

# Bewertung der potenziellen Wassererosion mit R und SAGA-GIS

Markus Möller<sup>1</sup> & Martin Volk<sup>2</sup>

<sup>1</sup> Julius Kühn-Institut, Institut für Pflanzenbau und Bodenkunde, Braunschweig

<sup>2</sup> Helmholtz-Zentrum für Umweltforschung GmbH – UFZ, Department Landschaftsökologie, Leipzig

## 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:

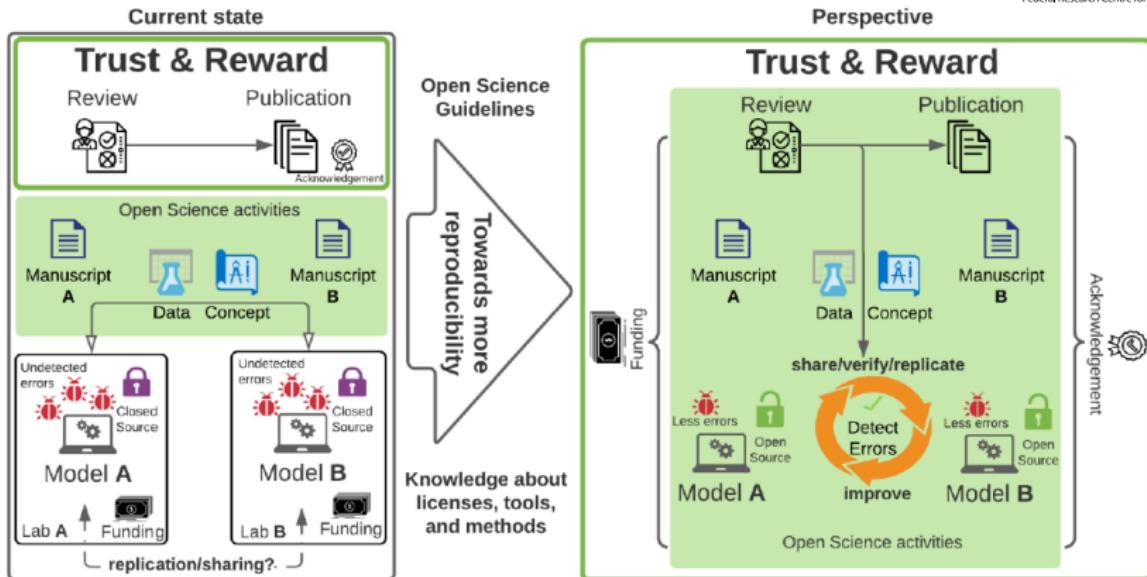
- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

### 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

# Reproducibility/Provenance



(1) Only modeling concepts and data are shared. Models and their code are seldomly shared and results not verified or replicated. This affects the trust in their results and also leads to undetected 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.

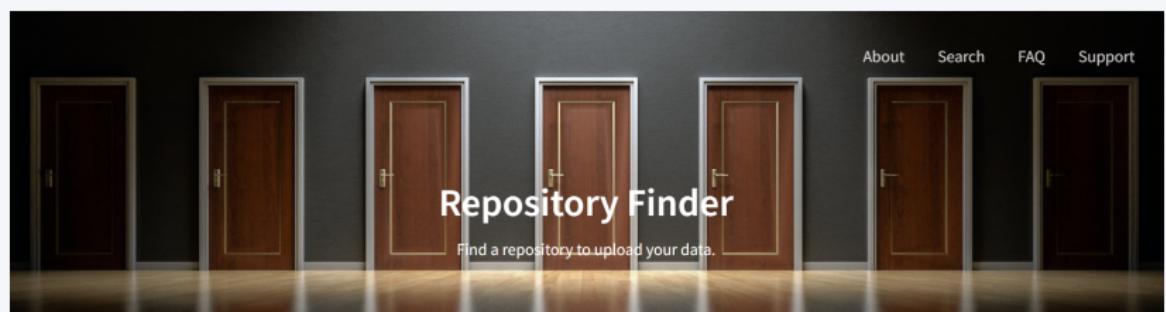
(3) This leads to an interactive process of verification and replication that reduces errors, increases research output, and fosters trust in the model results. However, this needs to be enabled by funding opportunities and an acknowledgement of efforts. Likely, this also requires additional qualified 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/>).

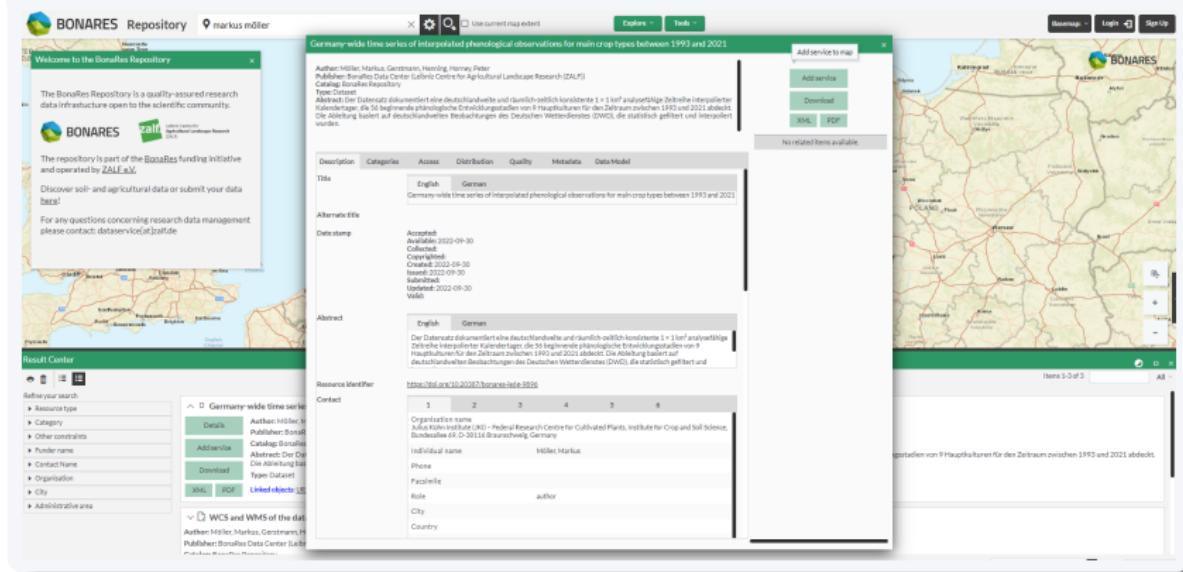
## Repository finder



<https://repositoryfinder.datacite.org>

# Repositorien: Daten

## BonaRes



The screenshot displays the BonaRes Repository interface. On the left, there's a search bar and a sidebar with navigation links like 'Resource type', 'Category', 'Other constraints', 'Provider name', 'Contact Name', 'Organization', 'City', and 'Administrative area'. The main content area shows a dataset detail page for 'Germany-wide time series of interpolated phenological observations for main crop types between 1993 and 2021'. The page includes tabs for Description, Categories, Access, Distribution, Quality, Metadata, and Data Model. The 'Description' tab is active, showing the title 'Germany-wide time series of interpolated phenological observations for main crop types between 1993 and 2021', a date stamp from 2022-09-30, and an abstract in English and German. The abstract describes the dataset as a 3x3 km grid of phenological observations for 9 main crop types between 1993 and 2021. Below the abstract is a resource identifier (DOI: <https://doi.org/10.20377/bonares-1674-5926>) and a contact section with a table for 'Organisational name', 'Individual name', 'Phone', 'Facsimile', 'Role', 'City', and 'Country'. On the right, there's a map viewer showing the spatial distribution of the data across Germany, with a legend indicating the 9 crop types. A modal window titled 'Add service to map' is open, showing options to 'Add service', 'Download', 'XML', and 'PDF'.

Möller, M., Gerstmann, H., Horney, P., 2022. Germany-wide time series of interpolated phenological observations for main crop types between 1993 and 2021.

# Repositorien: Daten

## Pangaea

 **PANGAEA.**  
Data Publisher for Earth & Environmental Science

Not logged in 

SEARCH SUBMIT HELP ABOUT CONTACT

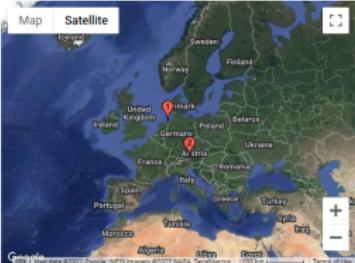
**Citation:** Preidl, Sebastian; Lange, Maximilian; Doktor, Daniel (2020): Land cover classification map of Germany's agricultural area based on Sentinel-2A data from 2016. PANGAEA,  <https://doi.org/10.1594/PANGAEA.910837>

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

[BIB Citation](#) [TeX 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](https://en.wikipedia.org/wiki/Software_repository).

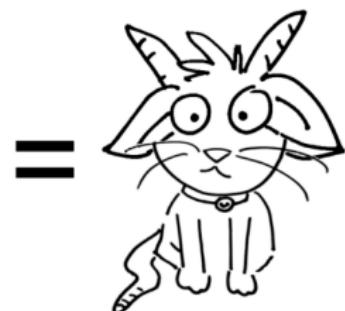
## Versionskontrolle

### Edit Conflict: Favourite Animal

Your version



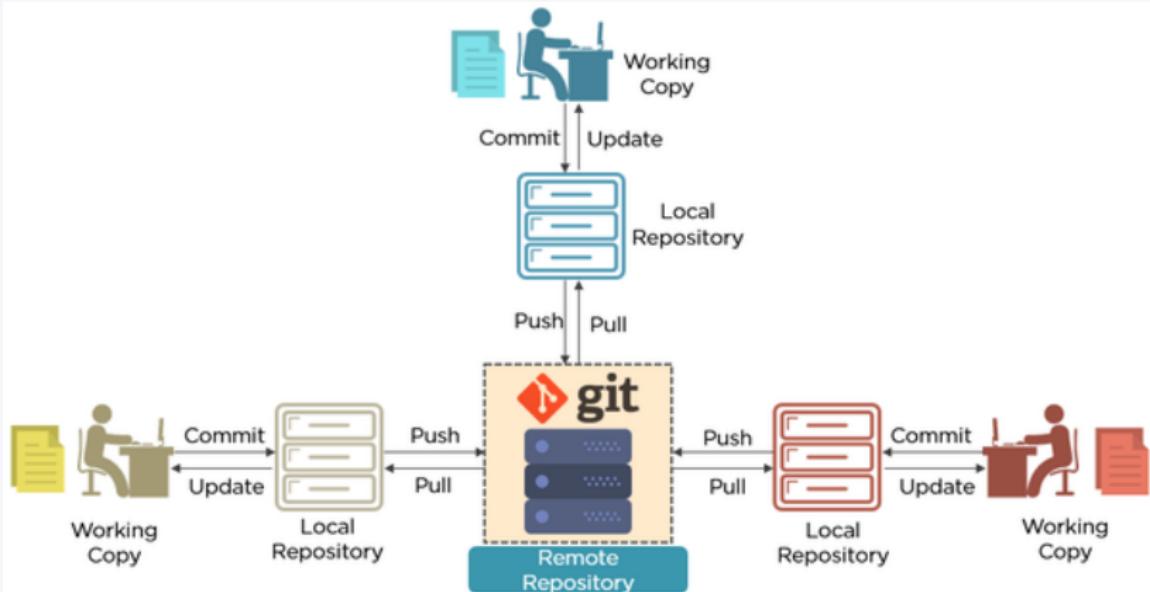
Cat Lover's version



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

## Repositorien: Code

Git

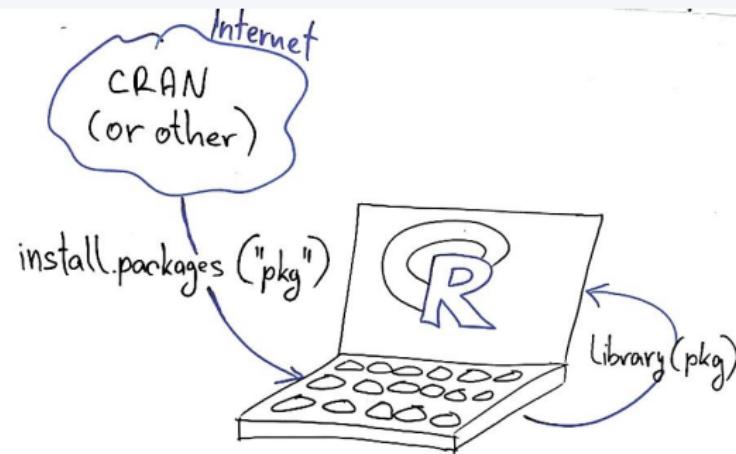


<https://www.gbnews.ch/version-control-system-a-non-technical-introduction/>

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

# Repositorien: Code

## Repository manager



[https://davidzeleny.net/wiki/doku.php/recoll:r\\_packages](https://davidzeleny.net/wiki/doku.php/recoll:r_packages)

## Zentrale Repositorien

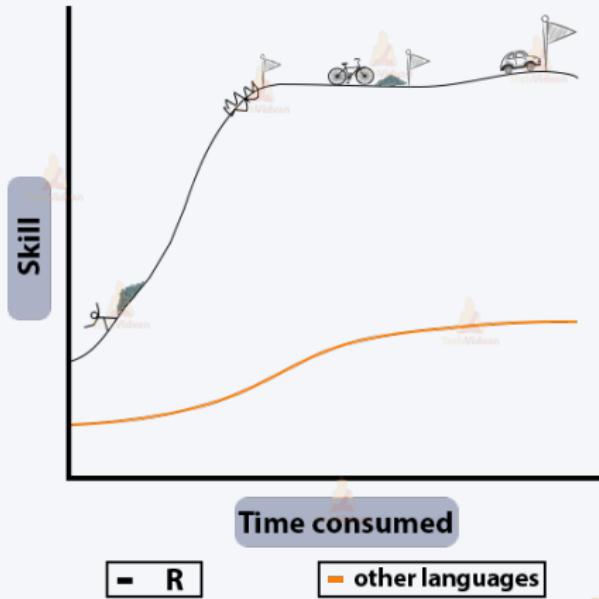
- CRAN
- Bioconductor

## Entwicklerrepositorien

- R-Forge
- GitHub

## Lingua franca for data science

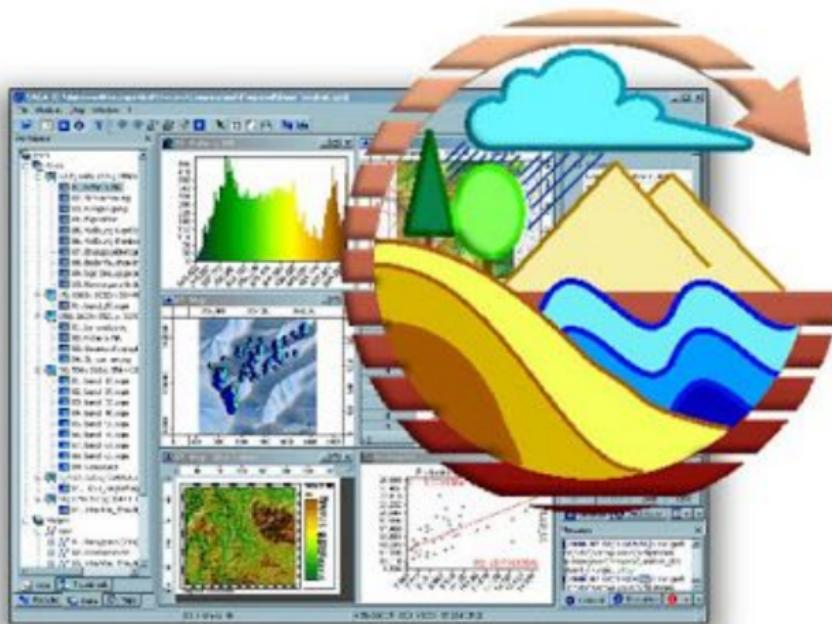
## Learning Curve of R Programming



## 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 ...

## System for Automated Geoscientific Analyses

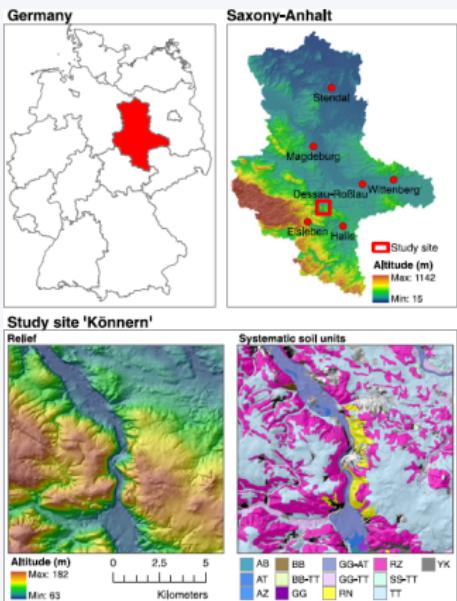


Conrad, O., Bechtel, B.,  
Bock, M., Dietrich, H.,  
Fischer, E., Gerlitz, L.,  
Wehberg, J., Wichmann,  
V., Böhner, J., 2015.  
System for Automated  
Geoscientific Analyses  
(SAGA) v. 2.1.4. Geosci.  
Model Dev. 8, 1991–2007

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

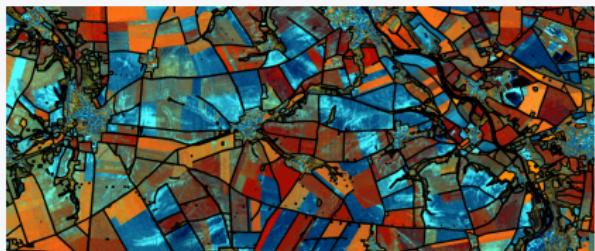
# Übung

## 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  $\mathbf{A}$  = langjähriger mittlerer Bodenabtrag bezogen auf eine Fruchtfolge,  $R$  = Niederschlagserosivitätsfaktor,  $K$  = Bodenerodierbarkeitsfaktor,  $LS$  = Hangneigungs- und Hanglängenfaktor

## Installation

- **R/Rstudio**

<https://posit.co/download/rstudio-desktop/>

- **git** <https://git-scm.com/download/win>

- **SAGA-GIS**

<https://sourceforge.net/projects/saga-gis/files/>

[https://github.com/JKI-GDM/LSS4\\_ABAG\\_SS2023](https://github.com/JKI-GDM/LSS4_ABAG_SS2023)

- Alle Funktionen und Daten sind im Github-Repositorium abgelegt.
- Über die Datei `WrappeR.R` können alle Funktionen aufgerufen werden.

## Gruppen

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

- ① [DGM10\\_EPSG31468.asc](#)
- ② [DGM40\\_EPSG31468.asc](#)
- ③ [DGM90\\_EPSG31468.asc](#)

## Aufgaben

- ① Erstellen Sie jeweils für eine DGM-Variante vier ABAG-Varianten!
- ② Vergleichen Sie ABAG-Varianten mittels linearer und nichtlinearer Regression!
- ③ Erstellen Sie eine Quarto-Dokumentation!

Für die LS-Faktorenberechnung existieren verschiedene Varianten!

## LS-Faktor: *Fließakkumulation*

- Deterministic 8
- Multiple flow direction

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

## LS-Faktor: *Berechnungsvarianten*

- Moore et al. 1991
- Desmet and Govers 1996

1. Berechnen Sie zwei *LS*-Faktorenvarianten mit und ohne Berücksichtigung von Feldblockgrenzen!

Function [fLSrsagacmd.R](#)

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

Function [fZonalStatistics.R](#)

- zwei *LS*-Faktorenvarianten (nB-LS, wB-LS)
- zwei *K*-Faktorenvarianten (KBS, KBK)
- eine R-Faktorenvariante (R) ([Auerswald et al. 2019](#))