavaScript-Bibliotheken, mit denen eine Vielzahl geeigneter Visualisierungen für Umweltmassendaten mit hohem Interaktivitätsgrad umsetzbar ist. Diese Werkzeuge sind sehr gut dokumentiert, unterliegen ständiger Weiterentwicklung und besitzen eine aktive Nutzergemeinschaft. Beide Visualisierungstools unterstützen die Arbeit mit Stylesheets (.css), womit die Anpassung des Layouts an spezielle Fragestellungen sowie nutzergruppen- und projektspezifische Vorhaben erleichtert wird. Die Verarbeitung von ASCII-Formaten ist in Form von CSV-Dateien möglich.

Ein großer Vorteil von Highcharts ist der einfach implementierbare Grafik- und Datenexport. Vor allem letzterer ist bei D3.js hingegen bislang noch problematisch. Dem gegenüber steht die Möglichkeit von D3.js, direkt Daten eines SOS abzurufen. Diese Kommunikation

kann im Fall von Highcharts bisher nur indirekt ermöglicht werden. Weiterer Handlungsbedarf besteht in der effektiven Darstellung der Visualisierungsergebnisse, denn diese sind derzeit bei beiden Tools abhängig vom gewählten Browser, den variierenden Bildschirmgrößen des Ausgabemediums (Desktop oder mobile Endgeräte) und der ansteigenden Ladezeiten bei zunehmenden Datenmengen.

5 Fazit

Die ausgewählten Visualisierungstools bieten zahlreiche Funktionalitäten, die der Optimierung von webbasierten Zeitreihendarstellungen dienen. Die mit ihnen umgesetzten Visualisierungen erfüllen die Kriterien der Effektivität, Expressivität und Angemessenheit in sehr hohem Maße. Auch bieten sie verschiedene Interaktionsmöglichkeiten. Ungeachtet dessen weisen sowohl Highcharts als auch D3.js individuelle Stärken und Schwächen auf. Eine Kombination beider Tools, zur Minimierung der jeweiligen Nachteile und zur Beibehaltung der inhärenten Stärken, ist sinnvoll.

Literatur

- 52°NORTH (2014), Sos-js. <http://52north.org/communities/sensorweb/> (03.12.2014).
- BOSTOCK, M. (2013), D3 – d3.js. <http://d3js.org/> (01.11.2014).
- HEINRICH, J. & WEISKOPF, D. (2013), State of the Art of Parallel Coordinates. In: EUROPEAN ASSOCIATION FOR COMPUTER GRAPHICS, Eurographics 2013.
- HIGHCHARTS (2014), Download. <http://www.highcharts.com/download> (01.11.2014).
- INSELBERG, A. (1985), The Plane with Parallel Coordinates. In: Visual Computer, 1 (4), 69-91.
- SCHUMANN, H. & MÜLLER, W. (2000), Visualisierung. Grundlagen und allgemeine Methoden. Springer-Verlag, Berlin/Heidelberg.
- SYNTAGMATIC (2015), Parallel-coordinates. <https://github.com/syntagmatic/parallel-coordinates> (18.02.2015).
- TURKAY, C., FILZMOSE, P. & HAUER, H. (2011), Brushing dimensions a dual visual analysis model for high-dimensional data. In: IEEE Transactions on Visualization and Computer Graphics, 12, 2591-2599.
- WEGMANN, E. J. (1990), Hyperdimensional Data Analysis Using Parallel Coordinates. Journal of the American Statistical Association, 85 (411), 664-675.