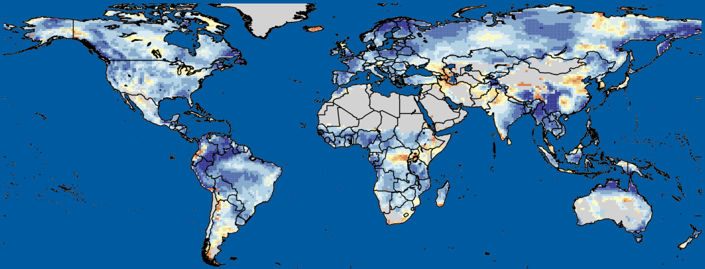
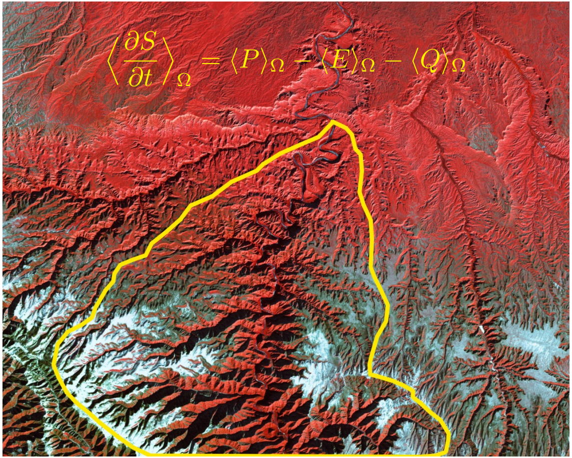
Multi-scale global reconstruction of water fluxes and states with mHM

Luis Samaniego, Maren Kaluza, Stephan Thober, Oldrich Rakovec



D165, EGU General Assembly
7th May 2020

Grand challenges in Land Surface Hydrology

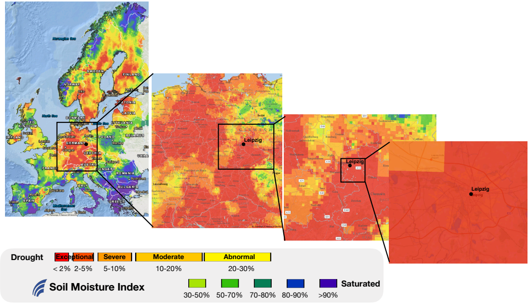


Source © NASA

"To develop the ability to globally monitor and predict the movement of water on the landscape at scales less than 1 km"
Wood et al. WRR 2011

"Everywhere and locally relevant"
Bierkens et al. HP 2014

Grand challenges in Land Surface Hydrology



Drought index SMI
based on mHM with E-OBS v18
for Aug. 2018

"To develop the ability to globally monitor and predict the movement of water on the landscape at scales less than 1 km"
Wood et al. WRR 2011

"Everywhere and locally relevant"
Bierkens et al. HP 2014

Grand challenges in Land Surface Hydrology

Nested ICON Model
seamless local mesh refinement

Source © DWD (Reinert et al.), MPI-M(Giorgetta et al.), 2016

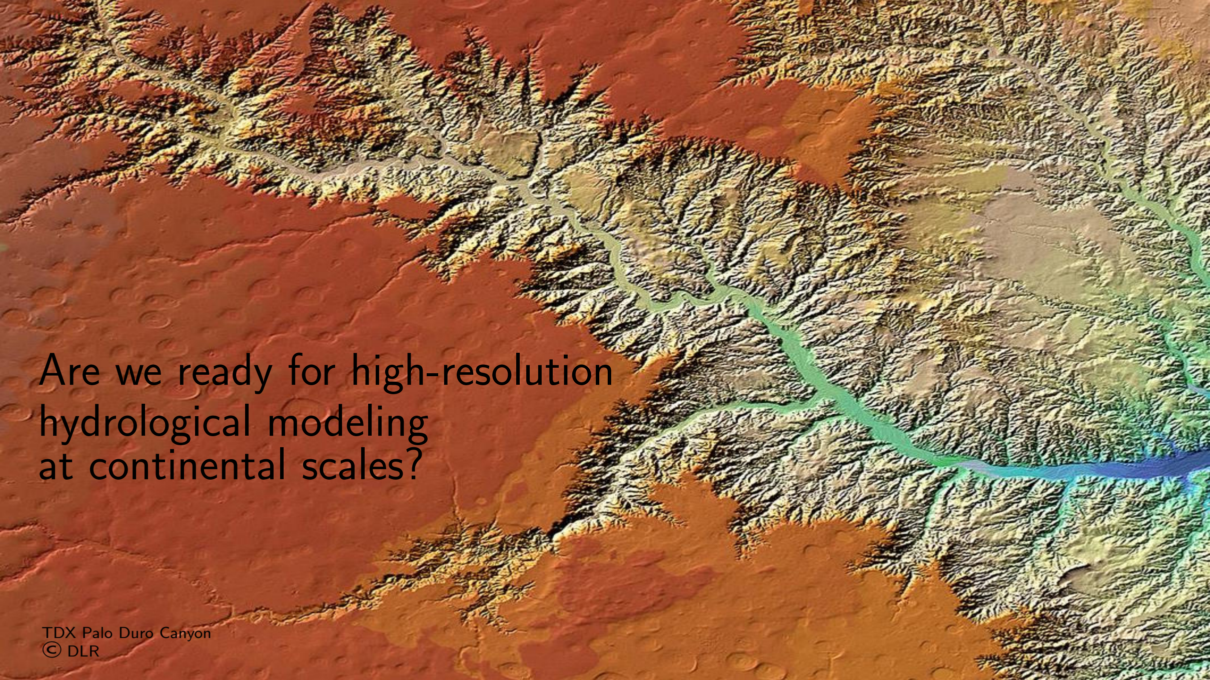
"Develop scale-independent land surface scheme for climate models"
IPCC AR5, 2014
Bauer et al, Nature, 2015

Grand challenges in Land Surface Hydrology

"Seamless prediction of the earth system, from minutes to months"

Source © WMO-1176, Shutterstock 2015

"Reduce uncertainties in the representation of processes in numerical models relevant to the improvement of predictive skill"
G. Brunet, S. Jones and B. Mills, 2015



Are we ready for high-resolution hydrological modeling at continental scales?

TDX Palo Duro Canyon
© DLR

Performance of the state-of-the-art GHMs

Table 8. Median Scores for Widely Used Performance Metrics Obtained for the Evaluation Catchments ($n = 1113$)[a]

Performance Metrics	HBV With Regionalized Parameters (10 Most Similar Donors)	HBV With Spatially Uniform Parameters	HTESSEL	JULES	LISFLOOD	ORCHIDEE	PCR-GLOBWB	SURFEX	SWBM	W3RA	WaterGAP3	Ensemble Mean
NSE daily	**−0.02**	−0.03	−0.59	−0.38	−0.55	−0.45	−1.67	−0.24	**0.01**	−0.59	−0.11	−0.35
NSE 5 day	**0.08**	**0.05**	−0.49	−0.44	−0.26	−0.53	−1.51	−0.21	0.05	−0.34	−0.11	−0.25
NSE monthly	**0.17**	**0.15**	−0.32	−0.39	−0.03	−0.67	−1.16	−0.02	0.14	−0.10	−0.10	−0.09

Beck et al. WRR 2016

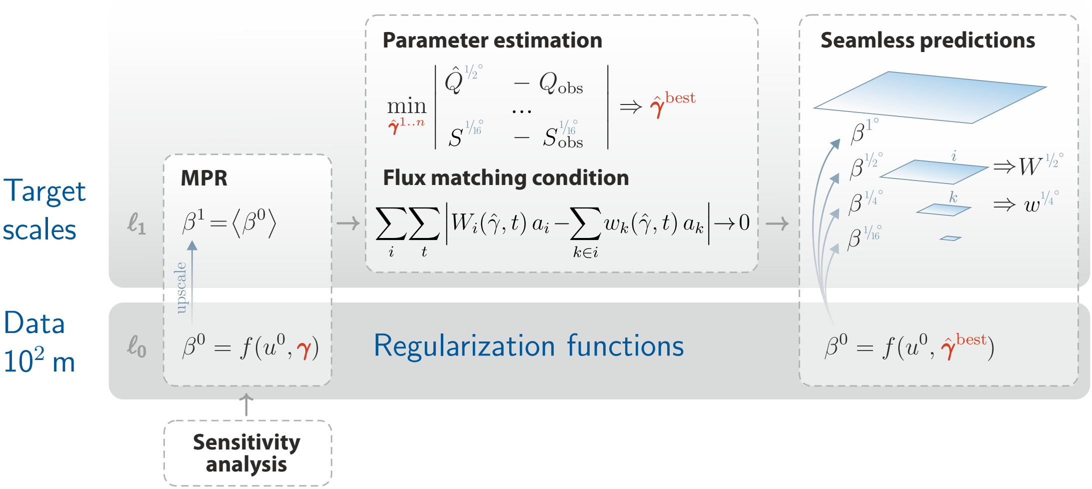
Goal:
Improve the performance of the global mHM model above the current state-of-the-art
www.ufz.de/mhm

Parameterization at meso/macro scales

development of empirical relationships. Accordingly, it may be said that the parameterization of hydrologic processes to the grid scale of general circulation models is a problem that has not been tackled, let alone solved.

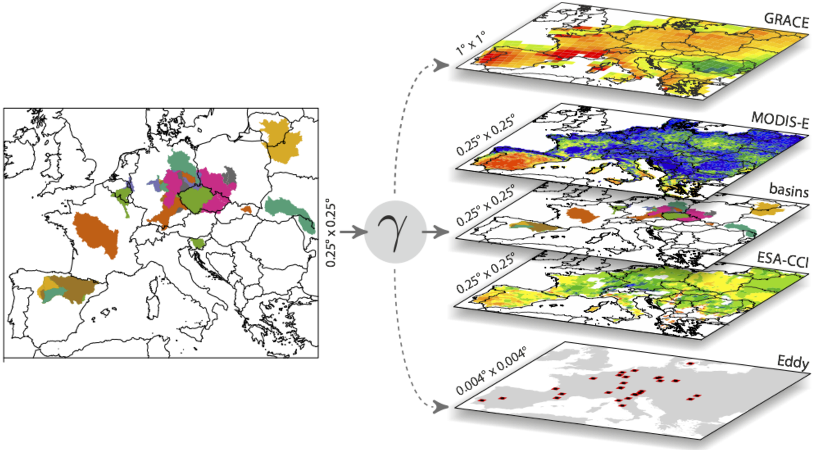
Dooge, 1982 p. 269 (Eagleson edt.)

Multiscale data verification & parameter estimation



Samaniego et al. WRR 2010, Rakovec et al. JHM 2016, Samaniego et al. HESS 2017

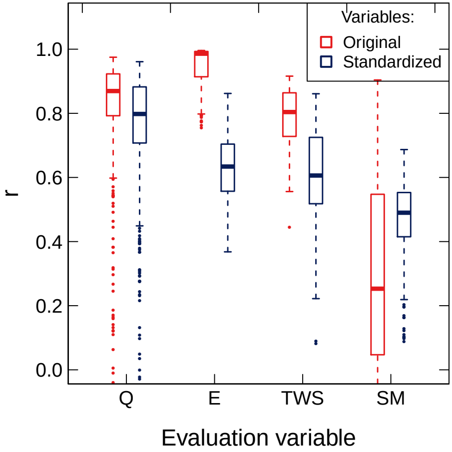
Nested model evaluation



Rakovec et al. JHM, 2016

Parameter estimation
streamflow only

Inverse modelling based on streamflow



Rakovec et al., JHM 2016
$n = 400$ Pan-EU basins

Constraining parameters against streamflow is a necessary, but not a sufficient condition to get robust estimates of E, TWS, SM.

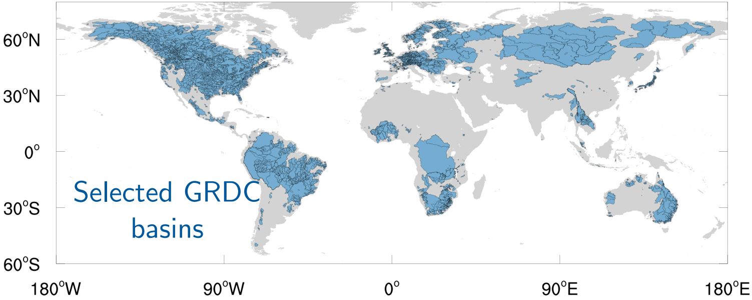
- **TWS**: GRACE ($1^\circ \times 1^\circ$)
- **E**: FLUXNET ($0.5^\circ \times 0.5^\circ$)
- **SM**: ESA-CCI ($0.25^\circ \times 0.25^\circ$)

Development of a Massive hybrid MPI/OpenMP scheme for mHM

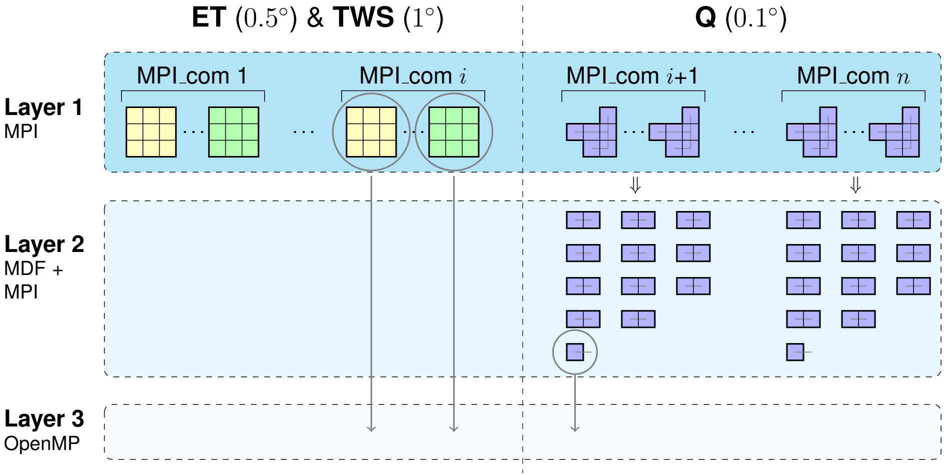
Research goals

mHM
- Setup a nested models for Q, TWS, and ET at their native resolutions
- Estimate $\hat{\gamma}^{\text{best}}$
  - 5500 GRDC streamflow series
  - FLUXNET ET
  - GRACE TWS anomaly

© JUWELS
www.fz-juelich.de



Multi-layer hybrid parallelization scheme



M. Kaluza et al GMD (2020, to be submited).

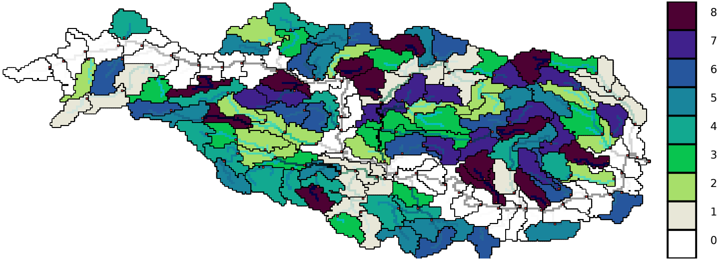
Parallelization of streamflow routing: a hard problem

mHM streamflow simulation on July 2002 European floods

- A river network is a graph.
- Performing time dependent operations in a graph is hard to parallelize.
- Even if it can be paralellized, it is even harder to get linear scaling.
- Existing algorithms are not efficient for large networks and HPC environments with thousands of cores.

Domain decomposition for streamflow routing

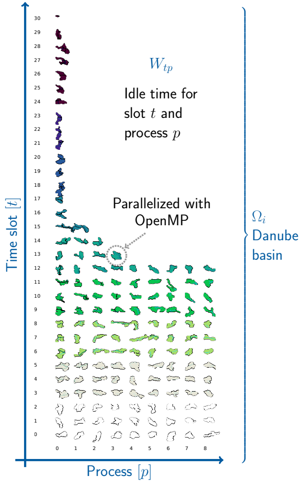
- Random allocation
- Li et al. (2011)
- Pfafstetter coding (Clark et al.)

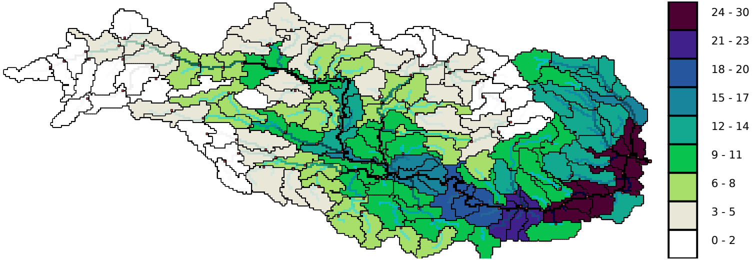


Process sequence
Danube: 26507 cells of 5×5 km^2

- Key variables:
  - Process Id p
  - Time slot t
- Our proposdsal: MDF ...

MPI decomposition of "forests" (MDF)

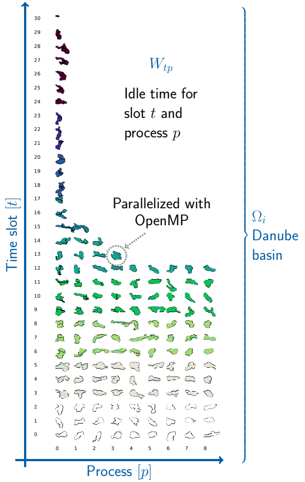




Slot time t

M. Kaluza et al GMD, 2020.

MPI decomposition of “forests” (MDF)



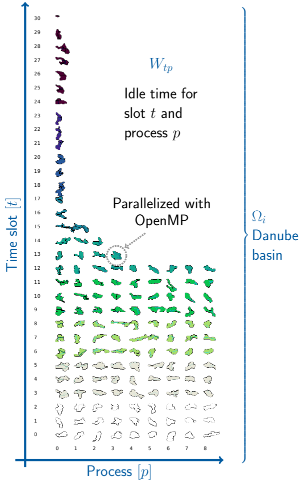
“Combinatorial scheduling NP-problem” aiming at:

$$\min_{P,T}\left[T\left(\bigcup_{i=1}^{n}\Omega_i, a, b, P\right), \sum_{p=1}^{P}\sum_{t=1}^{T} W_{tp}\right]$$

Subject to $cTP < B$

M. Kaluza et al GMD, 2020.

MPI decomposition of "forests" (MDF)



M. Kaluza et al GMD, 2020.

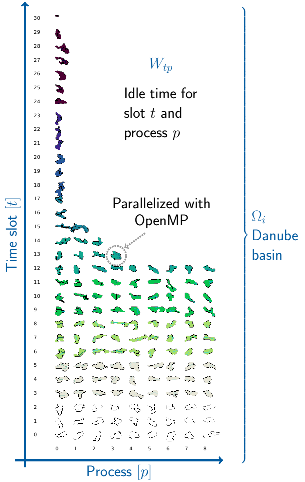
"Combinatorial scheduling NP-problem" aiming at:

$$\min_{P,T}\left[T\left(\bigcup_{i=1}^{n}\Omega_i, a, b, P\right), \sum_{p=1}^{P}\sum_{t=1}^{T} W_{tp}\right]$$

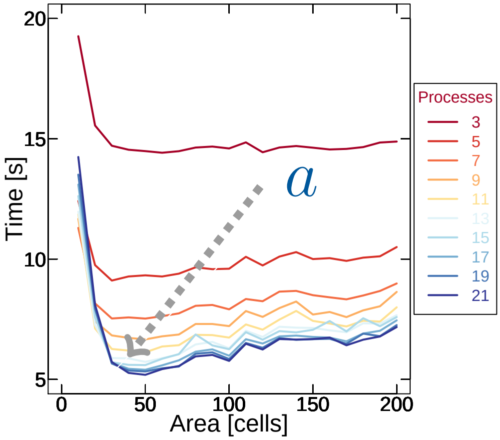
Subject to $cTP < B$

T	optimal number of time slots
B	budget in core-h
	$T \geq \max(\Delta_i(a))$
P	optimal number of processes
Δ_i	network depth of basin (Ω_i)
a, b	optimal sub-basin size, maximum buffer size (hardware)
n, t	indices for # basins and time slots
c	nr. cores per process (hardware)

MPI decomposition of "forests" (MDF)



M. Kaluza et al GMD, 2020.



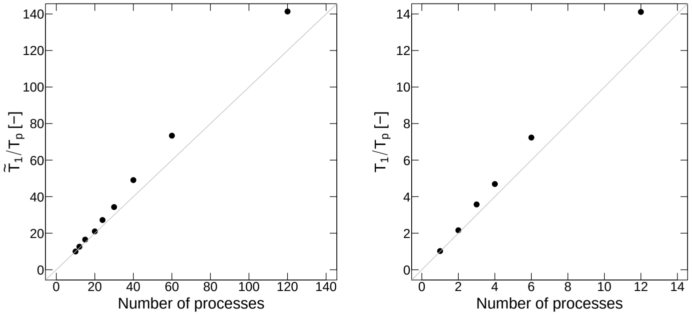
Sub-basin domain scaling

Layer 1: Speedup of a strong-scaling test (ET & TWS)

$$S = \frac{\widetilde{T}_1}{T_p}$$

$$\widetilde{T}_1 \approx 10\, T_{10}$$

- $\rightarrow$ **super-linear**
- Cache-effect induced by reducing the problem to fit in the "fast" RAM instead of the "slow" RAM
- Shorter distances in memory access (4 CPUs$\rightarrow$8 GB RAM)

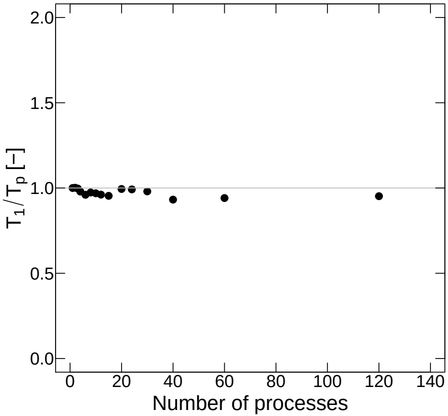


1 Process uses 96 CPUs

Layer 1: Parallel efficiency of a weak-scaling test

$$E = \frac{T_1}{T_p}$$

- $\rightarrow$ **optimal**
- NOTE: The problem size per processor stays fixed as more processors are added.

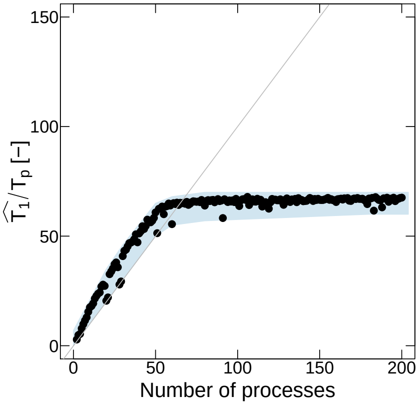


1 Process uses 96 CPUs

Layer 2: Speedup of a strong-scaling test (Q)

$$S = \frac{\hat{T}_1}{T_p}$$

- Routing → **is the limiting factor**
- MDF-v1 scales up to 5760 CPUs
- Test has 307 000 links
- Globe at 0.1°: ≈ 1 350 000 links

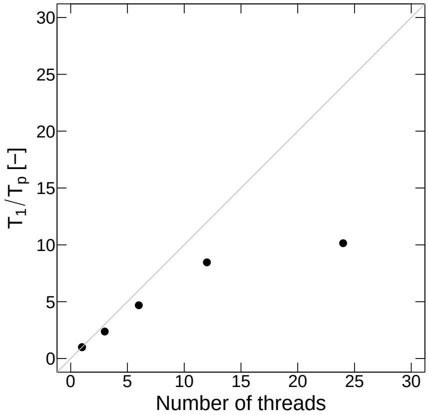


1 Process uses 96 CPUs

Layer 3: Speedup of a strong-scaling test (OpenMP)

$$S = T_1/T_p$$

- $\rightarrow$ **sub-linear**
- Almost linear up to 12 cores
- Inefficient for many CPUs becuase of fast-RAM limitations.
- But, not critical compared to Layers 1 and 2

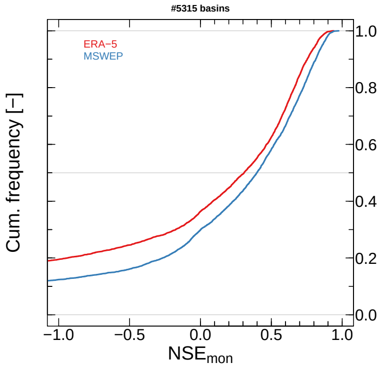


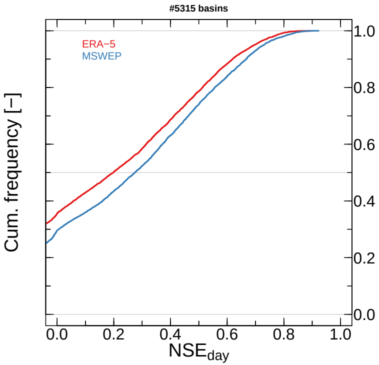
1 thread per CPU (max 2)

Model evaluation based on

GRDC streamflow gauges, GRACE TWS, FLUXNET ET

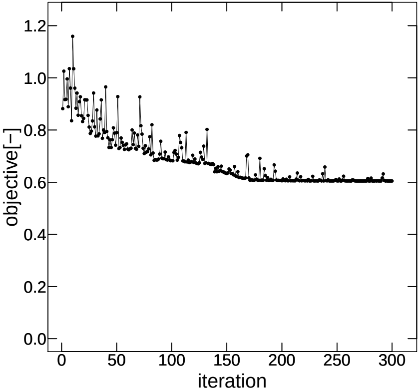
Global results (no calibration)





Median NSE (mHM, uncalibrated) =	**0.40**
Median NSE (other HMs) =	-0.09

Evolution of the combined objective function (OF)

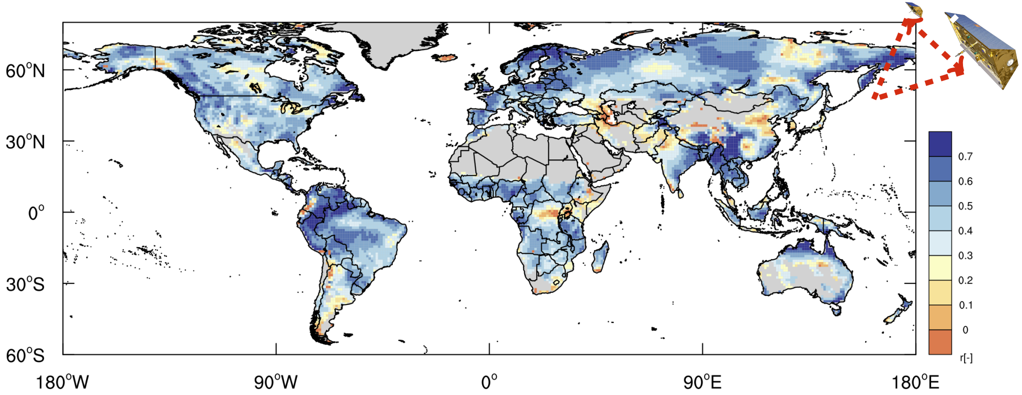


OF related with efficiency of global fields of ET and TWSA

Next steps:

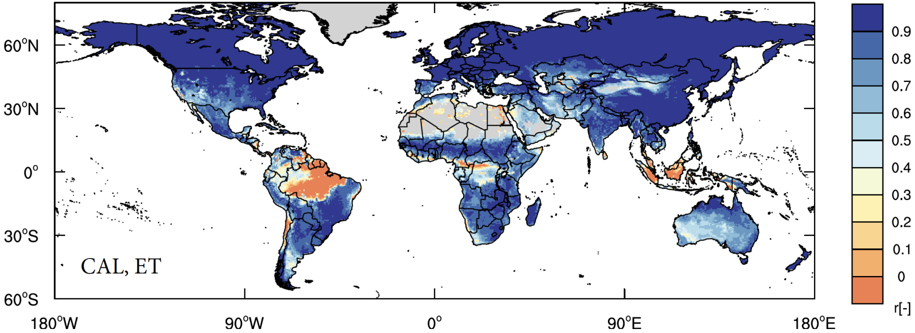
- Include 50 samples (random but fulfilling water balance related criteria) of 250 GRDC basins in the combined OF.
- Perform cross-validation.
- Expected optimization budget: 2.5 million core-hours.

Correlation between simulated TWSa (mHM) & GRACE



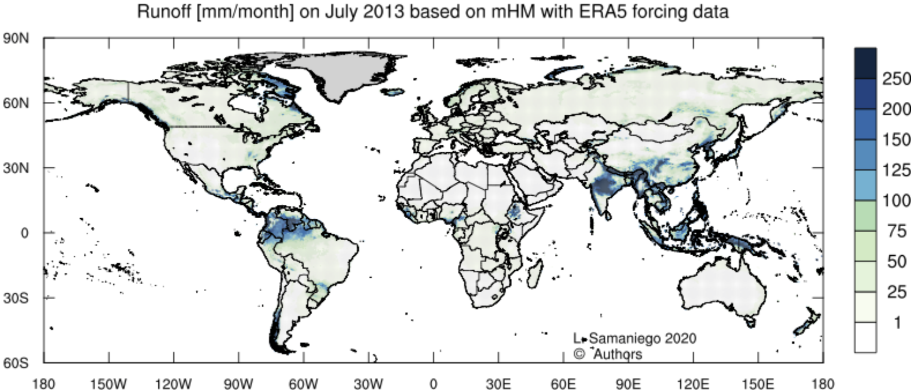
Rakovec et al. 2020 (in prep.).

Correlation between simulated ETa (mHM) & FLUXNET



Rakovec et al. 2020 (in prep.). The Poor poorperformance of ET in the Amazon and other tropical regions in under investigation. Plausible reasons: model, data, both?

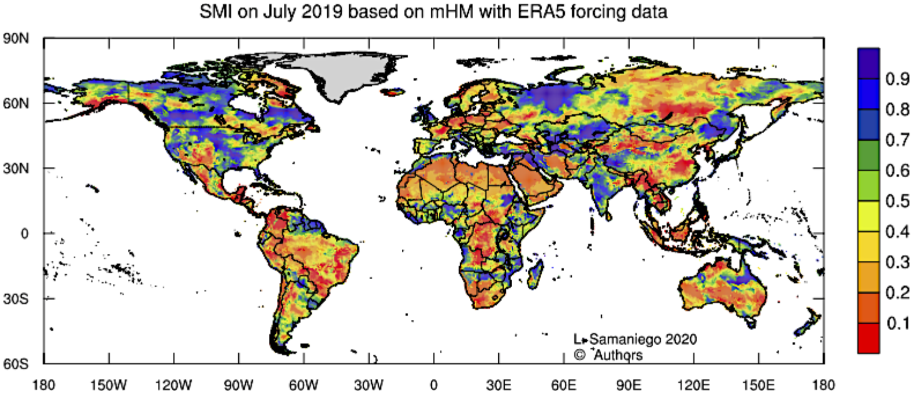
Estimated runoff → flood event



Runoff mm/month during July 2013.

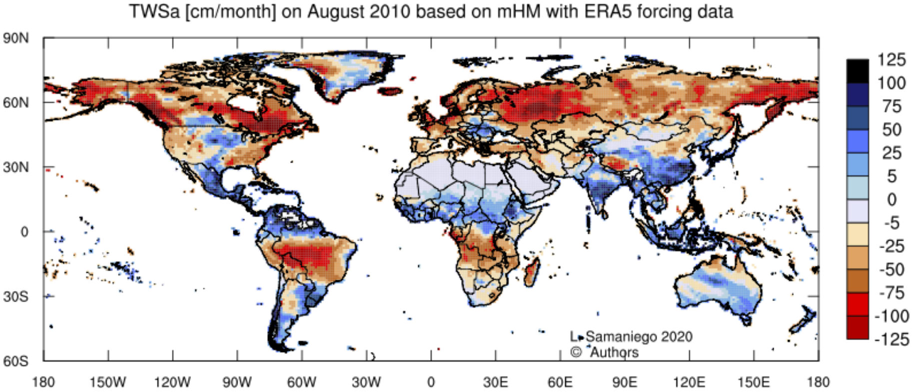
Shows the extrem increase of runoff generation during 2013 India Flood

Estimated soil moisture → drought events



Soil Moisture Index (simulated SM quantiles) during July 2019.
Shows the extent of the 2019 EU-drought (SMI<0.2)

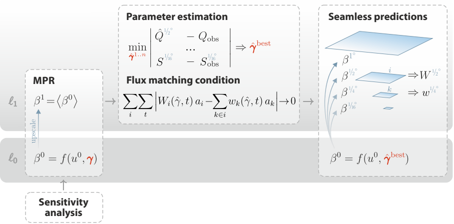
Estimated TWS anomaly → drought event



Simulated TWS anomaly during August 2010.
Shows the extent of the drought over Rusia

Conclusions and Outlook

- Nested multiscale parameter estimatons are possible only if the model exhibits a scale-invariant parameterization (mHM+MPR)
- Uncalibrated mHM+MPR are already better than the state-of-the-art model results reported in Beck et al. WRR
- MDFv1 is scalling well but river routing remains as the limiting factor
- Ongoing steps:
  - Conclude scaling tests
  - Perform the parameter estimation with 2.5e6 core-h budget @ the JUWELS supercomputer
  - Estimate uncertainties of water fluxes



Acknowledgment

GRACE: NASA - GFZ
MSWEP: Eric Wood, Hylke Beck
JUWELS: Julich IT team
UFZ: mHM developers
ESM team

Thank you!

www.ufz.de/mhm