A spatiotemporal ensemble machine learning framework for predictive mapping: One model to rule them all?

Tom Hengl

OpenGeoHub foundation

ESA UNCLASSIFIED – For ESA Official Use Only


BECAUSE IN THE END
THERE CAN BE ONLY ONE
memegenerator.net

My talk in a nutshell: simple evolution of things

Evolution of technology:

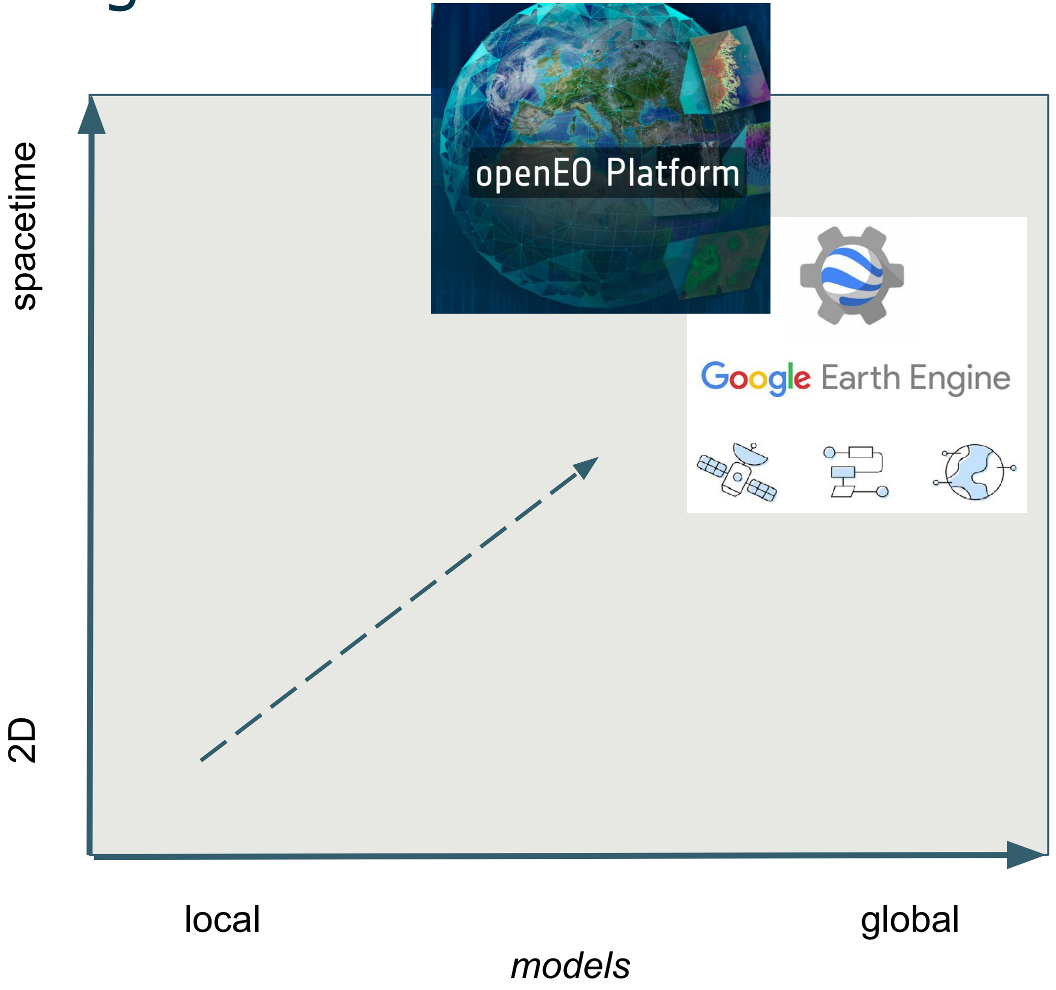
- Statistics -> Machine Learning;
- Field work -> Cloud-based sensor networks;
- Small data -> Big geospatial data (VVV);
- Printed maps -> AR/VR;

Social / organizational evolution:

- Local studies -> regional / global datasets;
- Printed papers -> reproduc. research (docker);

Personal evolution:

- 2D -> 3D+T;
- Manual processing -> automated mapping;
- Statistician -> Data Scientist;



Assuming that we could collate ALL measurements of a particular variable of interest, we could fit **a single global model** that explain dynamics / variation of the variable over the whole globe ("one model to rule them all").

Imagine for example **precipitation [in mm]** (hourly, daily), if we had the data from **all stations all measurements ever made** (possibly Billions of measurements), if we use all computers in the world, fit a spacetime model to this dataset, then this could become "the best" the most accurate model to represent rainfall (hence all our knowledge of rainfall could be converted to a single ML/AI system). We could apply similar rules to basically any field in

About 20 years ago -> Geostatistics

3.4. *Description of the full data set*

Figure 5 presents the complete 467 rainfall data in terms of proportion symbols and summarised with their statistics and associated histogram (Figure 6).

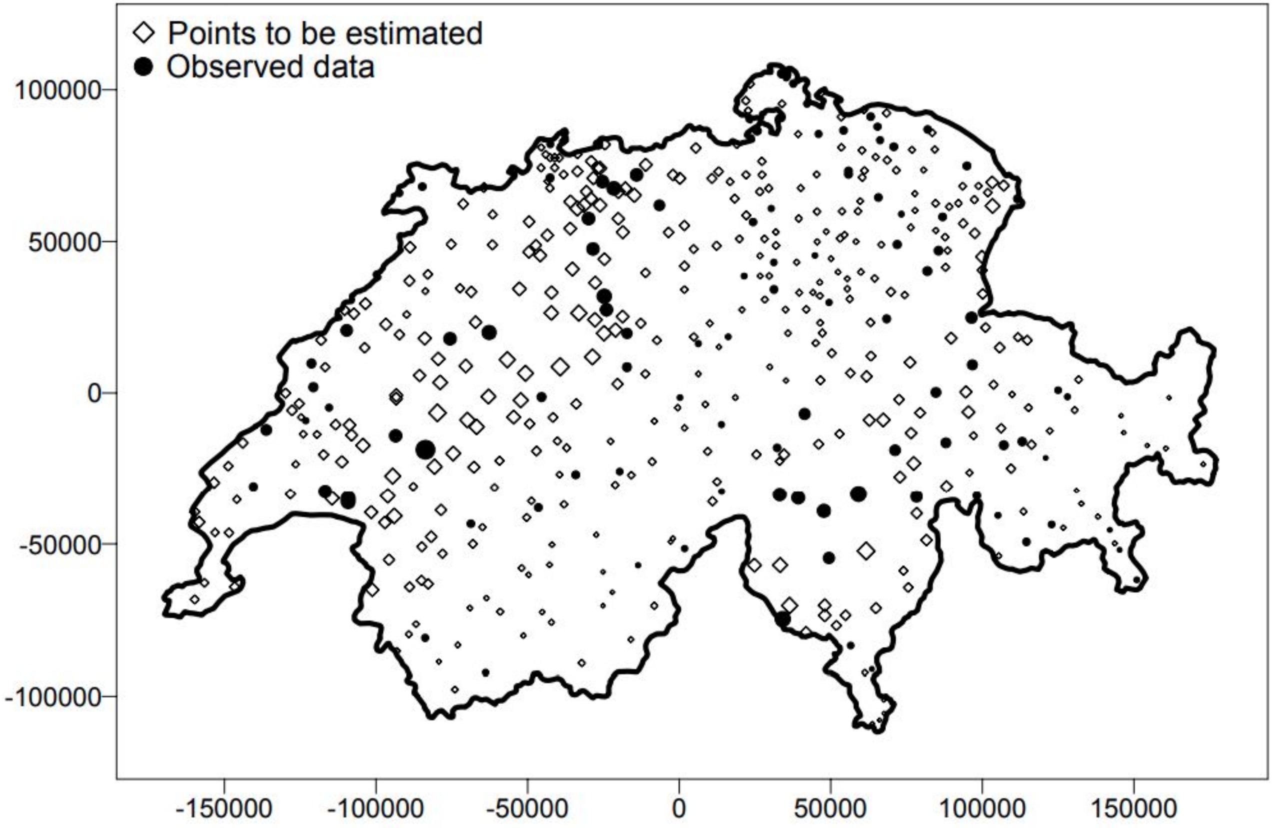
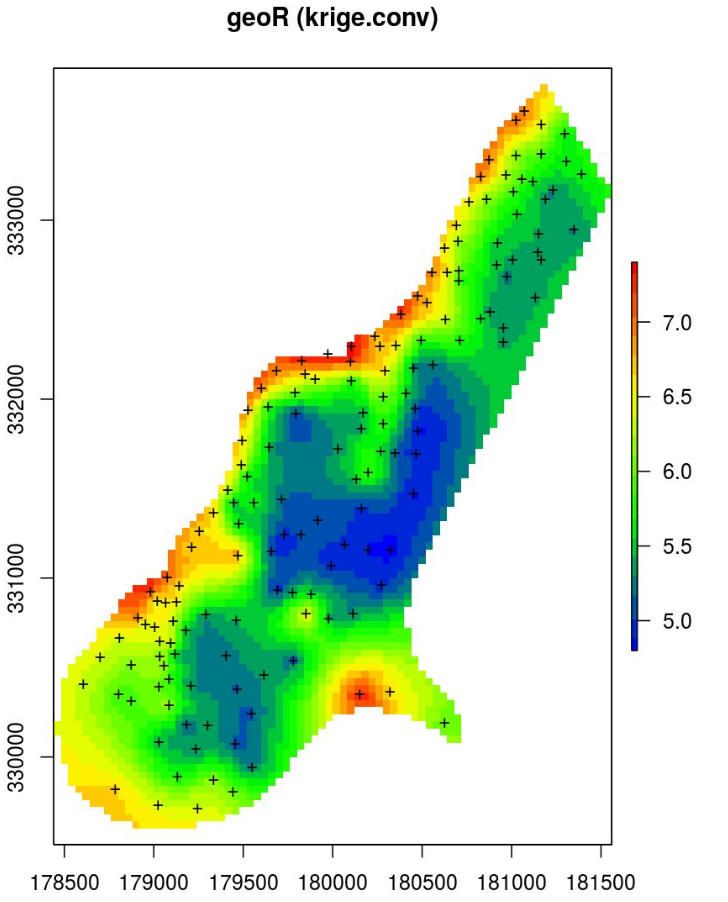


Figure 5. *Proportional symbols for the whole rainfall data set*

Source:
https://wiki.52north.org/AI_GEOSTATS/EventsSIC97

Meuse dataset (Zn concentration in soil)



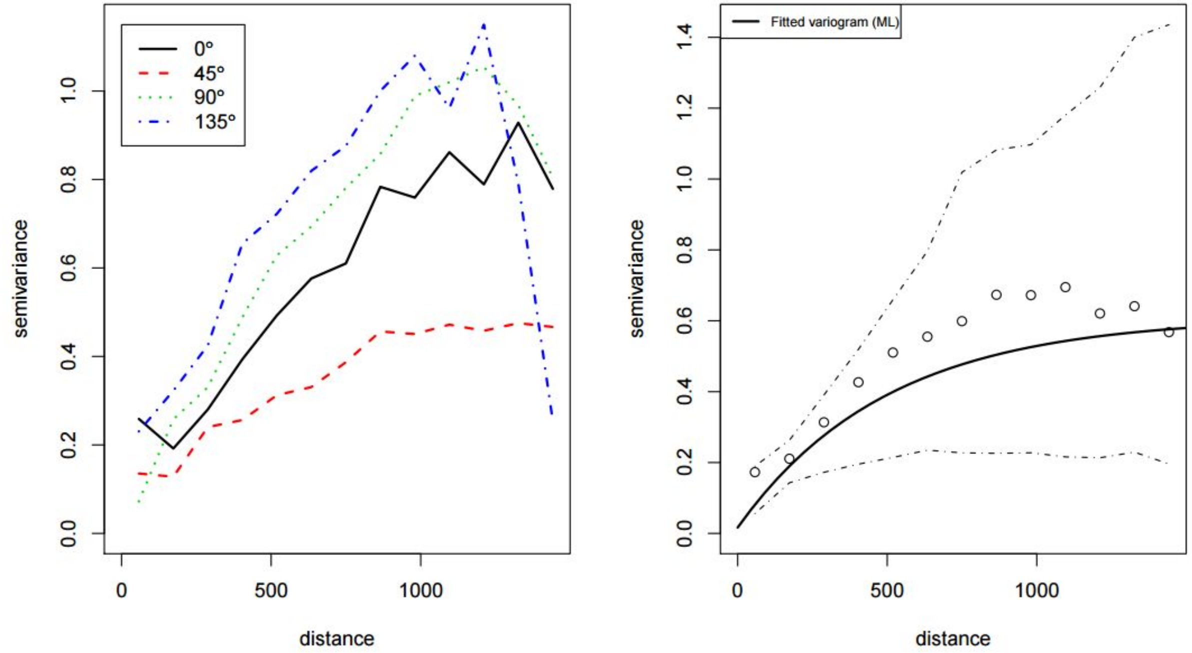


Fig. 5.15: Anisotropy (left) and variogram model fitted using the Maximum Likelihood (ML) method (right). The confidence bands (*envelopes*) show the variability of the sample variogram estimated using simulations from a given set of model parameters.

Ordinary kriging with log-normal distribution (geoR): needs many parameters to be set (manually)

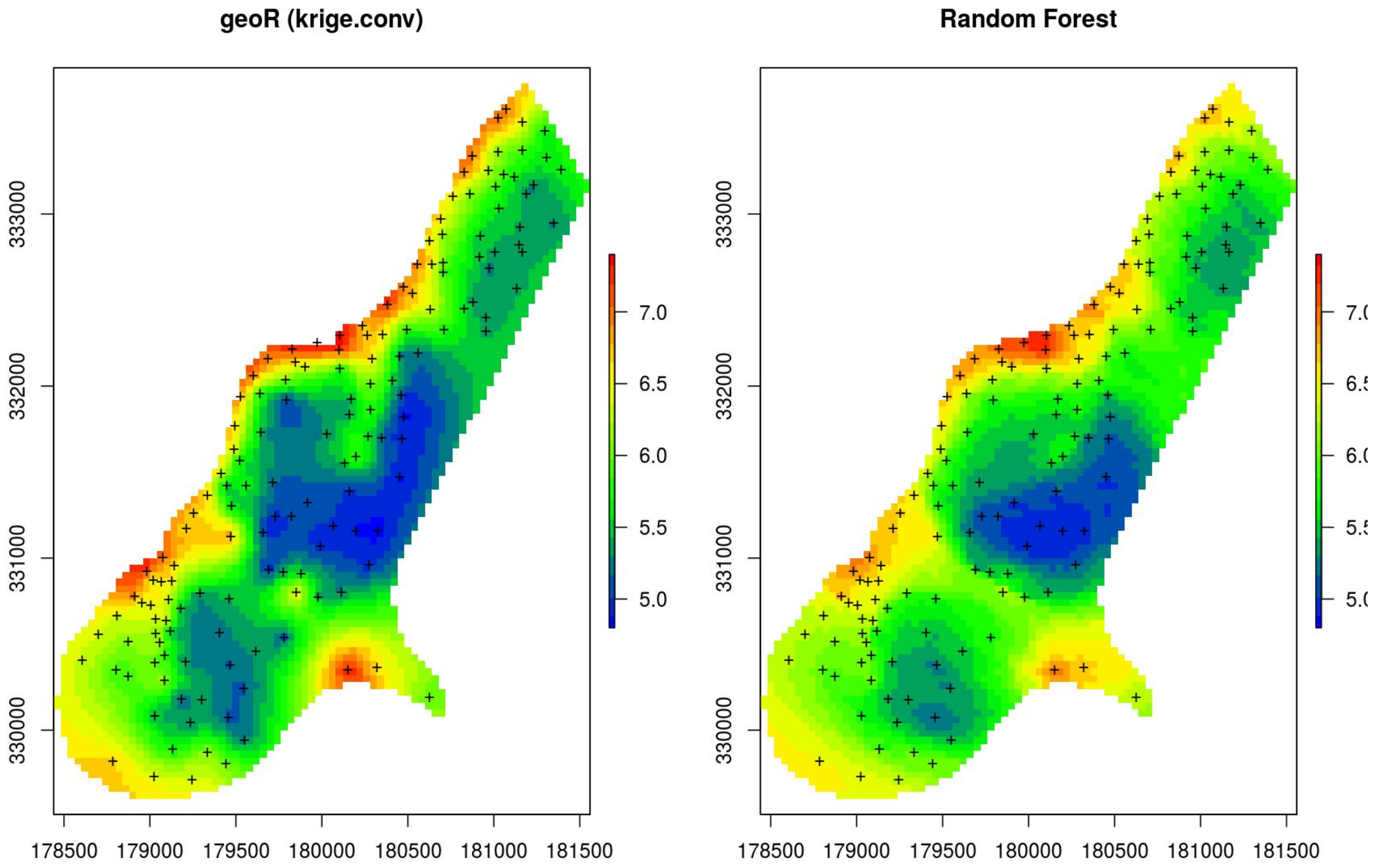
```
zinc.vgm <- likfit(zinc.geo, lambda=0,
ini=c(var(log1p(zinc.geo$data)),500), cov.model="exponential")
zinc.ok <- krige.conv(zinc.geo, locations=locs,
krige=krige.control(obj.m=zinc.vgm))

krige.conv: model with constant mean
krige.conv: performing the Box-Cox data transformation
krige.conv: back-transforming the predicted mean and variance
krige.conv: Kriging performed using global neighbourhood
```

Random Forest on buffer distances

```
grid.dist0 <- buffer.dist(meuse["zinc"], meuse.grid[1])
dn0 <- paste(names(grid.dist0), collapse="+")
fm0 <- as.formula(paste("zinc ~", dn0))
ov.zinc <- over(meuse["zinc"], grid.dist0)
m.zinc <- ranger(fm0, cbind(meuse@data["zinc"], ov.zinc))
zinc.rfd <- predict(m.zinc, grid.dist0@data)
```

Is ML the end of Geostatistics?



Conclusion: many modern Machine Learning techniques (especially tree-based) are universally applicable for general modeling purposes

PeerJ – Life & Environment About More AUTHORS SUBMIT ARTICLE Login Search

View 161 tweets
Related research
Share

< ENVIRONMENTAL SCIENCE

Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables

Research article Biogeography Soil Science Computational Science Data Mining and Machine Learning Spatial and Geographic Information Science

Tomislav Hengl[1], Madlene Nussbaum[2], Marvin N. Wright[3], Gerard B.M. Heuvelink[4], Benedikt Gräler[5] Tweet Authors

Published August 29, 2018

Note that a Preprint of this article also exists, first published March 14, 2018.

PubMed 30186691

179 Total citations
139 Recent citations
n/a Field Citation Ratio
1.99 Relative Citation Ratio

Enter your institution
To find colleagues at PeerJ
Enter to search

CALL FOR PAPERS
Call for Papers: Bio-inspired Deep Learning Image and Signal Processing Pipelines in Medical Oncology
View Call for Papers

Fully-funded PhD studentship at The Sainsbury Lab studying the virulence mechanisms of pathogen effectors
Norwich, united kingdom
View Job

Download

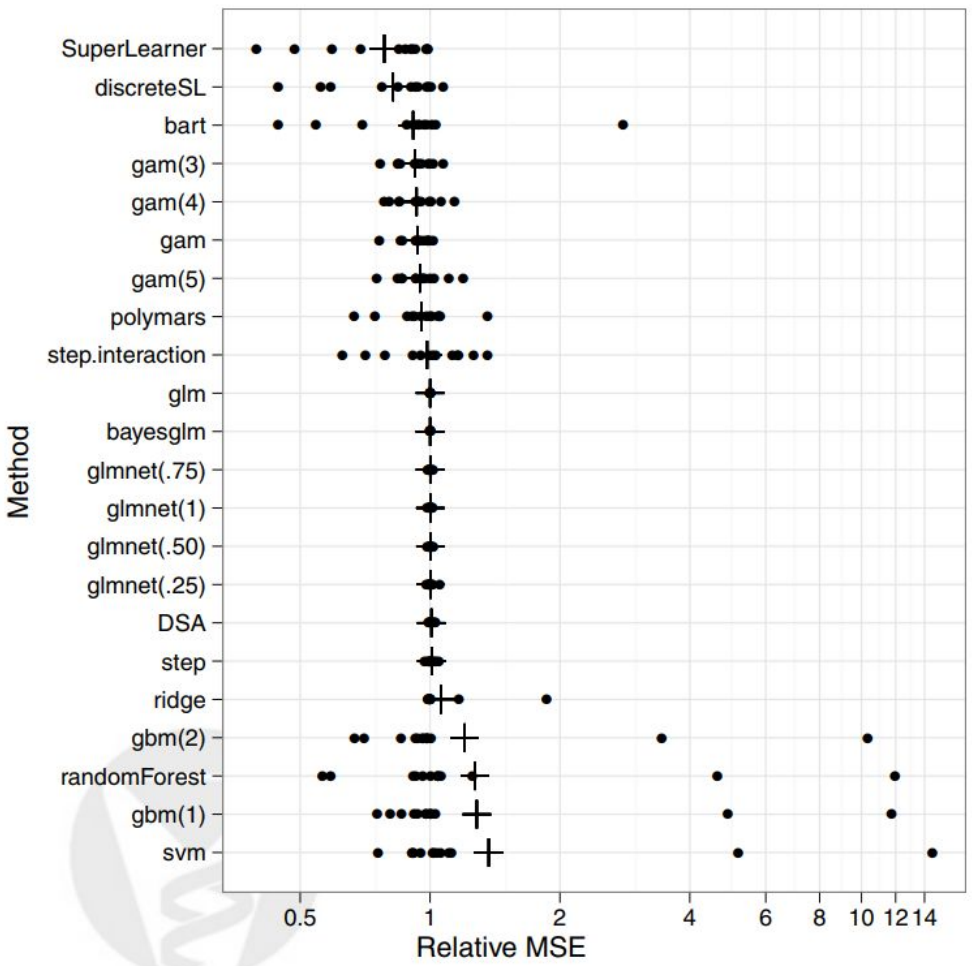
Author and article information

Abstract

Random forest and similar Machine Learning techniques are already used to generate spatial predictions, but spatial location of points (geography) is often ignored in the modeling process. Spatial auto-correlation, especially if still existent in the cross-validation residuals, indicates that the predictions are maybe biased, and this is suboptimal. This paper presents a random forest for spatial predictions framework (RFsp) where buffer distances from observation points are used as explanatory variables, thus incorporating geographical proximity effects into the prediction process. The RFsp framework is illustrated with examples that use textbook datasets and apply spatial and spatio-temporal prediction to numeric, binary, categorical, multivariate and spatiotemporal variables. Performance of the RFsp

From a single model -> **Ensemble Machine Learning**

Figure 3: 10-fold cross-validated relative mean squared error compared to glm across 13 real datasets. Sorted by the geometric mean, denoted with the plus (+) sign.



scikit learn vs mlr

"Ensemble methods are meta-algorithms that combine several machine learning techniques into one predictive model in order to **decrease variance** (bagging), **bias** (boosting), or **improve predictions** (stacking)." https://blog.statsbot.co/ensemble-learning-d1dcd548e936

This however comes at costs:

- higher computational load,
- higher RAM requirements,

https://github.com/Envirometrix/landmap

possible replacement for kriging methods (Hengl et al. 2018). Automation comes, however, at the high computing and RAM usage costs.

In the following example we use somewhat larger data set from the SIC1997 exercise.

```
data("sic1997")
X <- sic1997$swiss1km[c("CHELSA_rainfall","DEM")]
mR <- train.spLearner(sic1997$daily.rainfall, covariates=X, lambda=1)
rainfall1km <- predict(mR)
```

The processing is now much more computational because the data set consists from 467 points (hence 467 buffer distance maps need to be produced). This will make the regression matrix becoming extensive, and also 5x3 models need to be fitted. At the moment, using `train.spLearner` for point data set with >>1000 points should be done with caution.

The final results also shows quite similar results to universal kriging in geoR. The model error map above, however, shows more spatial contrast and helps detect areas of especially high errors.

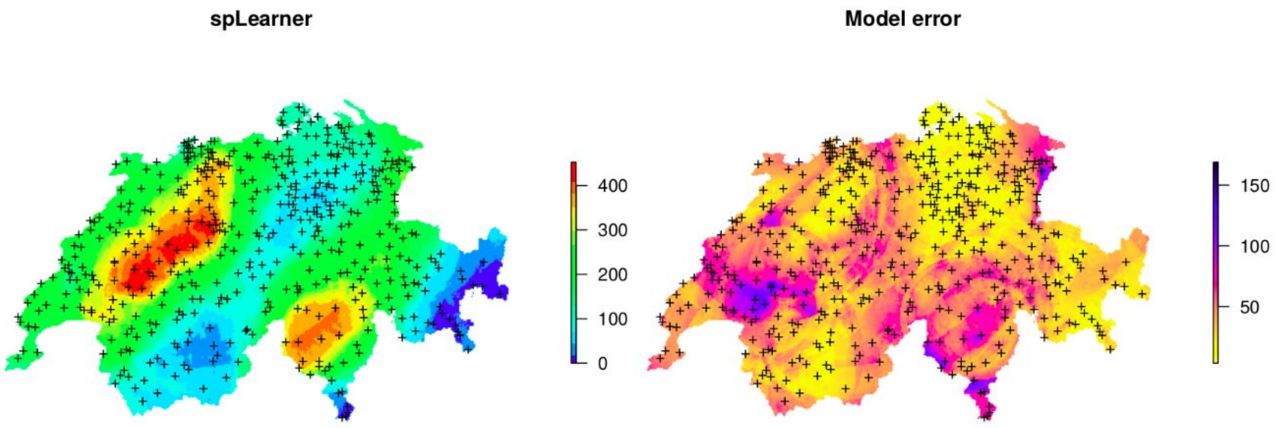
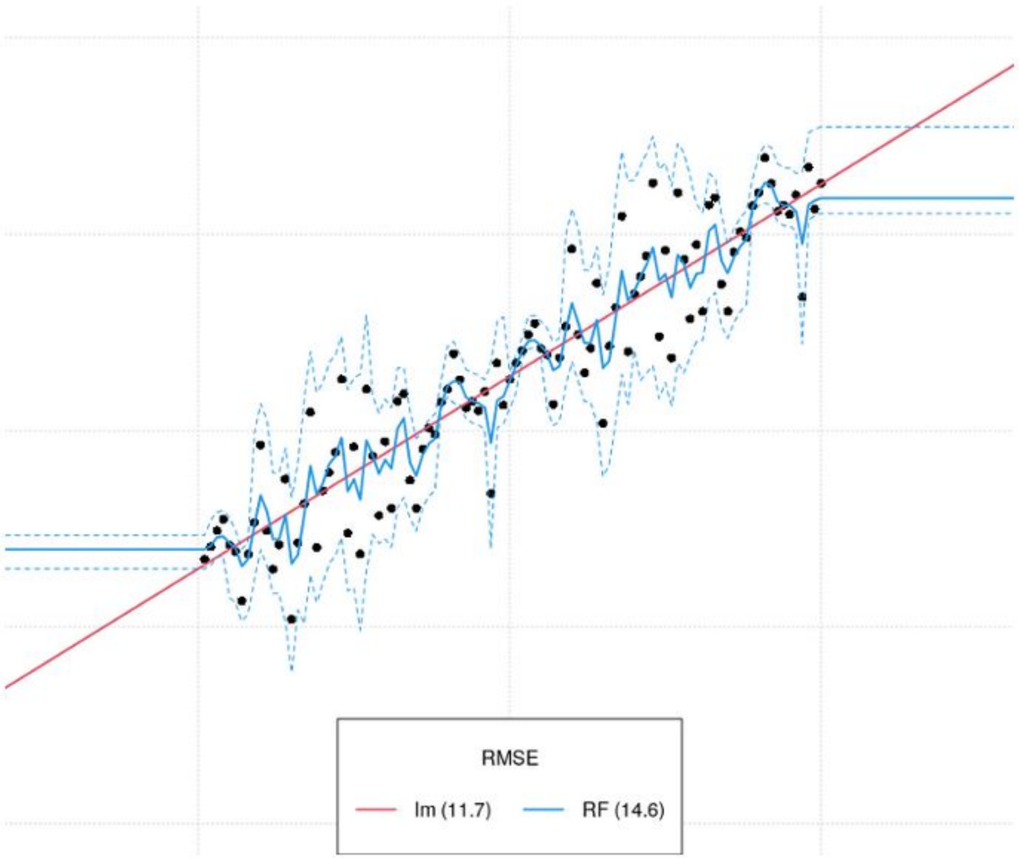


Figure: Predicted daily rainfall for the SIC1997 data set.

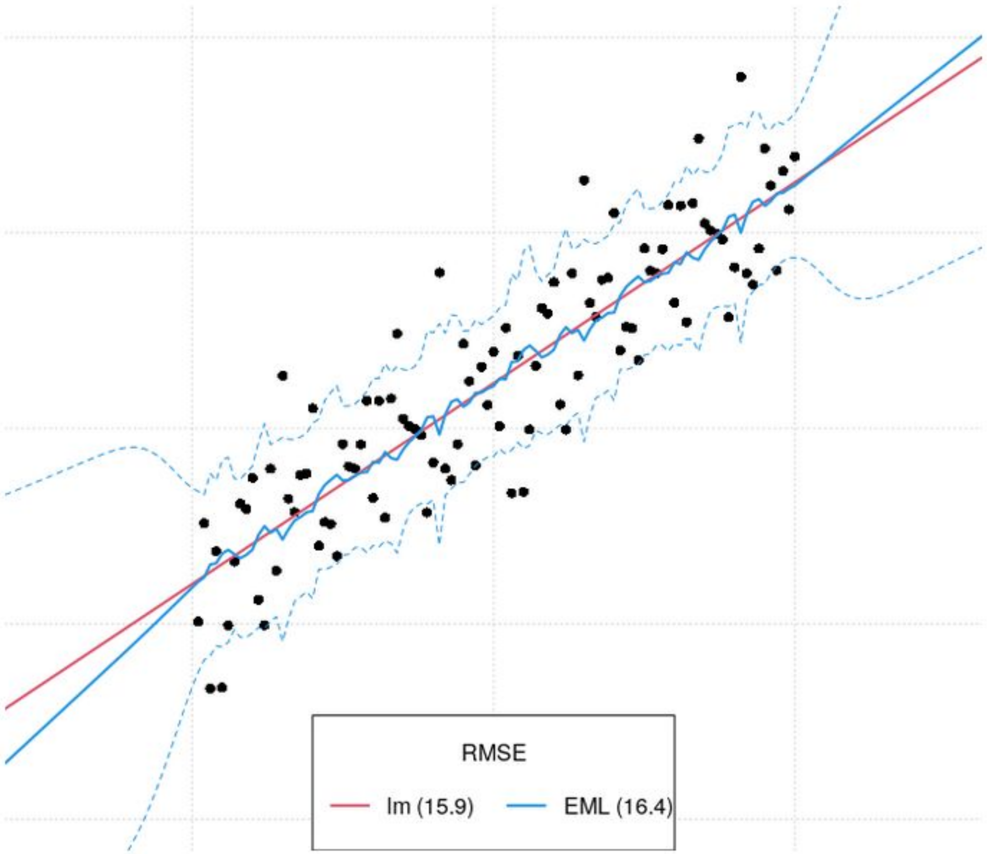
Default learners: **ranger (RF), xgboost, cvglmnet, cubist, SVM**

Conclusion: many modern Machine Learning techniques are applicable for general spatial prediction purposes

RF vs EML



Synthetic data set and linear (red line) and Random Forest (blue line) model fits. The prediction intervals are derived using the forestError package.



Ensemble Machine Learning helps with extrapolation problems

Sign in Get started

PROGRAMMING AI SOFTWARE DEVELOPMENT TECHNOLOGY DATA SCIENCE JOIN US

Extrapolation is tough for trees (tree-based learners), combining learners of different type makes it less tough

OpenGeoHub Follow
Jul 14 · 11 min read

Prepared by: Tom Hengl (OpenGeoHub)

> Some popular tree-based Machine Learning (ML) algorithms such as Random Forest (RF) and/or Gradient Boosting have been criticized about over-fitting effects and prediction / extrapolation in feature space that can lead to serious blunders and artifacts. Extrapolation seems to be especially cumbersome for regression problems, and many at the order of magnitude less complex models seem to outperform RF when it comes to extrapolation. Serious artifacts due to extrapolation and over-fitting decreases confidence in RF, especially if the prediction intervals are also unrealistic. Here we demonstrate that an Ensemble approach, that combines both diverse learners (complex and simple, tree-based and linear and polynomial models) and robust cross-validation, can be used to find a

Going global


PUBLISH ABOUT BROWSE SEARCH

advanced search

PLOS ONE

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

SoilGrids250m: Global gridded soil information based on machine learning

Tomislav Hengl, Jorge Mendes de Jesus, Gerard B. M. Heuvelink, Maria Ruiperez Gonzalez, Milan Kilibarda, Aleksandar Blagotić, Wei Shangguan, Marvin N. Wright, Xiaoyuan Geng, Bernhard Bauer-Marschallinger, Mario Antonio Guevara, Rodrigo Vargas, Robert A. MacMillan, [...], Bas Kempen [view all]

Published: February 16, 2017 • https://doi.org/10.1371/journal.pone.0169748

1,664 Save | 1,172 Citation | 67,786 View | 32 Share

Article | Authors | Metrics | Comments | Media Coverage

Download PDF | Print | Share

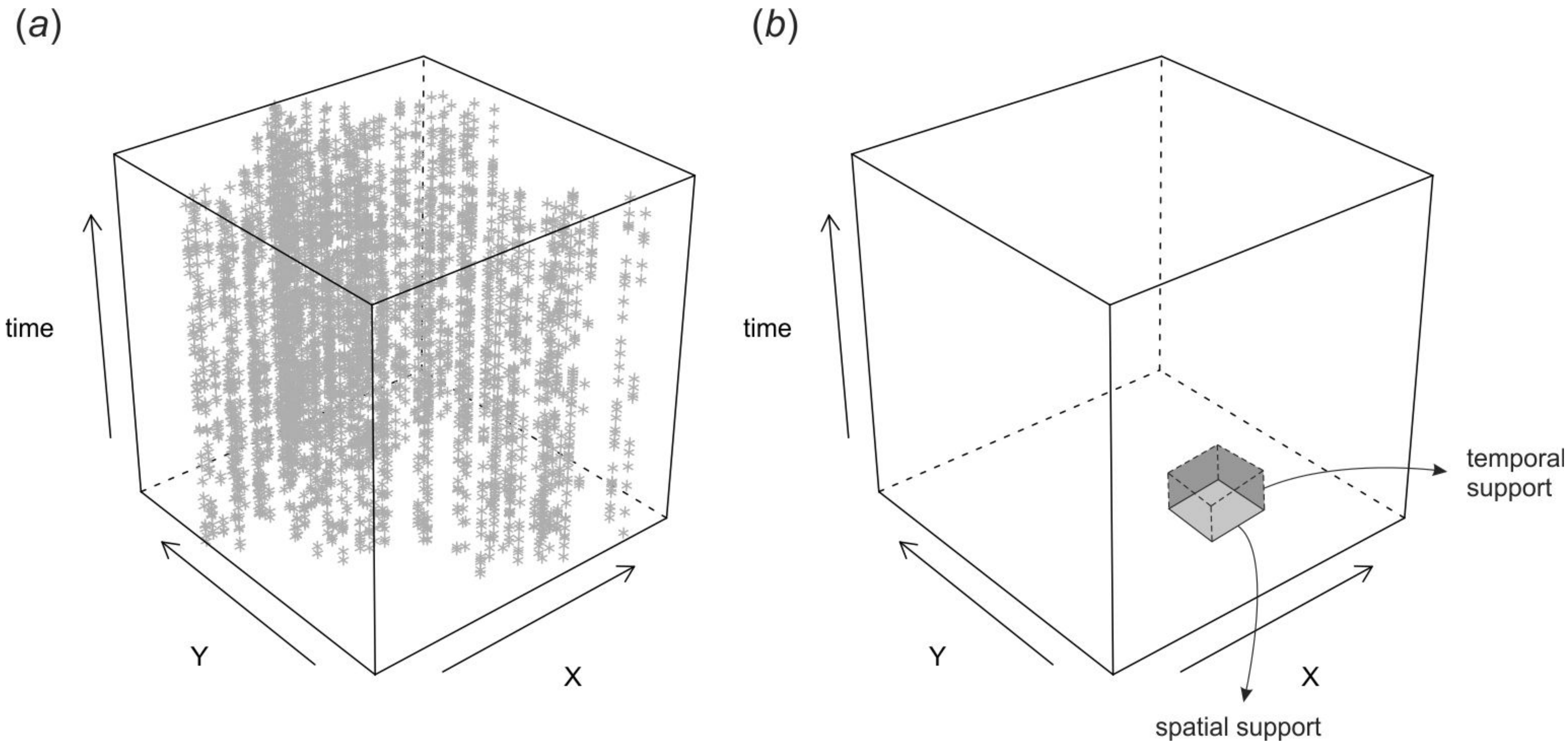
Check for updates

Abstract

This paper describes the technical development and accuracy assessment of the most recent and improved version of the SoilGrids system at 250m resolution (June 2016 update). SoilGrids provides global predictions for standard numeric soil properties (organic carbon, bulk density, Cation Exchange Capacity (CEC), pH, soil texture fractions and coarse fragments) at seven standard depths (0, 5, 15, 30, 60, 100 and 200 cm), in addition to predictions of depth to bedrock and distribution of soil classes based on the World Reference Base (WRB) and USDA classification systems (ca. 280 raster layers in total). Predictions were based on ca. 150,000 soil profiles used for training and a stack of 158 remote sensing-based soil covariates (primarily derived from MODIS land products, SRTM DEM derivatives, climatic images and global landform and lithology maps), which were used to fit an ensemble of machine learning methods —random forest and gradient boosting and/or multinomial logistic regression—as implemented in the R packages `ranger`, `xgboost`, `nnet` and `caret`. The results of 10–fold cross-validation show that the ensemble models explain between 56% (coarse fragments) and 83% (pH) of variation with an overall average of 61%. Improvements in the relative accuracy considering the amount of variation explained, in comparison to the previous version of SoilGrids at 1 km spatial resolution, range from 60 to 230%. Improvements can be attributed to: (1) the use of machine learning instead of linear regression, (2) to considerable investments in preparing finer resolution covariate layers and (3) to insertion of additional soil profiles. Further development of SoilGrids could include refinement of methods to incorporate input uncertainties and derivation of posterior probability distributions (per pixel), and further automation of spatial modeling so that soil maps can be generated for potentially hundreds of soil variables. Another area of future

ADVERTISEMENT
PLOS ONE COLLECTION Animal Welfare

Next frontier -> spacetime predictive mapping



Daily temperatures + MODIS LST

[Hengl et al. 2014]

Need for global c

nature > correspondence > article

a natureresearch journal

MENU

nature
International journal of science

Subscribe Search Login

CORRESPONDENCE · 26 FEBRUARY 2019

Soil pollution — speed up global mapping

Deyi Hou & Yong Sik Ok

Too few countries are investing in national surveys of soil pollution. A global map is urgently needed, not least to prevent international trading of contaminated produce and the migration of persistent organic pollutants across borders. We urge all member states at next month's fourth session of the UN Environment Assembly (UNEA) to speed up their assessments.

A global map of soil pollution will also guide policymakers on protecting soils; inform chemical and waste management (see Y. Geng *et al. Nature* **565**, 153–155; 2019); prevent further pollution by identifying sources and controlling polluter behaviour; and reduce risks to public health and the environment.

The World Health Organization and the United Nations Food and Agriculture Organization are among those required by the UNEA since December 2017 to report on the extent of global soil pollution, monitor future trends and identify associated risks and impacts. The results will be presented at the UNEA's fifth session in 2020. Many hurdles must be overcome before a global assessment can be made. Collaboration between developing and developed nations in allocating technical and financial resources is a priority.

PDF version

RELATED ARTICLES

See more letters to the editor

SUBJECTS

Environmental sciences Policy

WORLD VIEW | 17 July 2019

Sustainable development will falter without data

Unless governments establish competent monitoring systems, the world will not reach the UN Sustainable Development Goals, says Jessica Espey.

Jessica Espey

In 2013, I worked in Liberia's Ministry of Finance and Development Planning. My office was in a run-down beachside building with intermittent electricity and water. One day, the generator surged. Within seconds, we smelt singed plastic. Our computers, and other equipment the government could ill afford to replace, were ruined. The damage at the national statistics office next door was devastating. Reams of survey data typed in from paper reports were lost, along with tens of other data sets about educational outcomes, poverty rates and access to services. They had all been saved on just one computer.

Cash-strapped, infrastructure-limited national data systems run by staff who lack training and authority are common among poor countries. They are the biggest barrier to achieving the Sustainable Development Goals (SDGs) — covering everything from cleaner water to fairer societies — set by 193 countries and the United Nations in 2015, meant to put the world on a path to a sustainable future by 2030. As a forum to consider progress on SDGs meets this week, it must consider this fact: none of these goals can be met without a data revolution.

Many national statistics are compiled on paper, manually inputted to old computers, and unavailable or inconsistently accessible online. Thus, government statistics are not referred to for day-to-day (or even week-to-week) decisions. Those data that are available are usually out of date: only 35% of sub-Saharan countries have poverty data that were updated since 2015.

Download PDF

Related Articles

Development goals should enable decision-making

Map the interactions between Sustainable Development Goals

Sustainability: Map the evidence

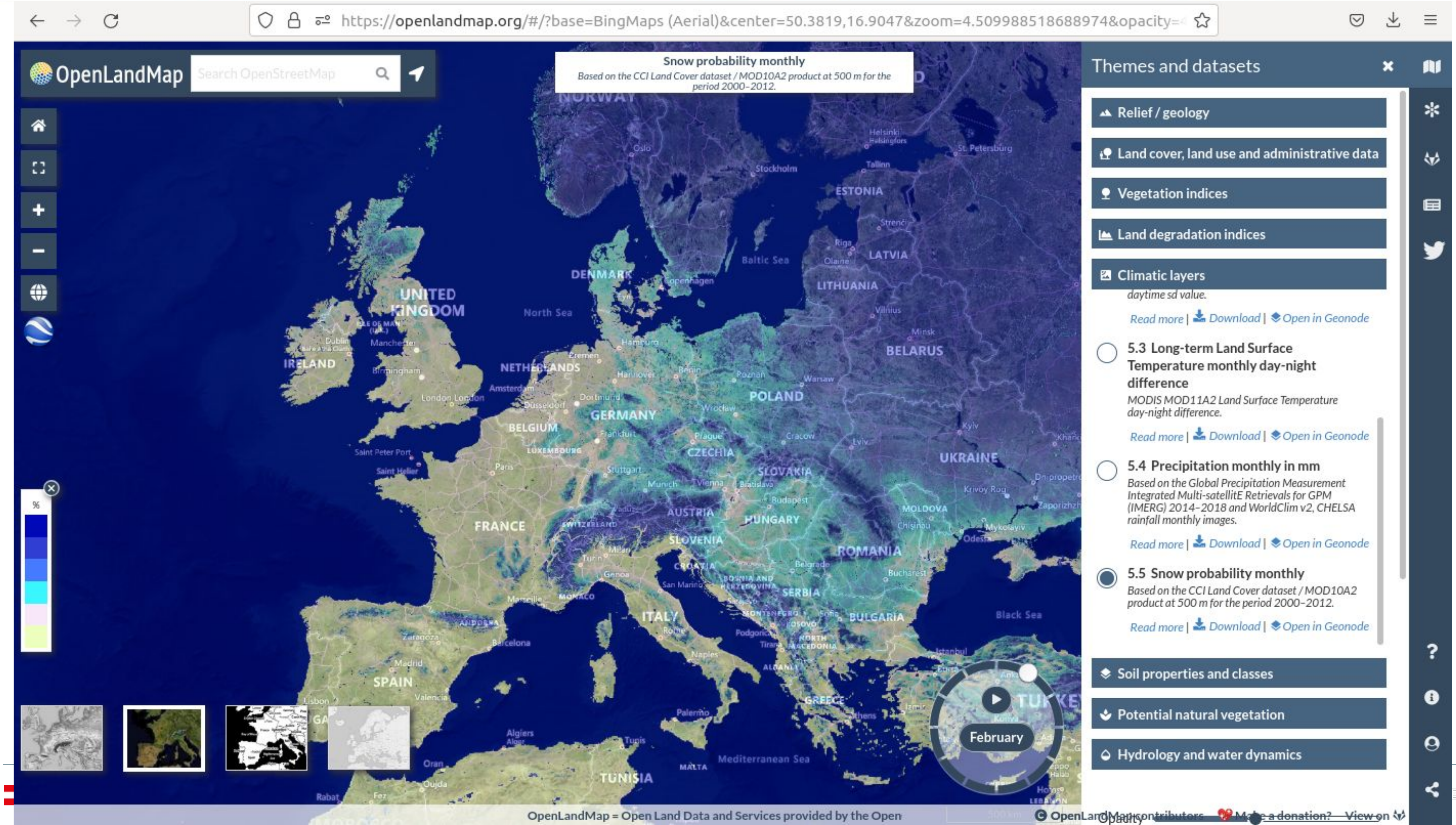
Subjects

Databases Developing world

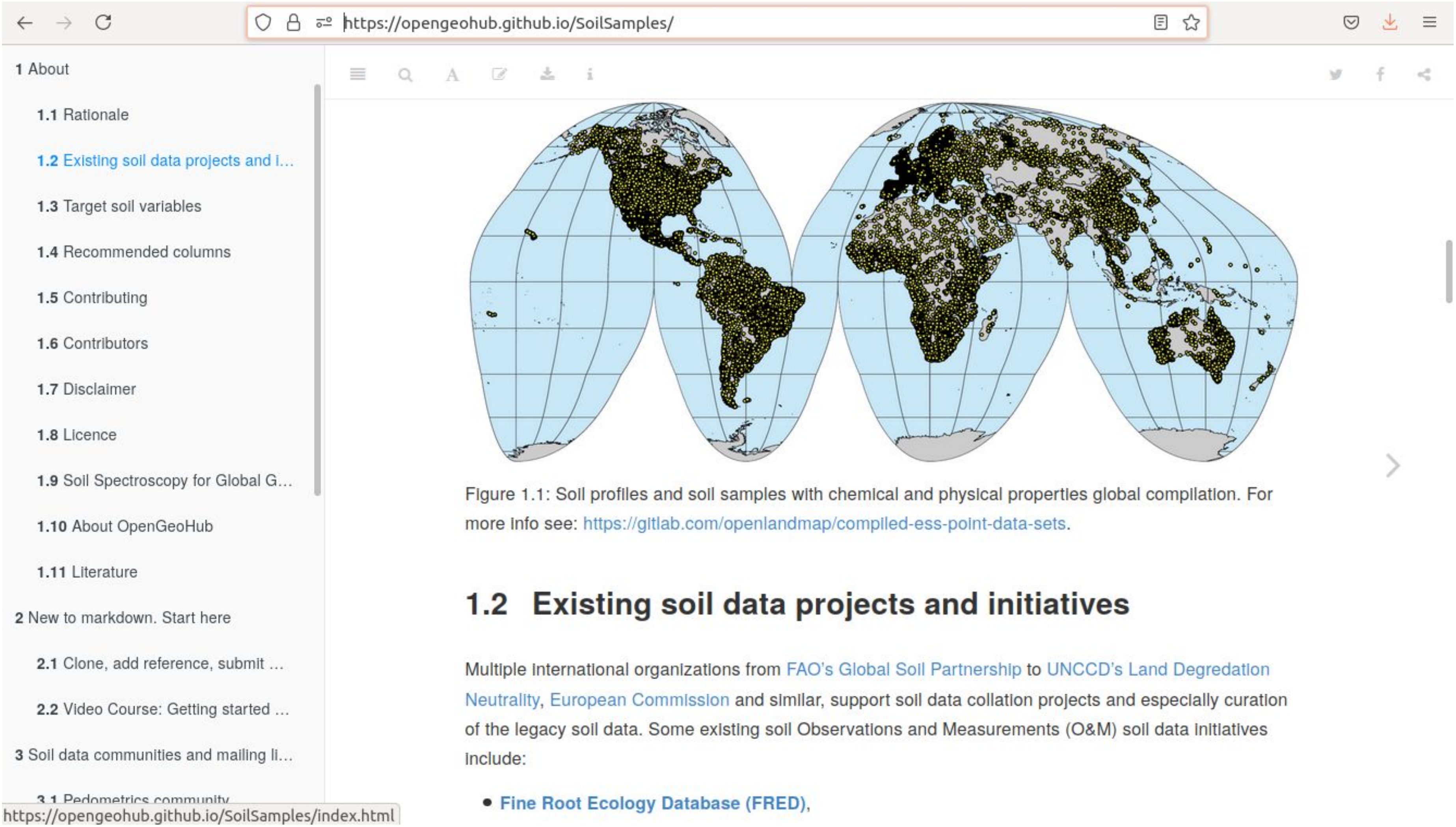
Sign up to Nature Briefing

An essential round-up of science news.

ESA-ECMWF Workshop
2021



https://gitlab.com/openlandmap/compiled-ess-point-data-sets/



https://opengeohub.github.io/SoilSamples/

Figure 1.1: Soil profiles and soil samples with chemical and physical properties global compilation. For more info see: https://gitlab.com/openlandmap/compiled-ess-point-data-sets.

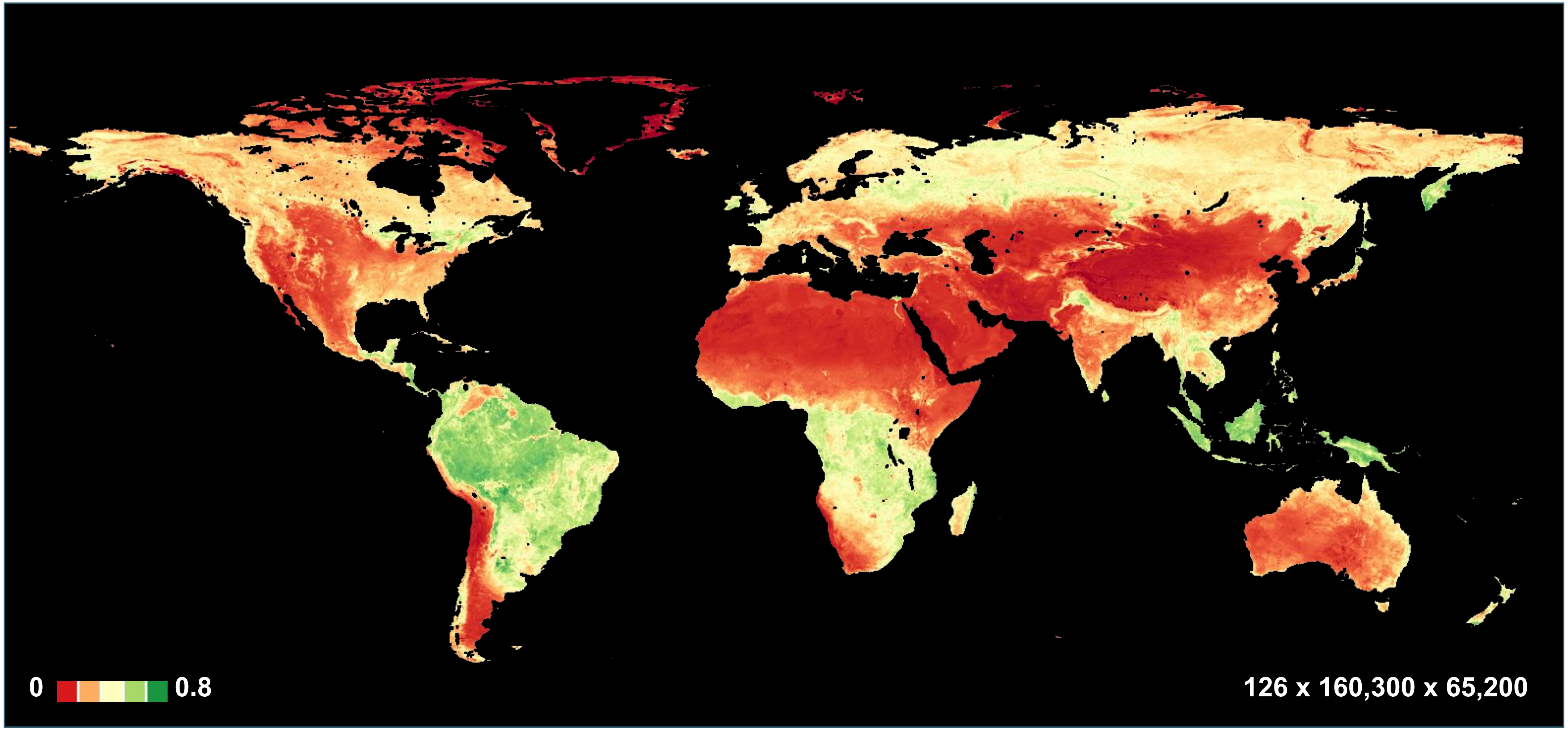
1.2 Existing soil data projects and initiatives

Multiple international organizations from FAO's Global Soil Partnership to UNCCD's Land Degradation Neutrality, European Commission and similar, support soil data collation projects and especially curation of the legacy soil data. Some existing soil Observations and Measurements (O&M) soil data initiatives include:

- Fine Root Ecology Database (FRED),

https://opengeohub.github.io/SoilSamples/index.html

MOD13Q1 EVI - Aggregated (2 months) and gap-filled



Global spatiotemporal models

$$Y(\lambda,\phi,t) = f\left[X_S(\lambda,\phi), X_A(\lambda,\phi,t_C), X_D(\lambda,\phi,t_C), X_{\mathrm{CL}}(\lambda,\phi,t_A), \varepsilon'\right] \tag{2}$$

where λ,ϕ are the longitude and latitude, t is time of observation or prediction e.g. date, X_S are the *"static"* covariates i.e. elevation, parent material, land form, X_A are the accumulation-based covariates such as cumulative primary productivity representing litter accumulation, X_D are the deposition variables such as cumulative erosion, X_{CL} is the past climate on a moving window of 5–10 years, t_C is the cumulative time, t_A is the moving window size for the past climate / environmental conditions, and ε' is the remaining unexplained part of variation.

Global spatiotemporal models

AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE

10.1002/2013JD020803

Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution

Milan Kilibarda[1], Tomislav Hengl[2], Gerard B. M. Heuvelink[3], Benedikt Gräler[4], Edzer Pebesma[4], Melita Perčec Tadić[5], and Branislav Bajat[1]

[1]Department of Geodesy and Geoinformatics, Faculty of Civil Engineering, University of Belgrade, Belgrade, Serbia, [2]ISRIC-World Soil Information, Wageningen, Netherlands, [3]Soil Geography and Landscape Group, Wageningen University, Wageningen, Netherlands, [4]Institute for Geoinformatics, University of Münster, Münster, Germany, [5]Meteorological and Hydrological Service, Zagreb, Croatia

Key Points:

- Global spatio-temporal regression-kriging daily temperature interpolation
- Fitting of global spatio-temporal models for the mean, maximum, and minimum temperatures
- Time series of MODIS 8 day images as explanatory variables in regression part

Correspondence to:
M. Kilibarda,
kili@grf.bg.ac.rs

Citation:
Kilibarda, M., T. Hengl, G. B. M. Heuvelink, B. Gräler, E. Pebesma, M. Perčec Tadić, and B. Bajat (2014), Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution, *J. Geophys. Res. Atmos.*, *119*, 2294–2313,

Abstract Combined Global Surface Summary of Day and European Climate Assessment and Dataset daily meteorological data sets (around 9000 stations) were used to build spatio-temporal geostatistical models and predict daily air temperature at ground resolution of 1 km for the global land mass. Predictions in space and time were made for the mean, maximum, and minimum temperatures using spatio-temporal regression-kriging with a time series of Moderate Resolution Imaging Spectroradiometer (MODIS) 8 day images, topographic layers (digital elevation model and topographic wetness index), and a geometric temperature trend as covariates. The accuracy of predicting daily temperatures was assessed using leave-one-out cross validation. To account for geographical point clustering of station data and get a more representative cross-validation accuracy, predicted values were aggregated to blocks of land of size 500 × 500 km. Results show that the average accuracy for predicting mean, maximum, and minimum daily temperatures is root-mean-square error (RMSE) = ±2°C for areas densely covered with stations and between

Global spatiotemporal models

perature is a function of geometric position of a particular location on Earth and day of the year. We call this a *geometric temperature trend*. The geometric temperature trend for the mean temperature was modeled as a function of the day of year and latitude (ϕ):

$$t_{\mathrm{geom}} = 30.4 \cos \phi - 15.5(1 - \cos \theta) \sin |\phi|, \tag{3}$$

where θ is derived as

$$\theta = (\mathrm{day} - 18) \frac{2\pi}{365} + 2^{1-\mathrm{sgn}(\phi)} \pi. \tag{4}$$

The number 18 represents the coldest day in the Northern and warmest day in the Southern Hemisphere and was derived empirically by graphical inspection of mean daily temperature plots from stations in the Northern and Southern Hemispheres. The sgn denotes the signum function that extracts the sign of a real number. Parameters 30.4°C and 15.5°C of the geometric temperature trend were calculated by least squares fitting on circa 44 million daily temperature observations from 2000 to 2011. These two numbers are, in fact, similar to the mean annual temperature on the equator and the mean global Earth temperature.

The linear model for the minimum daily temperature uses the same covariates as the linear model for mean daily temperature. The geometric temperature trend for minimum daily temperature was

$$t_{\mathrm{geom}} = 24.2 \cos \phi - 15.7(1 - \cos \theta) \sin |\phi|, \tag{5}$$

The geometric temperature trend for maximum daily temperature was

$$t_{\mathrm{geom}} = 37 \cos \phi - 15.4(1 - \cos \theta) \sin |\phi|, \tag{6}$$

Global spatiotemporal model = geometric temperature

```
temp.from.geom <- function(fi, day, a=30.419375,
                 b=-15.539232, elev=0, t.grad=0.6) {
  f = ifelse(fi==0, 1e-10, fi)
  costeta = cos( (day-18 )*pi/182.5 + 2^(1-sign(fi) ) *pi)
  cosfi = cos(fi*pi/180 )
  A = cosfi
  B = (1-costeta ) * abs(sin(fi*pi/180 ) )
  x = a*A + b*B - t.grad * elev / 100
  return(x)
}
```

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

a natureresearch journal

MENU

nature
International journal of science

Subscribe Search Login

Article | Published: 18 October 2017

Mastering the game of Go without human knowledge

David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel & Demis Hassabis

Nature **550**, 354–359 (19 October 2017) | Download Citation

666 Citations | 2572 Altmetric | Article metrics »

Editorial Summary

AlphaGo Zero goes solo

To beat world champions at the game of Go, the computer program AlphaGo has relied largely on supervised learning from millions of...

show more

Associated Content

Nature | News & Views

Artificial intelligence: Learning to play Go from scratch

Satinder Singh, Andy Okun & Andrew Jackson

Abstract

A long-standing goal of artificial intelligence is an algorithm that learns, *tabula rasa*, superhuman proficiency in challenging domains. Recently, AlphaGo became the first program to defeat a world champion in the game of Go. The tree search in AlphaGo evaluated positions and selected moves using deep neural networks. These neural networks were trained by supervised learning from human expert moves, and by reinforcement learning from self-play. Here we introduce an algorithm based solely on reinforcement learning, without human data, guidance or domain knowledge beyond game rules. AlphaGo becomes its own teacher: a neural network is trained to predict AlphaGo's own move selections and also the winner of AlphaGo's games. This neural network improves the strength of the tree search, resulting in higher quality move selection and stronger self-play in the next iteration. Starting *tabula rasa*, our new

Environmental Data Cube for Europe: https://maps.opendatascience.eu

All layers published and served as Cloud-Optimized GeoTIFFS (Wasabi S3),

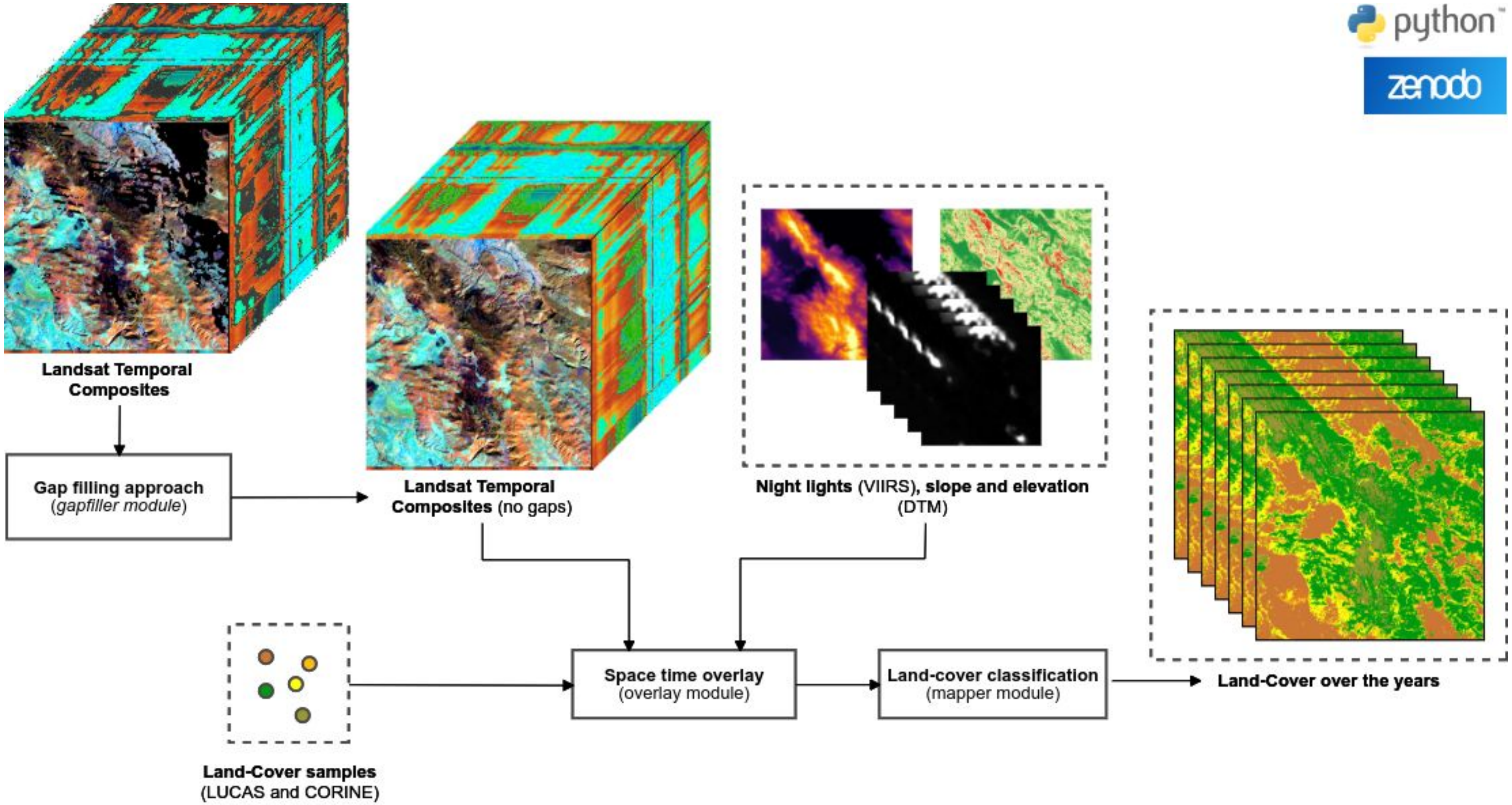
Metadata available via Geonetwork (still working on the STAC version),

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

CEF Telecom project 2018-EU-IA-0095.

This project is co-financed by the European Union.

eumap package for python (http://eumap.readthedocs.org/)



Research Square | Search preprints | Browse | Tools & Services | About | Sign In | + Submit a Preprint

This is a preprint, a preliminary version of a manuscript that has not completed peer review at a journal. Research Square does not conduct peer review prior to posting preprints. The posting of a preprint on this server should not be interpreted as an endorsement of its validity or suitability for dissemination as established information or for guiding clinical practice.

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/v2 Download PDF

LICENSE: This work is licensed under a CC BY 4.0 License. Read Full License

DECLARATIONS: View author declarations.

Abstract

A seamless spatiotemporal machine learning framework for automated prediction, uncertainty assessment, and analysis of long-term LULC dynamics is presented. The framework includes: (1) harmonization and preprocessing of high-resolution spatial and spatiotemporal input 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 consists of a random forest, gradient boosted tree classifier, and a artificial neural network, with a logistic regressor as meta-learner. The results show that the most important variables for mapping LULC in Europe are: seasonal aggregates of Landsat green and near-infrared bands, multiple Landsat-derived spectral indices, long-term surface water probability, and elevation. Spatial cross-validation of the model indicates consistent performance across multiple years with overall accuracy (weighted F1-score) of 0.49, 0.63, and 0.83 when predicting 44 (level-3), 14 (level-2), and 5 classes (level-1). The spatiotemporal model outperforms spatial models on known-year classification by 2.7% and unknown-year classification by 3.5%. Results of the accuracy assessment using 48,365 independent test samples shows 87% match with the validation points. Results of time-series analysis (time-series of LULC probabilities and NDVI images) suggest forest loss in large parts of Sweden, the Alps, and Scotland. An advantage of using spatiotemporal ML is that the fitted model can be used to predict LULC in years that were not included in its training dataset, allowing generalization to past and future periods, e.g. to predict land cover for years prior to 2000 and beyond 2020. The generated land cover time-series data stack (ODSE-LULC), including the training points, is publicly available via the Open Data Science (ODS)-Europe Viewer. Functions used to prepare data and run

BADGES

Prescreen

HISTORY

CURRENT STATUS: **POSTED**

Version 2
Posted 10 Nov, 2021
No community comments so far

Version 1
Posted 27 May, 2021
View this version

METRICS

Comments: 0
PDF Downloads: 200
HTML Views: 1002

scite_
0
0
0
0

SUBJECT AREAS

Artificial Intelligence and Machine Learning
Geographic Information Systems

MORE FROM RESEARCH SQUARE

Time series analysis and forecasting of China's energy production during Covid-19: statistical models vs machine

Conclusions I

Harmonizing, integrating (binding) global Observation & Measurement data (spacetime points) make sense (at least to me) as it **enables global research / could help with global monitoring projects**.

Ensemble Machine Learning is a universally applicable framework for modeling spacetime phenomena, spatiotemporal interpolation and uncertainty assessment.

Yes, potentially, we could put **ALL Earth System Science O&M and ALL EO data cubes together**, and build spacetime models that basically put all our measurements together into a single and/or few HUGE model (something like IPCC reports, just models / code!). The advantages of such model are:

- It would help produce an unbiased picture;
- It would probably give highest accuracy + could be used to predict future;

Some disadvantages of such models:

- They are super large and would take enormous computing power to update and share;
- They need all data to be consistent and to be available for all pixels of interest;
- Many researchers would need to work together;

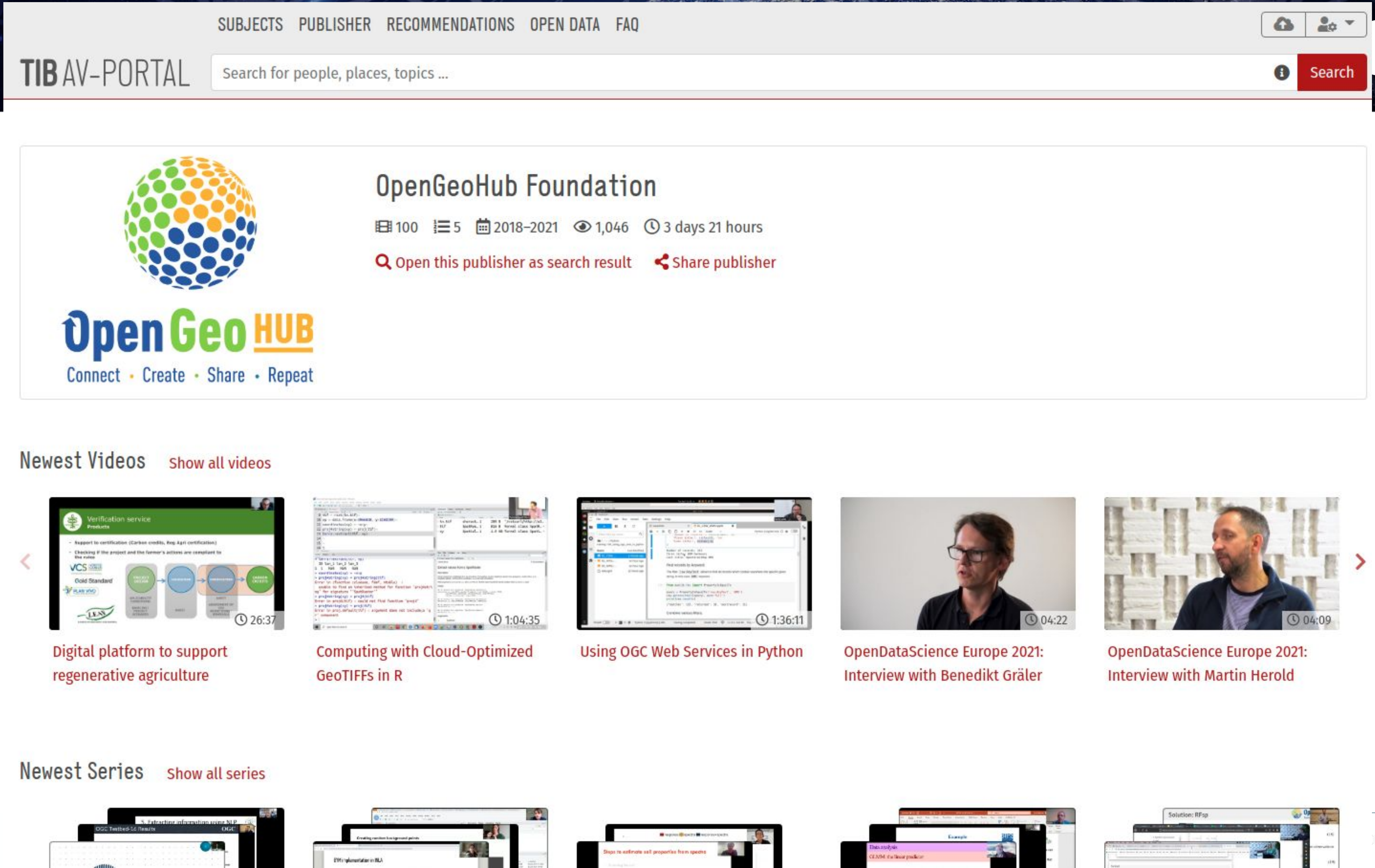
Conclusions II

Unfortunately, we (researchers) are “pushed” by the current profit-oriented economic system to compete with each other (research funding, publications, diverging national programmes).

It is difficult to do research without funding (of course) and majority of funding sources for research are national or regional (we even compete within the same institutions!).

For example, ESA and NASA collaborate, but if the researchers could have decided about the amount of collaboration, **90% of missions would have probably been = joint missions**! We researchers do NOT necessarily need competition -> we are primarily interested in understanding, educating, creating and enjoying complexity.

Think about how you can help produce better OpenStreetMap, OpenLandMap, Open Source software, Open Data projects, because competition is a good/healthy thing, but collaboration is even better!


SUBJECTS PUBLISHER RECOMMENDATIONS OPEN DATA FAQ
TIB AV-PORTAL
Search for people, places, topics ...
Search
OpenGeoHub Foundation
100 5 2018–2021 1,046 3 days 21 hours
Open this publisher as search result
Share publisher
Open Geo HUB
Connect · Create · Share · Repeat
Newest Videos
Show all videos
26:37
1:04:35
1:36:11
04:22
04:09
Digital platform to support regenerative agriculture
Computing with Cloud-Optimized GeoTIFFs in R
Using OGC Web Services in Python
OpenDataScience Europe 2021: Interview with Benedikt Gräler
OpenDataScience Europe 2021: Interview with Martin Herold
Newest Series
Show all series