State of R in Hydrological Modelling

Jairo A. Torres
Edzer Pebesma

Institute for Geoinformatics, ifgi, University of Münster, Germany

2nd OpenWater symposium and workshops
September 16*th*, 2013

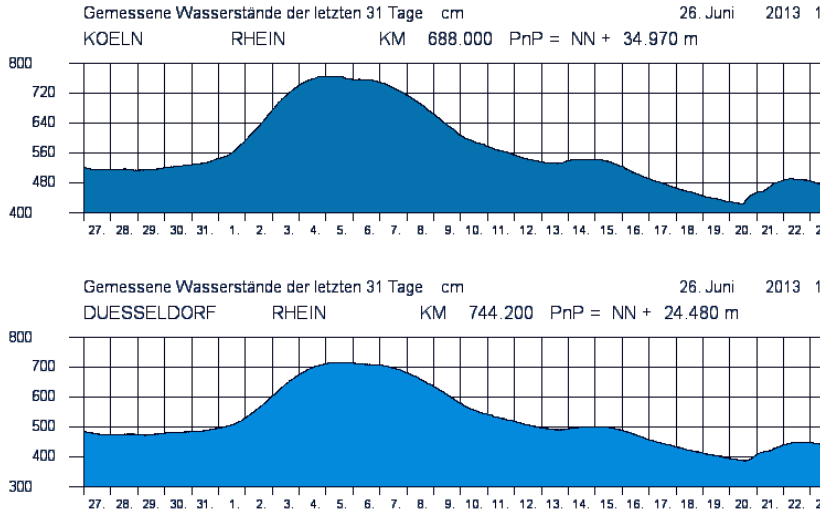
Justification

ifgi
Institute for Geoinformatics
University of Münster

- Material flows to coastal waters: complex interaction
  - catchment boundaries,
  - human activities therein,
  - trade of materials.
- Trans-boundary processes
  - residence time,
  - transport and fate of WQ determinants
    - physical,
    - chemical,
    - microbiological.
- WQ status: global implications

Purpose

- Models for real-time flow prediction
  - improve decision making [Beven, 2012]
    - water resources planning
    - flood protection
    - mitigation of contamination



German Federal Institute of Hydrology - BfG. 26.06.2013 19:40

Purpose

ifgi
Institute for Geoinformatics
University of Münster

- model comparison in structure, calibration methods and simulation events is essential for choosing objectively the suitable model configuration in hydrological modelling

- a novel, versatile, and open source language is provided by the R Project for Statistical Computing
  [Ihaka and Gentleman, 1996, R Development Core Team, 2013]

Motivation

ifgi
Institute for Geoinformatics
University of Münster

- Several packages have been developed in R code for developing, implementing and evaluating different hydrological modelling tools and frameworks
- R language is an emerging environment where research and development are the primary purpose
- R is mostly being developed by its users, i.e. domain experts, not by programmers without hydrology knowledge (this implies commitment and long-term involvement, but also at times messy code)

Motivation

ifgi
Institute for Geoinformatics
University of Münster

- Comprehensive modelling framework
  - Uncertainty analysis
  - Identify primary physical controls
  - Coupling inland hydrological models with coastal
  - Regional, transboundary and global scales

Motivation

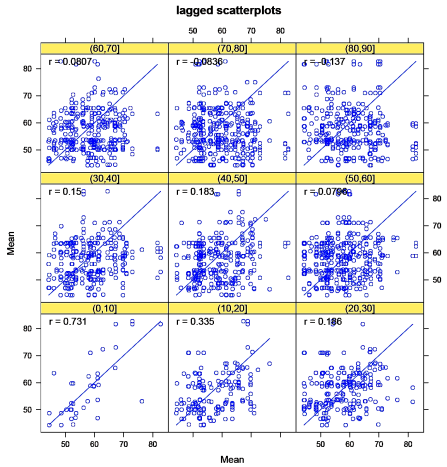
ifgi
Institute for Geoinformatics
University of Münster

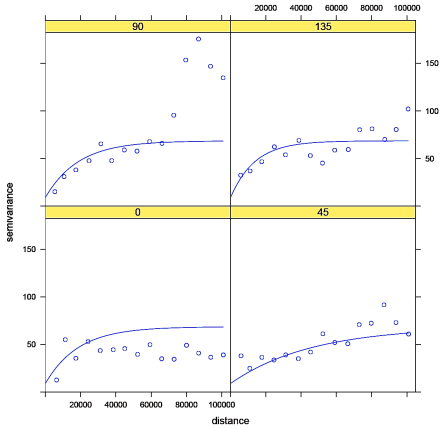
- Open source data and software in R, for:
  - statistical analysis and graphics
  - suitable and reproducible modelling framework
  - basic data manipulation (import, export, selection, joining of tables)
  - spatial interpolation
  - calibration and simulation
  - uncertainty analysis
  - geospatial capabilities for querying, updating, sharing and visualization of data, methods and results
  - general tools as support in hydrological analysis

Useful functionality in R

ifgi
Institute for Geoinformatics
University of Münster

- R as statistical computing and graphics environment, provides a large amount of generic functionality useful in hydrological modelling:
  - MCMC outputs
  - print quality visualising
  - time series
  - spatial and spatio-temporal data
  - geostatistics, downscaling, upscaling,
  - aggregation and disaggregation





Useful functionality in R

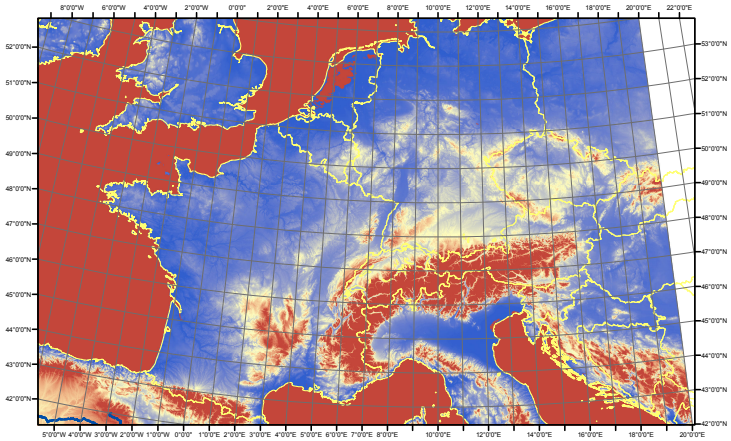
ifgi
Institute for Geoinformatics
University of Münster

- Higher level functionality:
  - cross platform
  - easy dissemination and new extension
  - reproducible research
  - support
    open mailing lists, forums, companies, books
  - open and active community

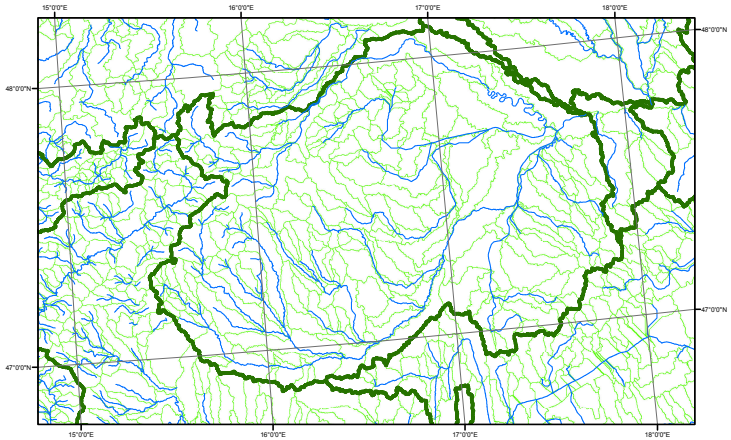
Data management and R

ifgi
Institute for Geoinformatics
University of Münster

- reproducibility in research:
  - share a common data model
  - independence of modelling framework



- standardised open data for research:
  - ECRINS dataset
  - EOBS dataset
  - web services



Database access and manipulation

ifgi
Institute for Geoinformatics
University of Münster

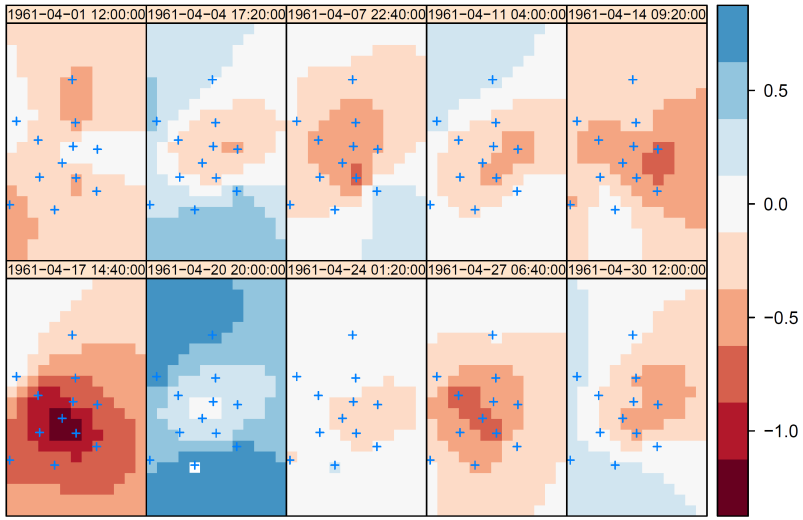
- `foreign` package [R Core Team et al., 2013]
  - `read.dbf`: importing .dbf file $\longrightarrow$ R dataframe
- `sp` package [Pebesma and Bivand, 2012]
  - ESRI polygon shapefile $\longrightarrow$ `SpatialPolygons`
  - `SpatialPoints` & `SpatialLines`
- geospatial data abstraction library (GDAL) $\longrightarrow$ `rgdal` [Bivand et al., 2013]
  - multiple classes of `sp`
  - projection/transformation operations $\longrightarrow$ PROJ.4 library
  - ESRI Shapefiles and PGDBs .mdb files $\longrightarrow$ OGR library
- `igraph: network analysis and visualization (Csardi, 2013)`

Spatial and temporal analysis

ifgi
Institute for Geoinformatics
University of Münster

`spacetime` package

- container for spatio-temporally referenced
- time series of points
- spatio-temporal fields
- catchment time series



[Pebesma, 2012]

Web services and interoperability

ifgi
Institute for Geoinformatics
University of Münster

- `RObsDat` by Reusser and Liersch (this Symposium)

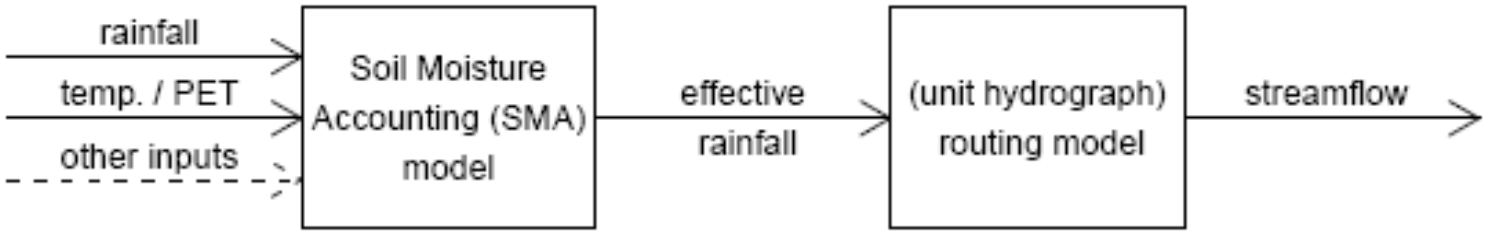
- R provides generic interfaces to `curl` and `libxml`
- `RCurl` package [Lang, 2013a]
- `XML` package [Lang, 2013b]
  - access web services
  - `sos4R` [Nüst et al., 2011]
  - retrieve sensor data from an OGC sensor observation service
  - retrieve data in `waterML`.
- `Rserve` package $\longrightarrow$ web service interface [Urbanek, 2013]
- WPS4R [Hinz et al., 2013]: higher-level, generic OGC WPS interface

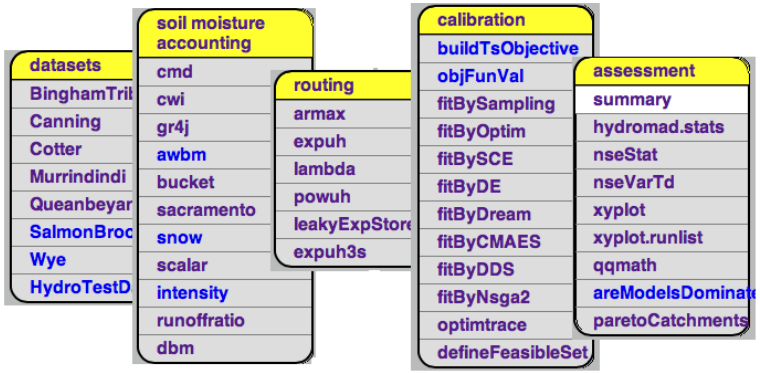
  52°North WPS

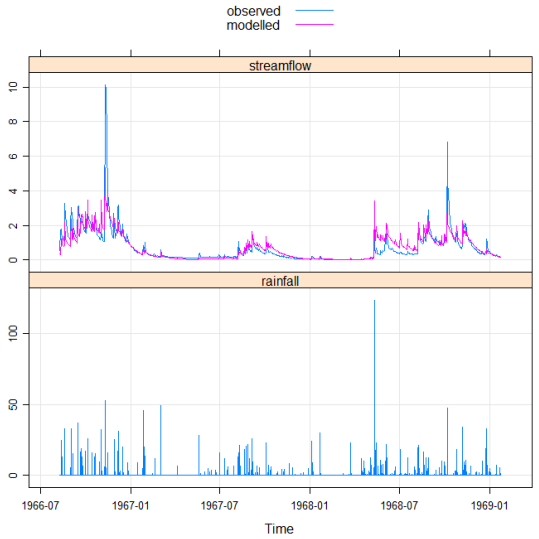
Hydrological Model Assessment and Development, hydromad

ifgi
Institute for Geoinformatics
University of Münster

[Andrews et al., 2011]







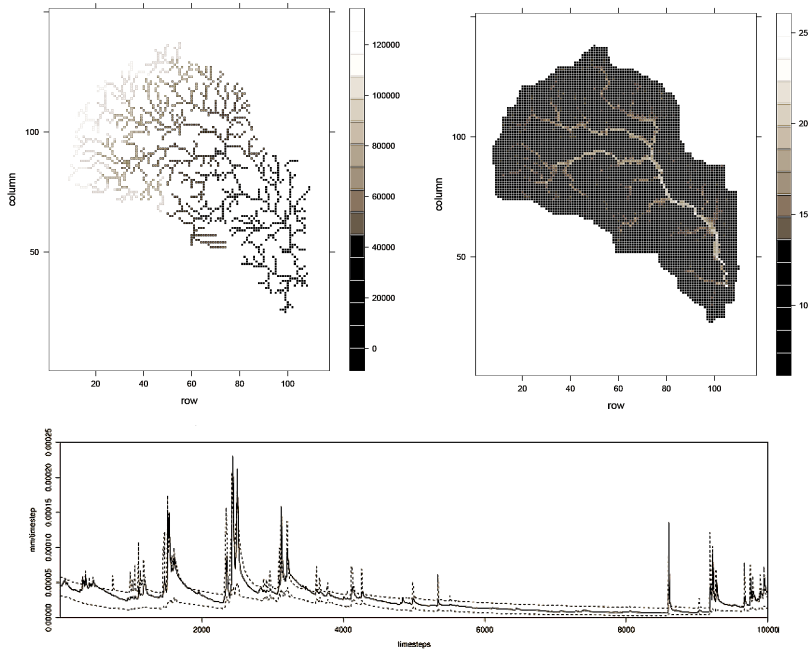
IHACRES CWI model with exponential unit hydrograph

`topmodel`, semi-distributed model

ifgi
Institute for Geoinformatics
University of Münster

[Beven et al., 1995, Beven, 1997]

- 1995 Fortran version by Keith Beven
- R version by [Buytaert, 2012]
- new functionality developed as part of the `RHydro` package on R-Forge



Further hydrological functionality

ifgi
Institute for Geoinformatics
University of Münster

- there are nearly 5000 packages on CRAN
- several additional packages published at CRAN in the area of hydrology
  - `HydroMe` package
  - `hydroTSM` package
  - hydroGOF
  - hydroPSO
  - EcoHydRology
  - wasim
  - seacarb
  - StreamMetabolism
  - oce
  - nsRFA
  - boussinesq
  - rtop

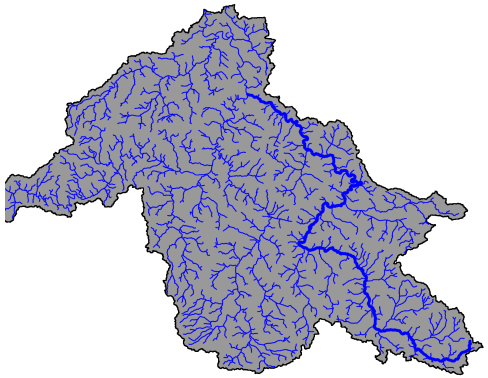
`http://cran.r-project.org/web/views/Environmetrics.html`

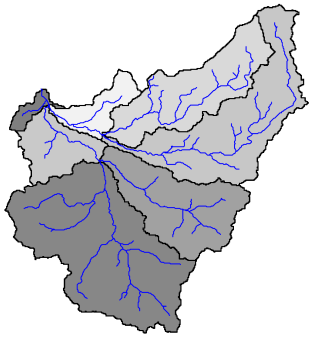
Current contribution

ifgi
Institute for Geoinformatics
University of Münster

`watersheds` package

- spatial analysis for watersheds aggregation
- spatial drainage networks analysis
- based on `sp` and `rgeos` packages (Pebesma & Bivand, 2005-2012)





Looking forward:

- Modelling framework for real-time flow prediction
  - flood assessment
  - mitigation of contamination

Conclusions

ifgi
Institute for Geoinformatics
University of Münster

- we present an introduction to the state-of-art of hydrological modelling as is developed in the open source software R.
- this constitutes a starting point for research because it grants a suitable programming platform where standardised tests and comparisons of models is possible, searching for reproducibility of methods and results as is often required in the context of science and research.

Thank you!

Questions?

arturo.torres@uni-muenster.de

Bibliography I

Andrews, F. T., Croke, B. F. W., and Jakeman, A. J. (2011).
An open software environment for hydrological model assessment and development.
Environmental Modelling and Software, 26.

Beven, K., Lamb, R., Quinn, P., Romanowicz, R., and Freer, J. (1995).
TOPMODEL.
In Sing, V. P., editor, *Computer Models of Watershed Hydrology. Water Resources Publications, Colorado. pp. 627-668.*

Beven, K. J. (1997).
Distributed Hydrological Modelling: Applications of the TOPMODEL Concept.
Wiley.

Beven, K. J. (2012).
Rainfall-Runoff Modelling: The Primer.
Wiley-Blackwell.
Lancaster University, UK.

Bivand, R., Keitt, T., Rowlingson, B., Pebesma, E., Sumner, M., and Hijmans, R. (2003-2013).
Package "rgdal": Bindings for the Geospatial Data Abstraction Library.
The Comprehensive R Archive Network, CRAN, 0.8-9 edition.

Buytaert, W. (2012).
Package 'topmodel'.
The Comprehensive R Archive Network, CRAN.

Bibliography II

Hinz, M., Nüst, D., Proß, B., and Pebesma, E. (2013).
Spatial Statistics on the Geospatial Web.
short paper, Agile 2013 proceedings.

Ihaka, R. and Gentleman, R. (1996).
R: a languague for data analysis and graphics.
Journal of Computational and Graphical Statistics, 5:299–314.

Lang, D. T. (2000-2013b).
Package "XML": Tools for parsing and generating XML within R and S-Plus.
The Comprehensive R Archive Network, CRAN, 3.96-1.1 edition.

Lang, D. T. (2004-2013a).
Package "RCurl": General network (HTTP/FTP/...) client interface for R.
The Comprehensive R Archive Network, CRAN, 1.95-4.1 edition.

Nüst, D., Stasch, C., and Pebesma, E. J. (2011).
Connecting R to the Sensor Web.
In Geertman, S., Reinhardt, W., and Toppen, F., editors, *Advancing Geoinformation Science for a Changing World*, Proceedings of AGILE, pages 227 – 246. Springer Lecture Notes in Geoinformation and Cartography.

Pebesma, E. (2012).
spacetime: Spatio-temporal data in r.
Journal of Statistical Software, 51, Issue 7.
http://www.jstatsoft.org/.

Bibliography III

Pebesma, E. and Bivand, R. (2005-2012).
Package "sp": classes and methods for spatial data.
The Comprehensive R Archive Network, CRAN, 0.9-99 edition.

R Core Team, Bivand, R., Carey, V. J., DebRoy, S., Eglen, S., Guha, R., Lewin-Koh, N., Myatt, M., Pfaff, B., Warmerdam, F., Weigand, S., and Free Software Foundation, Inc. (1999-2013).
Package "foreign": Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, dBase,...
The Comprehensive R Archive Network, CRAN, 0.8-54 edition.

R Development Core Team (2013).
R: A Language and Environment for Statistical Computing.

Urbanek, S. (2006-2013).
Package "Rserve": Binary R server.
The Comprehensive R Archive Network, CRAN, 0.6-8.1 edition.