Karlsruher Institut für Technologie

Dynamical clustering: A new approach to make distributed (hydrological) modeling more efficient by dynamically detecting and removing redundant computations

Uwe Ehret, Rik van Pruijssen, Marina Bortoli
Ralf Loritz, Elnaz Azmi, Erwin Zehe

Karlsruhe Institute of Technology

Motivation

- Distributed modelling is attractive as it accepts distributed observables and yields distributed predictions, but it is computationally expensive.
- Possible solutions for the computation problem are massive parallel computing, adaptive time stepping, or adaptive gridding
- In adaptive gridding, only neighboring model elements can be merged
- We could make distributed modeling more efficient by if we could also merge non-neighboring model elements

Hypotheses

- Similarity of model elements,
  - is a precondition for merging them,
  - is a function of their forcing, structure, states and fluxes,
  - and varies over time.
- If we can identify groups of similar model elements,
  - we can select a few representatives from each group,
  - calculate the dynamical evolution of these few representatives,
  - and assign their results to all model elements in the group.
- As similarity varies over time, re-grouping will be required from time to time
- If the computational extra costs for grouping are low, we will save computation time

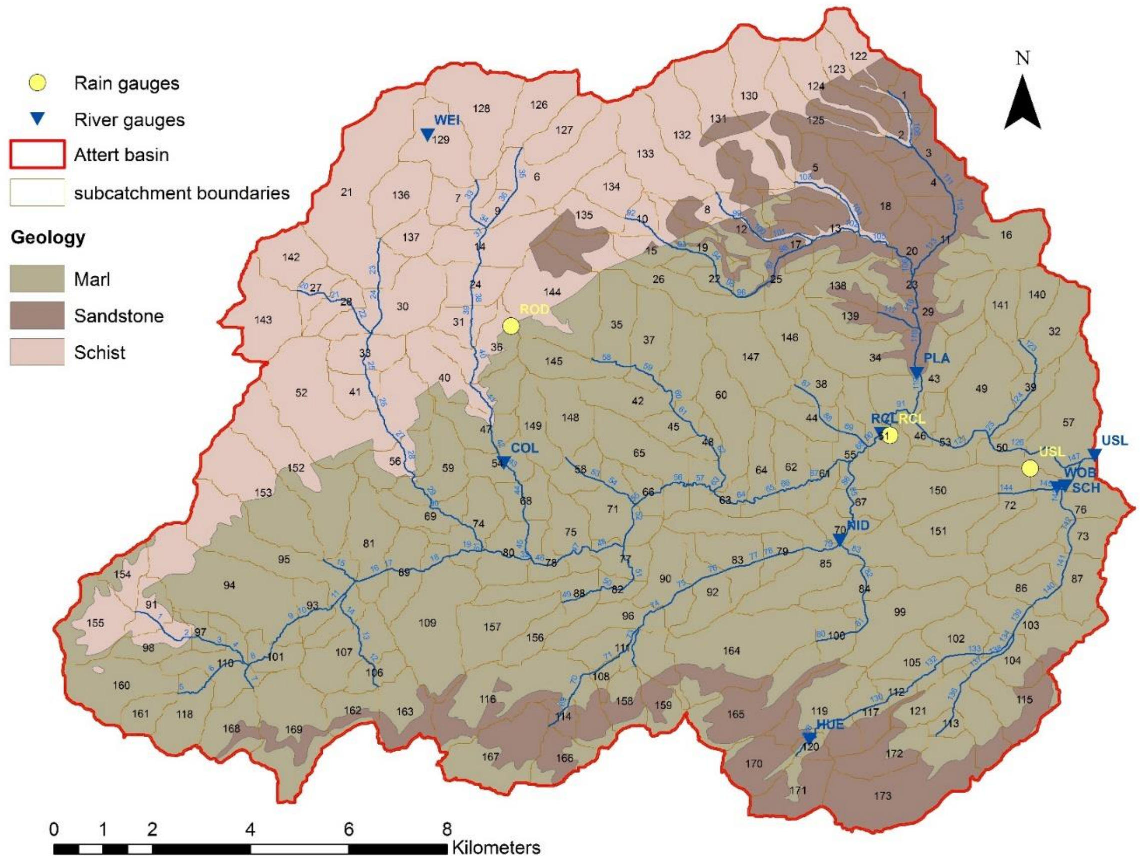
- Similarity: Models states are normalized and discretized (binned). Two model elements are similiar if their normalized states fall into the same bin.

Dynamical clustering

- Main steps
  - Find groups of similar model elements
  - From each group, randomly select a few representatives
  - Run the model for these representatives only
  - Assign results from the representatives to remaining group members
  - If representatives of a group become dissimilar → new grouping

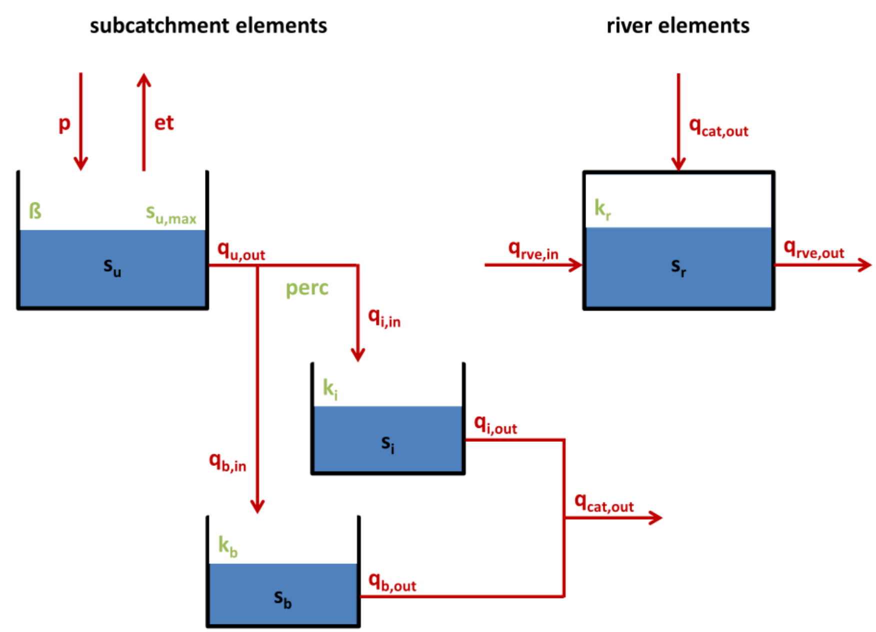
Test application

- Real-world system: Attert basin in Luxembourg (288 km²)



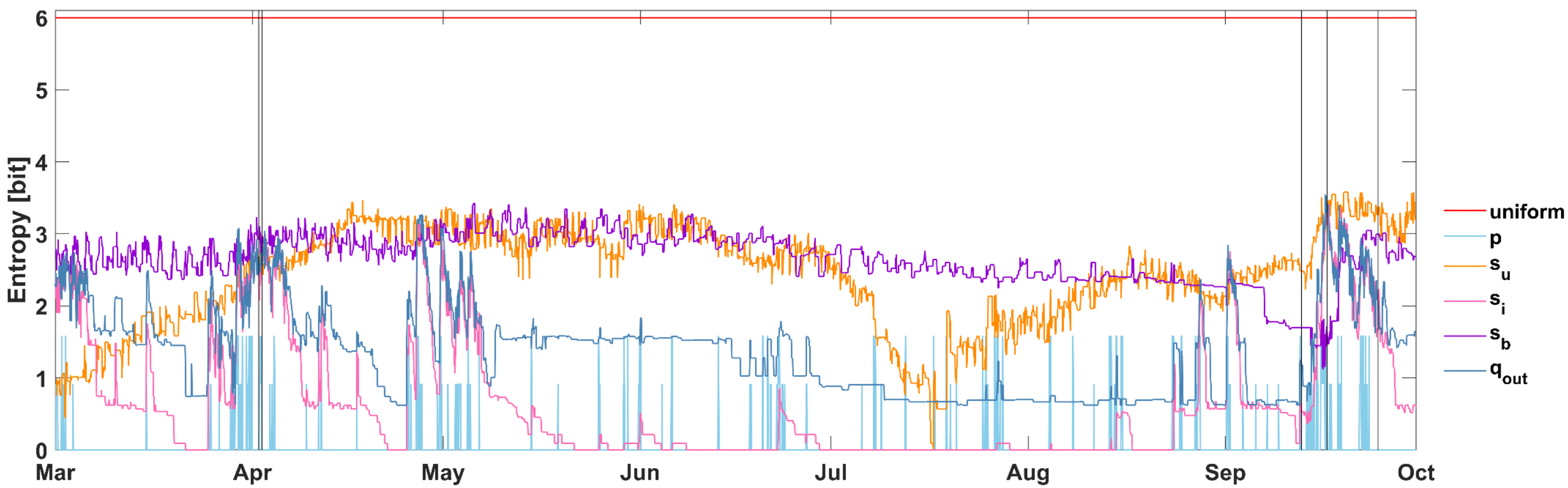
Test application

- Hydrological model: SHM
  - Distributed (173 sub catchments)
  - Conceptual water balance model (HBV-like)



Results

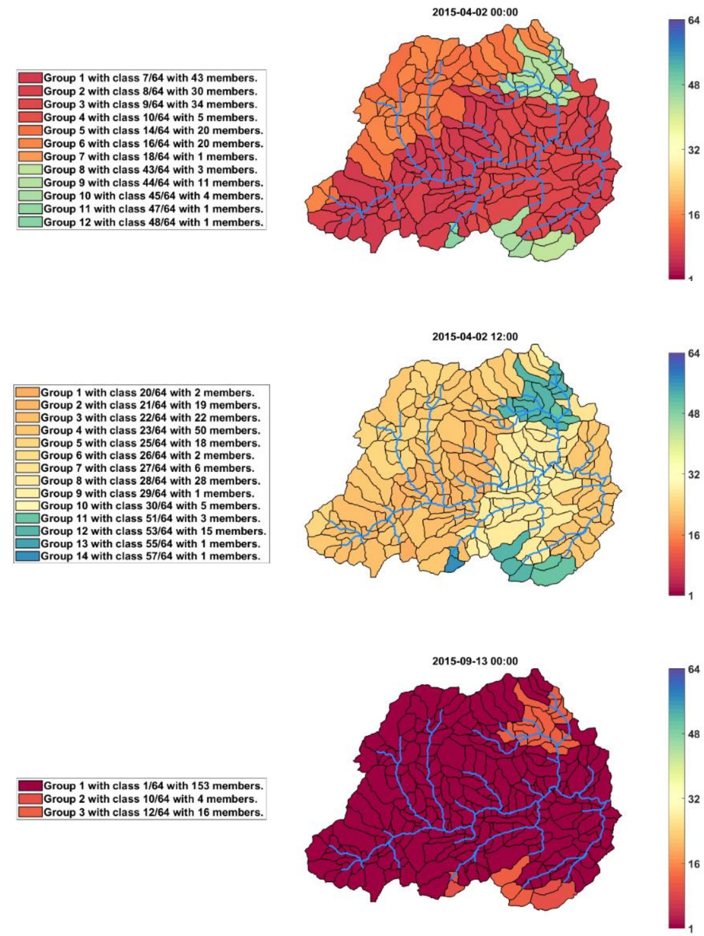
- Dynamical similarity among sub catchments
  - Is measured by the entropy of the distribution of sub catchment states



- → There is considerable similarity (entropies are lower than the benchmark entropy of the uniform distribution)
- → Similarity varies with time

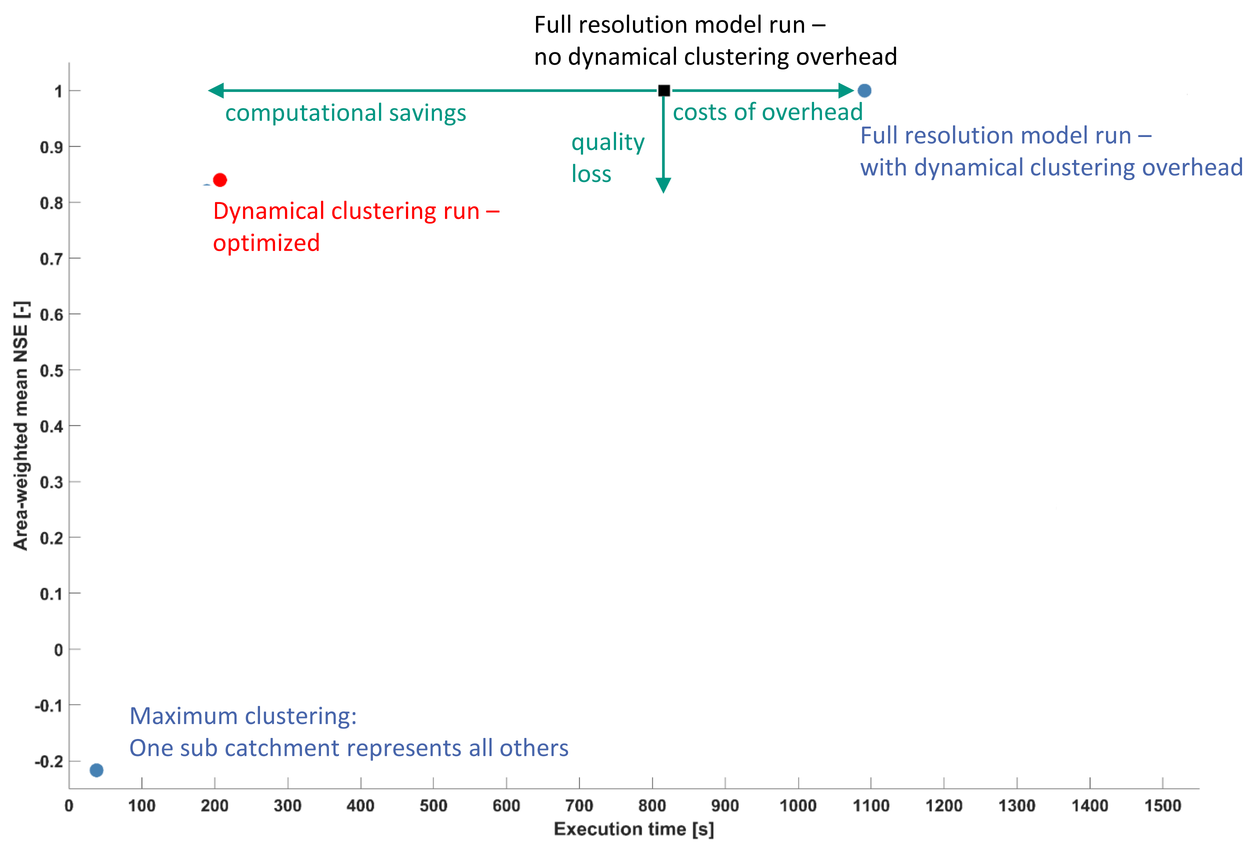
Results

- Dynamical similarity among sub catchments
  - The strength and the dominant controls of the spatial patterns of similarity vary with time
  - Geology and rainfall pattern are the main controls



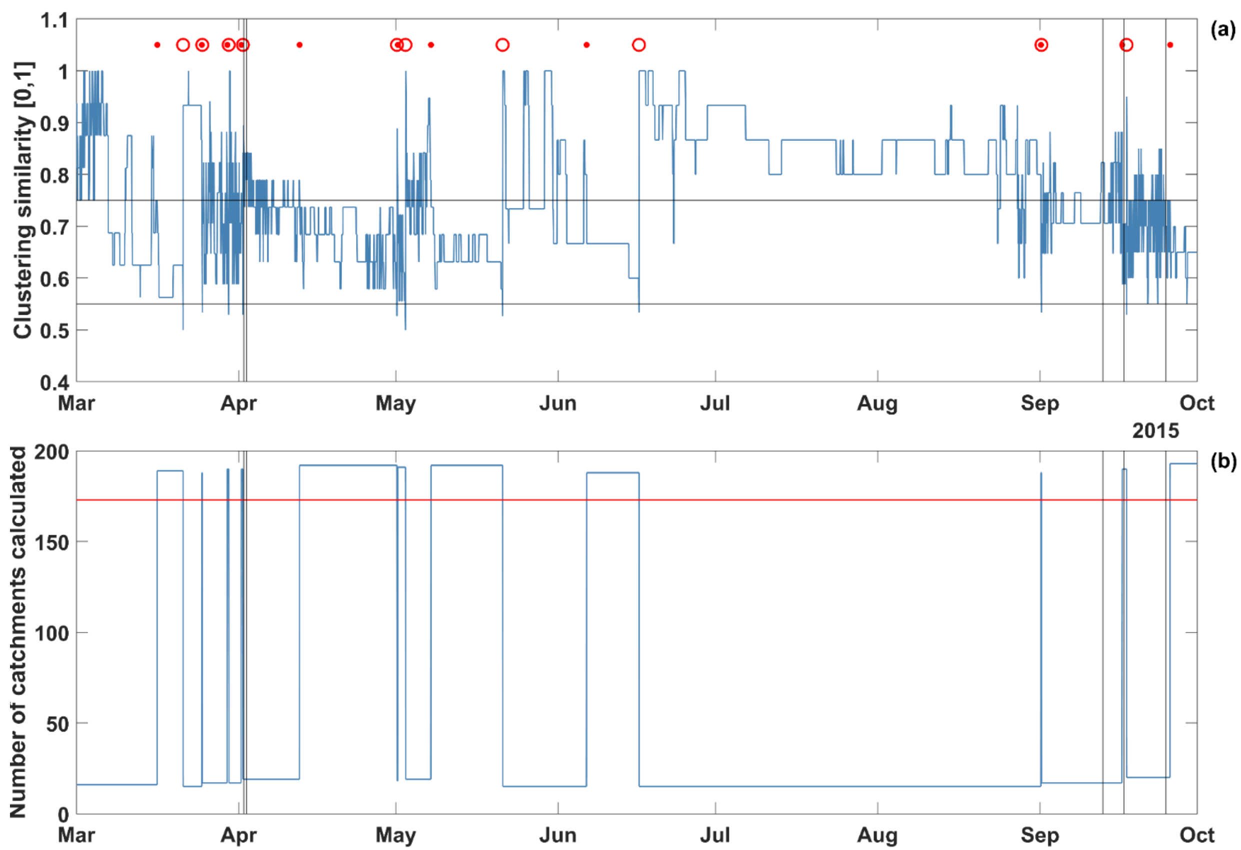
Results

- Dynamical clustering – computational savings vs. quality loss
  - Computational costs are measured by model execution time [s]
  - Model quality is measured by Nash-Sutcliffe efficiency



Results

- Dynamical clustering – Reclustering
  - Red circles show times of re-clustering (upper plot)
  - The required number of cluster representatives varies with time (lower plot)



Further information

- For details, please see the manuscript in HESS discussions at https://www.hydrol-earth-syst-sci-discuss.net/hess-2020-65/

Hydrology and Earth System Sciences
An interactive open-access journal of the European Geosciences Union

| EGU.eu | EGU Publications | EGU Highlight Articles | Contact | Imprint | Data protection |

Submit a manuscript | Manuscript tracking

About
Editorial board
Articles
Special issues
Highlight articles
Subscribe to alerts
Peer review
For authors
For reviewers
Imprint
Data protection

User ID
Password
New user? | Lost login?

Follow @EGU_HESS

Journal metrics
IF 4.936
IF 5-year 5.615
CiteScore 4.94
SNIP 1.612
IPP 4.70

https://doi.org/10.5194/hess-2020-65

Discussion papers

Abstract | Discussion | Metrics

Submitted as: research article — 28 Feb 2020

Review status
This preprint is currently under review for the journal HESS.

Adaptive clustering: A method to analyze dynamical similarity and to reduce redundancies in distributed (hydrological) modeling

Uwe Ehret[1], Rik van Pruijssen[1], Marina Bortoli[1], Ralf Loritz[1], Elnaz Azmi[2], and Erwin Zehe[1]
[1]Institute of Water Resources and River Basin Management, Karlsruhe Institute of Technology – KIT, Karlsruhe, Germany
[2]Steinbuch Centre for Computing, Karlsruhe Institute of Technology – KIT, Karlsruhe, Germany

Received: 08 Feb 2020 – Accepted for review: 25 Feb 2020 – Discussion started: 28 Feb 2020

Abstract. In this paper we propose adaptive clustering as a new way to analyse hydrological systems and to reduce computational efforts of distributed modelling, by dynamically identifying similar model elements, clustering them and inferring dynamics from just a few representatives per cluster. It is based on the observation that while hydrological systems generally exhibit large spatial variability of their properties, requiring distributed approaches for analysis and modelling, there is also redundancy, i.e. there exist typical and recurrent combinations of properties, such that sub systems exist with similar properties, which will exhibit similar internal dynamics and produce similar output when in similar initial states and when exposed to similar forcing. Being dependent on all these factors, similarity is hence a dynamical rather than a static phenomenon, and it is not necessarily a function of spatial proximity.

We explain and demonstrate adaptive clustering at the example of a conceptual, yet realistic and distributed hydrological model, fit to the Attert basin in Luxembourg by multi-variate calibration. Based on normalized and binned transformations of model states and fluxes, we first calculated time series of Shannon information entropy to measure dynamical similarity (or redundancy) among sub systems. This revealed that indeed high redundancy exists, that its magnitude differs among variables, that it varies with time, and that for the Attert basin the spatial patterns of similarity are mainly controlled by geology and precipitation. Based on these findings, we integrated adaptive clustering into the hydrological model. It constitutes a shell around the model hydrological process core and comprises: Clustering of model elements, choice of cluster representatives, mapping of results from representatives to recipients, comparison of clusterings over time to decide when re-clustering is advisable. Adaptive clustering, compared to a standard, full-resolution model run used as a virtual reality "truth", reduced computation time to one fourth, when accepting a decrease of modelling quality, expressed as Nash-Sutcliffe efficiency of sub catchment runoff, from 1 to 0.84.

We suggest that adaptive clustering is a promising tool for both system analysis, and for reducing computation times of distributed models, thus facilitating applications to larger systems and/or longer periods of time. We demonstrate the potential of adaptive clustering at the example of a hydrological system and model, but it should apply to a wide range of systems and models across the earth system sciences. Being dynamical, it goes beyond existing static methods used to increase model performance, such as lumping, and it is compatible with existing dynamical methods such as adaptive time-stepping or adaptive gridding. Unlike the latter, adaptive clustering does not require adjacency of the sub systems to be joined.

How to cite: Ehret, U., van Pruijssen, R., Bortoli, M., Loritz, R., Azmi, E., and Zehe, E.: Adaptive clustering: A method to analyze dynamical similarity and to reduce redundancies in distributed (hydrological) modeling, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-65, in review, 2020.

Copernicus Publications
The Innovative Open Access Publisher

Search articles
Search
Author

Special issue
Linking landscape organisation and hydrological functioning:...

Download
PDF XML

Short summary
In this paper we propose adaptive clustering as a new way to analyse hydrological systems and to...
Read more

Citation
BibTeX
EndNote

Share


Looking forward to see all of you in person again
at EGU 2021!