IO – Software for Terrestrial Systems Modelling Platform (TSMP)

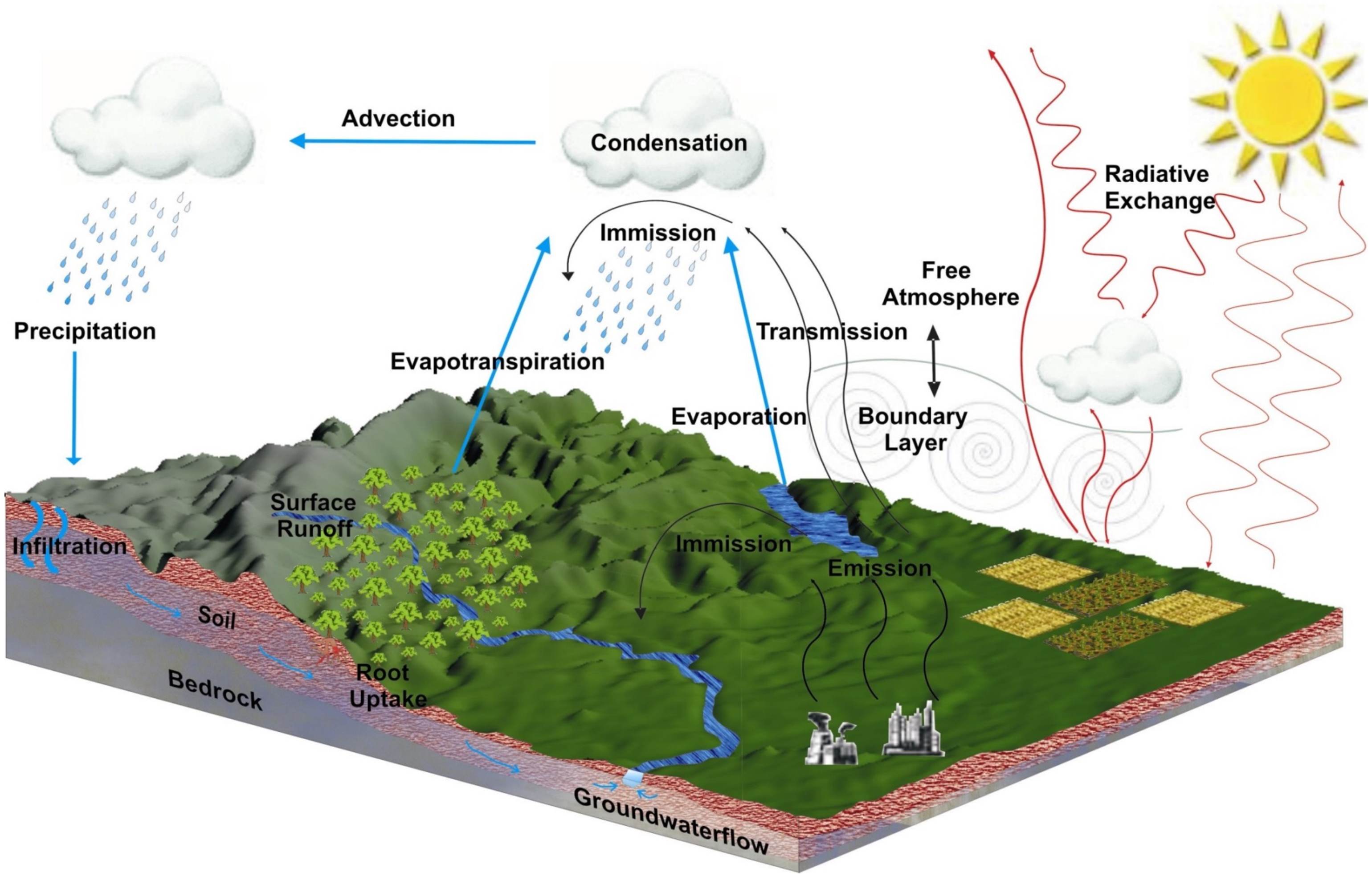
Ghazal **Tashakor**

Simulation and Data Lab Terrestrial Systems (JSC / JFZ)

This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 955811. The JU receives support from the European Union's Horizon 2020 research and innovation programme and France, the Czech Republic, Germany, Ireland, Sweden, and the United Kingdom.

CONCEPTUALISING THE TERRESTRIAL SYSTEM

Terrestrial water cycle and groundwater-to-atmosphere (G2A) interactions and feedbacks



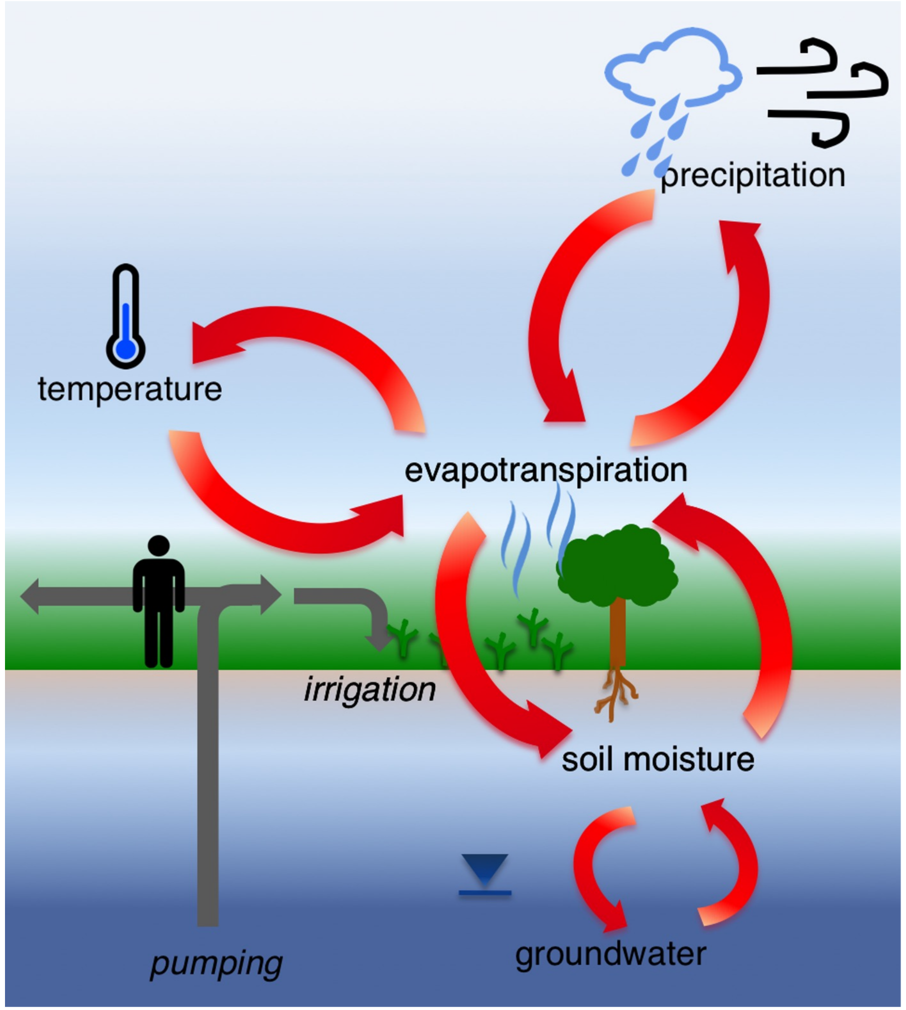
- Complex interactions and feedback between various sub-systems of the coupled geo-ecosystem have many drivers.
- Linkages through energy, mass, and momentum transfers Multiple Spatio-temporal scales.
- Anthropogenic physical system changes modify the land surface, ecosystem processes, and services with many socio-economic impacts.

Source: Goergen et al., ICRC-CORDEX Conference 2019, 2019

MOTIVATION FOR A MODELLING SYSTEM

Intensification of the hydrological cycle under climate change

- Global (climate, land use) change has an impact on the water as a resource, and its sustainable use, and affects water security.
- Better **understanding** and **prediction** of (increasing) **extreme hydroclimatic events** (e.g., droughts, heatwaves) and **related feedbacks** for informed adaptation (e.g., irrigation) or mitigation options, but:
  - Observations remain scarce/inconsistent
  - Climate models: Do not include or highly simplify land water
  - Hydrological models: Usually simplify surface-subsurface interactions and neglect two-way feedbacks with the atmosphere: terrestrial water cycle not closed
- **Human water use** has multiple local and non-local (climatic) effects (groundwater recharge/storage, discharge, ET/P recycling, etc.)

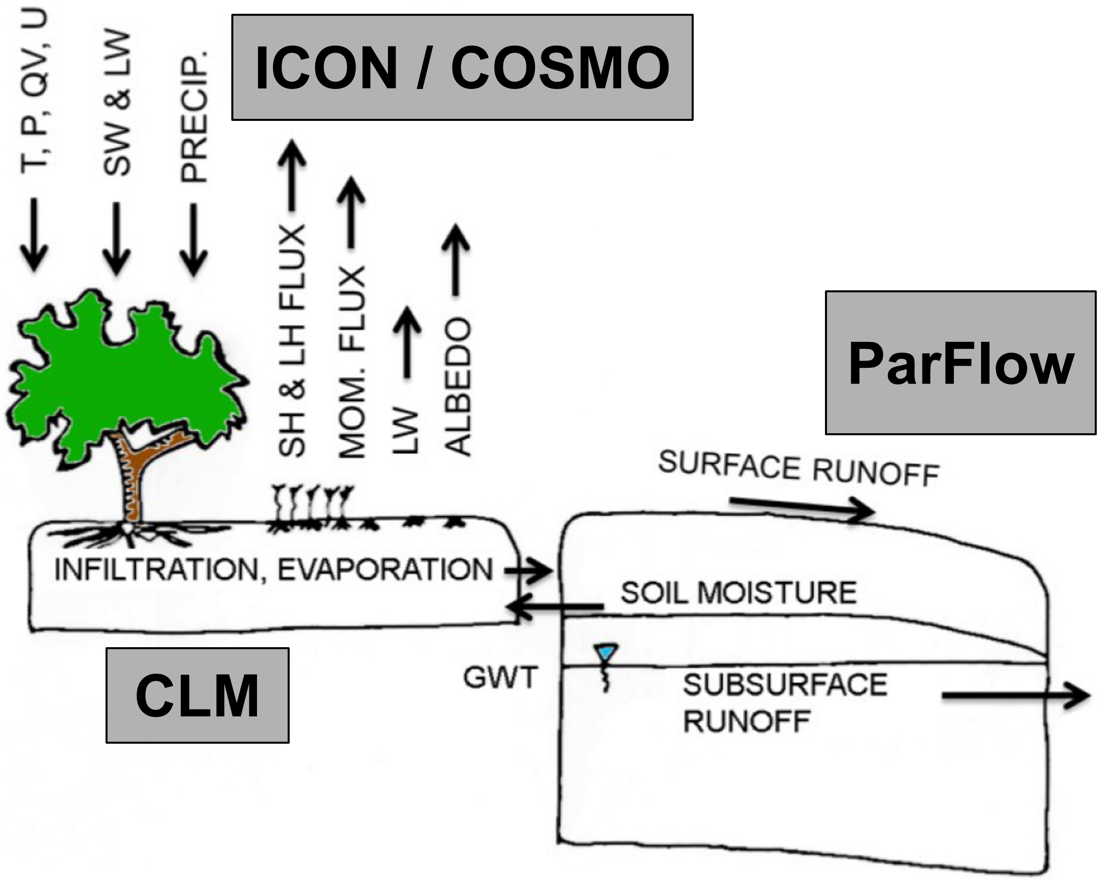


Adapted from Goergen et al., ICRC-CORDEX Conference 2019, 2019

TERRESTRIAL SYSTEM MODELLING PLATFORM (TSMP)

Solving the terrestrial water and energy cycle from groundwater to the atmosphere

- Represents process on soil, land, vegetation and atmosphere
- Scales consistent for each component
- Highly modular
- Numerical modeling system coupling COSMO, Community Land Model (CLM) and ParFlow
- Physically-based representation of transport processes of mass, energy and momentum
- Spanning sub-km resolutions to continental domains
- Explicit feedbacks between compartments
- Parallel Data Assimilation Framework (TSMP-PDAF)
- HPC-enabled, and working towards exascale

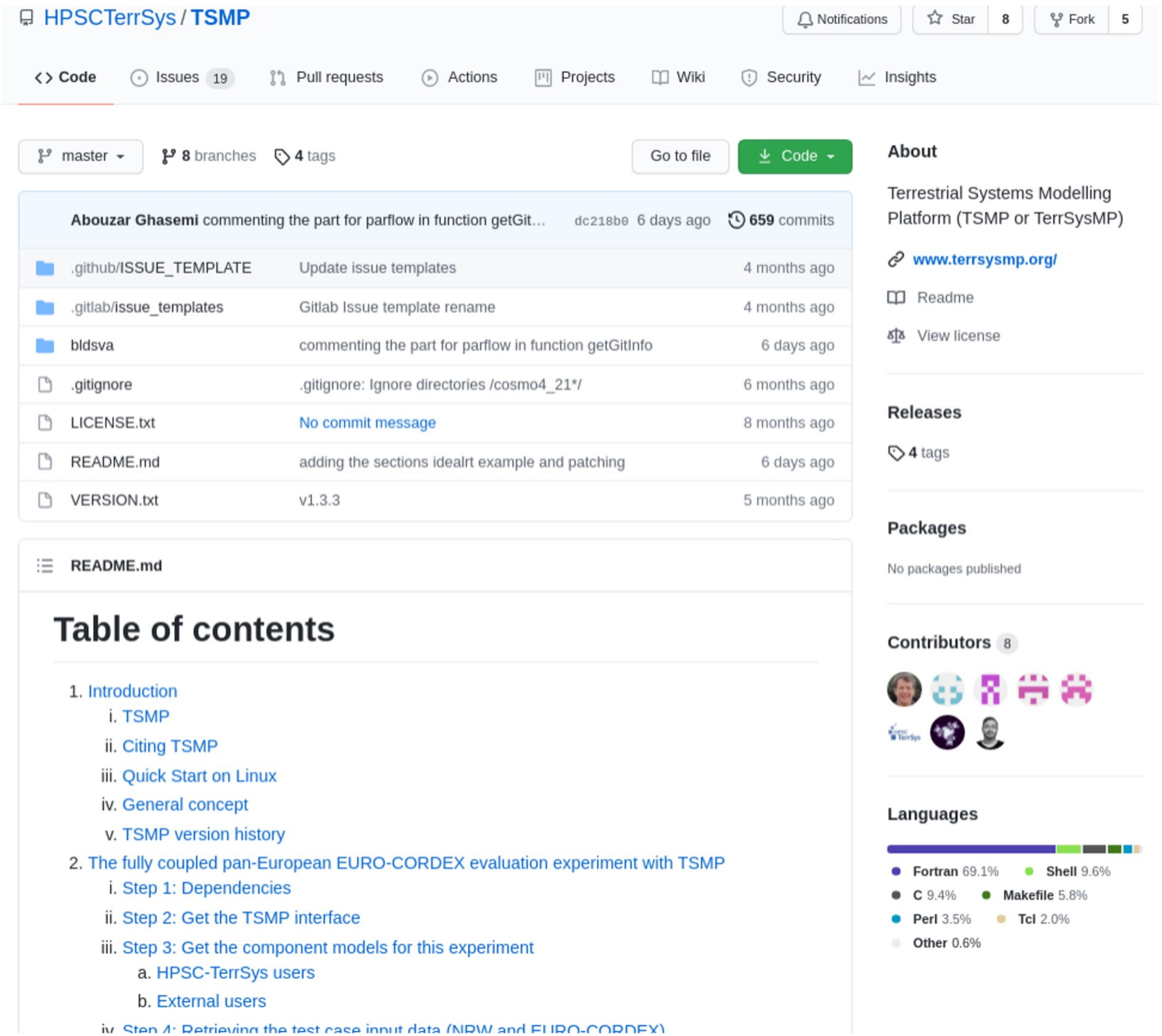


More information: https://www.terrsysmp.org/

TERRESTRIAL SYSTEM MODELLING PLATFORM (TSMP)

Code ownership and availability

- TSMP publicly available on GitHub, owned by SDLTS and HPSC-TerrSys
- ParFlow publicly available on GitHub (co-owned by IBG-3 / FZJ)
- COSMO available on request
- CLM available on request
- TSMP Software Sustainability Plan in development
- Typical cases available in DataPub
- Current release (tag) v1.2.3



https://github.com/HPSCTerrSys/TSMP

IO-SEA

TERRESTRIAL SYSTEM MODELLING PLATFORM (TSMP)

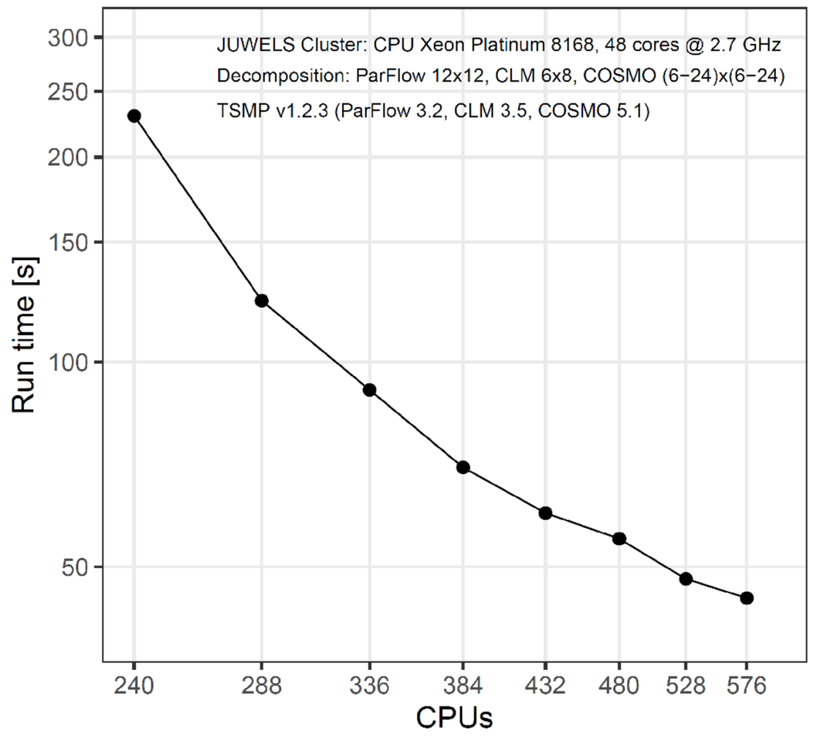
Software features

Source codes:

- TSMP: Interface and modelling framework
- COSMO/ICON (v4.21, v5.01): Atmospheric model
- CLM:(v3.5 and v5x): Land surface model
- ParFlow (v3.9): Surface and subsurface hydrologic model
- OASIS3-MCT2: Coupler
- PDAF: Parallel Data Assimilation Framework

Parallelism:

- Hybrid (MPI, MPI-CUDA, MPI-OpenMP), depends on component model
- Performance portability is an ongoing effort (e.g., ParFlow v3.9 eDSL)
- Heterogeneous computing enabled (ParFlow GPU + COSMO/CLM CPU)
- Parallel I/O through netCDF. ParFlow fully parallel, COSMO has a bottleneck
- Performance analysis (profiling/tracing): Exists



IO-SEA

TSMP – I/O features

Features concerning impact I/O

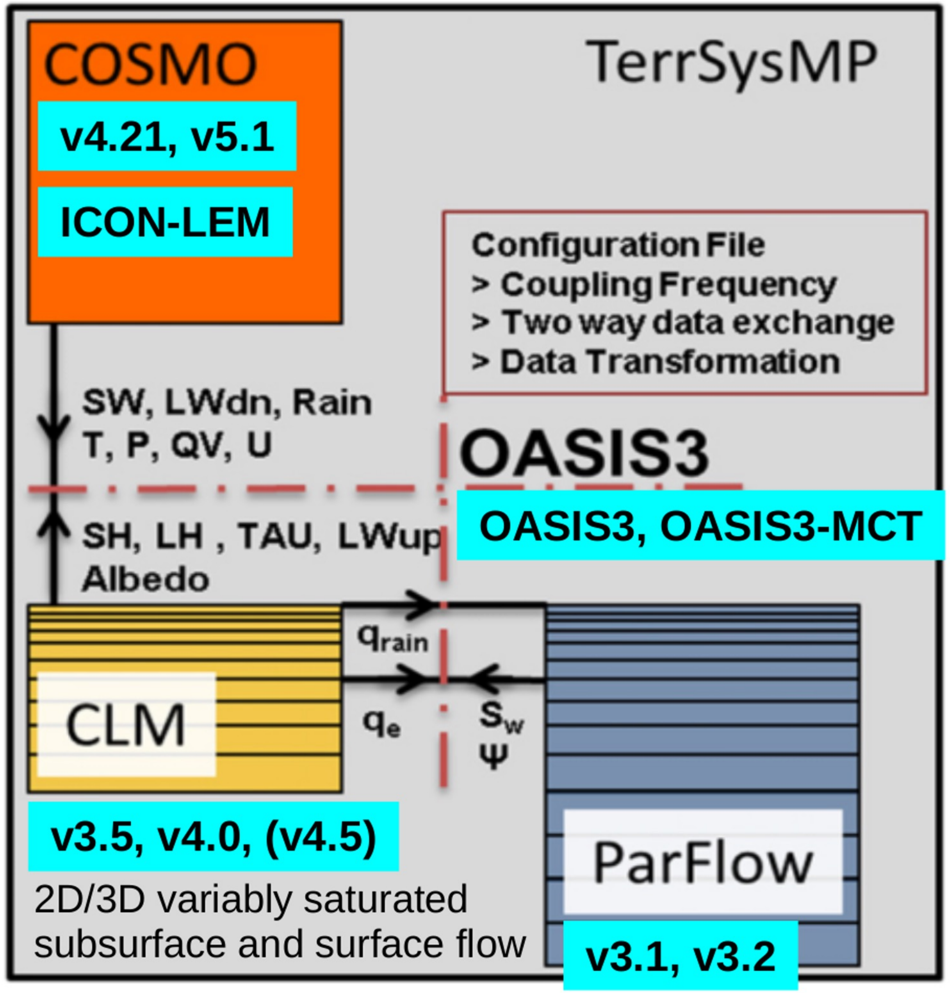
- Each model produces its own output
- Basically 3 groups of files (one per model) containing spatially distributed model output variables
- Mostly netCDF files (but also other standards are possible)

-must fetch BCs from global climate models (e.g., ECMWF)
- **I/O not really under control of TSMP**

Features affecting I/O

- Different model complexity
- Different model resolution
- Different time step

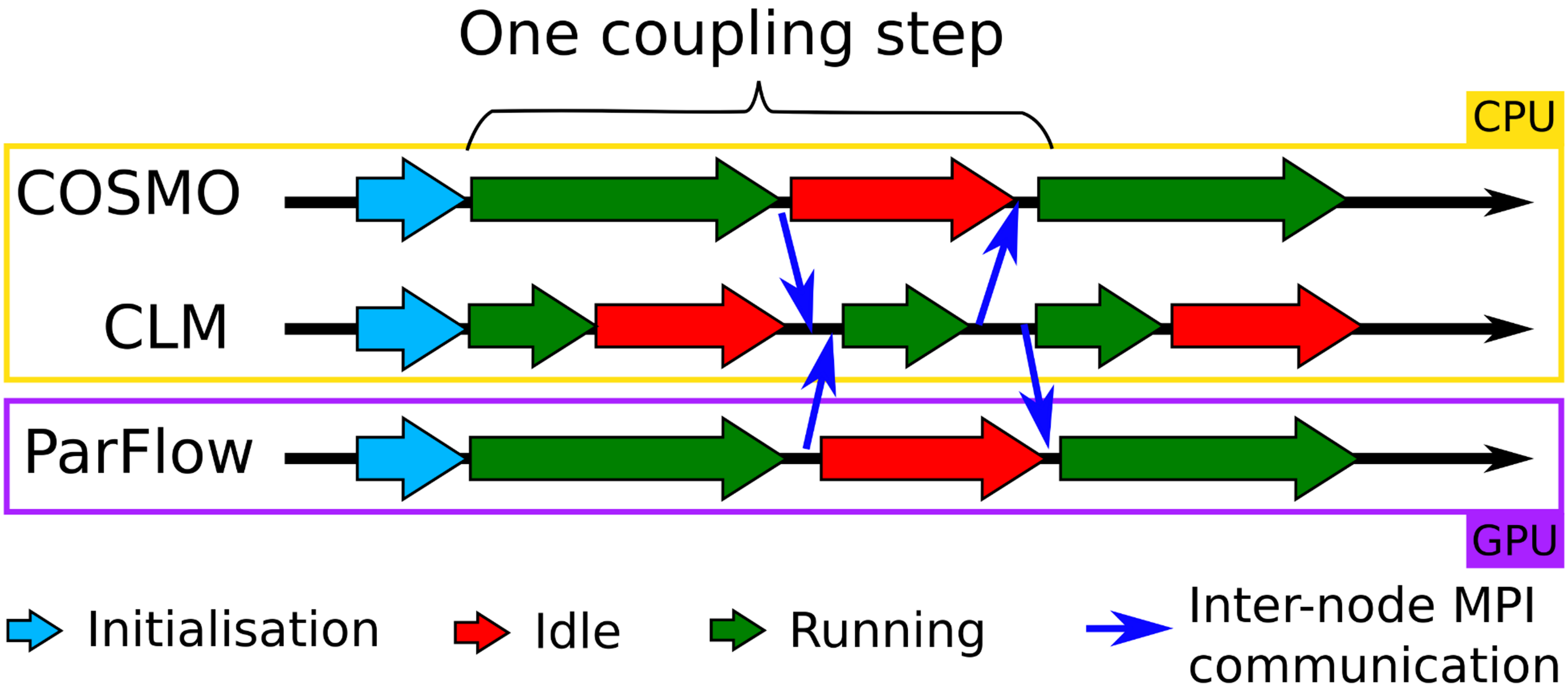
-consequently, **different output frequencies and volumes**



IO-SEA

TSMP COUPLING SCHEME

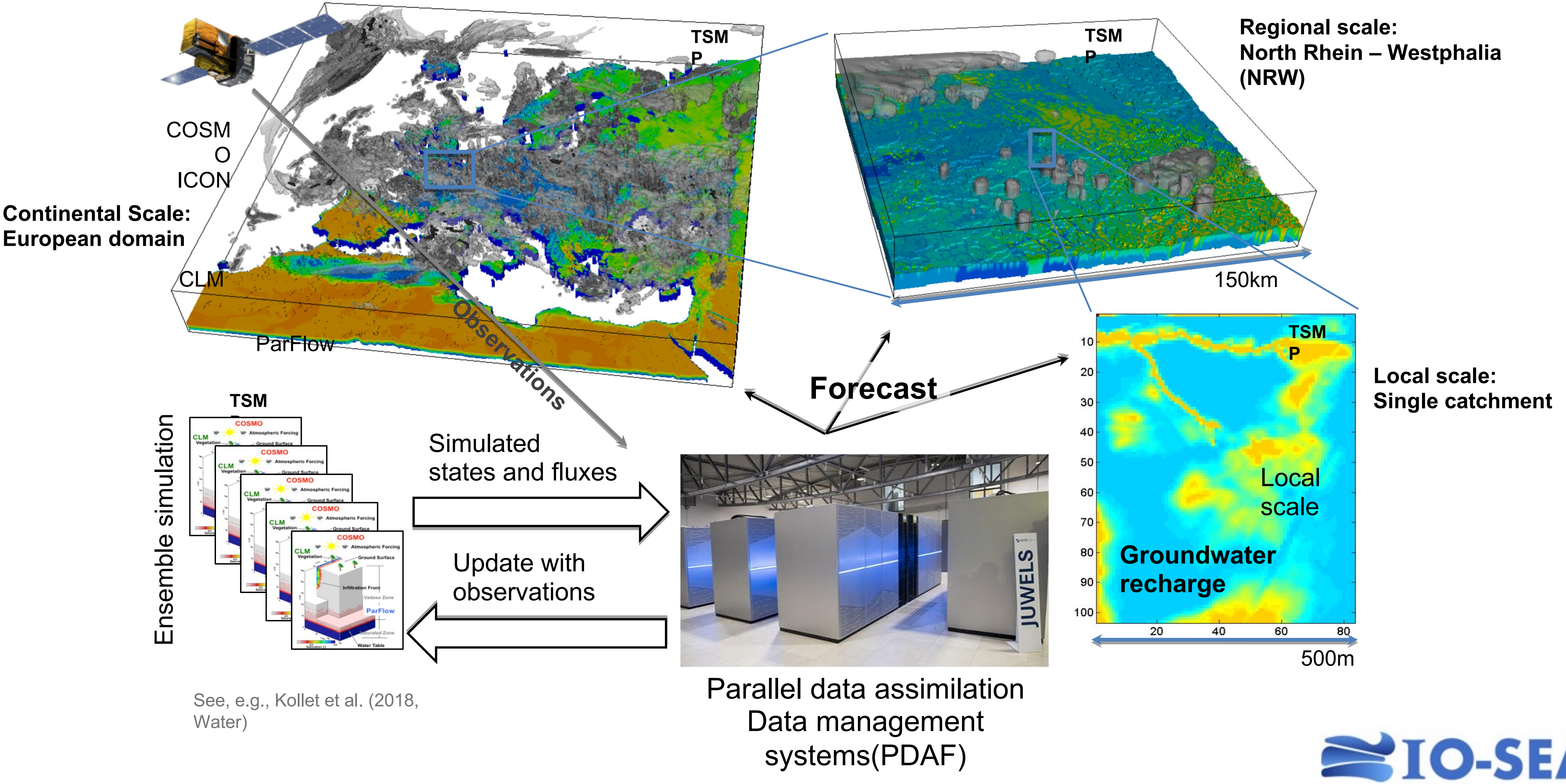
Communication pattern between components (programs)



IO-SEA

TSMP APPLICATIONS

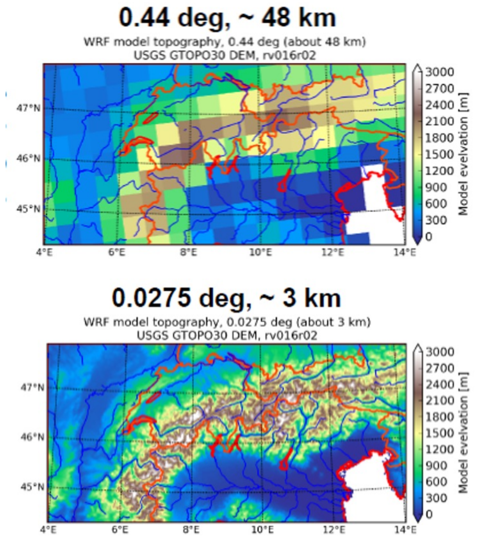
Spanning scales and Earth System compartments with HPC and big-data enabled solutions



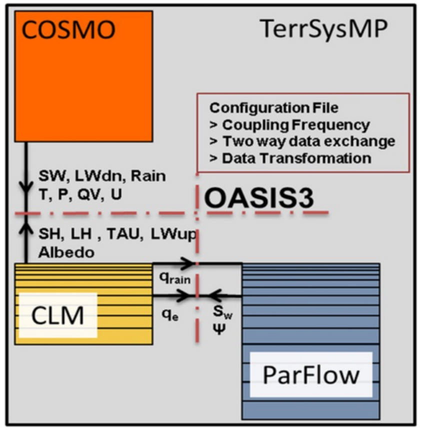
See, e.g., Kollet et al. (2018, Water)

CHALLENGES IN TERRESTRIAL SYSTEM MODELLING

Convection permitting resolution, short output intervals (added value)

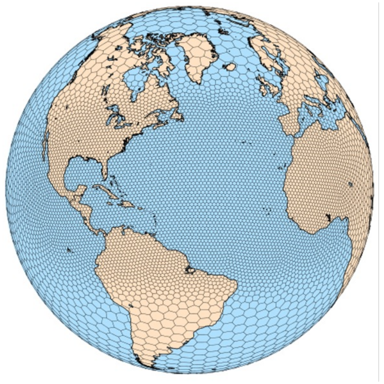


Multiphysics, **fully coupled** regional model systems (earth system modelling)



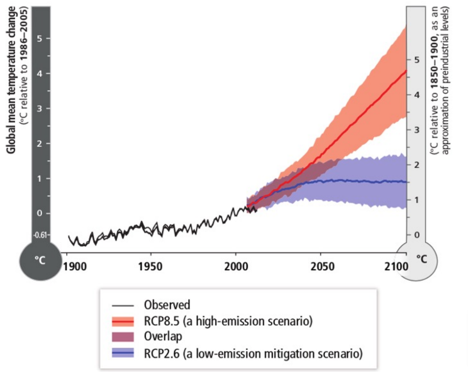
Shrestha et al. (2014, Mon Weather Rev)

Increasing model **domains** (multi-scale processes)



http://mpas-dev.github.io

Very **long integration** times and **ensemble** size (uncertainties, climate change)



IPCC WG2 AR5 SPM (2014)

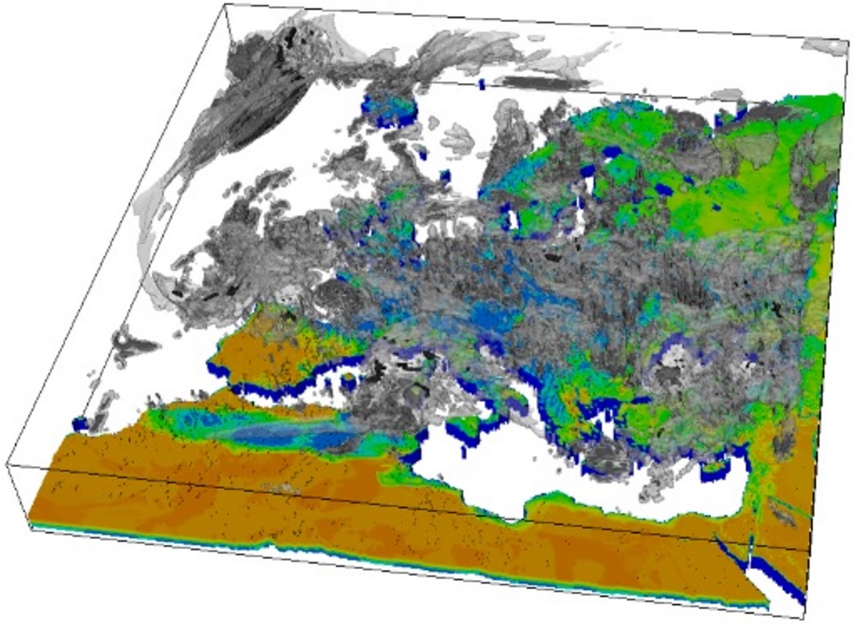
- Complex data-flow paths, modeling chains e.g. with impact models
- Data synthesis approaches, e.g. multiple (observational) (validation) datasets.
- Large legacy codes that need porting into accelerators.

IO-SEA

CHALLENGES IN TERRESTRIAL SYSTEM MODELLING

Challenges		Proposed solutions
Scalability	→	Deeper and more detailed (modular?) profiling
I/O bottlenecks	→	Improved parallel I/O strategies and data staging (IO-SEA!)
Load balancing/ Communication	→	Adaptive mesh refinement / node-to-node communication
Large problems	→	GPUs / heterogeneous / modular computing
Post-processing	→	In-situ visualization and analytics

TSMP Typical case application : EURO CORDEX 11



The EURO-CORDEX 11 case

Long term simulations (spanning 1989 to present time)
High spatial resolution (0.11° ~12.5 km), 436x424 grid points
Resources for production job in Juwels Cluster (JSC): 12 nodes x 48 CPUs
Typical runtime: 3 h / month (1.5 day/year) → 45 walltime days

IO requirements (and estimation on a typical job in Juwels)

COSMO output: 544 MB / sim.hour (~128 GB / walltime hour)
ParFlow output: 22 MB / sim.day (~220 MB / walltime hour)
CLM output: 68 MB/ hour + 107 MB / sim.day (~17 GB / walltime hour)
TSMP total output = 14.5 GB/sim.day (~145 GB/walltime hour) → 165.4 TB / simulation
Input of boundary forcing data ~3 GB / sim.hour (~720 GB / walltime hour) → 32 TB / simulation

Upcoming I/O challenges towards exascale

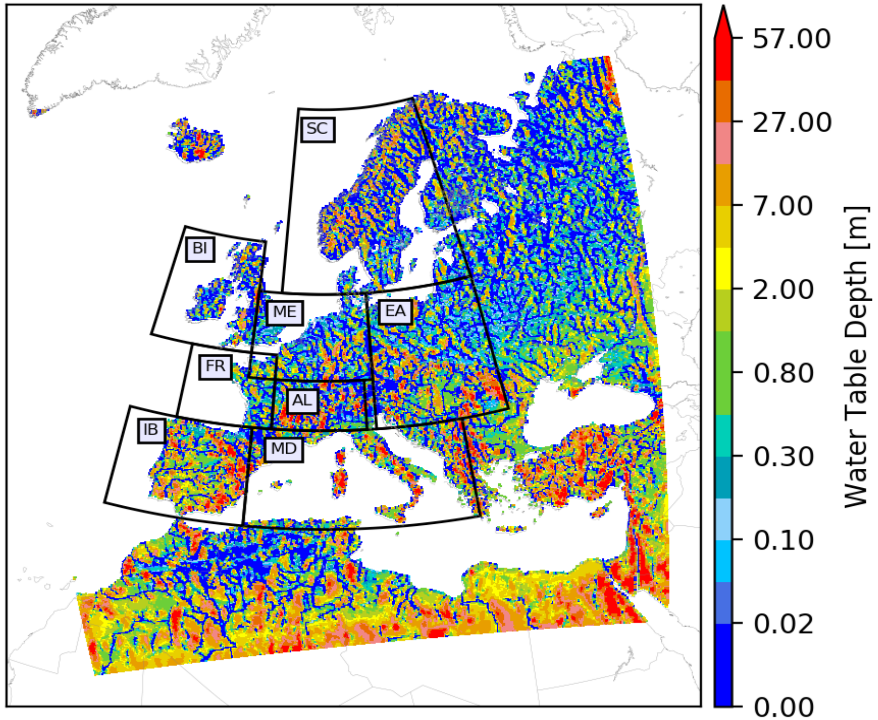
Improved performance (goal 1 SYPD) → ~ 50% speedup on current resources → higher I/O throughput
Increase in spatial resolution (< 2 km) → significantly increased volume per output.
Convection-permitting scales → possibly more output files (higher frequency).
GPU acceleration → asynchronous I/O and data staging
Expected new application over Africa, roughly three times the volume
Data Assimilation (parallel replicates!), increased runtime, input volume, data staging

IO-SEA

TYPICAL PROBLEM: TSMP PAN-EUROPEAN MODEL SETUP

Consistent with WCRP Coordinated Regional Downscaling Experiment (CORDEX) project

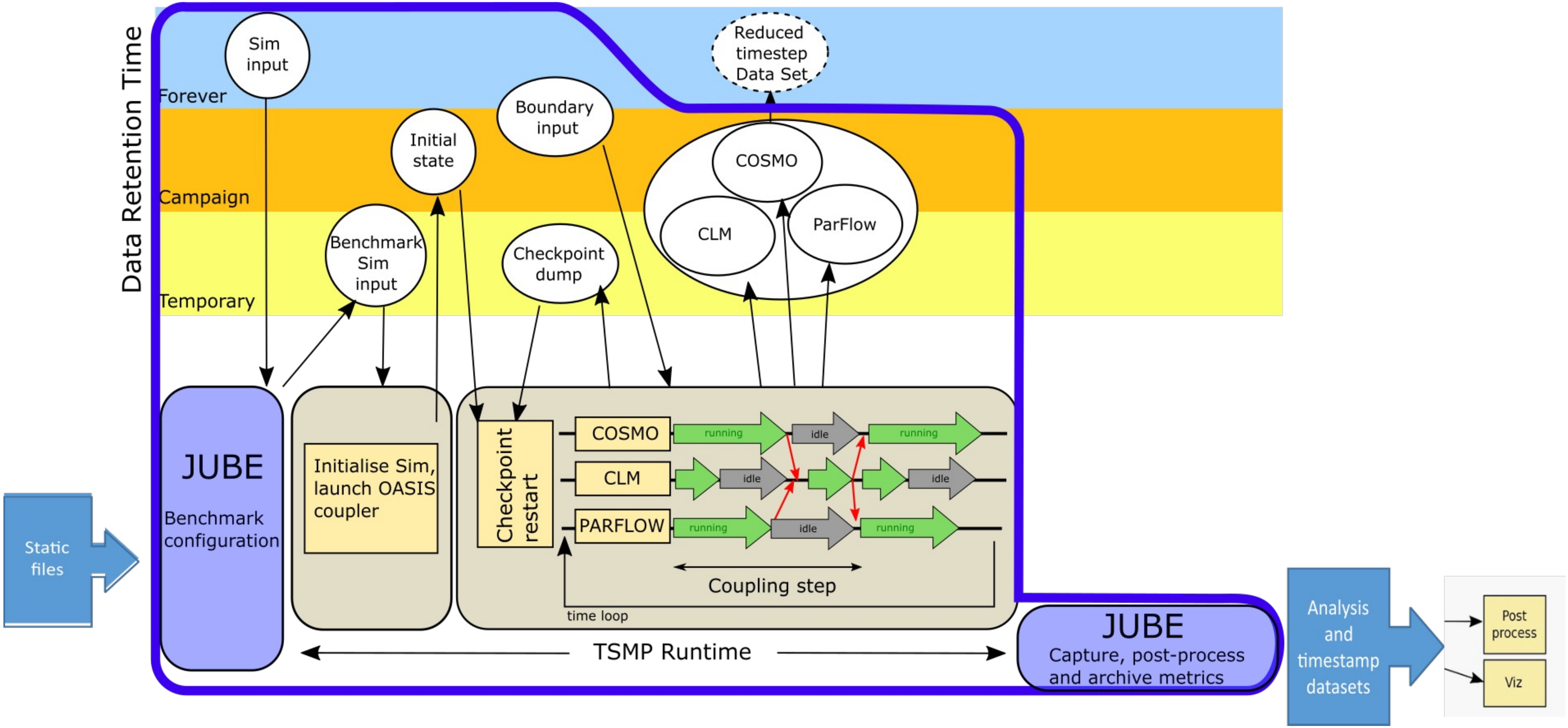
- CORDEX EUR-11 Gutowski et al. (2016, GMD)
  - Resolution: 0.11° (about 12km), 436 x 424 gridpoints
  - Vertical levels: 50 (COSMO), 10 (CLM), 15 (ParFlow)
  - Time steps: 60s (COSMO), 180s (CLM), 180s (ParFlow)
- Input data Keune et al. (2016, JGR)
  - Atmosphere: ERA-5
  - Land surface: MODIS data (4 plant functional types / grid cell)
  - Subsurface: FAO soil types (and Gleeson data base)



- Experiments
  1. Sensitivity studies, year 2003 (European heat wave) 1D vs 3D groundwater physics Keune et al. (2016, JGR)
  2. EURO-CORDEX evaluation: 1989-1995 spinup, 1996-2018 analysis, pristine conditions Furusho-Percot et al. (revision)
  3. Probabilistic water resources prediction, heatwave and drought 2018 impacts on 2018/19 Hartick et al. (revision)

IO-SEA

TSMP-JUBE Workflow for IOSEA



Different scenarios for I/O workflows

TSMP-M: the NRW forecasting system (TSMP-M)

Short term simulations: 14 days forecast

Daily job, very hot data

Currently operational in Juwels Cluster (JSC)

Requires global → regional → local data input workflow

Post-simulation data is pushed to web services

Typical case: EURO-CORDEX 11

Long term simulation (30+ years)

High spatial resolution (0.11° ~12.5 km)

436x424 grid points

Juwels @ JSC: 12 nodes x 48 CPUs

Runtime: 3 h / month (1.5 day/year)/45 walltime days

Upcoming plans for the future: Africa

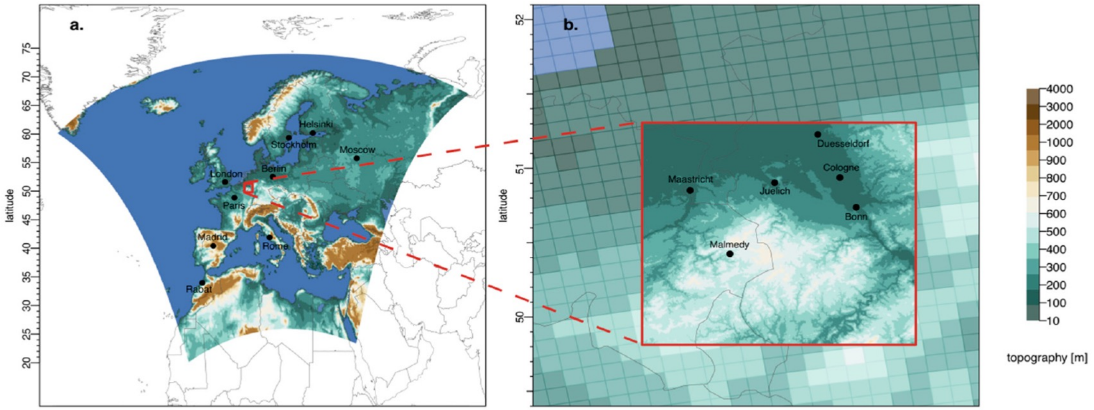
African domain is roughly 30.4 x 10^6 km^2

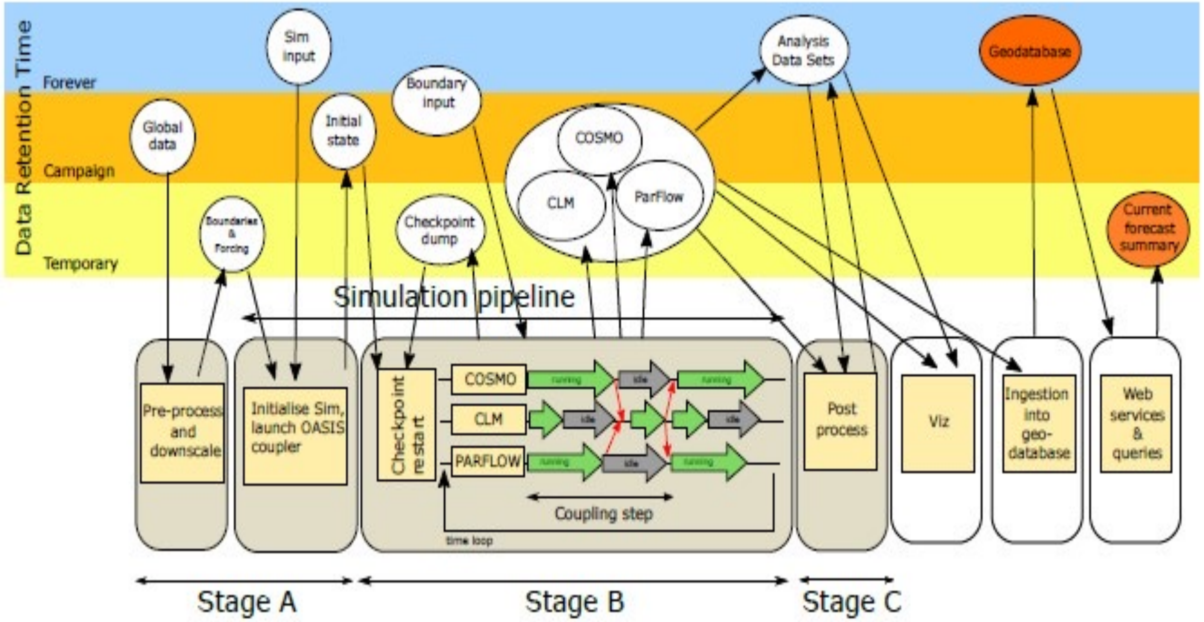
African domain is **three times larger** than Europe

Blunt extrapolation from Europe to Africa:

TSMP total output = 43.5 GB/sim.day → 496.2 TB / simulation

Input of boundary forcing data ~9 GB / sim.hour → 96 TB / simulation





EURO-CORDEX Pre-processing datasets and Output Volume

Input dataset : size and file count

File	Size(MB)	Number
Clm input (.nc)	645	4
ParFlow input(.pfsol) (.sa)	156	4
oasis*(.nc) and rmp files	213	6

Total: 1014 MB

Output dataset : size and file count

File	Size(MB)	Number
COSMO output (.nc)	463	15
CLM output r (.nc)	0.41	13
CLM output h0 (.nc)	67	13
ParFlow output press (.pfb)	21.2M	14
ParFlow output satur (.pfb)	21.2M	14
Other (lumped size)	746	72

Total: 9161 MB

IO-SEA

Ephemeral Services & Simulation Pipeline Description

The simulation pipeline includes the following phases for each of the components:

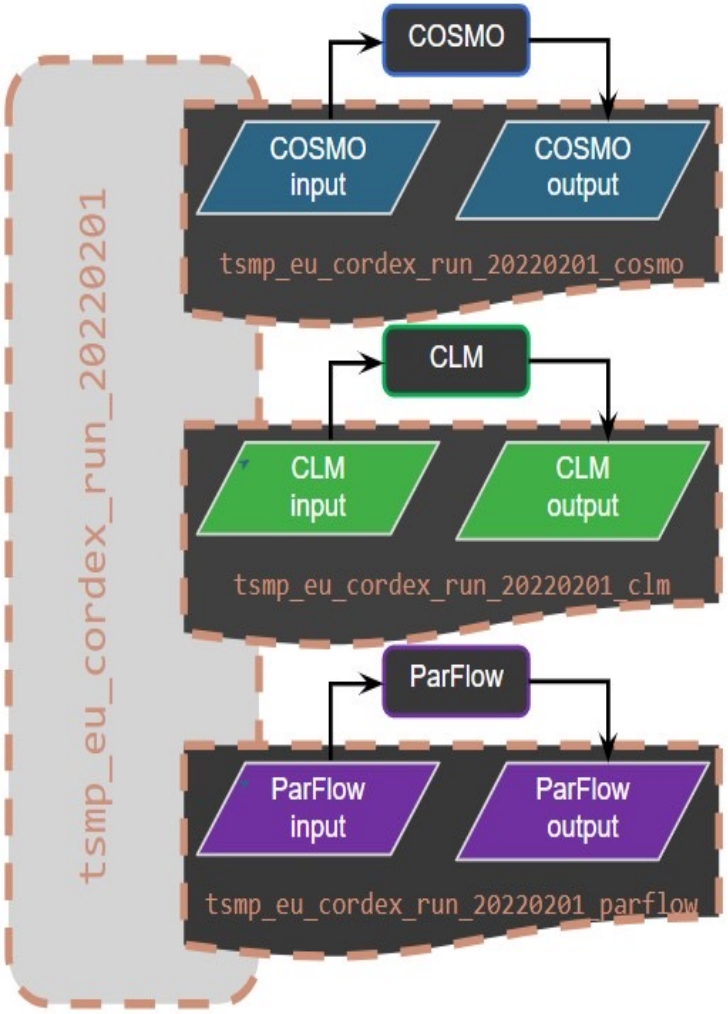
Model initialization and definition which includes model grid, model partition and I/O, constructing the coupled system via the OASIS layer which handles the coupling sequence and frequency, the names of the coupling fields, the spatial grid of the coupling fields, and finally the type of transformations of the 2D coupled fields.

Solving the systems of partial differential equations.

Sending–receiving of coupling fields between components through MPI communications.

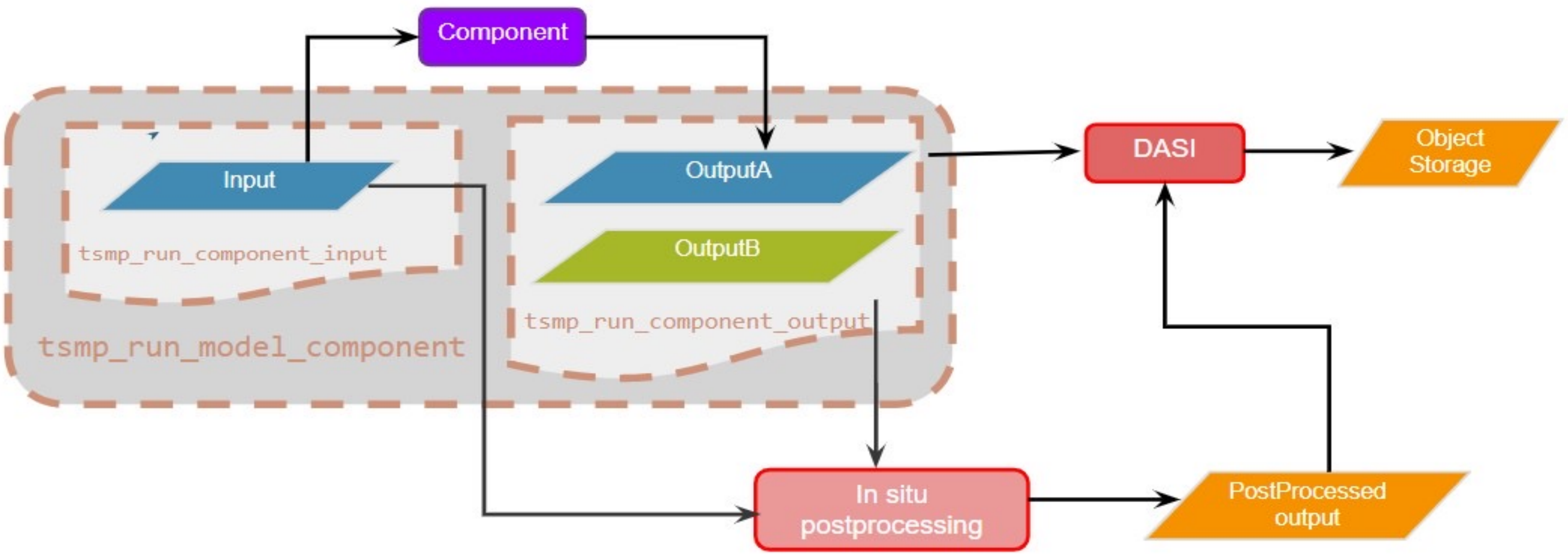
Check pointing and restarts of simulation (due to reaching maximum job duration as enforced by the scheduler).

Termination of simulation of each component model.



IO-SEA

TSMP Post-Processing



Stage C could be performed based on the DASI-ingested data, or on the POSIX TSMP output. Presumably, the choice will depend on the complexity and computational cost of the post-processing workflows and will be linked to the aforementioned concurrent or sequential relationship between Stages B and C. Therefore, in this stage heavy use of the POSIX-to-DASI and DASI-to-POSIX applications is foreseeable.

Launching the TSMP IO-SEA workflow

Stage A: TSMP Pre-processing step

Stage B: TSMP simulation pipeline and ingestion steps

Stage C: TSMP post-processing step

```
#!/bin/bash

# create namespaces & datasets
iosea-ns create --auto-create-dataset tsmp_eu_cordex_pre_run_20220201

# start session
#session name : Cordex_Session

iosea-wf start WORKFLOW=TSMP_workfkow.yaml SESSION=Cordex_Session

# Run Step A
iosea-wf run SESSION=Cordex_Session STEP=step_preprocessing

# Run main simulation
iosea-wf run SESSION=Cordex_Session STEP=step_simulation

# run to ingest POSIX into DASI
iosea-wf run SESSION=Cordex_Session STEP=step_ingestion

# Run Step C
iosea-wf run SESSION=Cordex_Session STEP=step_postprocessing

# display status
iosea-wf status SESSION=Cordex_Session

# finish workflow
iosea-wf stop SESSION=Cordex_Session
```