OPEN DATA SCIENCE EUROPE

WORKSHOP

Introduction to spatial and spatiotemporal data in R

Sept 6, 2021: 9:00 - 10:30

Tom Hengl

tom.hengl@opengeohub.org

https://opengeohub.org

Co-financed by the Connecting Europe Facility of the European Union

Programme

Day 1 Sept 6, 2021

9:00 - 10:30

Introduction to spatial and spatiotemporal data in R

By **Tomislav Hengl** Spatiotemporal Ensemble ML in R / computing with Cloud-Optimized GeoTIFFs

11:00 - 12:30

Modeling with spatial and spatiatemporal data in R

By **Tomislav Hengl** Spatiotemporal Ensemble ML in R / computing with Cloud-Optimized GeoTIFFs

13:30 - 15:00

Spatiotemporal Ensemble ML in R

By **Tomislav Hengl** Spatiotemporal Ensemble ML in R / computing with Cloud-Optimized GeoTIFFs

15:30 - 17:00

Computing with Cloud-Optimized GeoTIFFs

By **Tomislav Hengl** Spatiotemporal Ensemble ML in R / computing with Cloud-Optimized GeoTIFFs

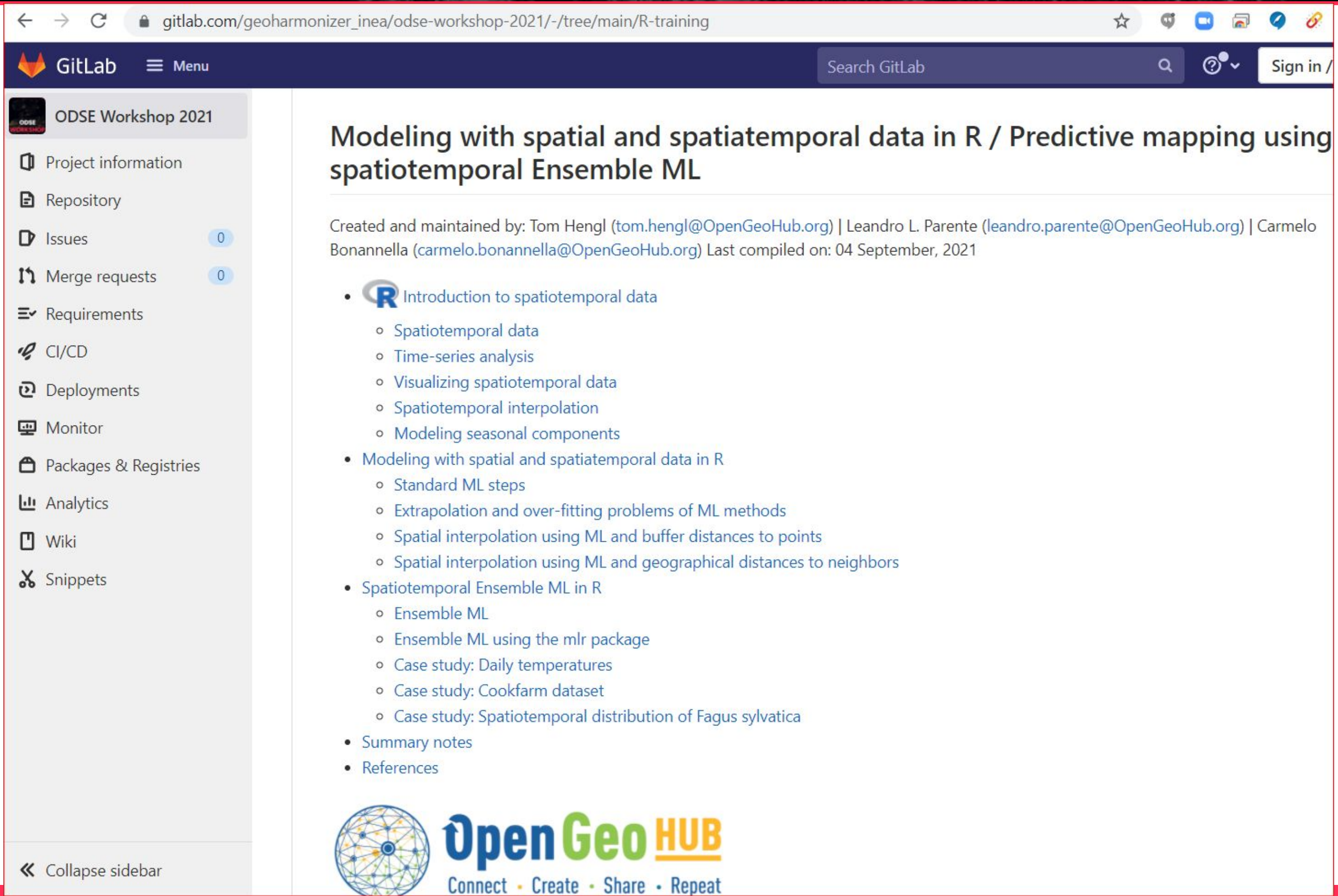
Day 2 Sept 7, 2021

9:00 - 10:30

Data visualization: from R to Google Earth and QGIS

By **Tomislav Hengl** Spatiotemporal Ensemble ML in R / computing with Cloud-Optimized GeoTIFFs

Rmarkdown notebook



gitlab.com/geoharmonizer_inea/odse-workshop-2021/-/tree/main/R-training

GitLab Menu Search GitLab Sign in /

ODSE Workshop 2021

- Project information
- Repository
- Issues 0
- Merge requests 0
- Requirements
- CI/CD
- Deployments
- Monitor
- Packages & Registries
- Analytics
- Wiki
- Snippets

Collapse sidebar

Modeling with spatial and spatiatemporal data in R / Predictive mapping using spatiotemporal Ensemble ML

Created and maintained by: Tom Hengl (tom.hengl@OpenGeoHub.org) | Leandro L. Parente (leandro.parente@OpenGeoHub.org) | Carmelo Bonannella (carmelo.bonannella@OpenGeoHub.org) Last compiled on: 04 September, 2021

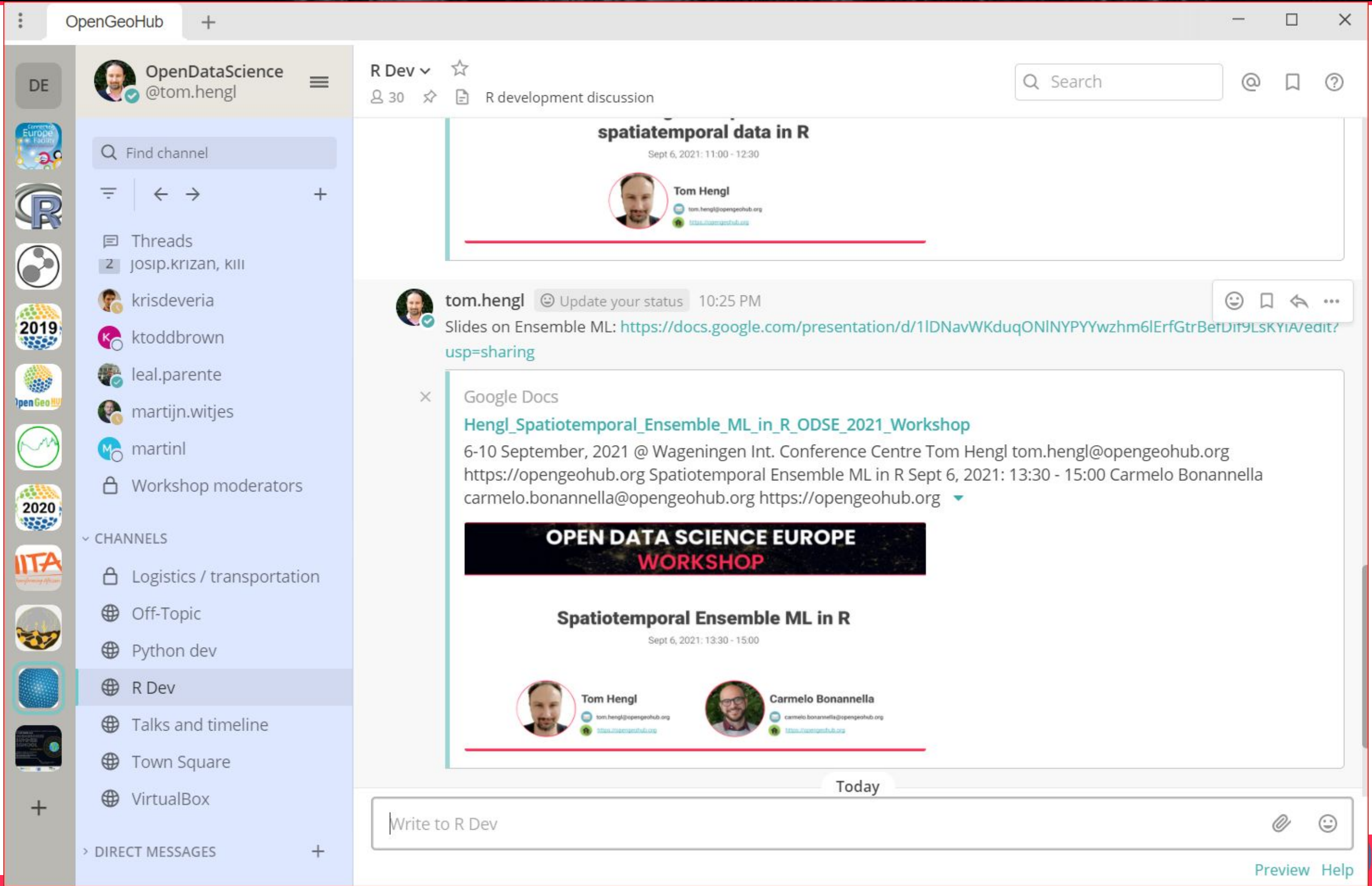
- Introduction to spatiotemporal data
  - Spatiotemporal data
  - Time-series analysis
  - Visualizing spatiotemporal data
  - Spatiotemporal interpolation
  - Modeling seasonal components
- Modeling with spatial and spatiatemporal data in R
  - Standard ML steps
  - Extrapolation and over-fitting problems of ML methods
  - Spatial interpolation using ML and buffer distances to points
  - Spatial interpolation using ML and geographical distances to neighbors
- Spatiotemporal Ensemble ML in R
  - Ensemble ML
  - Ensemble ML using the mlr package
  - Case study: Daily temperatures
  - Case study: Cookfarm dataset
  - Case study: Spatiotemporal distribution of Fagus sylvatica
- Summary notes
- References

OpenGeoHUB
Connect · Create · Share · Repeat

A lot of code, a lot of real-life datasets and examples

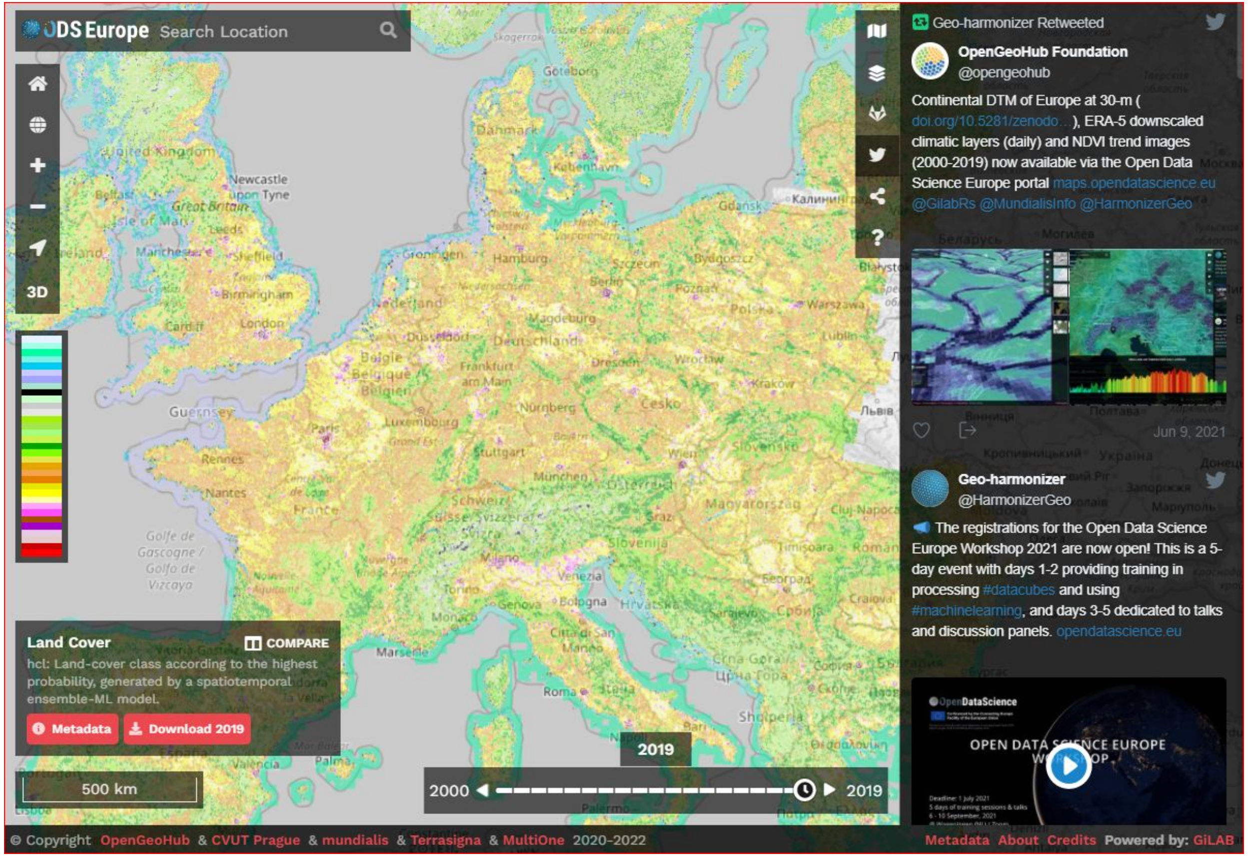
OpenDataScience

https://mattermost.opengeohub.org/opendatascience/channels/r-dev



With the right motivation, you can learn R in 1 day!

An Open Environmental Data Cube for Europe



https://maps.opendatascience.eu

Currently almost 10TB data available as Cloud-Optimized GeoTIFFs

Outline

- Spatial vs spatiotemporal data;
    - Key difference between the two,
    - Time dimension -> easy to plot, difficult to model (especially to predict the future!),
- Time-series analysis;
- Visualizing spatiotemporal data (some examples);

Spatiotemporal data

- geographic location (longitude and latitude or projected X, Y coordinates);
- spatial location accuracy or size of the block / volume in the case of bulking of samples;
- height above the ground surface (elevation);
- start and end time of measurement (year, month, day, hour, minute etc.);

Spatiotemporal data

Analysis of spatiotemporal data is somewhat different from pure spatial analysis.

Time is not just another spatial dimension i.e. it has specific properties and different statistical assumptions and methods have been developed for spatio- temporal data than for spatial data.

For an introduction to spatiotemporal data in R please refer to the **spacetime package** tutorial (Pebesma & others, 2012).

Spatiotemporal data

spacetime: Spatio-Temporal Data in R

ifgi
Institute for Geoinformatics
University of Münster

52north
exploring horizons

Edzer Pebesma

Abstract

This document describes classes and methods designed to deal with different types of spatio-temporal data in R implemented in the R package **spacetime**, and provides examples for analyzing them. It builds upon the classes and methods for spatial data from package **sp**, and for time series data from package **xts**. The goal is to cover a number of useful representations for spatio-temporal sensor data, and results from predicting (spatial and/or temporal interpolation or smoothing), aggregating, or subsetting them, and to represent trajectories. The goals of this paper are to explore how spatio-temporal data can be sensibly represented in classes, and to find out which analysis and visualisation methods are useful and feasible. We discuss the time series convention of representing time intervals by their starting time only. This vignette is the main reference for the R package `spacetime`; it been published as Pebesma (2012), but is kept up-to-date with the software.

Pebesma & Bivand: Spatial Data Science

keen-swartz-3146c4.netlify.app

Code

Spatial Data Science

with applications in R

Edzer Pebesma, Roger Bivand

2021-07-20

Preface

Data science is concerned with finding answers to questions on the basis of available data, and communicating that effort. Besides showing the results, this communication involves sharing the data used, but also exposing the path that led to the answers in a comprehensive and reproducible way. It also acknowledges the fact that available data may not be sufficient to answer questions, and that any answers are conditional on the data collection or sampling protocols employed.

Spacetime reference example in KML

```
<Placemark>
  <name>Bilogora</name>
  <TimeStamp>
    <begin>2008-02-17T00:00:00Z</begin>
    <end>2008-02-18T00:00:00Z</end>
  </TimeStamp>
  <Point>
   <coordinates>17.2057,45.8851,0</coordinates>
  </Point>
</Placemark>
```

Spatiotemporal data

Spatio-Temporal Data Types:

An Approach to Modeling and Querying

Moving Objects in Databases*

Martin Erwig [1]
Ralf Hartmut Güting [1]
Markus Schneider [1]
Michalis Vazirgiannis [1+2]

1) Praktische Informatik IV
Fernuniversität Hagen
D-58084 Hagen
GERMANY

2) Computer Science Division
Dept. of Electr. and Comp. Engineering
National Tech. University of Athens
Zographou, Athens, 15773
GREECE

Abstract: Spatio-temporal databases deal with geometries changing over time. In general, geometries cannot only change in discrete steps, but continuously, and we are talking about moving objects. If only the position in space of an object is relevant, then *moving point* is a basic abstraction; if also the extent is of interest, then the *moving region* abstraction captures moving as well as growing or shrinking regions. We propose a new line of research where moving points and moving regions are viewed as three-dimensional (2D space + time) or higher-dimensional entities whose structure and behaviour is captured by modeling them as abstract data types. Such types can be integrated as base (attribute) data types into relational, object-oriented, or other DBMS data models; they can be implemented as data blades, cartridges, etc. for extensible DBMSs. We expect these spatio-temporal data types to play a similarly fundamental role for spatio-temporal databases as spatial data types have played for spatial databases. The paper explains the approach and discusses several fundamental issues and questions related to it that need to be clarified before delving into specific designs of spatio-temporal algebras.

Erwig, M., Gu, R. H., Schneider, M., Vazirgiannis, M., & others. (1999). Spatio-temporal data types: An approach to modeling and querying moving objects in databases. *GeoInformatica*, *3*(3), 269–296. http://dx.doi.org/10.1023/A:1009805532638

Spatiotemporal data

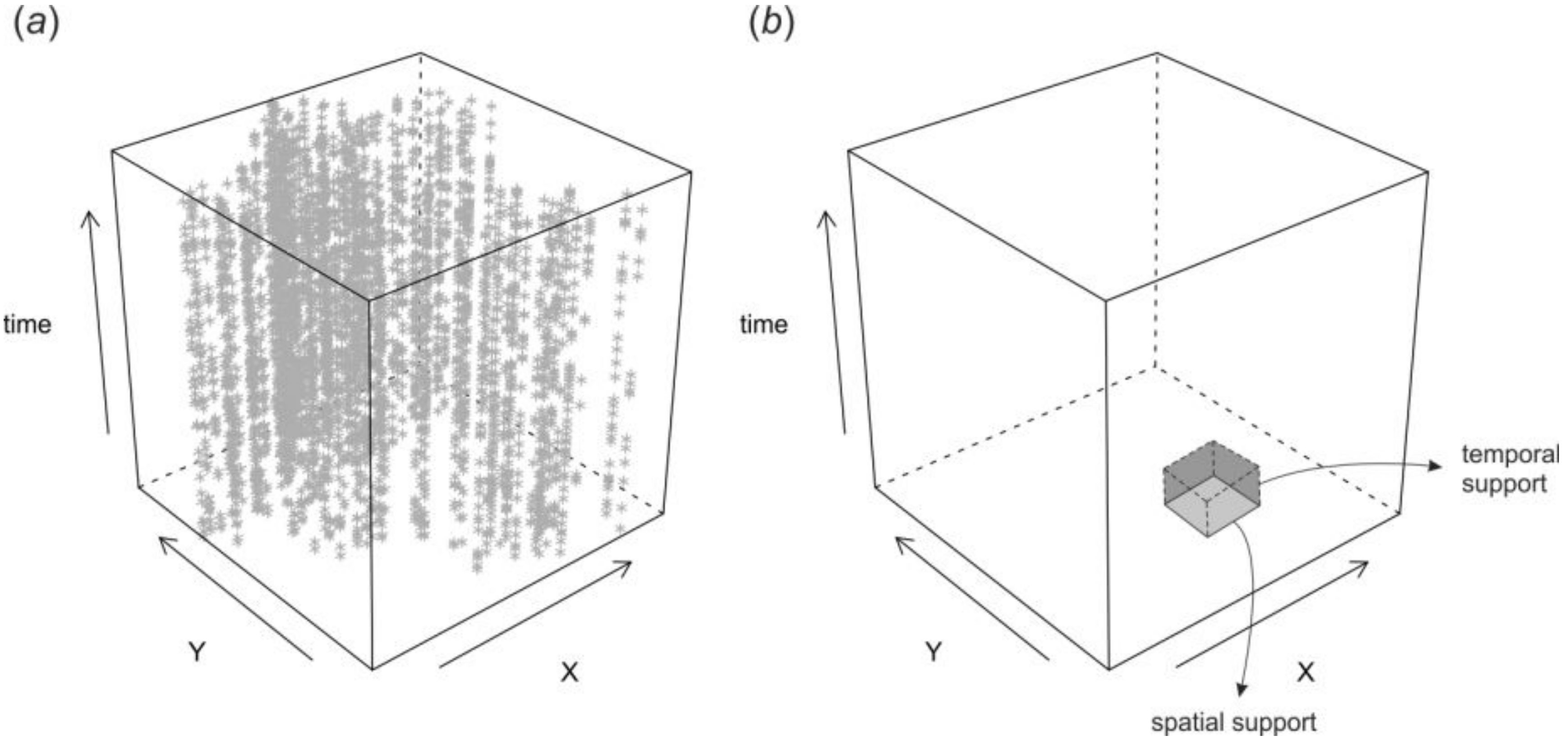
For ERWIG et al. (1999) spatio-temporal data sets and corresponding databases can be matched with the two major groups of features: (1) ***moving or dynamic objects*** *(discrete or vector geometries), and (2)* ***regions*** *(fields or continuous features).*

Fields / regions:

Usually one of the three:

1. **quantity or density of some material or chemical element**,
2. **energy flux** or any similar physical measurements,
3. **probability of occurrence of some feature or object**,

Spacetime cube (2D+T)



Spatiotemporal kriging

Spatio-temporal prediction of daily temperatures 275

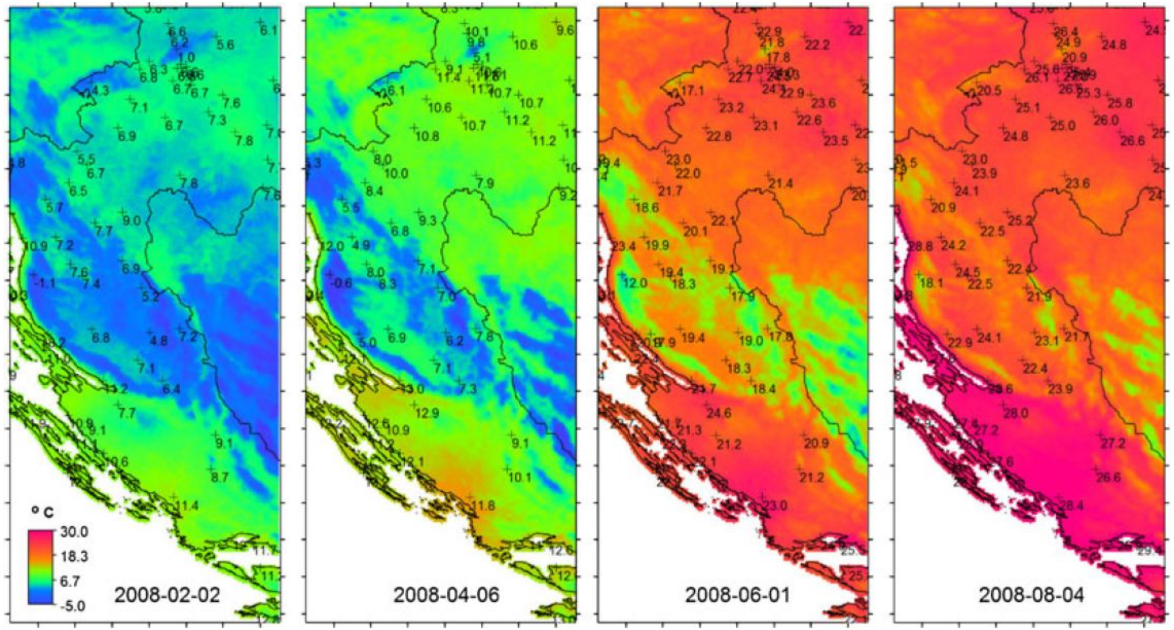


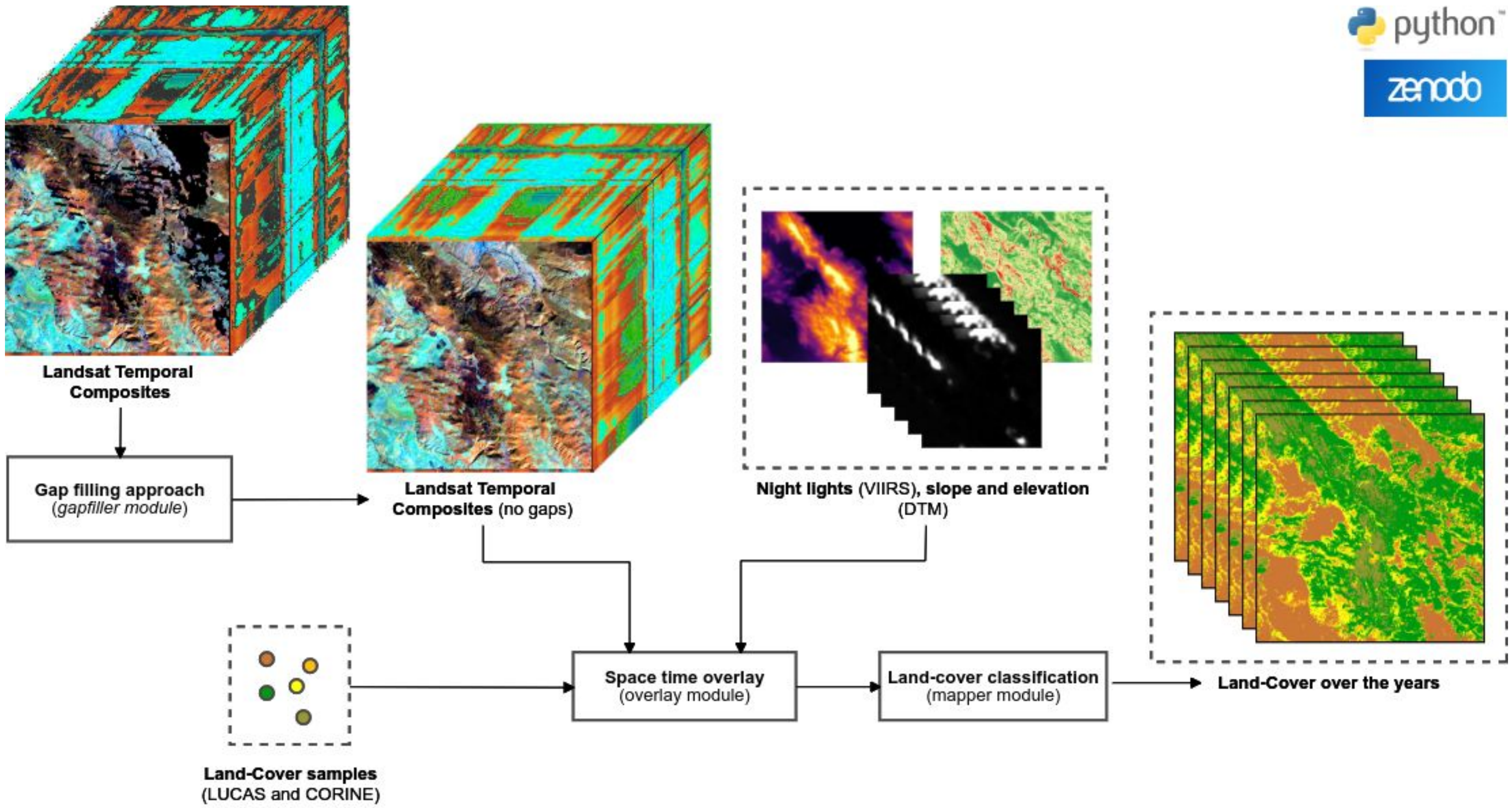
Fig. 8 Mean daily temperatures for four arbitrary dates predicted using spatio-temporal regression-kriging and actual observed values. Because the prediction model is significant (84% of variability explained by the model), it may be used to map space-time patterns in the neighboring countries

[krigeST function in the gstat package]() allows for producing spacetime predictions from point observations

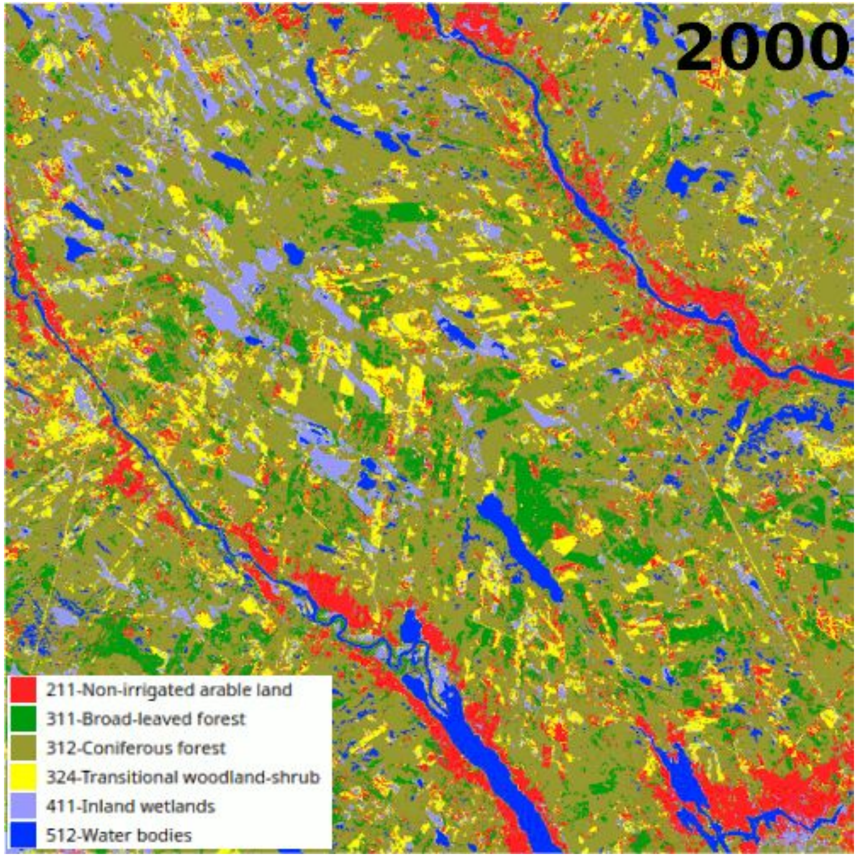
Source: [10.1007/s00704-011-0464-2]()

OpenDataScience

Spatiotemporal overlay



Spatiotemporal predictions



This animation shows the land-cover classes for an area located in Sweden (tile 22497) according to the space time predictions. This example is a small use case that used 680 point samples, obtained in different years, to train a single model and to predict the land-cover in the region over the time.

https://maps.opendatascience.eu

Spatiotemporal predictions

Research Square

Browse · Tools & Services · About · Sign In · + Submit a Preprint

METHOD ARTICLE

A spatiotemporal ensemble machine learning framework for generating land use / land cover time-series maps for Europe (2000 – 2019) based on LUCAS, CORINE and GLAD Landsat

Martijn Witjes, Leandro Parente, Chris J. van Diemen, Tomislav Hengl, Martin Landa, Lukas Brodsky, Lena Halounova, Josip Krizan, Luka Antonic, Codrina M Ilie, Vasile Craciunescu, Milan Kilibarda, Ognjen Antonijevic, Luka Glusica

DOI: 10.21203/rs.3.rs-561383/v1 Download PDF

DECLARATIONS: View author declarations.

Abstract

A seamless spatiotemporal machine learning framework for automated prediction, uncertainty assessment, and analysis of land use / land cover (LULC) dynamics is presented. The framework includes: (1) harmonization and preprocessing of high-resolution spatial and spatiotemporal covariate datasets (GLAD Landsat, NPP/VIIRS) including 5 million harmonized LUCAS and CORINE Land Cover-derived training samples, (2) model building based on spatial k-fold cross-validation and hyper-parameter optimization, (3) prediction of the most probable class, class probabilities and uncertainty per pixel, (4) LULC change analysis on time-series of produced maps. The spatiotemporal ensemble model was fitted by combining random forest, gradient boosted trees, and artificial neural network, with logistic regressor as meta-learner. The results show that the most important covariates for mapping LULC in Europe are: seasonal aggregates of Landsat green and near-infrared bands, multiple Landsat-derived spectral indices, and elevation. Spatial cross-validation of the model indicates

BADGES

Prescreen

HISTORY

CURRENT STATUS: **POSTED**

Version 1

Posted 27 May, 2021

No community comments so far

METRICS

SUBJECT AREAS

Artificial Intelligence and Machine Learning

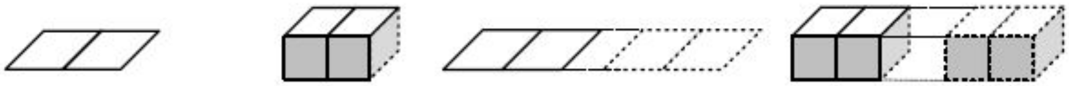
Geographic Information Systems

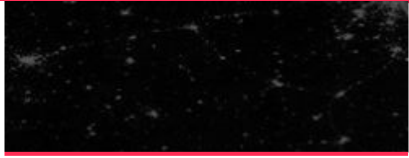
MORE FROM RESEARCH SQUARE

Jenny 5 - the robot

Multi Expression Programming

OpenDataScience

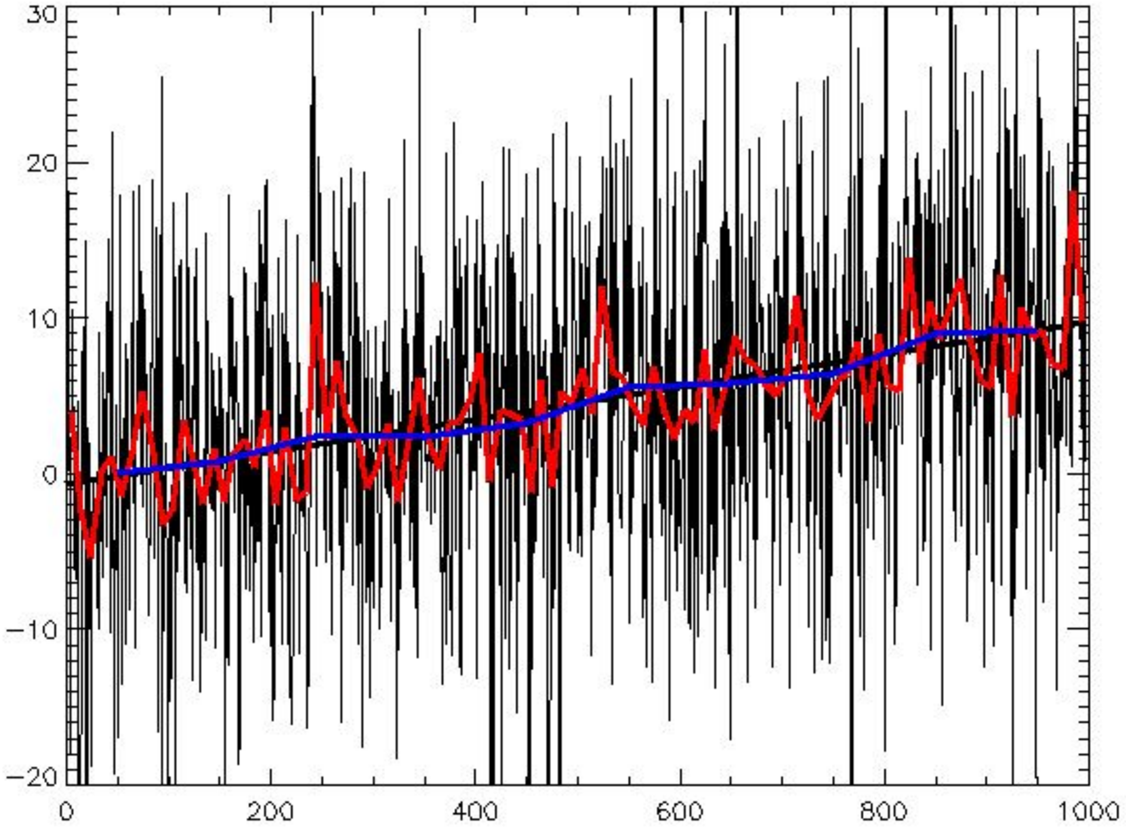




	2D	3D	2D+T	3D+T
longitude	✓	✓	✓	✓
latitude	✓	✓	✓	✓
altitude		✓		✓
time			✓	✓
minimum # of variogram parameters	3	4	4*	5*
# of prediction locations	N_s	$N_s \times N_d$	$N_s \times N_t$	$N_s \times N_d \times N_t$

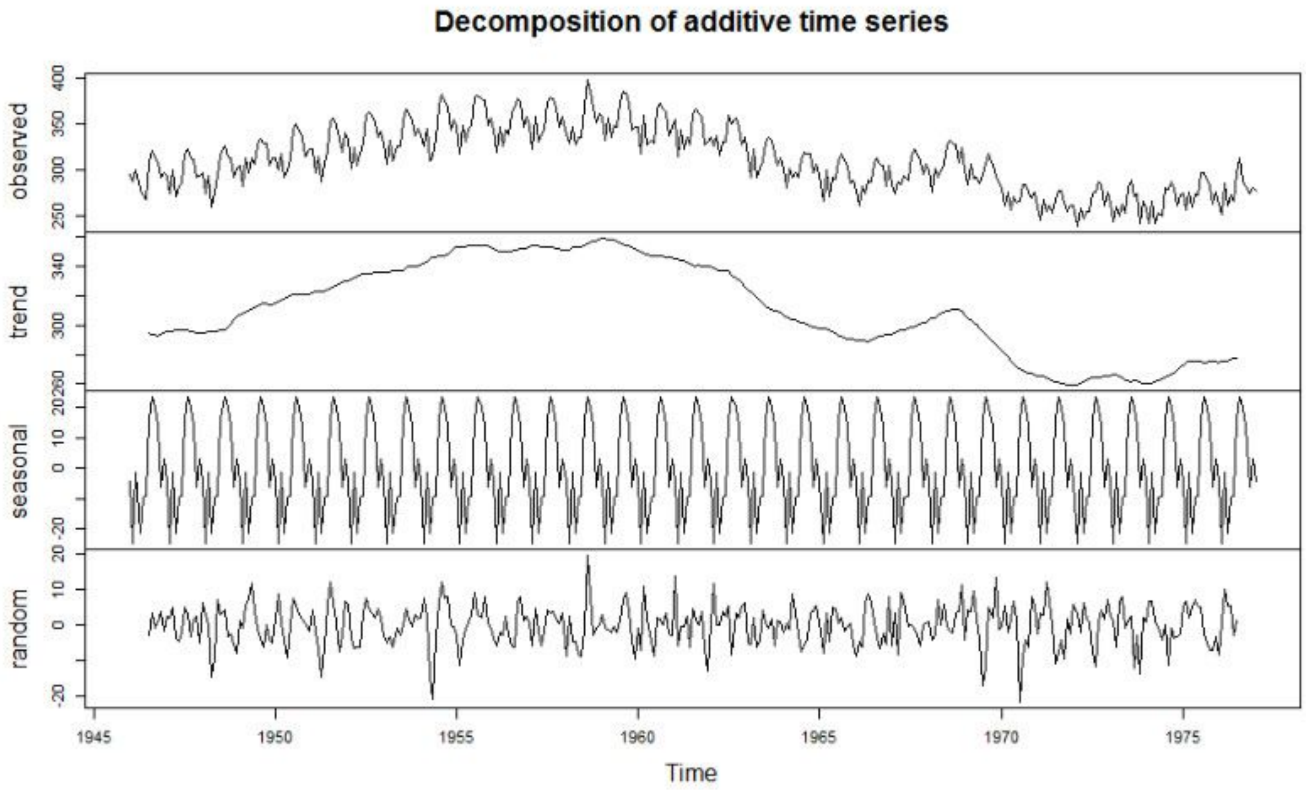
*Temporal anisotropy parameter

Time-series data



A time series is a series of data points indexed (or listed or graphed) in time order. Most commonly, a time series is a sequence taken at successive equally spaced points in time.
Source:
https://en.wikipedia.org/wiki/Time_series

Time-series components



Time-series (something very actual)

Number of confirmed cases over time (per 100,000 inhabitants)

This graph shows how many confirmed cases of coronavirus were reported per 100,000 inhabitants.

Show everything | Past 5 weeks

— Confirmed cases over time per 100

31 August 2020

All schools open again

All primary and secondary schools are fully open

ses per 100.000

O Uitgelichte gebeurtenis

Source: RIVM

Time-series components

Time-series (signal) can be decomposed into four groups of components:

1. Long-term component **(trend)** determined by long-term geological and extraterrestrial processes,
2. Seasonal monthly and/or daily component **(seasonality)** determined by Earth rotation and incoming sun radiation,
3. **Variation component** which can be due to semi-chaotic behavior and/or local factors (hence autocorrelated), and
4. **Pure noise** i.e. measurement errors and similar,

Long-term component: example global temperature

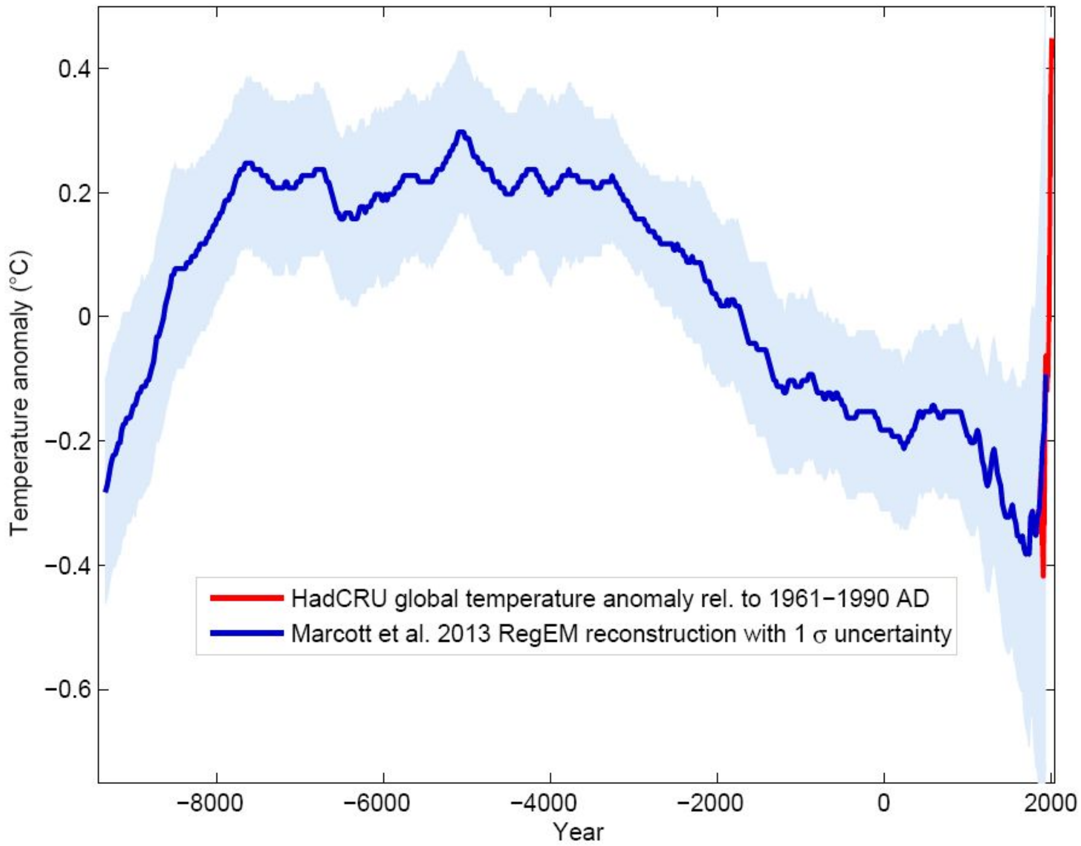


Figure: Global temperature reconstruction from proxy data of [Marcott, Shakun, Clark, & Mix (2013)](). This shows how global temperature varies on long-term term. Graph by: Klaus Bitterman.

Example of seasonality pattern

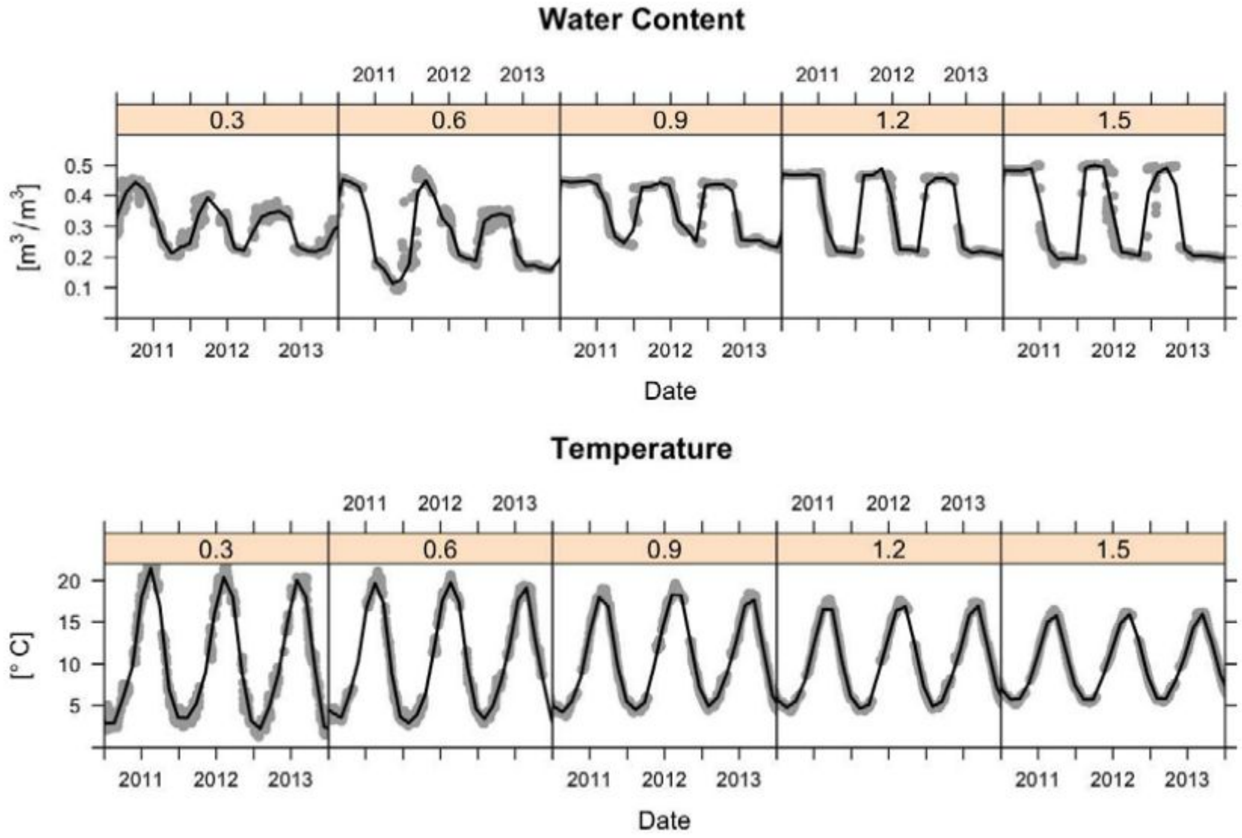


Figure: Sensor values from five depths (0.3, 0.6, 0.9, 1.2, and 1.5 m) at one station at Cook Agronomy Farm from January 2011–January 2014. The black line indicates locally fitted splines ([Gasch et al., 2015](#)).

Visualizing spatiotemporal data

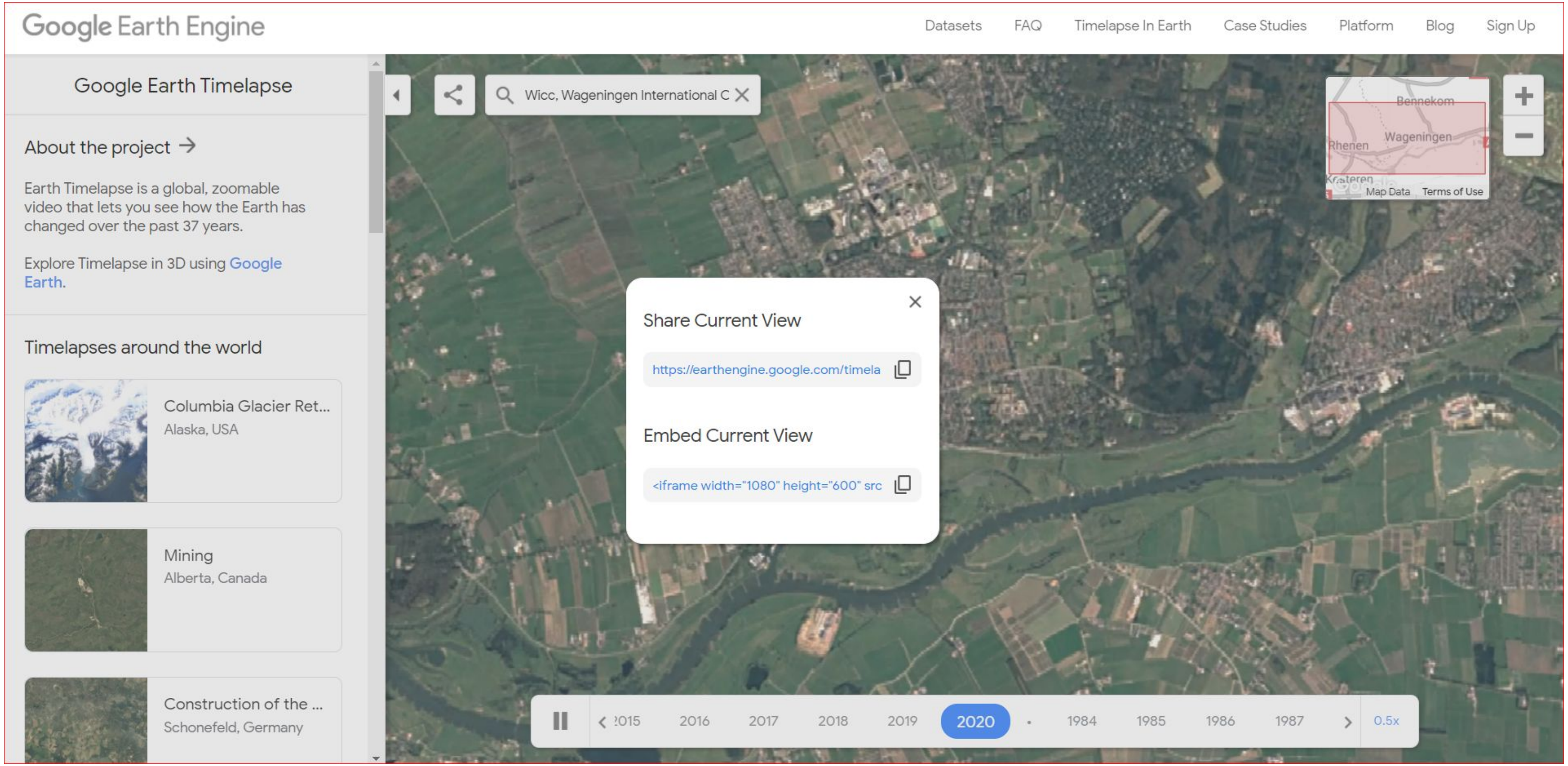
There are three possible groups of ways to visualize spatiotemporal data:

1. Using static images showing trend parameters together with time-series plots at selected representative point locations.
2. Using **time-slices** or series of visualizations of the same spatial domain but changing in time.
3. Using animations or interactive plots with time-sliders allowing users to choose speed and direction of animation.

Example: foot and mouth disease data

```r
library(plotKML)
data(fmd)
fmd0  <- data.frame(fmd)
coordinates(fmd0) <- c("X", "Y")
proj4string(fmd0) <- CRS("+init=epsg:27700")
fmd_sp <- as(fmd0, "SpatialPoints")
dates <- as.Date("2001-02-18") + fmd0$ReportedDay
library(spacetime)
fmd_ST <- STIDF(fmd_sp, dates, data.frame(ReportedDay=fmd0$ReportedDay))
data(SAGA_pal)
## Not run:
plotKML(fmd_ST, colour_scale=SAGA_pal[[1]], open.kml = FALSE)
```

Example: Google time-lapse



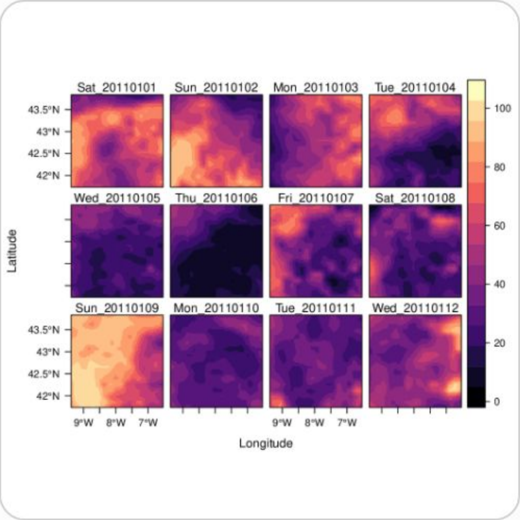
Visualizing spacetime data in R

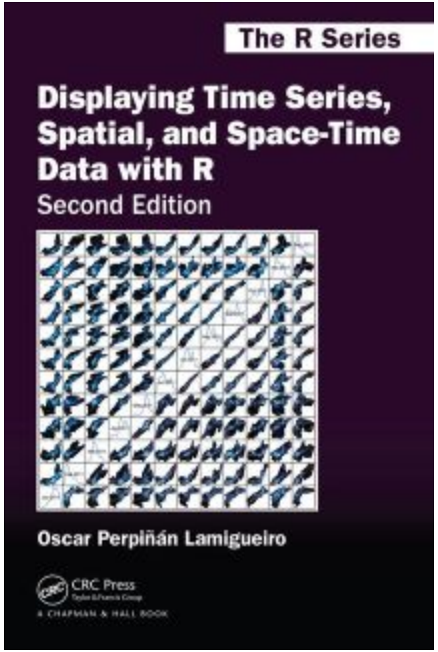
Table of Contents

Spatio-temporal data - Displaying time series, spatial and space-time data with R

Raster data

Level plots





Visualizing spacetime data in R: tmap

r-tmap.github.io/tmap-book/nutshell.html#quick-maps

Chapter 3 tmap in a nutshell

The **tmap** package allows the creation of thematic maps with great flexibility. It accepts spatial data in various formats shape objects (section 3.1) Next, the data can be used to create simple, quick maps (section 3.2) and more complex and expandable maps (section 3.3). These maps can be presented in two modes - as a static map and an interactive one. Additionally, **tmap** makes it possible to create small multiples map (section 3.5) and map animations (section 3.6).

3.1 Shape objects

As we established in chapter 2, spatial data comes in various file formats related to two main data models - vector and raster. There are also several spatial object classes in R, for example, `sf` from the **sf** package for vector data and `stars` from **stars** for raster data and spatial data cubes. Additionally, packages such as **sp**, **raster**, or **terra** offer their own classes, and this abundance of spatial object classes can be generally overwhelming. Gladly, **tmap** can work with all of the above objects - it treats all supported spatial data classes as so-called *shape objects*.

Animated maps (Geocomputation with R)

geocompr.robinlovelace.net/adv-map.html#animated-maps

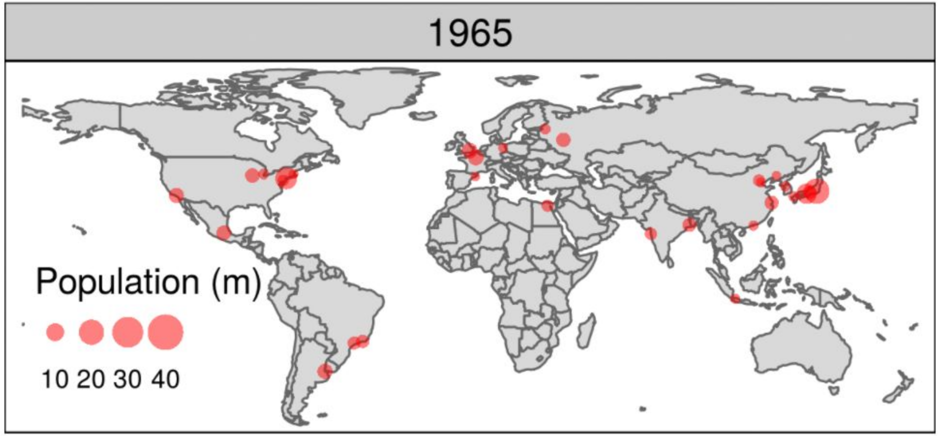
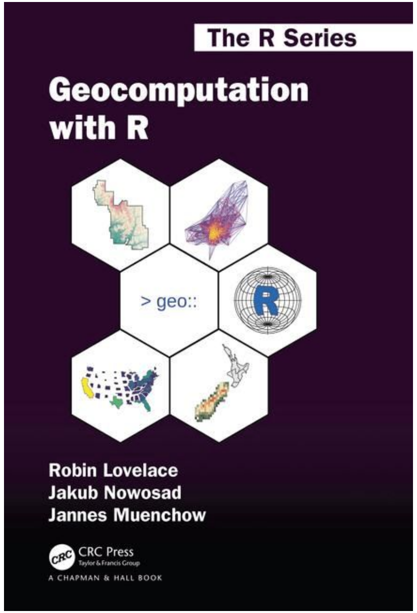


FIGURE 8.16: Animated map showing the top 30 largest urban agglomerations from 1950 to 2030 based on population projects by the United Nations. Animated version available online at:



Summary

- Spatiotemporal data (2D+T, 3D+T) is at the order of magnitude more complex than spatial data (larger data volumes, more complex interactions, ideally we would like to understand processes);
- For data to be usable for spatiotemporal modeling it needs to have known (at least): lon, lat, location error, begin, end-time of measurement;
- There are in principle four main groups of methods for modeling spacetime data: (1) geostatistics, (2) ML, (3) process-based modeling, (4) hybrid methods;