



# TIME SERIES DATA FORMATS AND LIBRARIES

KATHARINA HOVESTADT UND CAROLIN WORTMANN



# TIME SERIES

- Datenpunkte, die in einer bestimmten zeitlichen Ordnung vorliegen
- Unterscheidung zwischen diskret und kontinuierlich
  - kontinuierlich:  $y(t)$
  - diskret:  $y_t$  mit  $t$  aus  $\{1, \dots, n\}$  als zeitdiskreter stochastischer Prozess
- Zeitreihenanalyse :
  - Zweck:
    - Untersuchung und Verständnis der einer Datenreihe zugrunde liegenden Kräfte, Strukturen, Abhängigkeiten und Muster
    - Anpassung von Modellen an die Daten
    - Nutzung zu Vorhersage, Monitoring und Reaktion
  - Art der Analyse stark abhängig von zu untersuchenden Daten und Zweck der Analyse
  - Sehr unterschiedliche Anwendungsgebiete

→ z.B. Veranstaltung „Einführung in die Modellierung dynamischer räumlicher Prozesse“

# DATUMSFORMATE NACH XML-SCHEMA 1.1-2

- Perioden
  - `-?PnYnMnDTnHnMnS`
  - Beispiel: `PIY3M2DT3H4M20S` -> 1 Jahr, 3 Monate, 2 Tage, 3 Stunden, 4 Minuten, 20 Sekunden
- Zeitpunkte
  - Datum
  - Uhrzeit
  - `YYYY-MM-DDTHH:MM:SSZTimezone`
  - Beispiel: `2018-01-01T12:25:45+01.00` -> 1.1.2018 12 Uhr 25 Minuten 45 Sekunden (MEZ) Winterzeit in Deutschland

# VERSCHIEDENE DATENFORMATE ZUR SPEICHERUNG VON ZEITREIHEN

- Tabellarische Darstellung von Datums-/Zeitformaten
  - csv
- Orthogonale mehrdimensionale Arrays zur Darstellung von Zeitformaten
  - NetCDF
- OGC Geopackage
  - SQLite mit SpatiaLite
- ...

# Beispiel DATUMSANGABEN IN CSV DATEIEN (TABELLARISCHE DARSTELLUNG)

- Speicherung der Daten in Tabellen
- Erste Zeile beschreibt Tabellenkopf
- Trennzeichen zwischen den Zellen
- Datums- und Zeitdarstellung im xml-schema 11-2

Datum, Ort, Temperatur in Grad Celsius,  
2018-09-25T12:00, Rheine, 2.5,  
2018-09-25T12:00, Emden, 3.5,  
2018-09-27T12:00, Rheine, 3.5,  
2018-09-27T12:00, Emden, 5.0



Datum	Ort	Temperatur in Grad Celsius
25.09.2018 12:00 Uhr	Rheine	12.5
25.09.2018 12:00 Uhr	Emden	
25.09.2018 12:00 Uhr	Rheine	13.5
25.09.2018 12:00 Uhr	Emden	15.0

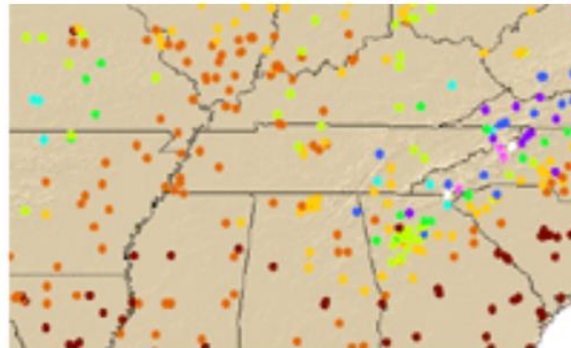
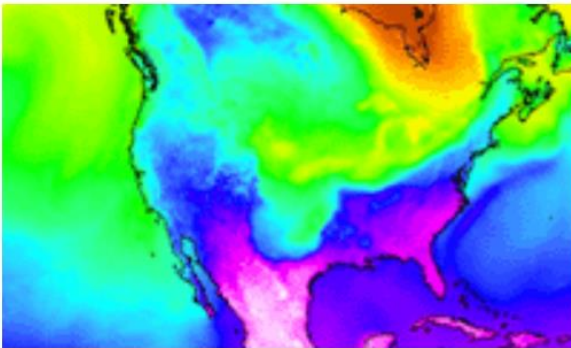
# NET CDF TIME SERIES

## ORTHOGONALE MEHRDIMENSIONALE ARRAYS ZUR DARSTELLUNG VON ZEITFORMATEN

- Standard im atmosphärischen und ozeanischen Bereich -> Windgeschwindigkeit, Wellenhöhe,...
- binäres Dateiformat
  - Header: Metadaten
  - Daten in Arrays
- Mehrdimensionale Arrays
  - Jede Achse enthält genau einen feature type
  - Nur eine Achse darf unendlich sein
- Zeitkoordinate wird in Sekunden seit einem Zeitpunkt gemessen
  - verschiedene Kalendertypen (näheres im Handout)
- praktisch, wenn die Daten entlang einer Achse dieselben Koordinaten aufweisen
  - Bsp: Temperaturmessungen zu fünf gleichen Zeitpunkten an drei verschiedenen Orten
    - Zeitachse hat dann fünf Elemente; es gibt 3 Instanzen (Orte)
- Wenn nicht zu allen Zeitpunkten Daten verfügbar sind, ist das Array unvollständig multidimensional

## NetCDF Time Series – Feature Types

- Verschiedene Feature Types:
  - **Point:** Einzelne Datenpunkte
  - **Time series:** Datenpunkte, gemessen am selben Ort, mit monoton steigender Zeit
  - **Trajectory:** Datenpunkt, gemessen entlang einer Strecke, mit monoton steigender Zeit
  - **Profile:** geordnetes Datenpunktpaket, gemessen entlang einer vertikalen Linie mit fixen horizontalen Positionen und einem bestimmten Zeitpunkt
  - Können kombiniert werden
  - Meistens: Rasterdaten entlang eines Zeitstrahls



*Beispiel für netCDF-Daten: Links – Temperatur; Rechts – Druck an bestimmten Punkten.*

Quelle: <https://pro.arcgis.com/de/pro-app/help/data/multidimensional/a-quick-tour-of-netcdf-data.htm>

# OGC GEOPACKAGE

- GeoPackages sind SQLite Datenbanken.
- Dateiendung .gpkg.
- enthalten verschiedene Tabellen
  - Metadaten
  - benutzerdefinierten Daten (Shapefiles)
- Features der Tabellen und Metadatatabellen können Zeitstempel haben (Zeitpunkt der letzten Änderung)
- Rasterdaten und Vektordaten können gespeichert werden



# VERGLEICH

- Arbeit mit CSV-Dateien in verschiedenen Kursen
  - leicht verständlich
- GeoPackage für Datenbanken-Fans hilfreich
  - SQLite kommt ohne Server aus
  - Ist darauf ausgelegt ist, wenig Ressourcen zu verbrauchen
    - Man kann durch die SQL Anweisungen auf fast jedem Endgerät flexibel Daten selektieren
  - Möglichkeit mittels Packages z.B. Daten aus Excel-Dateien zu speichern
  - Raster- und Vektordaten
- NetCDF arbeitet mit Arrays, Standard für atmosphärische und ozeanische Daten
  - In ArcGis allerdings immer nur Ausschnitt der Daten sichtbar
  - Verschiedene Messungen an verschiedenen Orte entlang einer Strecke zu verschiedenen Zeitpunkten in verschiedenen Höhen -> flexibel

# NUTZUNGSBEISPIELE FÜR DATUMS-/ZEITFORMATE

- Windenergiegewinnung in Deutschland zwischen 2000 und 2015

id	power	dt	hubheight	diameter	NUTS ID	lon	lat	N1
1	1500.0	01.12.2002	61.5	77.0	DE145	962.823.262	4.853.227.800. 000.000	DE1
2	1500.0	01.12.2002	61.5	77.0	DE145	963.619.308	4.853.290.663	DE1
3	750.0	07.07.1999	70.0	48.0	DE145	963.024.314	483.911.626	DE1
4	750.0	01.07.1999	70.0	48.0	DE145	963.005.164	4.839.274.454	DE1
5	2000.0	23.11.2012	138.38	82.0	DE145	97.591.964	4.844.863.777	DE1
6	2000.0	30.11.2012	138.38	82.0	DE145	976.413.263	4.844.806.099	DE1

- Daten wurden in CSV-Dateien gespeichert
- Stündliche Aufzeichnung

# LIBRARIES

- Nutzen: Vermeidung von Programmierung vieler Analyse- und Visualisierungsfunktionen
- Stellen Funktionen/Methoden zur Analyse und Verarbeitung von Zeitreihen zur Verfügung, zum Beispiel:
  - Generierung und Manipulation von Zeitreihen
  - Arbeiten mit Zeitpunkten und Perioden
  - Konversion und Anpassung von Daten an bestimmte Frequenzen
  - Statistische Analyse
  - Bestimmung von saisonaler und Trendkomponente sowie Zufallsschwankungen
  - Plotten
  - Sortierung
  - Bedingte Selektion von Daten
  - Für verschiedenste Programmiersprachen und Anwendungen

## Libraries

- Beispiele:
  - Python: Pandas
  - R: CRAN R – Time Series Analysis
  - C#: Cronos
  - Java: Morpheus

# PRAKTISCHE ERFAHRUNGEN

- SII: Assignment No. 5: Rest Services für Time series und queries → Abfragen von Daten eines REST Services in einem bestimmten Zeitraum
- Geostatistik I und II: Verarbeitung und Analyse von Zeitreihen und -daten in R
- Reference Systems: Temporal Data and Temporal Reference Systems (Sumrada, 2003) im Learnwebkurs verfügbar

DANKE FÜR EURE AUFMERKSAMKEIT!

- Gibt es Fragen?

# XML-SCHEMA I I-2

## VERSCHIEDENE DATENTYPEN

- duration
  - Zeitdauer
  - Konzept der Erfassung und Darstellung wurde von ISO8601 übernommen
  - Addition und Subtraktion werden unterstützt
  - Beispiel: P1Y2M3DT4H5M6S : 1 Jahr, 2 Monate, 3 Tage, 4 Stunden, 5 Minuten und 6 Sekunden
- yearMonthDuration
  - (-)PYYYYMM
  - P2018-10-25
- dayTimeDuration
  - (-)PHH:MM:SS
  - P12:02:05

## Xml-schema 1-2 verschiedene Datentypen

- `dateTime`
  - Zeitpunkt (mit Zeitzone)
  - Kein Wert, außer der Zeitzone darf fehlen
  - Beispiel: `2018-10-25T12:02:05+01:00` □ 25. Oktober 2018, 12:02Uhr und 5 Sekunden in der Zeitzone +1 □ MEZ (Winterzeit in Deutschland)
  - Angabe der Zeitzone ist freiwillig
  - `dateTimeStamp` → Zeitzone verpflichtend
- `time`
  - `HH:MM:SSZTimezone`
- `date`
  - `YYYY-MM-DDZTimezone`
- Extra Konventionen für Gregorianischen Kalender (wird jeweils ein g vorangestellt)
  - `gYear`, `gMonth`, `gDay`, `gMonthDay`, `gYearMonth`



# QUELLEN

- W3C Recommendations - Formats for Dates and Times: <https://www.w3.org/TR/2015/REC-tabular-data-model-20151217/#formats-for-dates-and-times>
- W3C - XML Schema Definition Language (XSD) - Datatypes: <https://www.w3.org/TR/2015/REC-tabular-data-model-20151217/#bib-xmlschema11-2>
- Orthogonal multidimensional array representation of time series: [http://cfconventions.org/cf-conventions/v1.6.0/cf-conventions.html#\\_orthogonal\\_multidimensional\\_array\\_representation\\_of\\_time\\_series](http://cfconventions.org/cf-conventions/v1.6.0/cf-conventions.html#_orthogonal_multidimensional_array_representation_of_time_series)
- OGC - GeoPackage Encoding Standard: <http://www.geopackage.org/spec120/>
- GeoPackage - Getting started with GeoPackage: <http://www.geopackage.org/guidance/getting-started.html>
- Wikipedia - GeoTIFF: <https://de.wikipedia.org/wiki/GeoTIFF>
- NetCDF - Features and Feature Type Conventions: [http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.html#\\_features\\_and\\_feature\\_types](http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.html#_features_and_feature_types)
- ArcGis - Beschreibung von NetCDF Dateien in ArcGIS: <https://pro.arcgis.com/de/pro-app/help/data/multidimensional/a-quick-tour-of-netcdf-data.htm>
- Engineering Statistics Handbook - Moving AVerage and Smoothing Techniques: <https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc42.htm>

## Quellen

- Time Series Analysis: <http://itfeature.com/time-series-analysis-and-forecasting/time-series-analysis-forecasting>
- Zeitreihenanalyse: <https://de.statista.com/statistik/lexikon/definition/144/zeitreihenanalyse/>
- W3 - Formats for Dates and Times: <https://www.w3.org/TR/2015/REC-tabular-data-model-20151217/#formats-for-dates-and-times>
- W3 – XML Schema 11-2: <https://www.w3.org/TR/2015/REC-tabular-data-model-20151217/#bib-xmlschema11-2>
- Windenergiegewinnung – Praxisbeispiel: [https://zenodo.org/record/1435091#.W6IGi\\_ZCQ2w](https://zenodo.org/record/1435091#.W6IGi_ZCQ2w)
- Pandas für Python: <http://pandas.pydata.org/>
- CRAN-R für R: <https://cran.r-project.org/web/views/TimeSeries.html>
- Cronos für C#: <https://archive.codeplex.com/?p=cronos>
- Morpheus für Java: <https://github.com/zavtech/morpheus-core/blob/master/README.md>