28th International Conference on Geoinformatics, Nanchang, Jiangxi, China | Nov. 3-5, 2021

A modular and parallelized modeling framework for distributed watershed modeling and scenario analysis

Liang-Jun Zhu (朱良君), Junzhi Liu, Cheng-Zhi Qin, A-Xing Zhu

State Key Laboratory of Resources and Environment Information System,

Institute of Geographic Sciences and Natural Resources Research,

Chinese Academy of Sciences

中国科学院
CHINESE ACADEMY OF SCIENCES

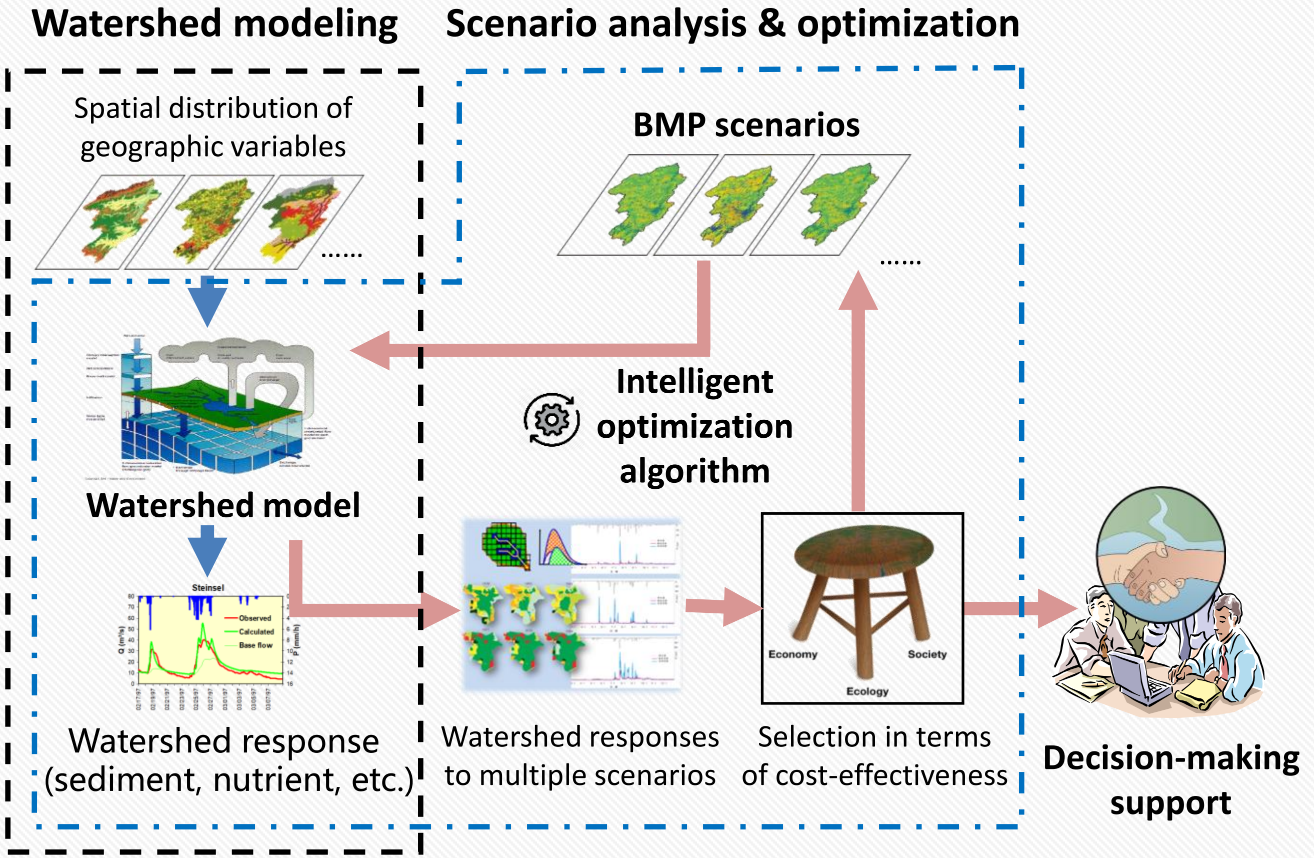
11/05/2021

Outline

1. Background and study issue
2. Basic idea and overall design
3. Case study
4. Conclusion and future work

1 Background and study issues

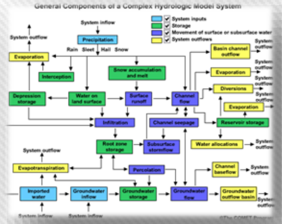
- Conflict between economic development and environmental conservation presents a huge challenge to watershed management.
- Integrated watershed modeling and scenario analysis provides a modern research paradigm to address this challenge.



朱阿兴, 朱良君, 史亚星, 秦承志, 刘军志. **2019**. 流域系统综合模拟与情景分析——自然地理综合研究的新范式? *地理科学进展*, 38(8): 1111–1122.

Key issues of integrated watershed modeling and scenario analysis

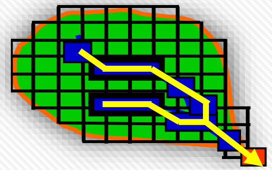
Systematization



- Physical geographic processes
- Human activity effects
- …

➔ Quantification of watershed response to management scenarios

Spatialization



- Spatial explicitly distribution
- Spatial interaction
- …

➔ Representation in process simulation and best management practice (BMP) configuration

Efficiency

- Reducing computation amount
- Parallel computing
- …

➔ High efficiency to answer

Watershed modeling & scenario analysis

Ease of use

- Intelligent inference
- User-friendly interface
- …

➔ Non-expert users

Decision-making support

➔ Reliable and effective

A flexible, extensible, and efficient modeling framework is needed!

朱阿兴, 朱良君, 史亚星, 秦承志, 刘军志. **2019**. 流域系统综合模拟与情景分析——自然地理综合研究的新范式? *地理科学进展*, 38(8): 1111–1122.

Existing modeling framework for watershed modeling

- **Environmental Modeling Framework (EMF)**
  - ✓ Standard interfaces **for coupling existing models**
  - ✓ **Parallel computing** support for common operations (e.g., regridding)
  - × **May not provide specific support for the parallelization of distributed watershed models**

ESMF

OpenMI

CSDMS
community surface dynamics modeling system

- **Watershed Modeling Framework**
  - ✓ **EMF specifically designed for watershed modeling**, e.g., OMS3 (David et al., 2013) and ECHSE (Kneis, 2015).
  - × **Shared-memory multithreaded programming** (e.g., **OpenMP**), **limited scalability** on distributed-memory platforms (e.g., **SMP cluster**).

Study issue

How to design a **flexible, extensible, and efficient watershed modeling framework** to promote research of integrated watershed modeling and scenario analysis?

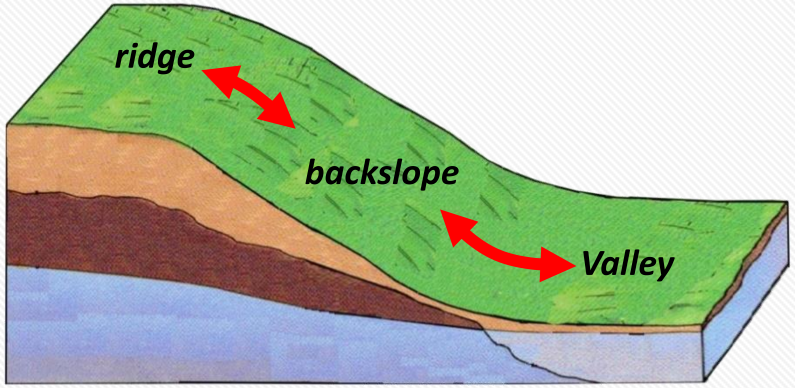
- **flexible and extensible**
- **Efficient**
- **Easy-to-use**

2 Basic idea and overall design

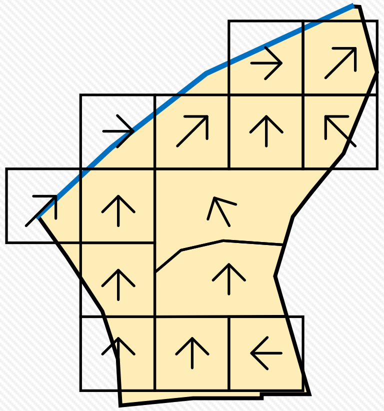
Hierarchical spatial discretization of a watershed from different perspectives:

- ✓ **Watershed processes**: Subbasin (including channel) – hillslope – slope position – patch (Band, 1999)
- ✓ **Distributed watershed simulation**: Subbasin (including channel), grid cell or patch, etc.
- ✓ **Management practice allocation**: Subbasin, hillslope, slope position, grid cell or patch, etc.
- ✓ …

Watershed and subbasins

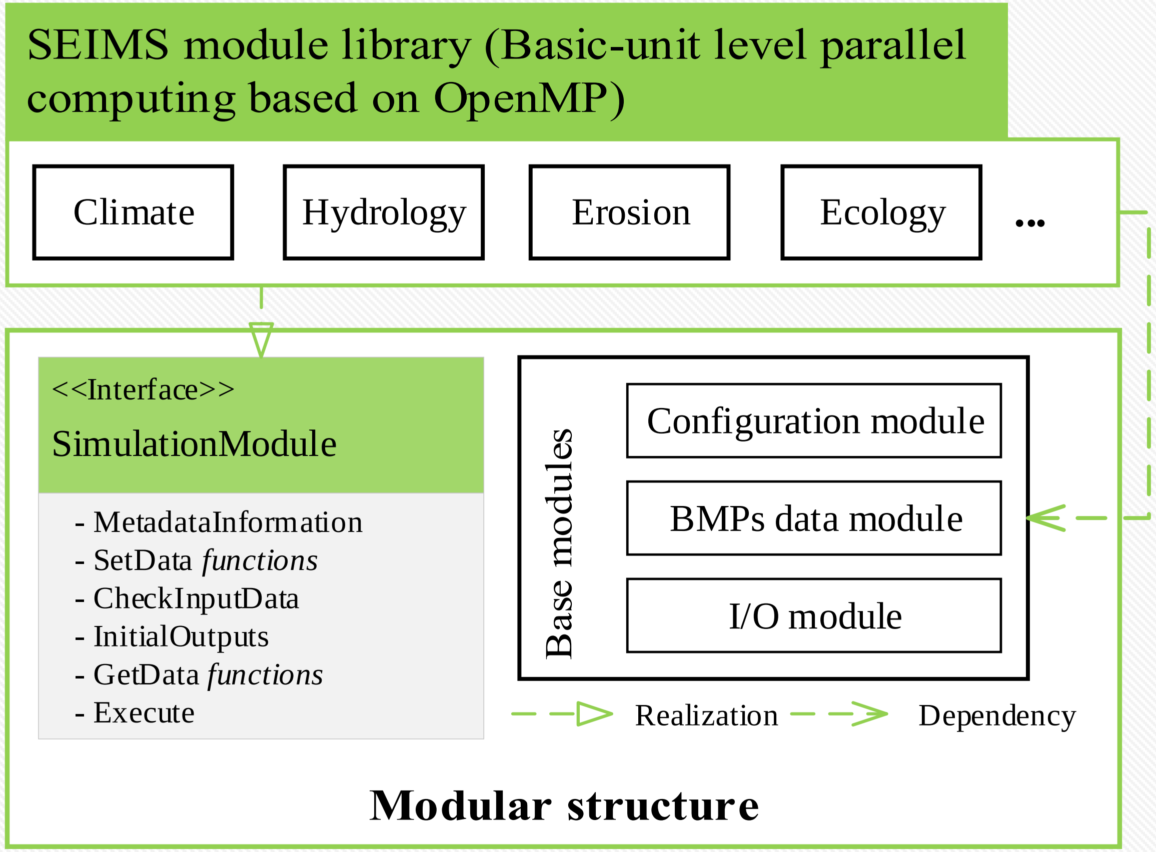


Hillslope and slope positions



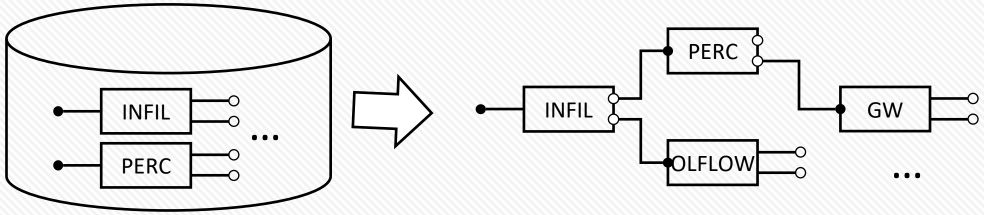
Grid cell or patch

Flexible and extensible modular structure



➢ For distributed watershed simulation

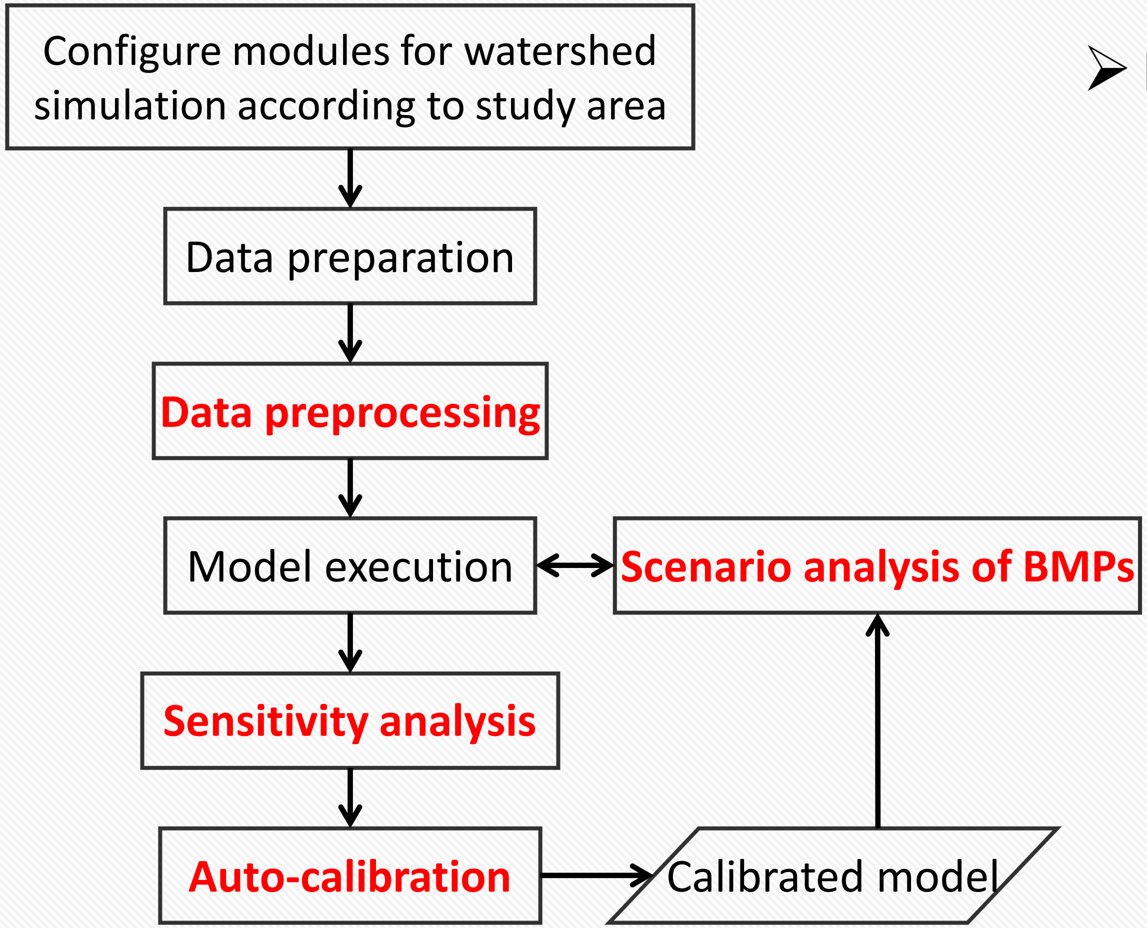
- **Each watershed subprocess is simulated on one type of simulation unit** by one module using a specific algorithm.
- Each module inherits from standard and concise interfaces which exposes IO information.
- User-configured modules are dynamically loaded and linked as a simulation workflow.



Module library **Simulation workflow**

Watershed modelers can focus on and contribute specific simulation algorithms!

Flexible and extensible modular structure

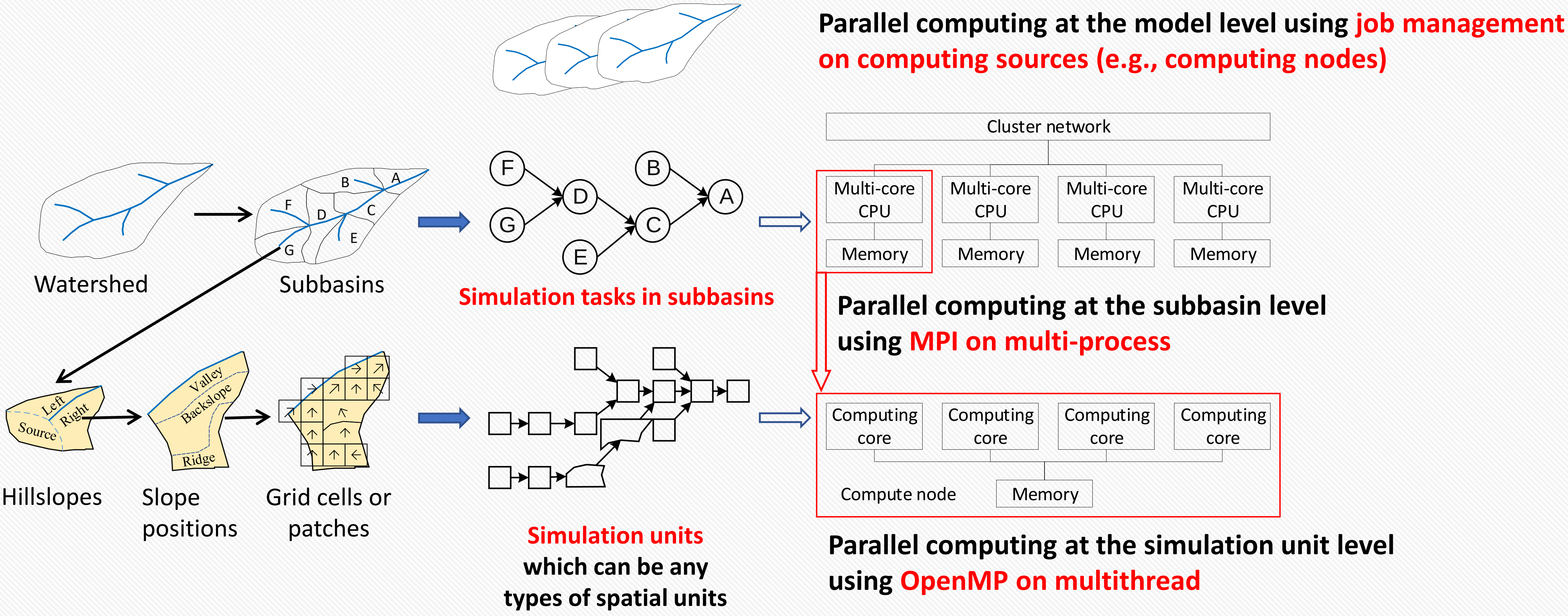


- **For watershed modeling and scenario analysis**
  - Several modules (or packages) for different steps.
  - **Each module (or package) defines a general, configurable, and extensible workflow**.
  - Generic independent functions are also summarized, e.g., repeatedly executing models and gathering outputs.

Watershed modelers can easily extend data for simulation and algorithms for model-level applications.

Efficient and easy-to-use multi-level parallel computing middleware

- **Model-level parallelization**: job management by workload manager, e.g., SLURM, SCOOP in Python
- **Inside-model parallelization**: two-level parallelization strategy (Liu et al., 2014, 2016) that exploit the parallelizability at both coarse-grained and fine-grained levels.



Technology selection and implementation

SEIMS, short for Spatially Explicit Integrated Modeling System

- **Programming languages:**
  - **C++:** SEIMS main programs and modules
  - **Python:** Utility tools of entire workflow, e.g., data preprocessing, sensitivity analysis, auto-calibration, and scenario analysis.
- **Data management:** MongoDB database, for its support of flexible data structure and high IO concurrency
- **Module library:** covering hydrology, erosion, nutrient cycling, and plant growth processes from WepSpa, SWAT, LISEM, etc.
- **Source code:** freely available in Github https://github.com/lreis2415/SEIMS

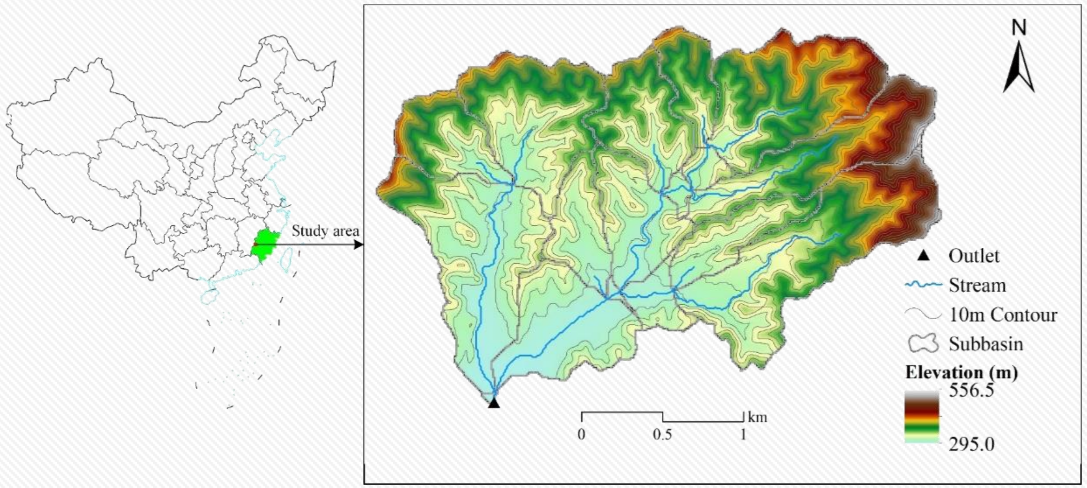
SEIMS aims to facilitate rapid development of parallelized watershed models and model-level application tools such as scenario analysis.

Zhu L-J, Liu J*, Qin C-Z*, Zhu A-X. **2019**. A modular and parallelized watershed modeling framework. *Environmental Modelling & Software*, 122: 104526.

3 Case study – scenario analysis of BMP for mitigating soil erosion

Study area: Youwuzhen watershed (~5.39 km^2, **53,933 grid cells with a 10 m resolution**), Fujian province, China

- ***Location***: in the upstream of Ting river, **the typical red-soil hilly region in southeastern China**
- ***Terrain***: low hills with steep slopes (average slope: 16.8°), broad alluvial valleys
- ***Climate***: under a mid-subtropical monsoon moist climate
- ***Landuse***: primarily, forest (59.8%), paddy field (20.6%), and orchard (12.8%)
- ***Soil***: red soil (dominant type, infertile, acidic, nutrient-deficient, poor in organic matter, low capacity for holding and supplying water) and paddy soil.
- **Representative BMPs** for mitigating soil erosion



(Chen et al. 2017; Qin et al., 2018)

Closing measures (CM)

Arbor-bush-herb mixed plantation (ABHMP)

Low-quality forest improvement (LQFI)

Orchard improvement (OI)

SEIMS-based Youwuzhen daily model

Systematization

Spatialization

Efficiency

Ease of use

Decision-making

```
### Driver factors, including meteorological data and precipitation
1 | TimeSeries | | TSD_RD
2 | Interpolation_0 | Thiessen | ITP
### Hillslope processes
3 | Soil temperature | Finn Plauborg | STP_FP
4 | PET | PenmanMonteith | PET_PM
5 | Interception | Maximum Canopy Storage | PI_MCS
6 | Snow melt | Snowpeak Daily | SNO_SP
7 | Infiltration | Modified rational | SUR_MR
8 | Depression and Surface Runoff | Linsley | DEP_LINSLEY
9 | Hillslope erosion | MUSLE | SERO_MUSLE
10 | Plant Management Operation | SWAT | PLTMGT_SWAT
11 | Percolation | Storage routing | PER_STR
12 | Subsurface | Darcy and Kinematic | SSR_DA
13 | SET | Linearly Method from WetSpa | SET_LM
14 | PG | Simplified EPIC | PG_EPIC
15 | ATMDEP | Atomosphere deposition | ATMDEP
16 | NUTR_TF | Transformation of C, N, and P | NUTR_TF
17 | Water overland routing | IUH | IUH_OL
18 | Sediment overland routing | IUH | IUH_SED_OL
19 | Nutrient | Attached nutrient loss | NUTRSED
20 | Nutrient | Soluble nutrient loss | NUTRMV
21 | Pothole | SWAT cone shape | IMP_SWAT
22 | Soil water | Water balance | SOL_WB
### Route Modules, including water, sediment, and nutrient
23 | Groundwater | Linear reservoir | GWA_RE
24 | Nutrient | groundwater nutrient transport | NUTRGW
25 | Water channel routing | MUSK | MUSK_CH
26 | Sediment routing | Simplified Bagnold eq. | SEDR_SBAGNOLD
27 | Nutrient | Channel routing | NutrCH_QUAL2E
```

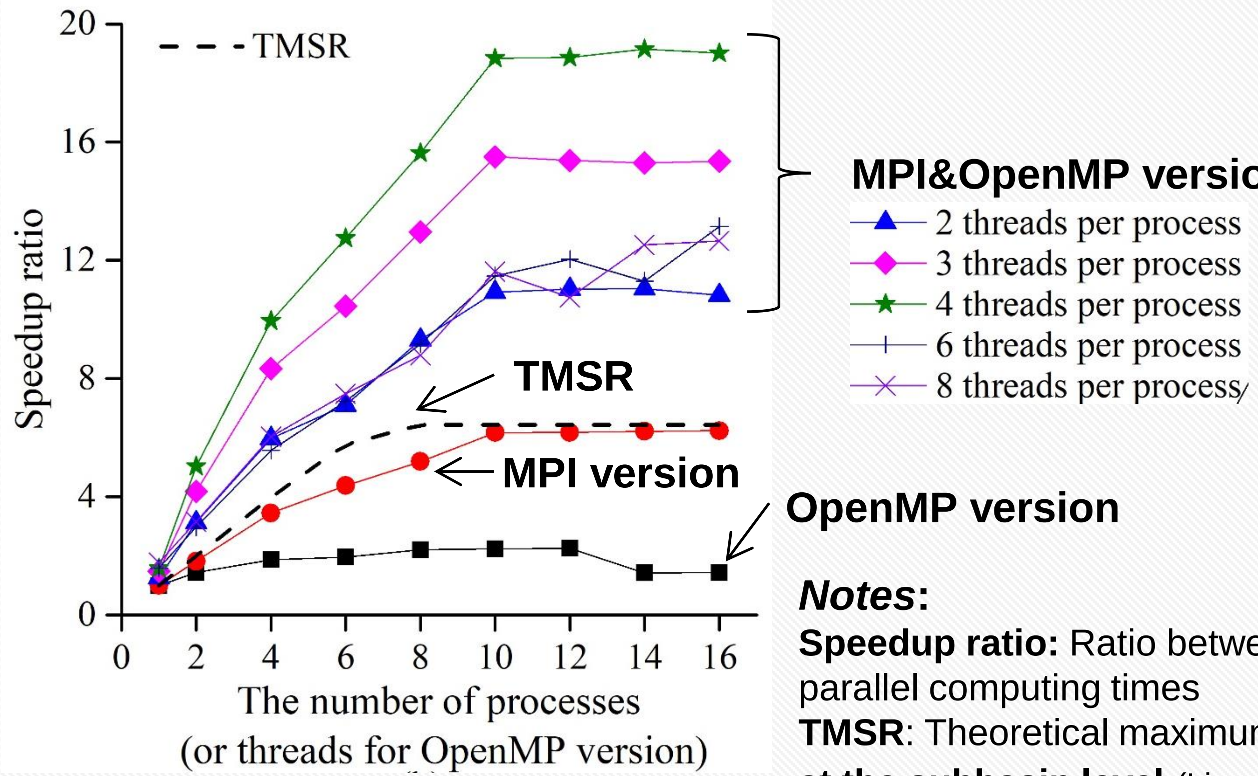
Loading and preprocessing driver factors, e.g., climate data

Hillslope processes, e.g., potential evapotranspiration, canopy interception, depression storage, surface runoff, percolation, interflow, plant growth, soil loss.

Channel routing processes of water, sediment, nutrient, etc.

Parallel performance of the two-level parallelization strategy

- Systematization
- Spatialization
- **Efficiency**
- Ease of use
- Decision-making



- 53,933 grid cells
- 17 subbasins
- A Linux cluster with 134 computing nodes
- Each node has 12 physical cores

Notes:
Speedup ratio: Ratio between serial and parallel computing times
TMSR: Theoretical maximum speedup ratio **at the subbasin level** (Liu et al., 2013)

- **Subbasin level parallelization (MPI version) is greater than that of basic simulation unit level (OpenMP version).**
- **The two-level parallelization (MPI&OpenMP version) is dramatically improved than any single level parallelization and greater than *TMSR*.**

Zhu L-J, Liu J*, Qin C-Z*, Zhu A-X. **2019**. A modular and parallelized watershed modeling framework. *Environmental Modelling & Software*, 122: 104526.

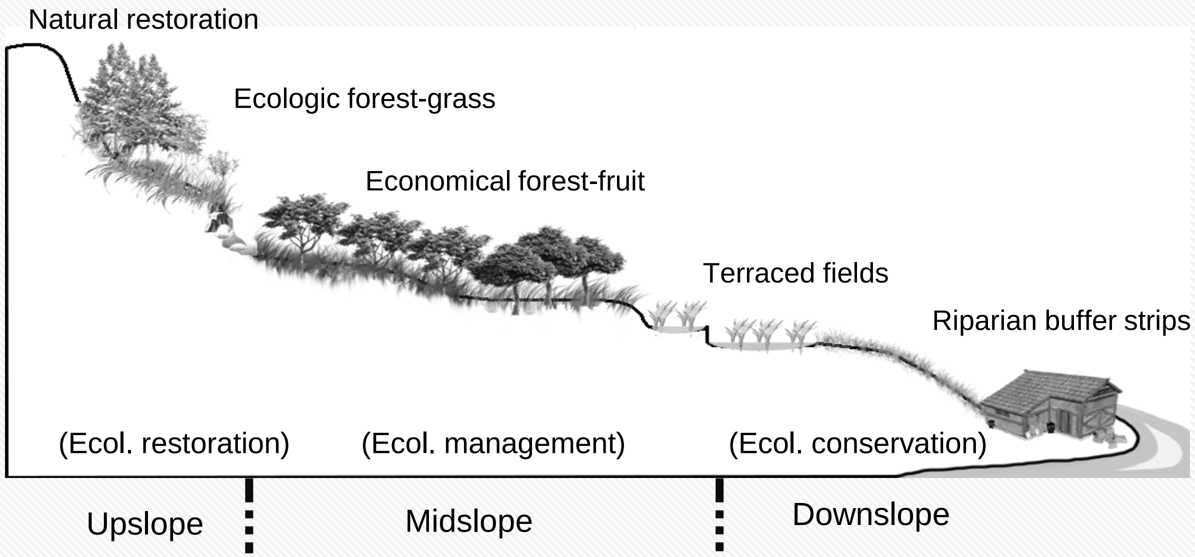
Consideration of spatial interaction of BMPs in scenario analysis

- Systematization
- **Spatialization**
- Efficiency
- Ease of use
- **Decision-making**

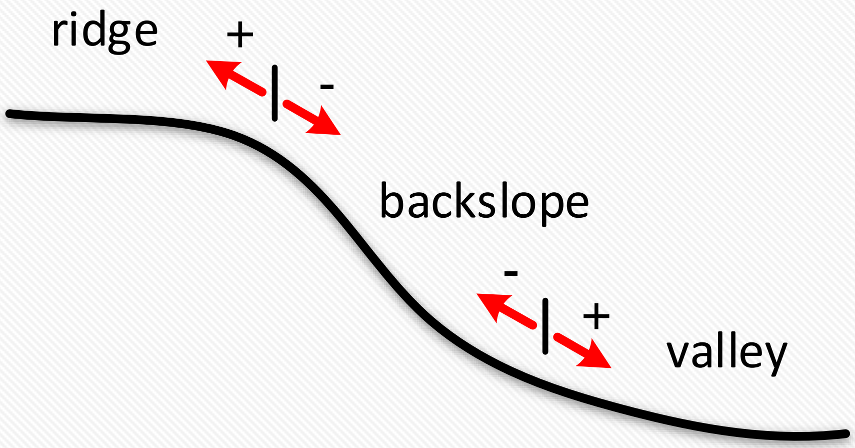
➢ **Spatial optimization of BMPs based on slope position units** **(Qin et al., 2018; Zhu et al., 2019):**

- Spatial interaction of BMPs configured along hillslope
- Domain knowledge such as integrated watershed management scheme in practice

➢ **Adjust boundaries of slope position units to consider the optimization of BMP areas** **(Zhu et al., 2021).**

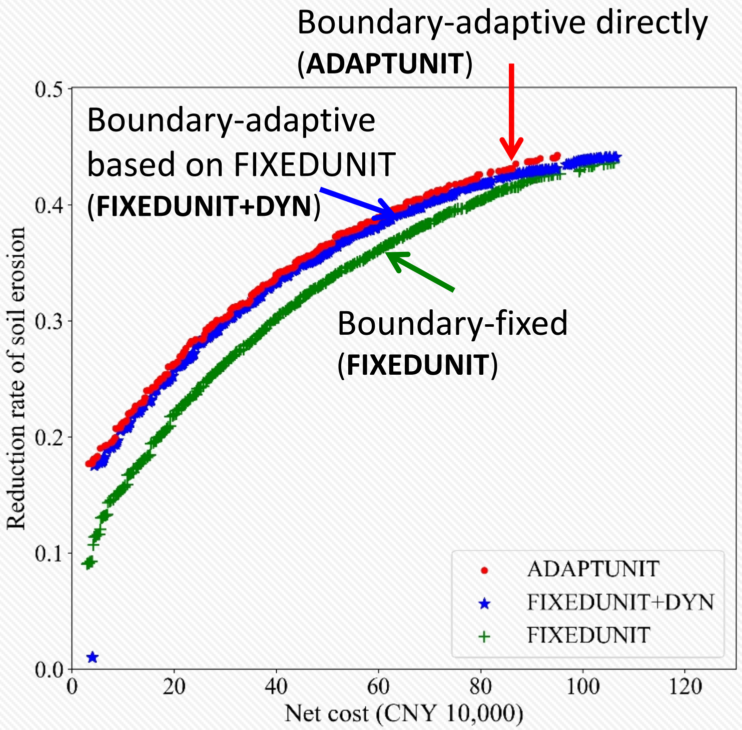


e.g., practical management experiences of soil and water conservation in SE China
(an integrated management scheme adapted from Cai et al. (2012))

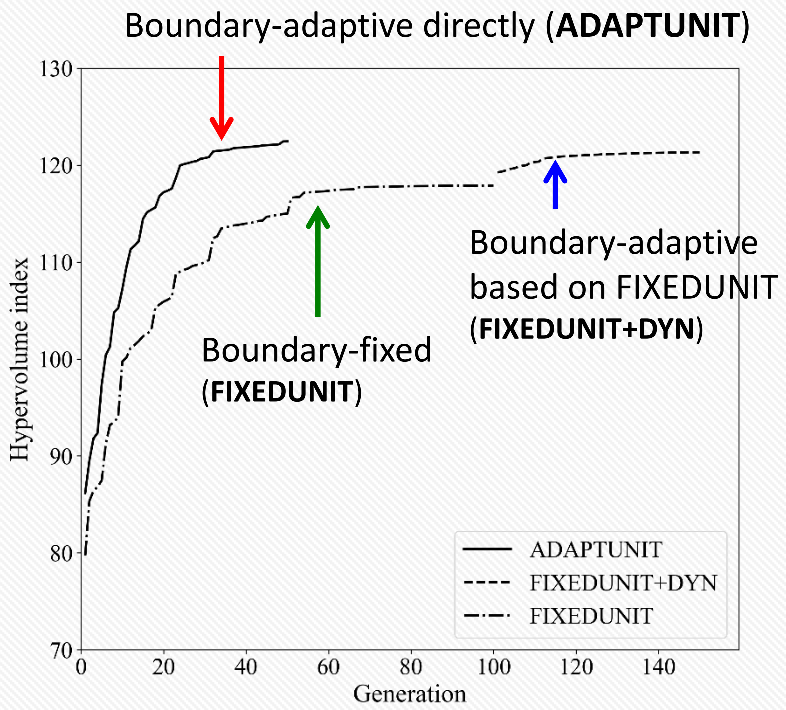


Dynamic boundary adjustment based on fuzzy slope positions

Results: Near-optimal Pareto solutions and hypervolume index



Near-optimal Pareto solutions for the 50th and 100th generations

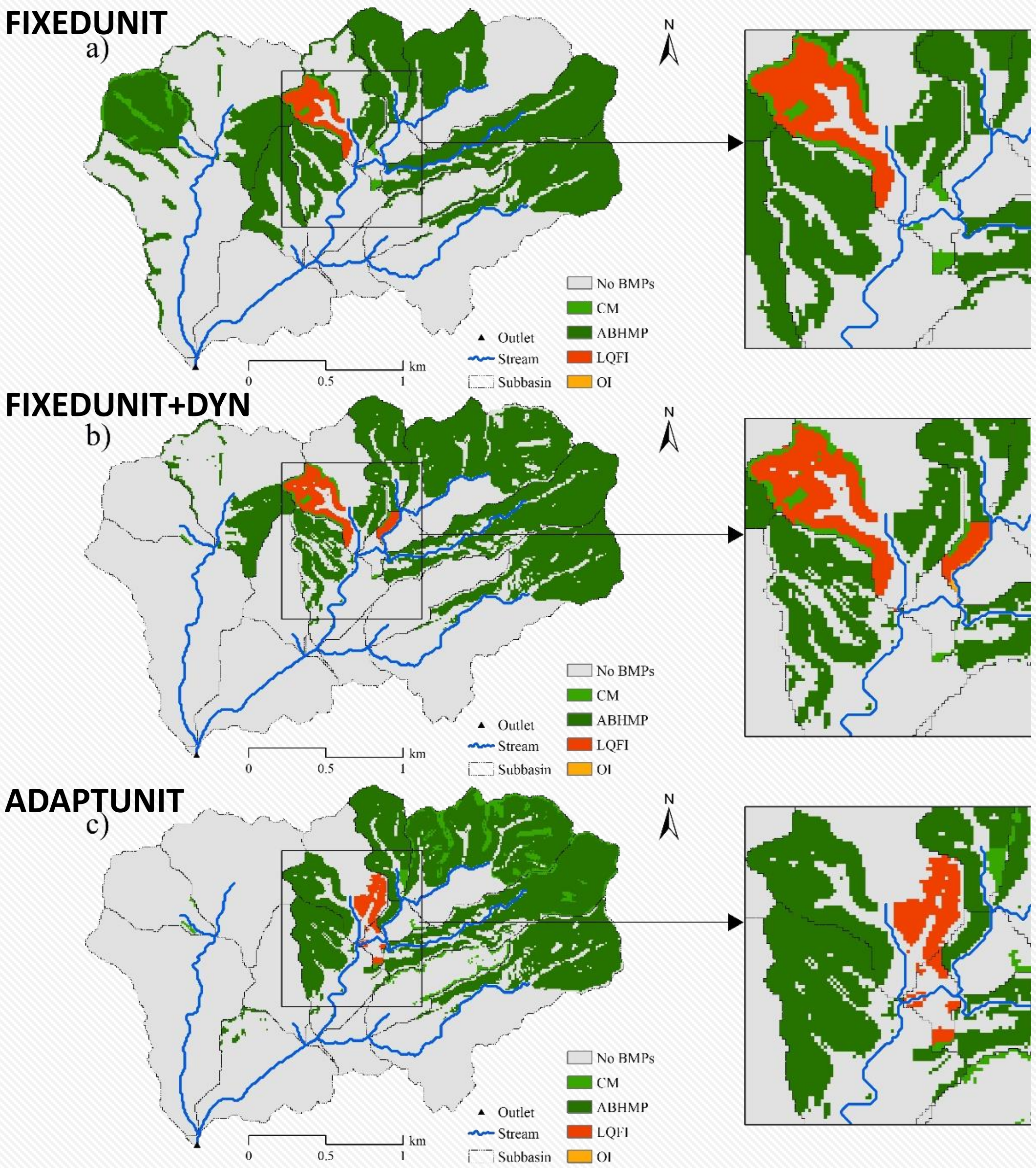


Hyervolume index changes with generations

- ✓ **Boundary-adaptive method performed the best**
- ✓ **Significantly enlarge the search space and obtain optimal BMP scenarios with better cost-effectiveness and higher optimization efficiency.**

Zhu L-J, Qin C-Z*, Zhu A-X. **2021**. Spatial optimization of watershed best management practice scenarios based on boundary-adaptive configuration units. *Progress in Physical Geography: Earth and Environment*, 45(2):207–227.

Results: spatial distribution of optimized BMP scenarios



- ✓ Compared with fixed boundary units, BMP scenarios based on **boundary-adaptive units showed more fragmented or even mosaic spatial distribution** (b and c compared to a).
- ✓ With more hillslopes underwent boundary adjustments, **utilizing boundary adjustment from the initialization of optimization produce better BMP scenarios** (c compared to b).

4 Conclusion

SEIMS: A modular and parallelized modeling framework for distributed watershed modeling and scenario analysis

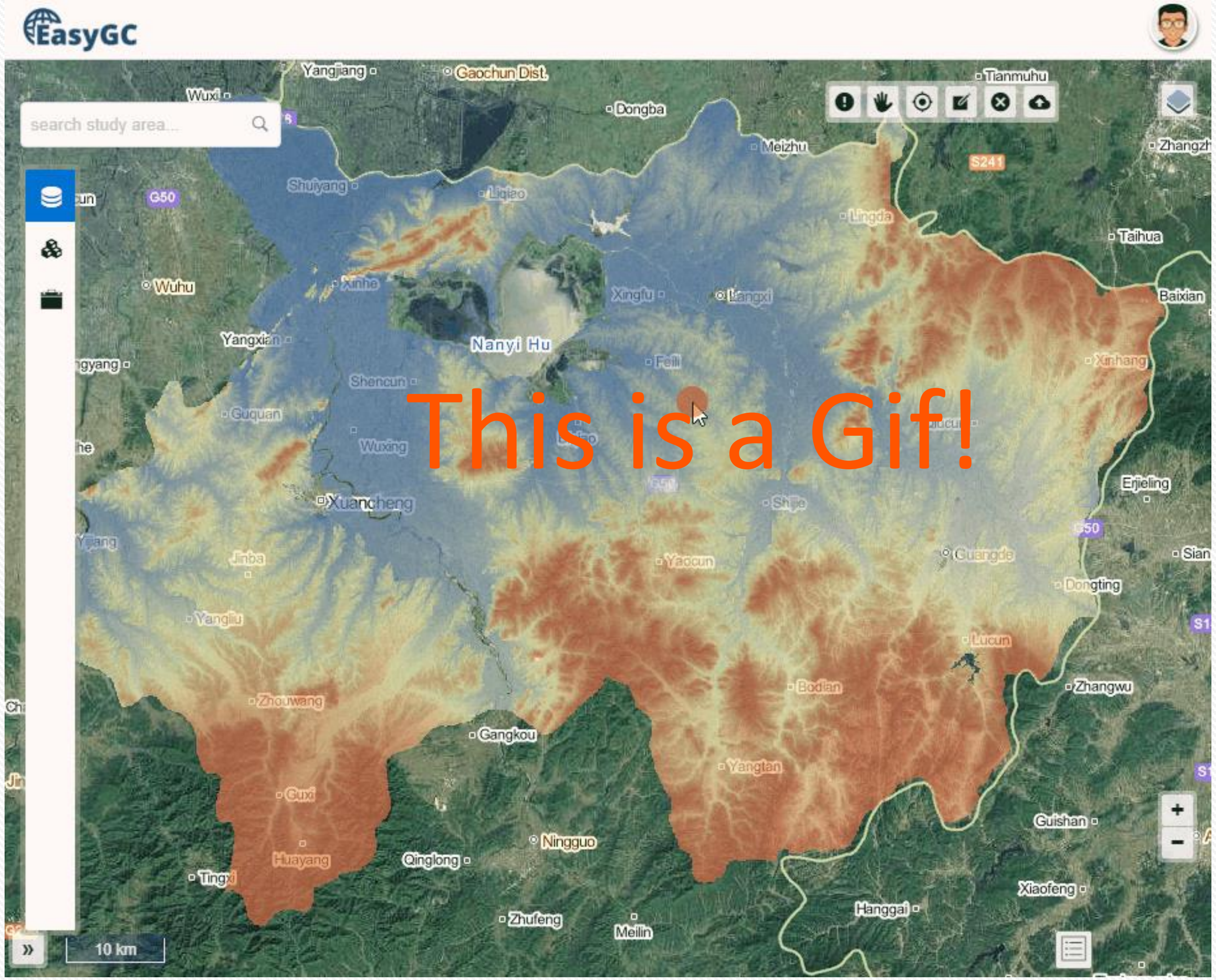
- **Systematization**: Flexible and extensible modular structure
- **Spatialization**: Spatially explicit modeling and scenario analysis
- **Efficiency**: Multi-level parallel computing middleware
- **Ease of use**: Transplant/rewrite/write new SEIMS modules in a nearly serial programming manner
- **Decision-making**: Knowledge-driven scenario analysis

Future direction – Intelligent modeling environment

- **Systematization**
- **Spatialization**
- **Efficiency**
- **Ease of use**
- **Decision-making**

EasyGC platform
(Easy Geo-Computation)
http://easygeoc.net:8090

Directed by *Prof. A-Xing Zhu*
and *Prof. Cheng-Zhi Qin*



- **Automatic data discovery and preparation**
- **Intelligent model construction**
- **Efficient model execution in the Cloud**
- **…**

Thanks for your attention!

Selected peer-reviewed papers:

Zhu L-J, Qin C-Z*, Zhu A-X. **2021**. Spatial optimization of watershed best management practice scenarios based on boundary-adaptive configuration units. *Progress in Physical Geography: Earth and Environment*, 45(2):207–227.

朱阿兴, 朱良君*, 史亚星, 秦承志, 刘军志. **2019**. 流域系统综合模拟与情景分析——自然地理综合研究的新范式? *地理科学进展*, 38(8): 1111–1122.

Zhu L-J, Liu J*, Qin C-Z*, Zhu A-X. **2019**. A modular and parallelized watershed modeling framework. *Environmental Modelling & Software*, 122: 104526.

Qin C-Z, Gao H-R, Zhu L-J*, Zhu A-X, Liu J-Z, Wu H. **2018**. Spatial optimization of watershed best management practices based on slope position units. *Journal of Soil and Water Conservation*, 73(5): 504–517.

Open-source software:

SEIMS (Spatially Explicit Integrated Modeling System): *https://github.com/lreis2415/SEIMS*

zlj@lreis.ac.cn

https://zhulj.net