Bewertung der potenziellen Wassererosion mit R und SAGA-GIS

Markus Möller[1] & Martin Volk[2]

[1] Julius Kühn-Institut, Institut für Pflanzenbau und Bodenkunde, Braunschweig

[2] Helmholtz-Zentrum für Umweltforschung GmbH – UFZ, Department Landschaftsökologie, Leipzig

FAIR

Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

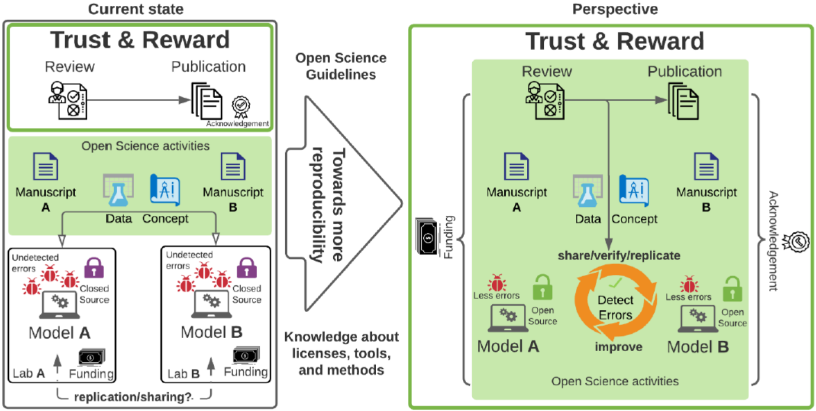
To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

Wilkinson, M.D. et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3

Reproducability/Provenance



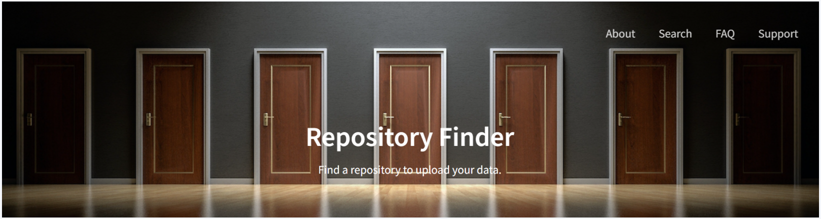
(1) Only modeling concepts and data are shared. Models and their code are seldomely shared and results not verified or replicated. This affects the trust in their results and also leads to undected software errors. Funding is focused on novel findings. Acknowledgement is focused on publications.	(2) Establishment of Open Source concepts and active integration of verification and replication in the review process.	(3) This leads to and interactive process of verification and replication that reduces errors, increases research output, and fosters trust in the model results. However, this needs to be enabled by funding opportunities and an acknowledgement of efforts. Likely, this also requires additonal qualified staff and increased teaching efforts.

Reinecke, R., Trautmann, T., Wagener, T., Schüler, K., 2022. The critical need to foster computational reproducibility. Environ. Res. Lett. 17, 041005

Repositorien: Daten

Digitale Forschungsdatenrepositorien sind Informationsinfrastrukturen, die digitale Forschungsdaten möglichst dauerhaft – anhand der Anforderungen der jeweiligen Nutzergruppe – speichern und organisieren um die Auffindbarkeit und Zugänglichkeit der Daten zu sichern (https://www.forschungsdaten.org/).

Repository finder



https://repositoryfinder.datacite.org

Repositorien: Daten

Pangaea

PANGAEA.

Data Publisher for Earth & Environmental Science

Not logged in

SEARCH SUBMIT HELP ABOUT CONTACT

Citation: **Preidl, Sebastian; Lange, Maximilian; Doktor, Daniel (2020):** Land cover classification map of Germany's agricultural area based on Sentinel-2A data from 2016. *PANGAEA*, https://doi.org/10.1594/PANGAEA.910837

Always quote citation above when using data! You can download the citation in several formats below.

RIS Citation | BibTeX Citation | Copy Citation | Facebook | Twitter | Show Map | Google Earth | 890 | 170

Abstract: Overcoming the obstacle of frequent cloud coverage in optical remote sensing data is essential for monitoring dynamic land surface processes from space. APiC, a novel adaptable pixel-based compositing and classification approach, is especially designed to use high resolution spatio-temporal space-borne data.

Here, pixel-based compositing is used separately for training data and prediction data. First, cloud-free pixels covered by reference data are used within adapted composite periods to compile a training dataset. The compiled training dataset contains samples of spectral reflectances for respective land cover classes at each composite period. For land cover prediction, pixel-based compositing is then applied region-wide. Multiple prediction models are used based on temporal subsets of the compiled training dataset to dynamically account for cloud coverage at pixel level. Thus we present a data-driven classification approach which is applicable in regions with different weather conditions, species composition and phenology.

The capability of our method is demonstrated by mapping 19 land cover classes across Germany for the year 2016 based on Sentinel-2A data. Since climatic conditions and thus plant phenology change on a large scale, the classification was carried out separately in six landscape regions of different biogeographical characteristics. The study drew on extensive ground validation data provided by the federal states of Germany.

For each landscape region, composite periods of different lengths have been established, which differ regionally in their temporal arrangement as well as in their total number, emphasising the advantage of a flexible regionalised classification procedure. Using a random forest classifier and evaluating outcomes with independent reference data, an overall accuracy of 88% was achieved, with particularly high classification accuracy of around 90% for the major land cover types. We found that class imbalances have significant influence on classification accuracy. Based on multiple temporal subsets of the compiled training dataset, over 10,000 random forest models were calculated and their performance varied considerably across and within landscape regions. The calculated importance of composite periods show that a high temporal resolution of the compiled training dataset is necessary to better capture the different phenology of land cover types.

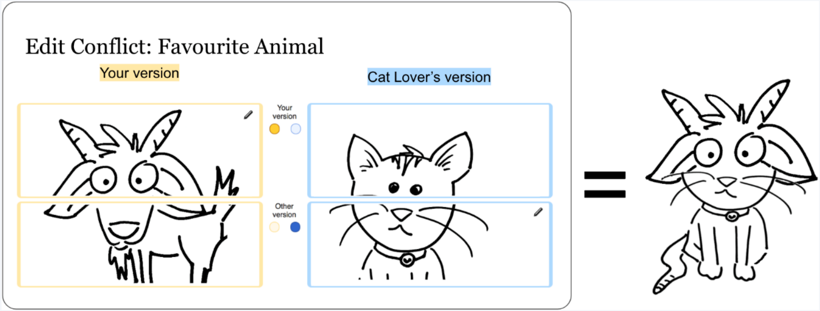
Preidl, S.; Lange, M.; Doktor, D., 2020, Land cover classification map of Germany's agricultural area based on Sentinel-2A data from 2016. PANGAEA

Repositorien: Code

Code

A software repository is a storage location for software packages and is typically managed by source control or repository managers. https://en.wikipedia.org/wiki/Software_repository.

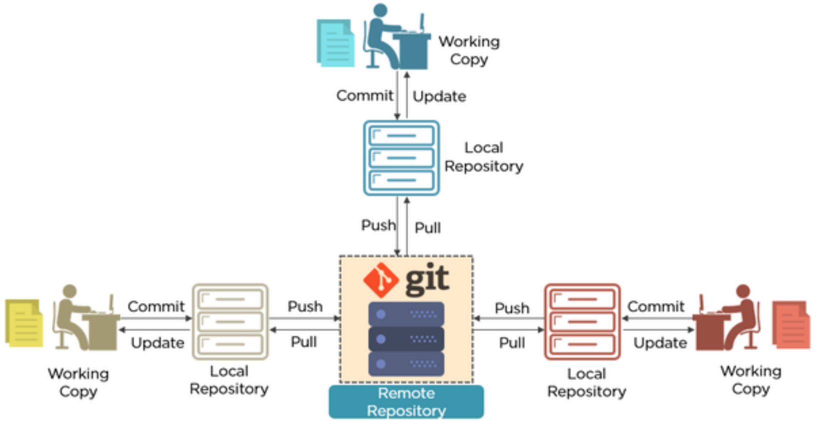
Versionskontrolle



https://www.gbnews.ch/version-control-system-a-non-technical-introduction/

Repositorien: Code

Git

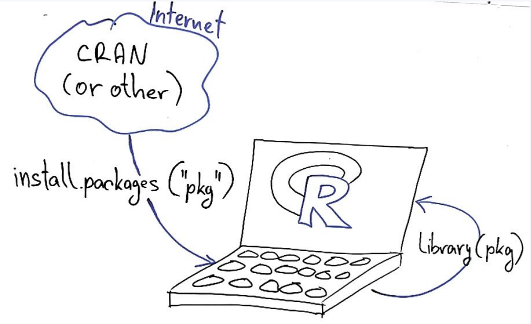


https://www.gbnews.ch/version-control-system-a-non-technical-introduction/

... I name all my projects after myself. First Linux, now Git. Linus Torwald

Repositorien: Code

Repository manager



https://davidzeleny.net/wiki/doku.php/recol:r_packages

Zentrale Repositorien

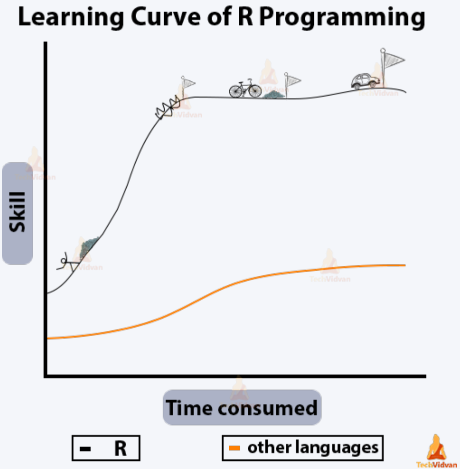
- CRAN
- Bioconductor

Entwicklerrepositorien

- R-Forge
- GitHub

R

Lingua franca for data science



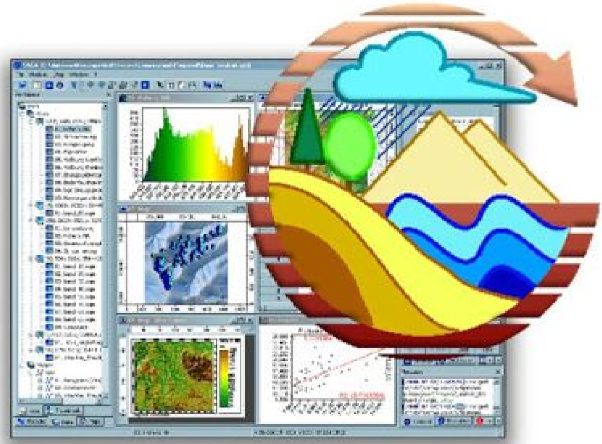
https://techvidvan.com/tutorials/r-tutorial/

Why R?

Open-source · Aktive Community · Umfassende und standardisierte Package-Bibliotheken (CRAN, Bioconductor) · plattform-übergreifende Kompatibilität · Machine Learning · Raster- und Vektordatenanalyse · Big Data · GIS-Operationen . . .

SAGA-GIS

System for Automated Geoscientific Analyses

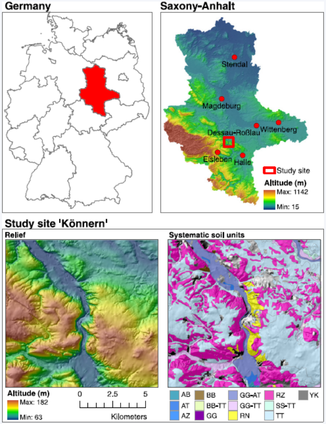


https://sourceforge.net/projects/saga-gis/

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geosci. Model Dev. 8, 1991–2007

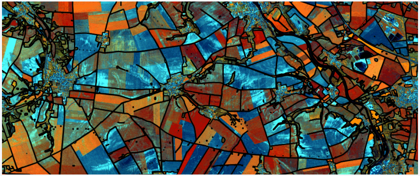
Übung

Messtischblatt Könnern



Möller, M., Koschitzki, T., Hartmann, K.-J., Jahn, R., 2012. Plausibility test of conceptual soil maps using relief parameters. CATENA 88, 57–67

Feldblöcke



Bewertung der potenziellen Wassererosion

$$\mathbf{A} = R \times K \times LS$$

mit **A** = langjähriger mittlerer Bodenabtrag bezogen auf eine Fruchtfolge, R = Niederschlagserosivitätsfaktor, K = Bodenerodierbarkeitsfaktor, LS = Hangneigungs- und Hanglängenfaktor

Übung

https://github.com/FLFgit/ABAG

- Alle Funktionen und Daten sind im Github-Repositorium abgelegt.
- Über die Datei WrappeR.R können alle Funktionen aufgerufen werden.
- Voraussetzung ist die Installation von SAGA-GIS.

Gruppen

Jede Gruppe führt den Workflow am Beispiel einer DGM-Variante aus!

1. DGM10_EPSG31468.asc
2. DGM40_EPSG31468.asc
3. DGM90_EPSG31468.asc

Übung

1. Berechnen Sie verschiedene Fließakkumulationsvarianten!

Function fLSrsaga.R/fLSrsagacmd.R (Parameter $M.CA =$ Fließakkumulations- methode)

- CA0 Deterministic 8 ($M.CA = 0$)
- CA4 Multiple flow direction ($M.CA = 4$)

Bircher, P., Liniger, H.P., Prasuhn, V., 2019. Comparing different multiple flow algorithms to calculate RUSLE factors of slope length (L) and slope steepness (S) in Switzerland. Geomorphology 346, 106850

2. Verwenden Sie die zwei Fließakkumulationsvarianten für die Berechnung von vier LS-Faktorenvarianten!

Function fLSrsaga.R/fLSrsagacmd.R (Parameter $M.LS1 =$ LS-Faktorenberechnungsvariante)

- LS10_CA0, LS10_CA4 ⇒ Moore et al. 1991 ($M.LS1 = 0$)
- LS11_CA0, LS11_CA4 ⇒ Desmet and Govers 1996 ($M.LS1 = 1$)

Übung

3. Berechnen Sie zwei *LS*-Faktorenvarianten mit Berücksichtigung von Feldblockgrenzen!

Function fLSrsaga.R/fLSrsagacmd.R (Parameter $M.LS2 =$ *LS*-Faktorenberechnungsvariante)

- LS20 ⇒ Moore et al. 1991 ($M.LS2 = 0$)
- LS21 ⇒ Desmet and Govers 1996 ($M.LS2 = 1$)

4. Führen Sie eine Zonal Statistics-Operation mit den Faktoren(varianten) durch

Function fZonalStatistics.R (Parameter $COL.NAME =$ zu vergebender Spaltenname)

- sechs *LS*-Faktorenvarianten (LS10_CA0, LS10_CA4, LS11_CA0, LS11_CA4, LS20, LS21)
- zwei *K*-Faktorenvarianten (KBS, KVBK)
- eine R-Faktorenvariante (R) (Auerswald et al. 2019)

Übung

4. Vergleichen Sie die ABAG-Ergebnisvarianten!

Function fPlotR.R

- Generieren Sie zwölf verschiedene ABAG-Varianten durch die Multiplikation der Faktoren.
- Die Funktion fPlotR.R berechnet Genauigkeitsmaße (R^2 / $RMSE$), die auf dem paarweisen Vergleich von Datenreihen (hier: ABAG-Spalten der Feldblock-Datei) basieren.

Erstellen und interpretieren Sie eine Tabelle mit den Genauigkeitsmaßen für alle Variantenkombinationen!

Variantenpaar	R^2	$RMSE$
LS10_CA0*KBS*R/LS10_CA0*KVBK*R	...	...
LS10_CA0*KBS*R/LS11_CA0*KBS*R	...	...
...	...	...