



Microsoft Research

Faculty
Summit

2014 15TH ANNUAL

Numerical Modeling of Ecohydrological Processes and Contaminant Transport Using Microsoft Azure Cloud

Chunmiao Zheng

Center for Water Research

Peking University

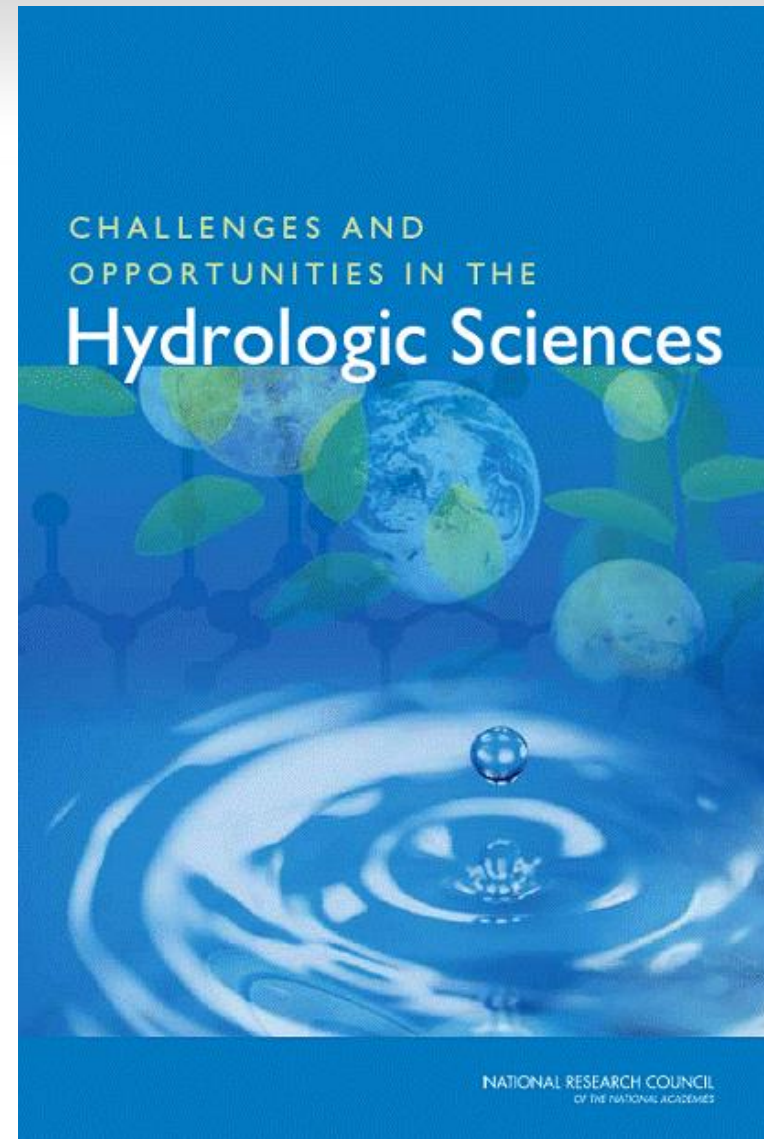
(<http://hydro.pku.edu.cn>)

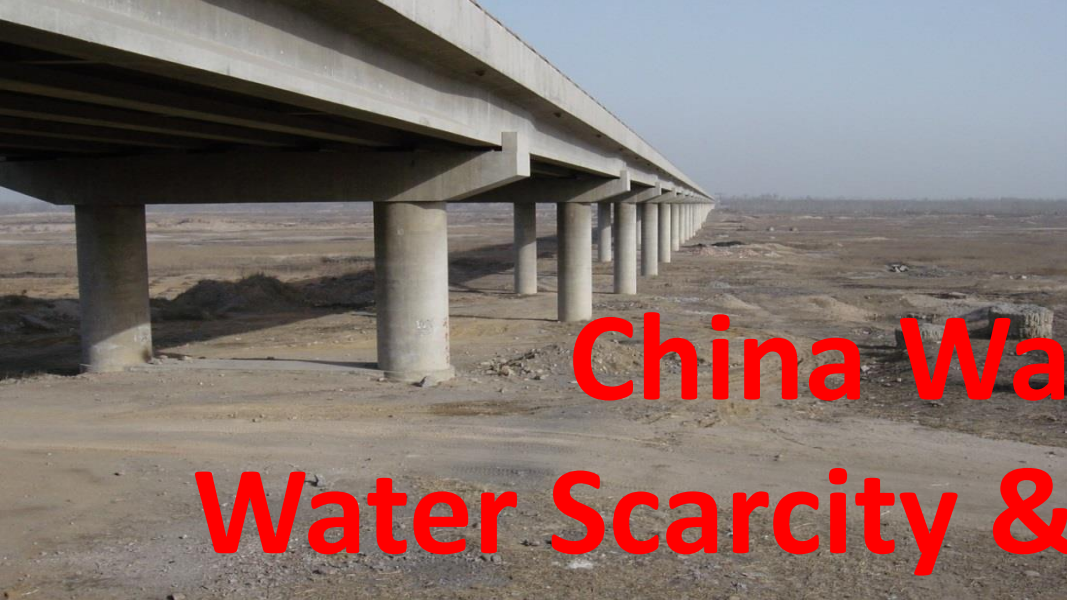


Three Grand Challenges in Hydrological Sciences

- Water Cycle:
An Agent of Change
- Water and Life
- Clean Water for People
and Ecosystems

– National Research Council (2012)





China Water Crisis: Water Scarcity & Water Pollution



Water Pollution Control Action Plan ~2 Trillion RMB

LETTERS

edited by Jennifer Sills

China's "Love Canal" Moment?

IN FEBRUARY 2013, A POSTING IN CHINA'S BURGEONING BLOGOSPHERE ACCUSED BUSINESS owners in Shandong Province of disposing waste water through injection wells and contaminating shallow groundwater (1). This seemingly innocuous message, describing a practice that is not uncommon, ignited a firestorm on the Internet (2). The outburst of condemnation and concern caught many observers by surprise and reached the uppermost echelon of the Chinese government (3). Du Ying, Vice Chair of the powerful National Development and Reform Commission, declared that "China needs a law specifically designed for groundwater protection" (4).

Groundwater provides about 20% of total water supplies for China, and 50 to 80% of water in water-scarce north and northwest regions of the country (5). However, the outlook for groundwater quality is bleak. According to the latest round of water well sampling in 2011 in more than 200 cities and administrative regions by China's Ministry of Land and Resources, fully 55% of more than 4700 samples indicated groundwater of category IV or V [on a scale of I to V from the best to poorest quality (6)]. Still, no one knows the true extent and severity of groundwater pollution in China.

It is imperative that the Chinese government move aggressively and assertively to combat groundwater pollution. The challenges and action of the United States and other developed countries should serve as an example.

The United States alone has spent hundreds of billions of dollars on detecting, monitoring, assessing, and remediating contaminated groundwater since the 1970s (7), when groundwater

6. China Ministry of Environmental Protection, "National Quality Standard for Ground Water (GB/T 14848-93)" (1994).
7. National Research Council, "Alternatives for managing the nation's complex contaminated groundwater sites" (National Academies Press, Washington, DC, 2012).

The True Challenge of Giant Marine Reserves

THE NEWS FOCUS STORY "GIANT MARINE reserves pose vast challenges" (C. Pala, 8 February, p. 640) discusses the potential role of large marine reserves in conserving pelagic species. We highlight three points that are misrepresented in the article.

First, the story describes tropical tuna stocks in the Indian Ocean as "depleted," implying that they are severely overexploited (1). Fishery overcapacity is certainly worrisome, but these stocks are not currently overexploited (2). This clarification by no means precludes appropriate use of area-based management, but stock status is central to weighing different management options.

Second, the article claims that "mainstream marine biologists are more optimistic" about the efficacy of the Chagos Islands reserve because "tuna there don't necessarily swim vast distances." We disagree. Juvenile



Call to action. A child protests in the Love Canal neighborhood in 1978.

of the United States alone has spent hundreds of billions of dollars on detecting, monitoring, assessing, and remediating contaminated groundwater since the 1970s (7), when groundwater

Science 340
May 17, 2013

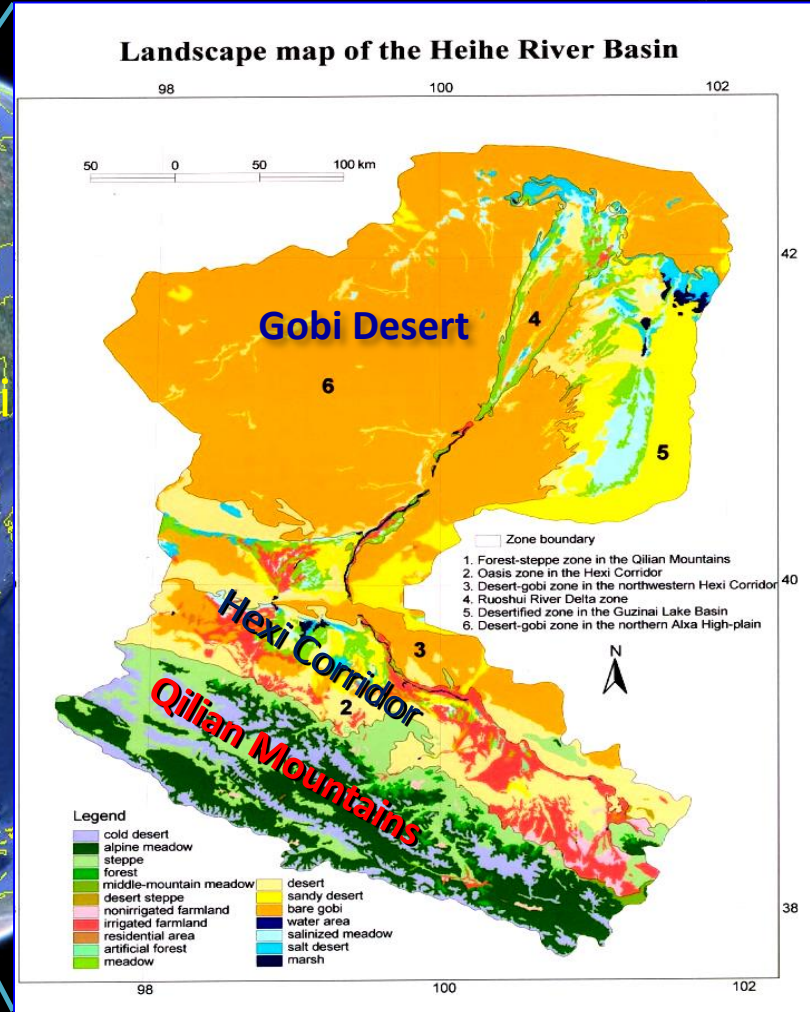
C. Zheng & J. Liu

Case Study 1: Numerical Modeling of Ecohydrological Processes in Heihe River Basin



The second largest inland river basin in China

Total Area:
130,000 km²

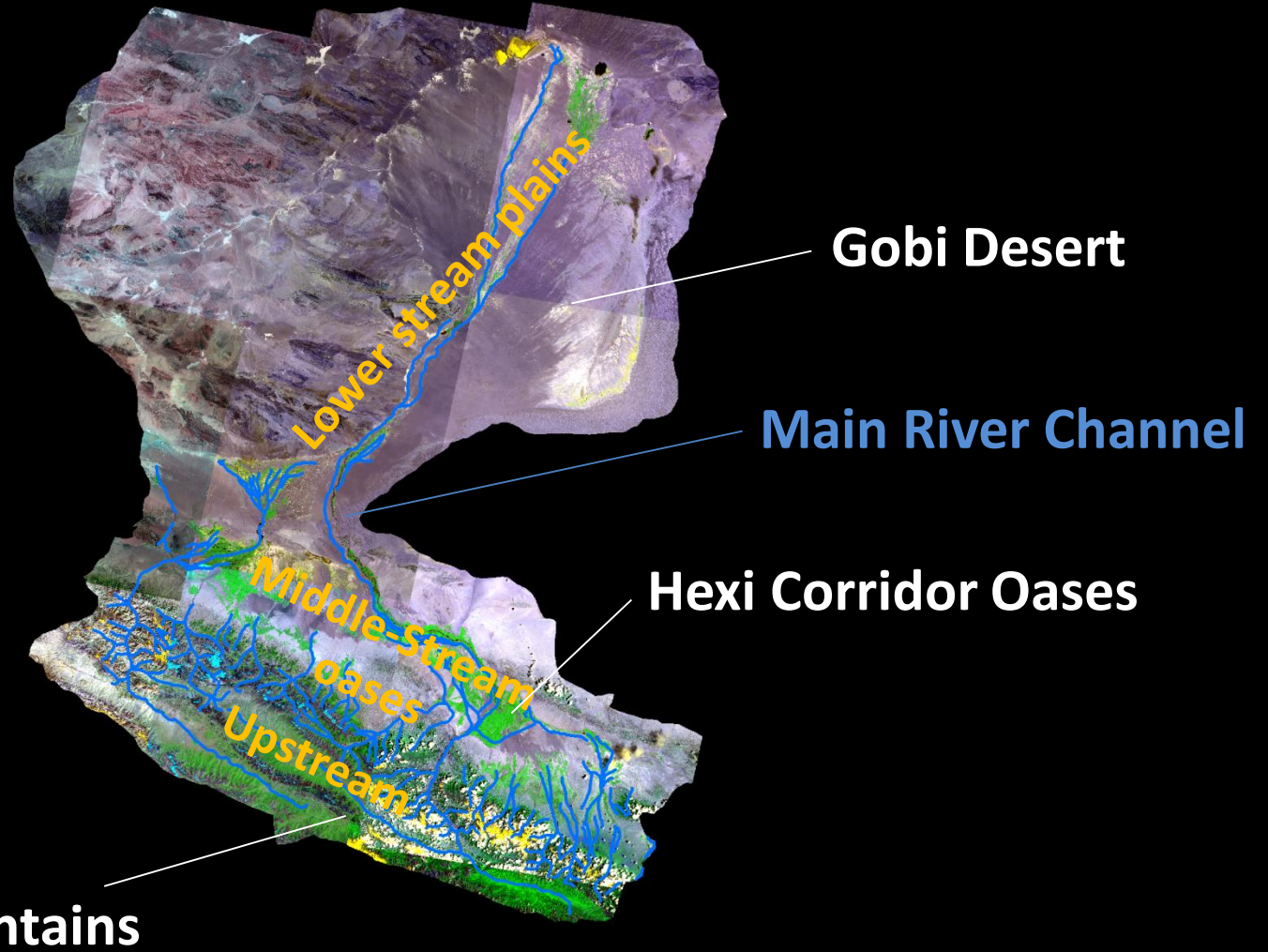




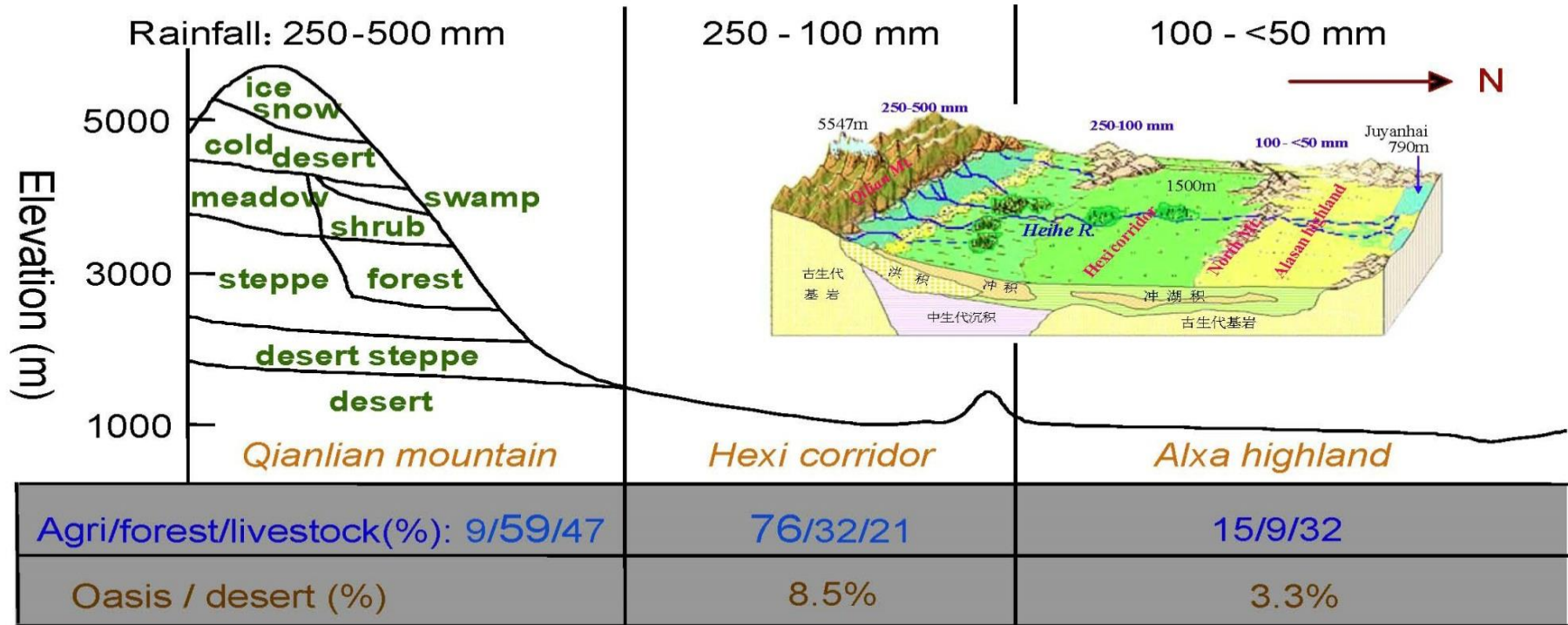
Mountain: 34%

Oasis: 9%

Desert: 57%



Landscape of Heihe River Basin



Vertical profile in the mountains



Ejina Basin



Basin

Field Study

Terminal Lake

Qinghai Country

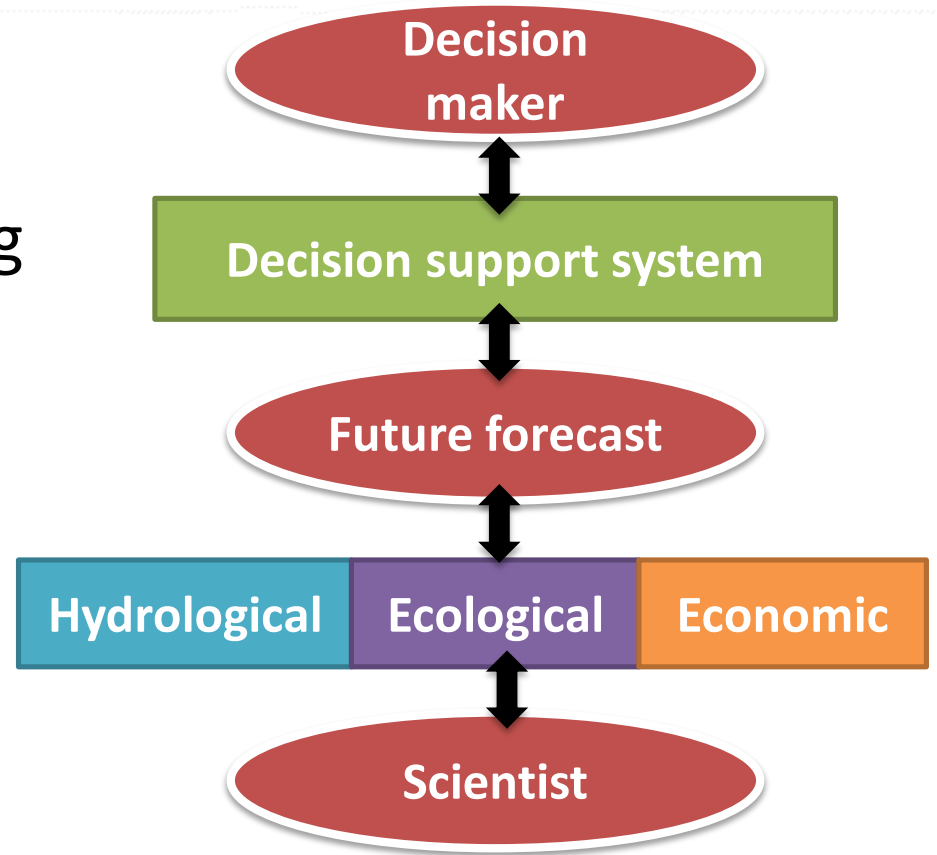
Heihe Research Program

- An on-going major research programme of National Natural Science Foundation of China (2011-2018)
- Conduct an **integrative study of ecological and hydrologic processes** in Heihe River Basin toward more sustainable water resources management
- Led by Prof. Cheng Guodong of Chinese Academy of Sciences and advised by an expert panel of multidisciplinary scientists
- 200 million RMB core funding (**~32 million USD**)

Overall Objectives

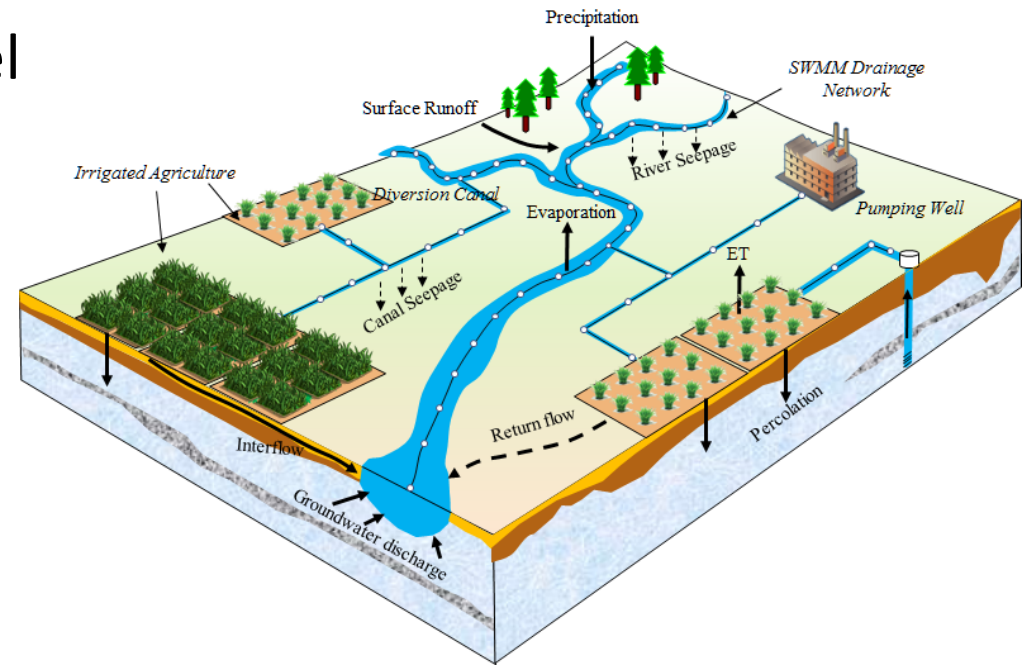
- Integrate observation, experimentation, and modeling
- Improve predictive capability
- Increase water use efficiency

Toward more sustainable water resources in arid ecosystems

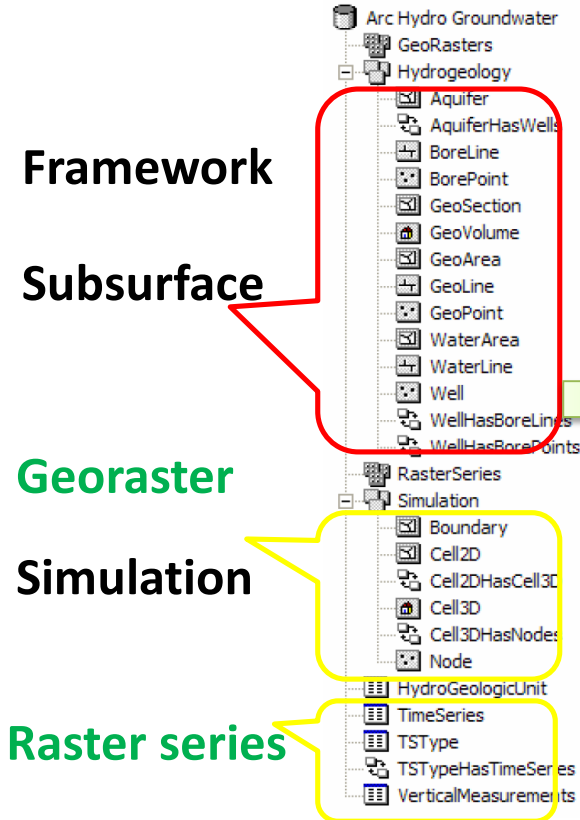


HEIFLOW (Hydrological-Ecological Integrated watershed-scale FLOW model)

- Physically-based distributed-parameter 3D numerical model
- Include all key components of the hydrological cycle
- 2D overland flow and river channel hydrodynamics
- Saturated/unsaturated zones
- 3D solute and heat transport
- Modules for agricultural crops and desert vegetation



Heihe River Basin Geodatabase



Surface water :(2-D)

Waterbody (Polygon feature class): lake, ponds, (swamps)

Waterline (Line feature class): streams, rivers

WaterPoint (Point feature class): springs, water withdrawal/discharge locations

Watershed (Polygen feature class): drainage areas

Groundwater: (2-D , 3-D)

Aquifer (Polygon feature class): confined and unconfined

Well (Point feature class): monitoring, water supply, and irrigation wells

Borehole (Boreholelog Table): vertical data

Boreline and BorePoint: 3D (z-enabled) line and point feature class

GeoArea, GeoPoint

“Digital Heihe” Datasets

DIGITAL RIVER BASIN

<http://westdc.westgis.ac.cn/>



Foundation data

DEM, Topographic, Hydrological map



Earth observation data

Landsat, ASTER, QuickBird——



Thematic data

Geology, Hydrology, Vegetation——



Observation data

Meteorology、 Hydrology、 Groundwater——



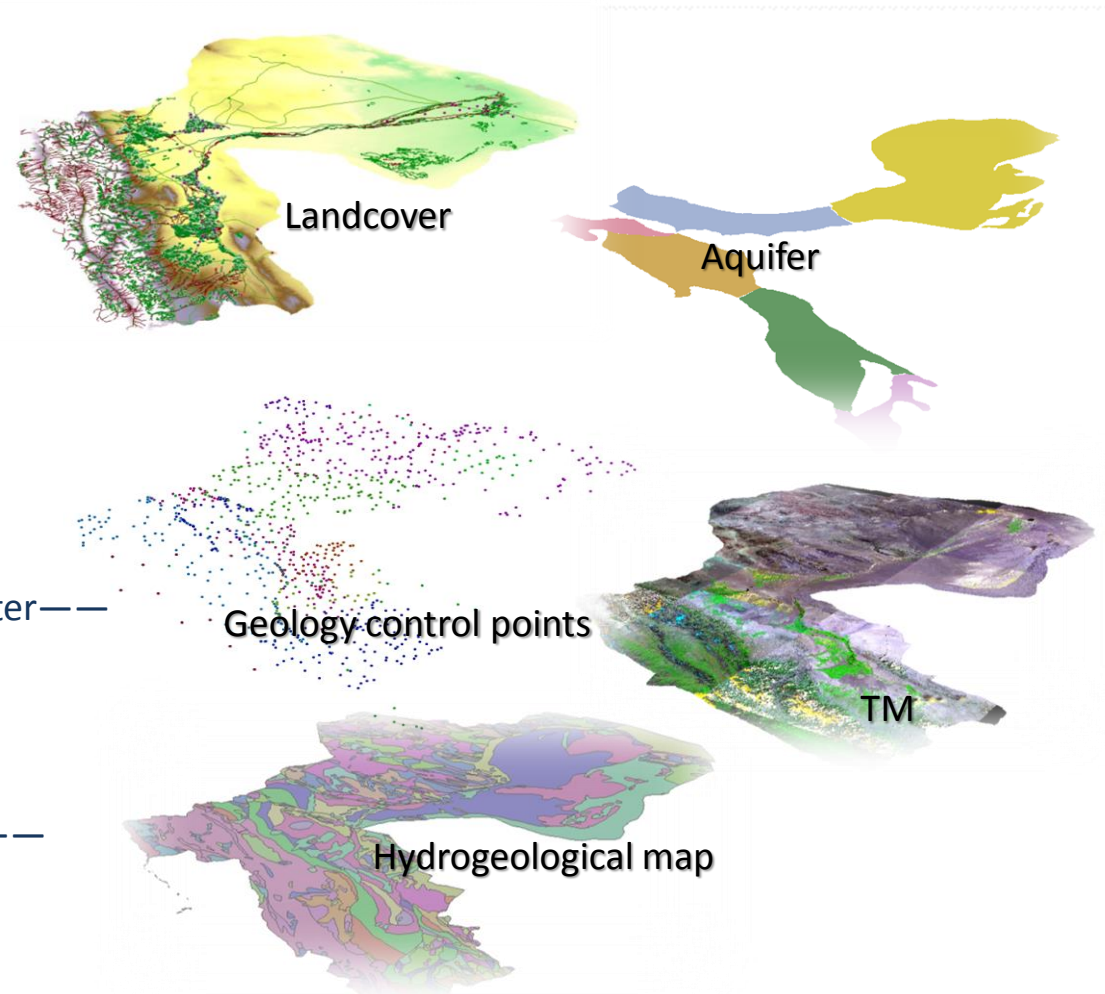
Experimental data

Field Survey, test——

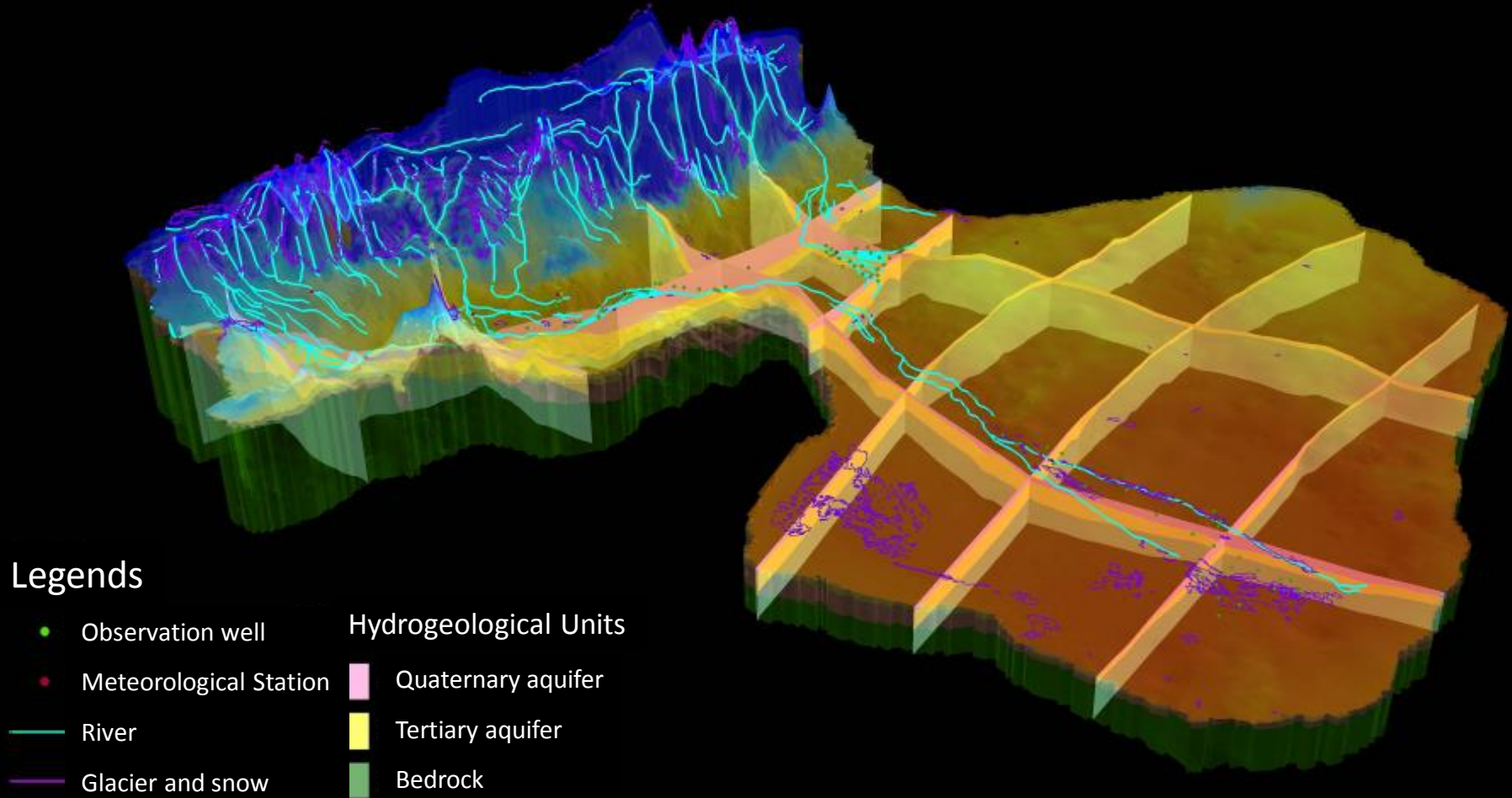


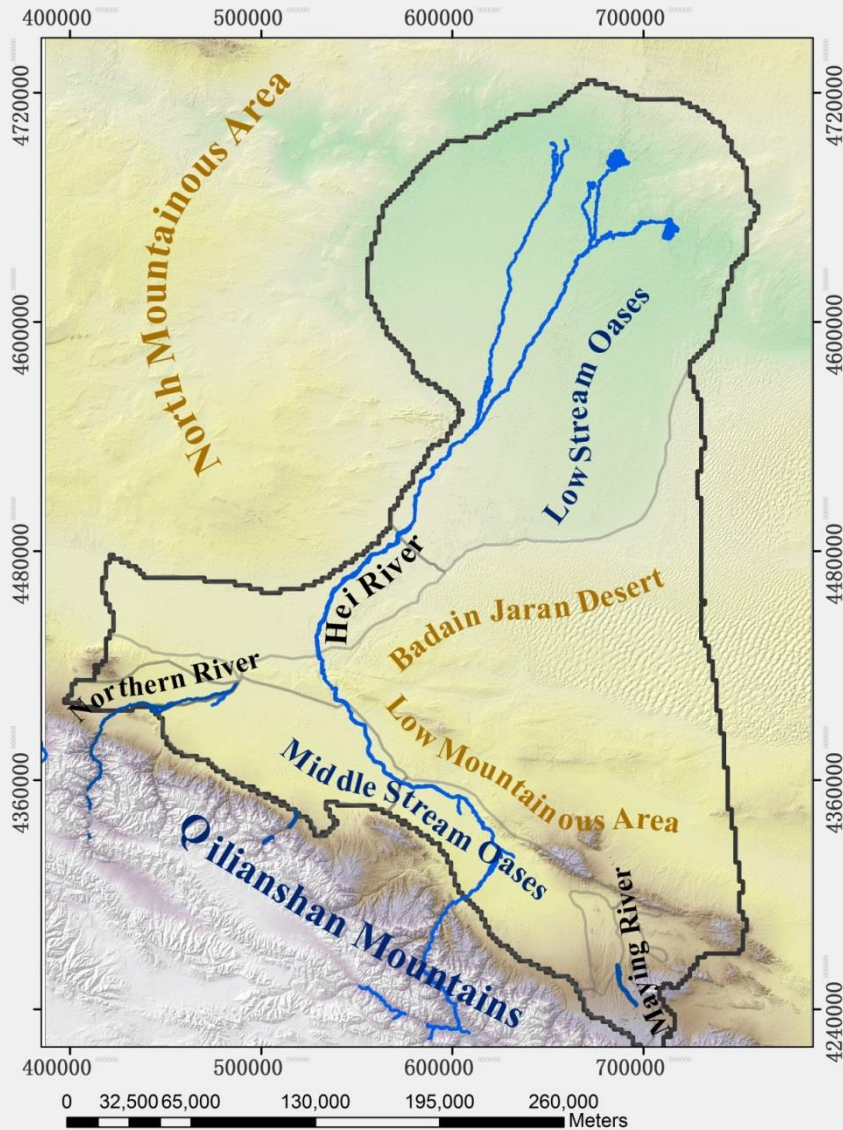
Model data

Radiation, Land assimilation, SWAT——



3D View of Subsurface with Cross Sections

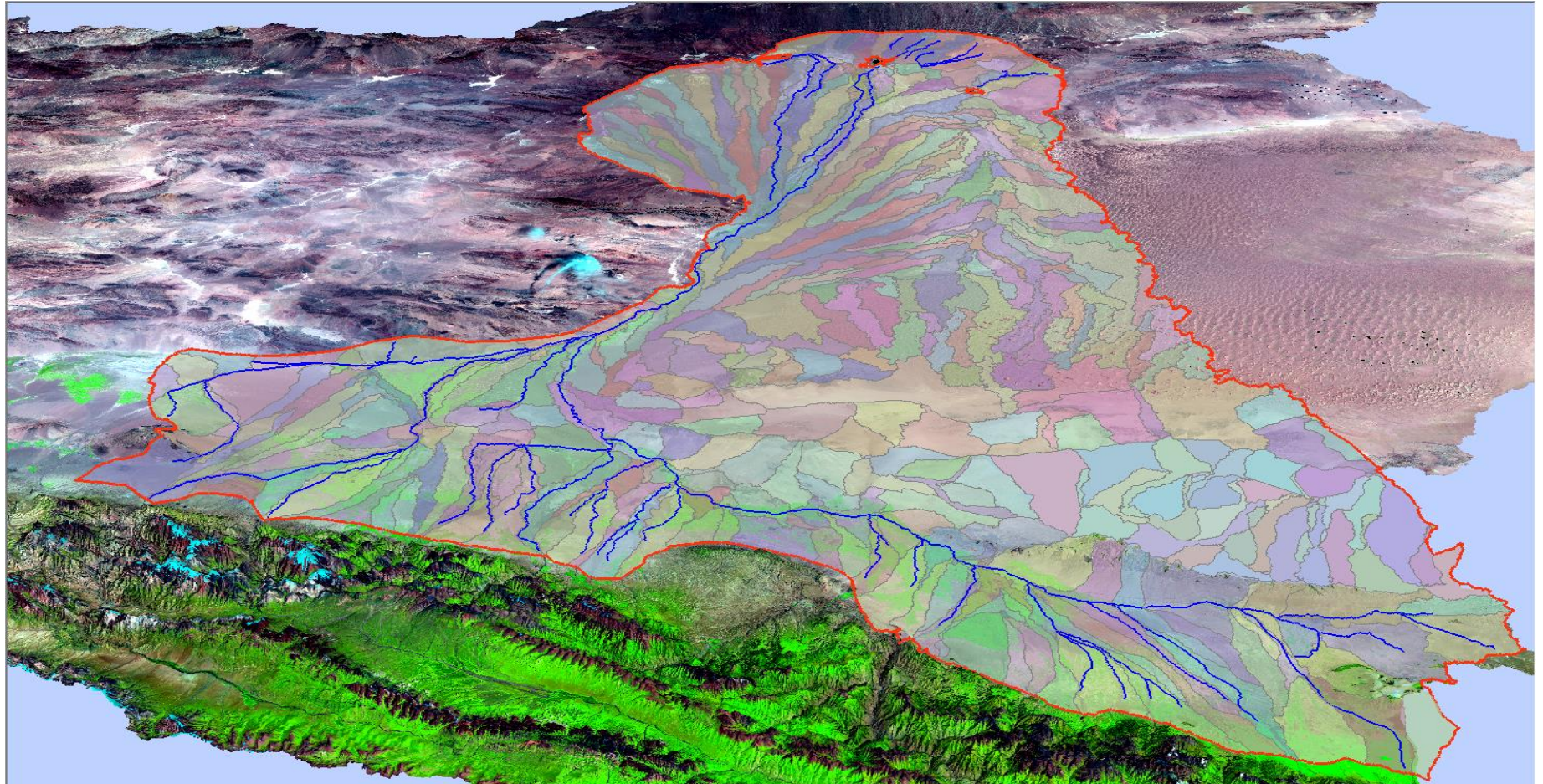




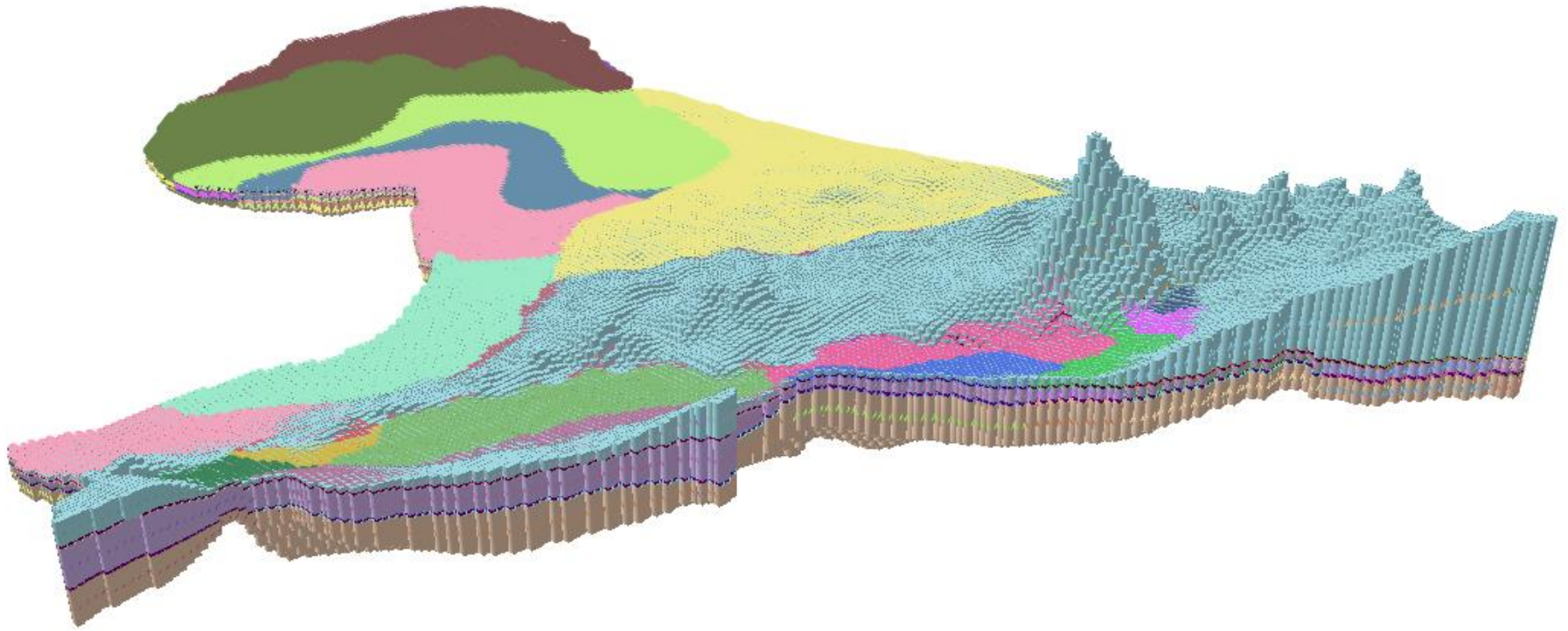
**Middle & Lower
Heihe River Basin
Model Domain:
~100,000 km²**

**Grid spacing:
1 km by 1 km
Rows: 548
Columns: 404
Layers: 5**

Surface Water Model for Middle/Lower Heihe

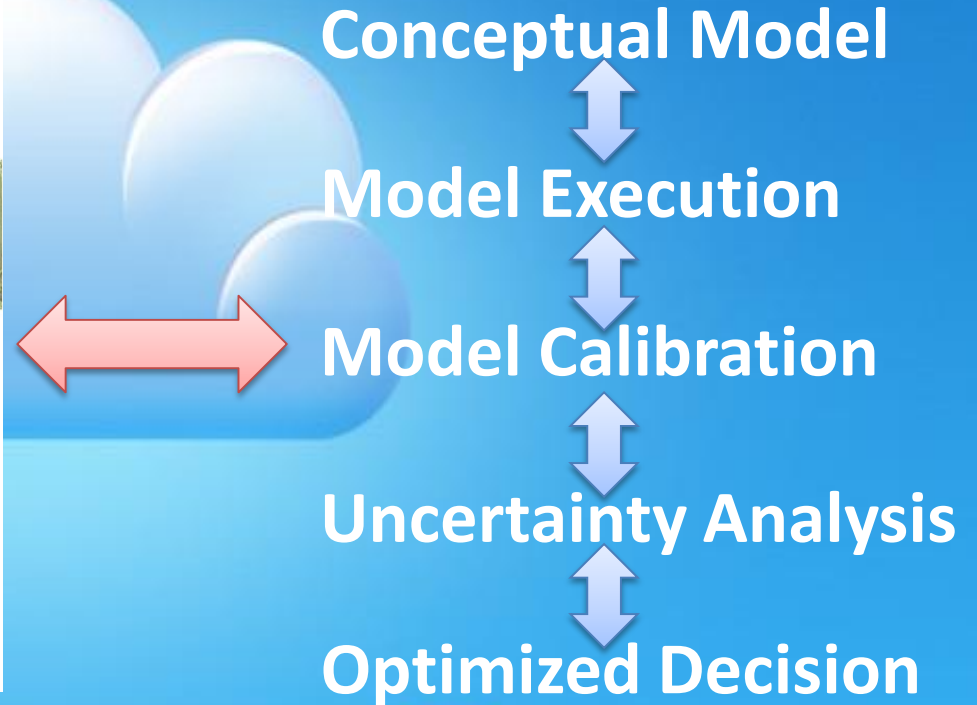
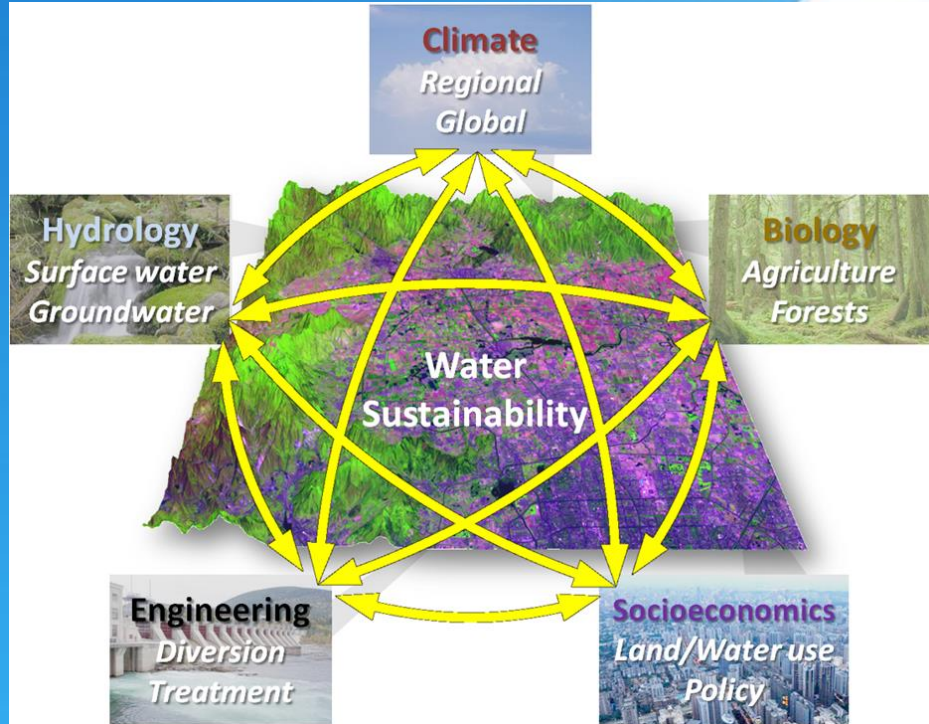


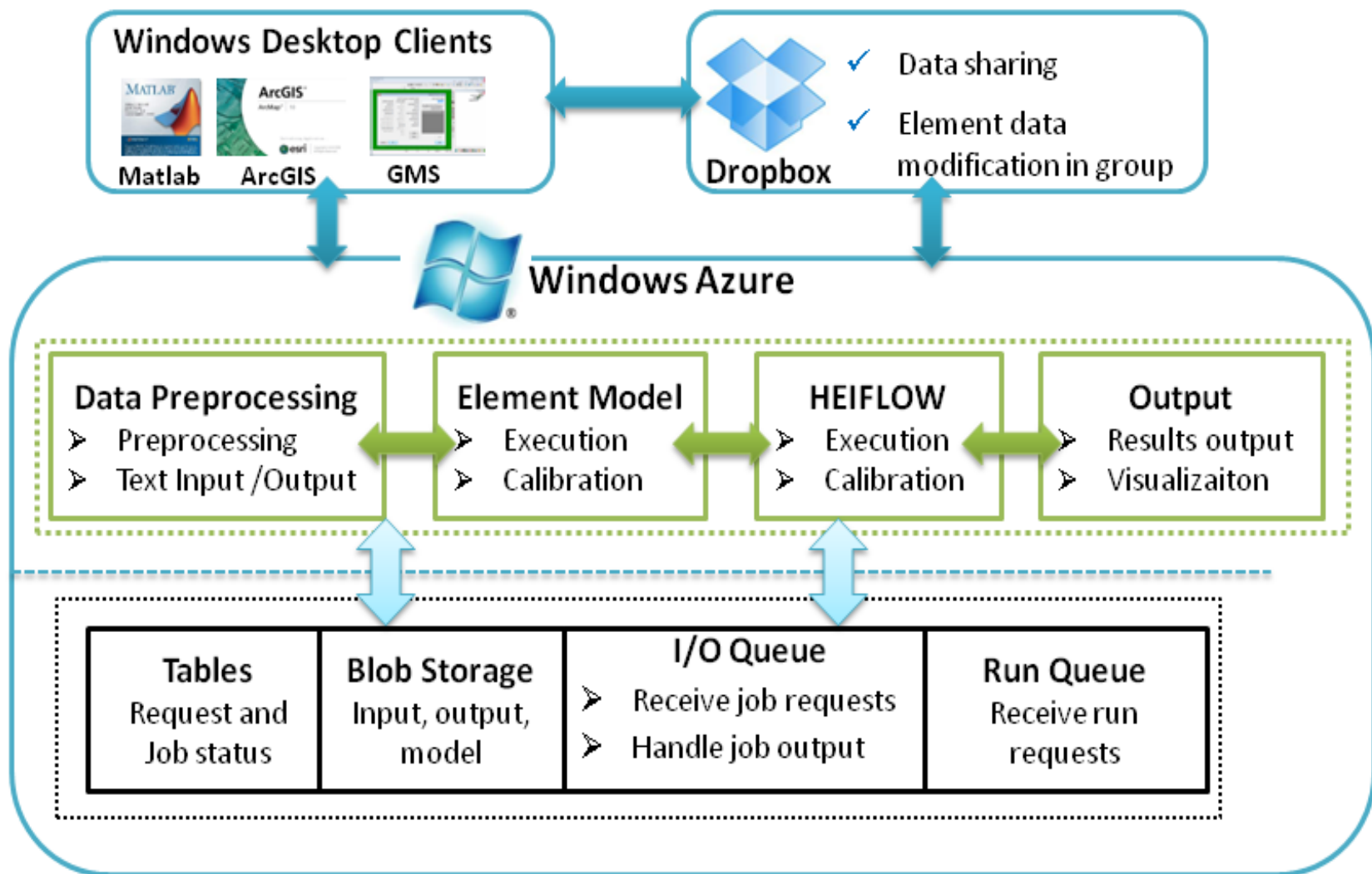
Discretization and Parameter Zonation for Subsurface Model



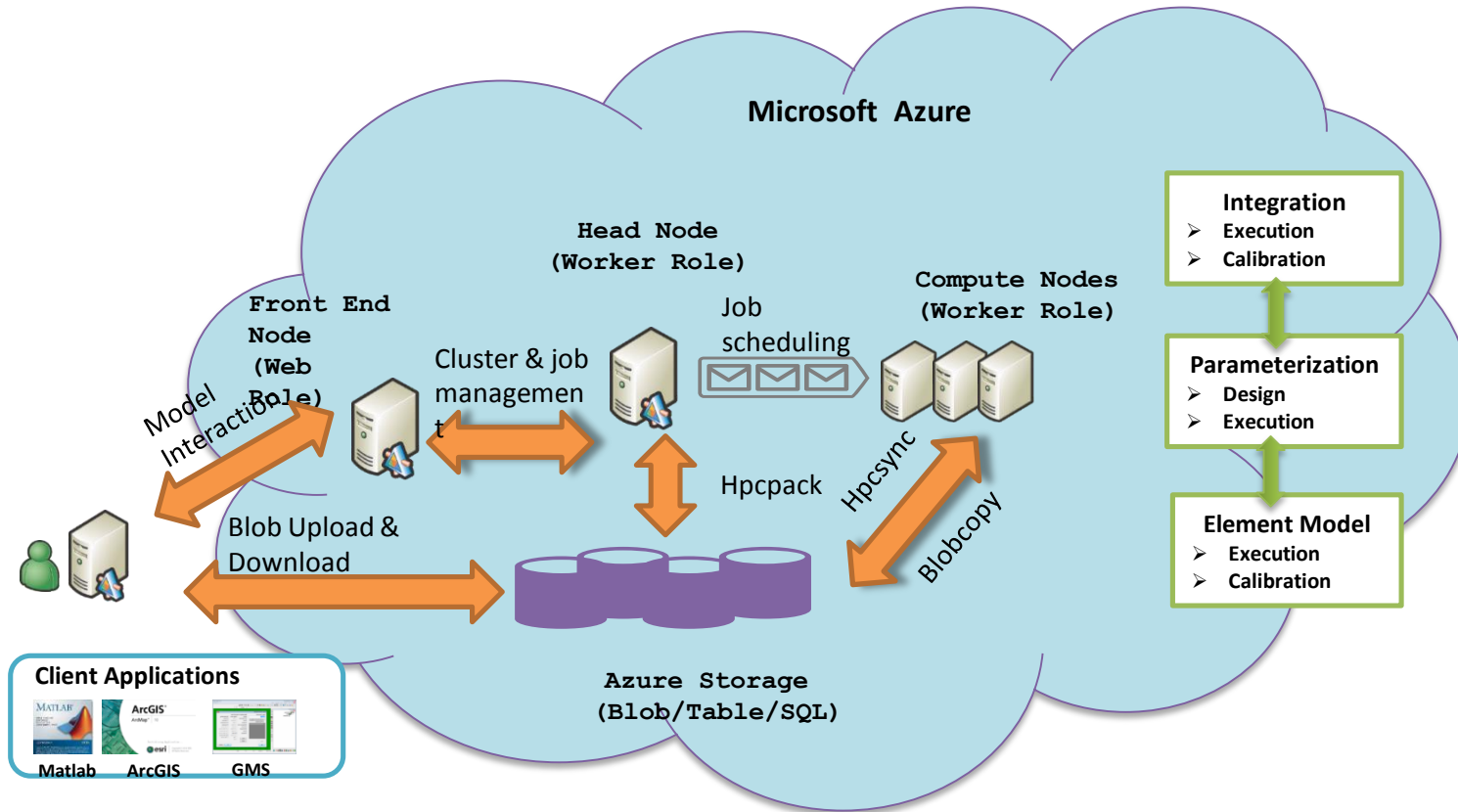
Layer 1	2	3	4	5
Unconfined aquifer	Aquitard 1	Shallow confined	Aquitard 2	Deep confined

HEIFLOW on the Cloud

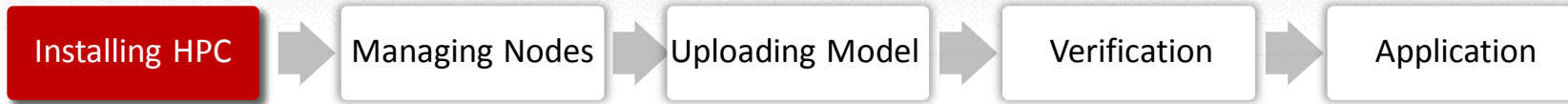




HPC on Microsoft Azure for ecohydrological modeling



Deployment of HPC on Azure for ecohydrological modeling



The screenshot displays the Microsoft HPC Pack 2012 R2 Cluster Manager interface. The window title is "Cluster PKUHPC - HPC Pack 2012 R2 Cluster Manager". The interface is divided into a left sidebar and a main content area.

Configuration

- Deployment To-do List
- Network
- Node Templates
- Images
- Job Templates
- Users
- Services
- iSCSI Deployment
- Resource Pools

Deployment To-do List

Required deployment tasks (Complete)

- Configure your network**
Choose one of five network topologies for your cluster.
- Provide installation credentials**
Specify the user name and password to use for system configuration, and when adding nodes to your cluster.
- Configure the naming of new nodes**
Specify the naming convention to use when generating names automatically for new nodes.
- Create a node template**
Create a template that defines the steps to follow when adding nodes to your cluster.

Optional deployment tasks

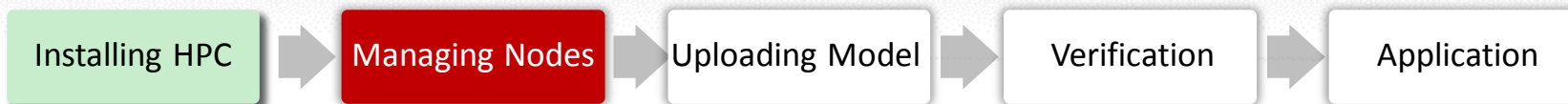
- Validate your environment before deploying nodes**
Run a set of short diagnostic tests to find common problems that can affect node deployment.
[Go online to get the latest set of tests](#)
- Add an operating system image**
Create a new image or load an existing image to use with your node templates when deploying nodes.
- Add or remove users**
Add or remove users or administrators for your cluster.
- Add nodes to your cluster**
Deploy nodes, import a node XML file, or add preconfigured nodes to your cluster.
- Manage drivers**
Add device drivers to the operating system images.
- Configure job scheduler policies and settings**
Customize policies, error handling, filters, and e-mail notifications for your cluster.
- Set Windows Azure storage connection string**
Set a Windows Azure Storage connection string for Service-Oriented Architecture (SOA) jobs that run on Windows Azure nodes.

Help

- [Overview of Microsoft HPC Pack](#)
- [Online Resources for Microsoft HPC Pack](#)

2 items selected

Deployment of HPC on Azure for ecohydrological modeling



Cluster PKUHPC - HPC Pack 2012 R2 Cluster Manager

File View Tasks Options Go Help

Back Forward Navigation Pane Actions Filter: By Group By Health Search: Node Name Clear All

Node Management

- Nodes (4)
 - By Node Health
 - OK (4)
 - Warning (0)
 - Error (0)
 - Transitional (0)
 - Unapproved (0)
 - By Node State
 - Online (4)
 - Offline (0)
 - Draining (0)
 - Provisioning (0)
 - Rejected (0)
 - Not-Deployed (0)
 - By Group
 - HeadNodes
 - ComputeNodes
 - WCFBrokerNodes
 - WorkstationNodes
- Configuration
- Node Management**
- Job Management
- Diagnostics
- Charts and Reports

Nodes (4)

Node Name	Node State	Node Health	Node Template	Location
AzureCN-0007	Online	OK	Default AzureNode Template	
AzureCN-0008	Online	OK	Default AzureNode Template	
AzureCN-0009	Online	OK	Default AzureNode Template	
PKUHPC	Online	OK	HeadNode Template	

Actions

Pivot To

- Jobs for the Selected Nodes
- Failed Diagnostics for the Nodes
- Operations for the Nodes

Node Actions

- Bring Online
- Take Offline
- Start
- Stop
- Reboot
- Shut Down
- Run Command ...
- Add Node ...**
- Reimage
- Maintain
- Change Role ...
- Delete
- Reject
- Assign Node Template ...
- Edit Properties...
- Export Node XML File ...
- Run Diagnostics ...
- View Performance Charts
- Remote Desktop

Data updated: 6/27/2014 8:17:53 AM

Deployment of HPC on Azure for ecohydrological modeling



Administrator: Windows PowerShell ISE

```
File Edit View Tools Debug Add-ons Help
```

```
PackingDeploy.ps1 X
1 if(!$args[0])
2 {
3     Write-Host "Input folder is null!"
4     return
5 }
6 if(!(Test-Path $args[0]))
7 {
8     Write-Host "$args[0] is invalid!"
9     return.
10 }
11 $sNameFolder=$args[0]
12 Write-Host "$sNameFolder is valid!"
13 $sNameTemplate="Default AzureNode Template"#$args[1]
14 $sNameZip="$sNameFolder.zip"
15 hpcpack create $sNameZip $sNameFolder
16 hpcpack upload $sNameZip /nodetemplate:$sNameTemplate /relativePath:$sNameFolder
17
```

PS C:\App> .\PackingDeploy.ps1 pest
pest is valid!

HpcPack completed.

IMPORTANT:
If the package contains a SOA service, the name of the package must be [service name].zip or [service name]_[service version].zip depending on whether the service is versioned.
If the package contains an XLL, the name of the package must be [XLL name].zip.

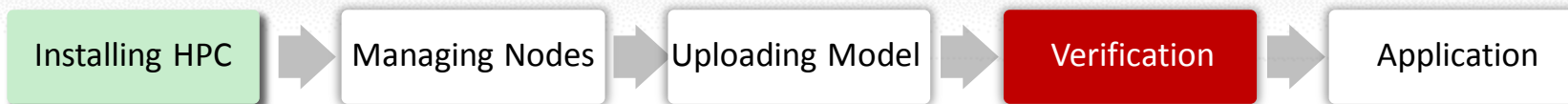
Connecting to head node: PKUHPC
Querying for Windows Azure Worker/VM Role node template "Default AzureNode Template"
Windows Azure Worker/VM Role node template found. Retrieving Azure storage account name and key.
Found storage account: pkucwr and key: *****
A relativepath is used while an OPC package is to be uploaded. Checking whether the relativepath already exist.
Transferring File pest.zipPackage pest.zip successfully uploaded to Windows Azure storage account.
File transfer(s) completed.

PS C:\App>

Completed

Ln 19 Col 12 100%

Deployment of HPC on Azure for ecohydrological modeling



```
D:\Windows\system32\cmd.exe - .\runslave.bat
Waiting for PEST....
PEST is alive: waiting for command to run model ....
(Press <Ct1-C> if you wish to stop PSLAUE execution.)
Running model ....

Model completed: waiting for command to run model again.
(Press <Ct1-C> if you wish to stop PSLAUE execution.)
Running model ....

Model completed: waiting for command to run model again.
(Press <Ct1-C> if you wish to stop PSLAUE execution.)
Running model ....

Model completed: waiting for command to run model again.
(Press <Ct1-C> if you wish to stop PSLAUE execution.)
Running model ....

Model completed: waiting for command to run model again.
(Press <Ct1-C> if you wish to stop PSLAUE execution.)
Running model ....
```

Applications | run | 4/25/2014 8:08 AM | Windows Batch File

```
D:\Windows\system32\cmd.exe - .\runslave.bat
Running model ....

Model completed: waiting for command to run model again.
(Press <Ct1-C> if you wish to stop PSLAUE execution.)
Running model ....

Model completed: waiting for command to run model again.
(Press <Ct1-C> if you wish to stop PSLAUE execution.)
Running model ....

Model completed: waiting for command to run model again.
(Press <Ct1-C> if you wish to stop PSLAUE execution.)
Running model ....

Model completed: waiting for command to run model again.
(Press <Ct1-C> if you wish to stop PSLAUE execution.)
Running model ....

Model completed: waiting for command to run model again.
(Press <Ct1-C> if you wish to stop PSLAUE execution.)
Running model ....
```

```
Administrator: Windows PowerShell - run.bat
Starting phi for this iteration: 277.79
Calculating Jacobian matrix: running model 10 times ....
10 runs completed.

Lambda = 20.000 ----->
running model .....
Phi = 238.69 ( 0.859 of starting phi)

Lambda = 10.000 ----->
running model .....
Phi = 239.88 ( 0.864 of starting phi)

Lambda = 40.000 ----->
running model .....
Phi = 249.55 ( 0.898 of starting phi)

No more lambdas: phi rising
Lowest phi this iteration: 238.69
Maximum factor change: 3.000 ["ro2"]
Maximum relative change: 2.000 ["ro2"]

OPTIMISATION ITERATION NO. : 4
Model calls so far : 30
Starting phi for this iteration: 238.69
Calculating Jacobian matrix: running model 10 times ....
10 runs completed.

Lambda = 20.000 ----->
running model .....
Phi = 78.025 ( 0.327 of starting phi)

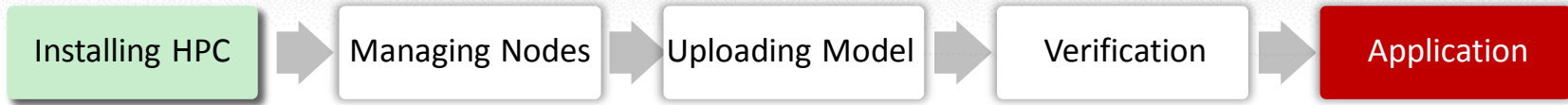
Lambda = 10.000 ----->
running model .....
Phi = 63.254 ( 0.265 of starting phi)

No more lambdas: phi is less than 0.3000 of starting phi
Lowest phi this iteration: 63.254
Maximum factor change: 3.000 ["ro2"]
Maximum relative change: 2.000 ["ro2"]

OPTIMISATION ITERATION NO. : 5
Model calls so far : 42
Starting phi for this iteration: 63.254
Calculating Jacobian matrix: running model 10 times ....
8 runs completed.
```

UKB

Deployment of HPC on Azure for ecohydrological modeling



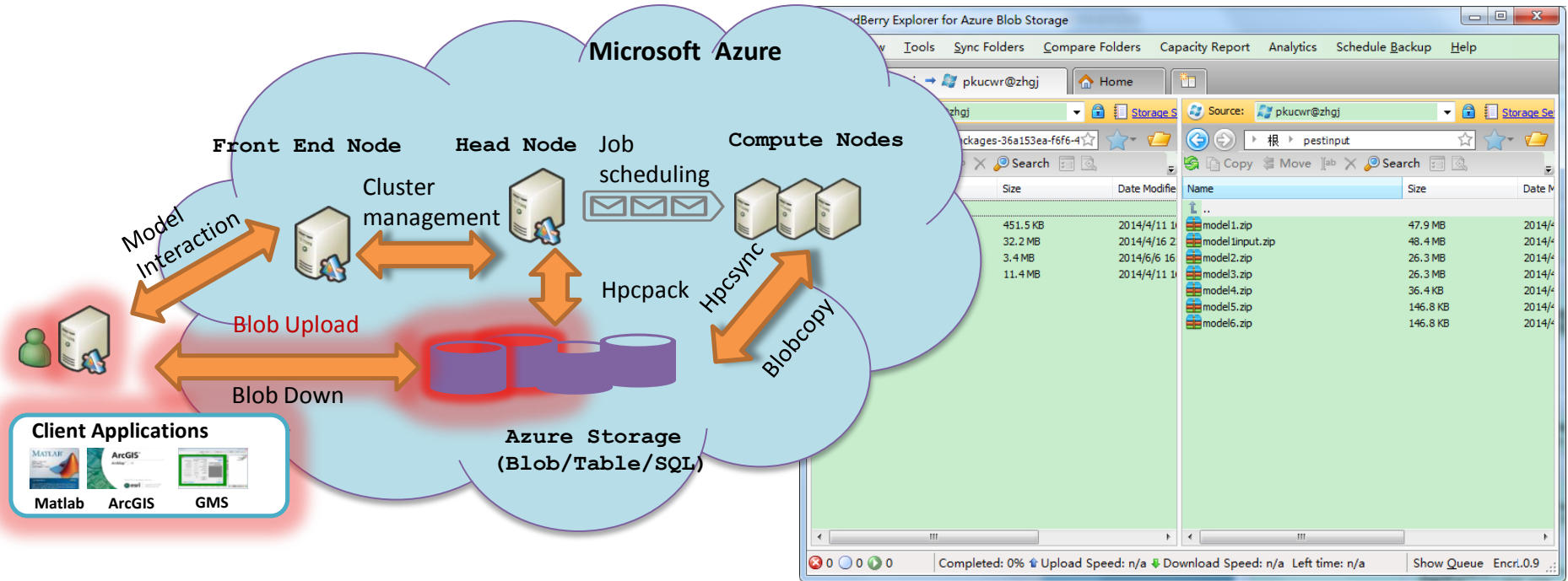
The screenshot shows a 'New Job' configuration window with a 'Parametric Sweep Task' dialog box open. The dialog box contains the following fields and options:

- Task name:** Eco-hydrologic model Task
- Step 1: Select the start and end values for the sweep task:**
 - Start value: 1
 - End value: 80
- Step 2: Select the amount to increment the value at each step of the sweep task:**
 - Increment value: 1
- Step 3: Enter the command line, working directory, and file locations for the sweep task.**
 - Command line: `runmodel .cmd *`
 - Working directory: `%CCP_PACKAGE_ROOT%\appst\`
 - Standard input: (empty)
 - Standard output: (empty)
 - Standard error: (empty)
- Step 4: Preview your sweep task:**

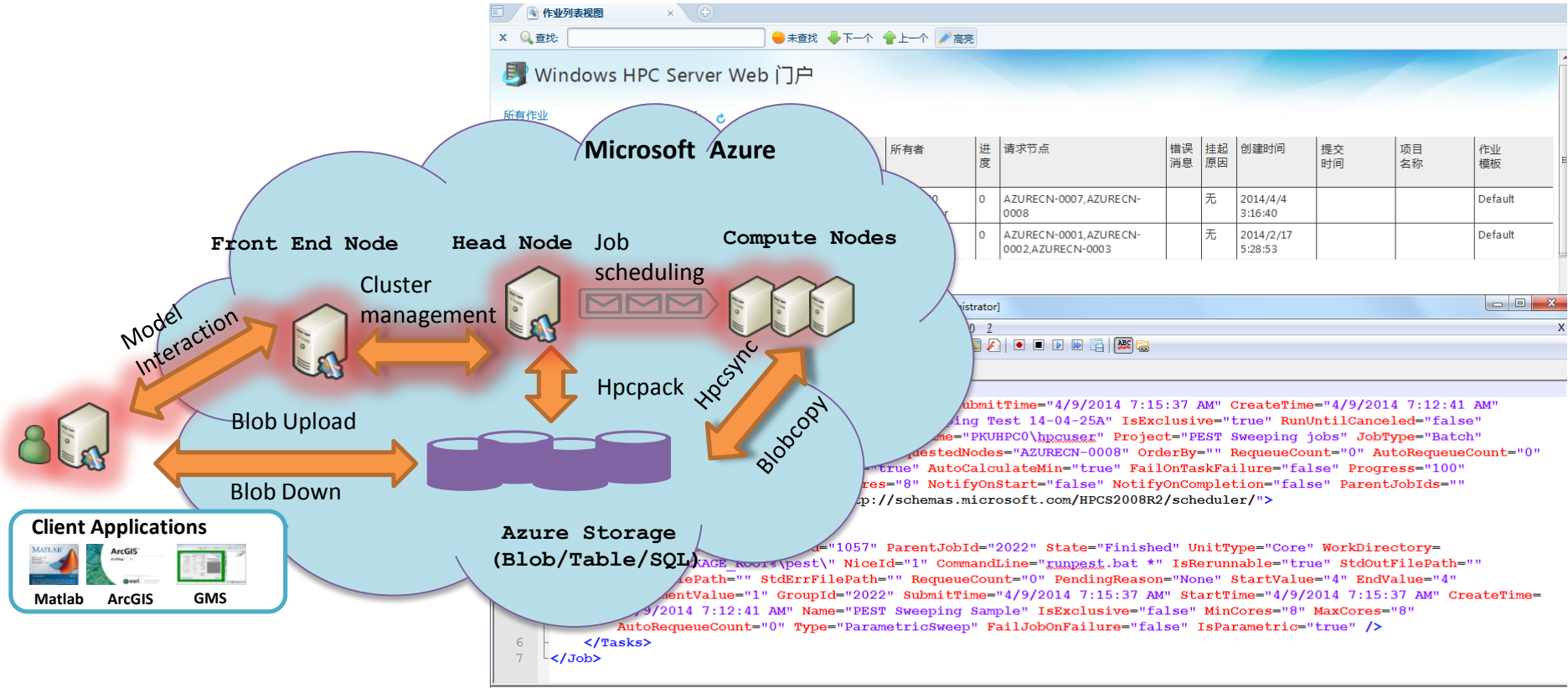
Command Line	Standard Output
runmodel.cmd 1	
runmodel.cmd 2	
...	
runmodel.cmd 100	

Buttons: OK, Cancel

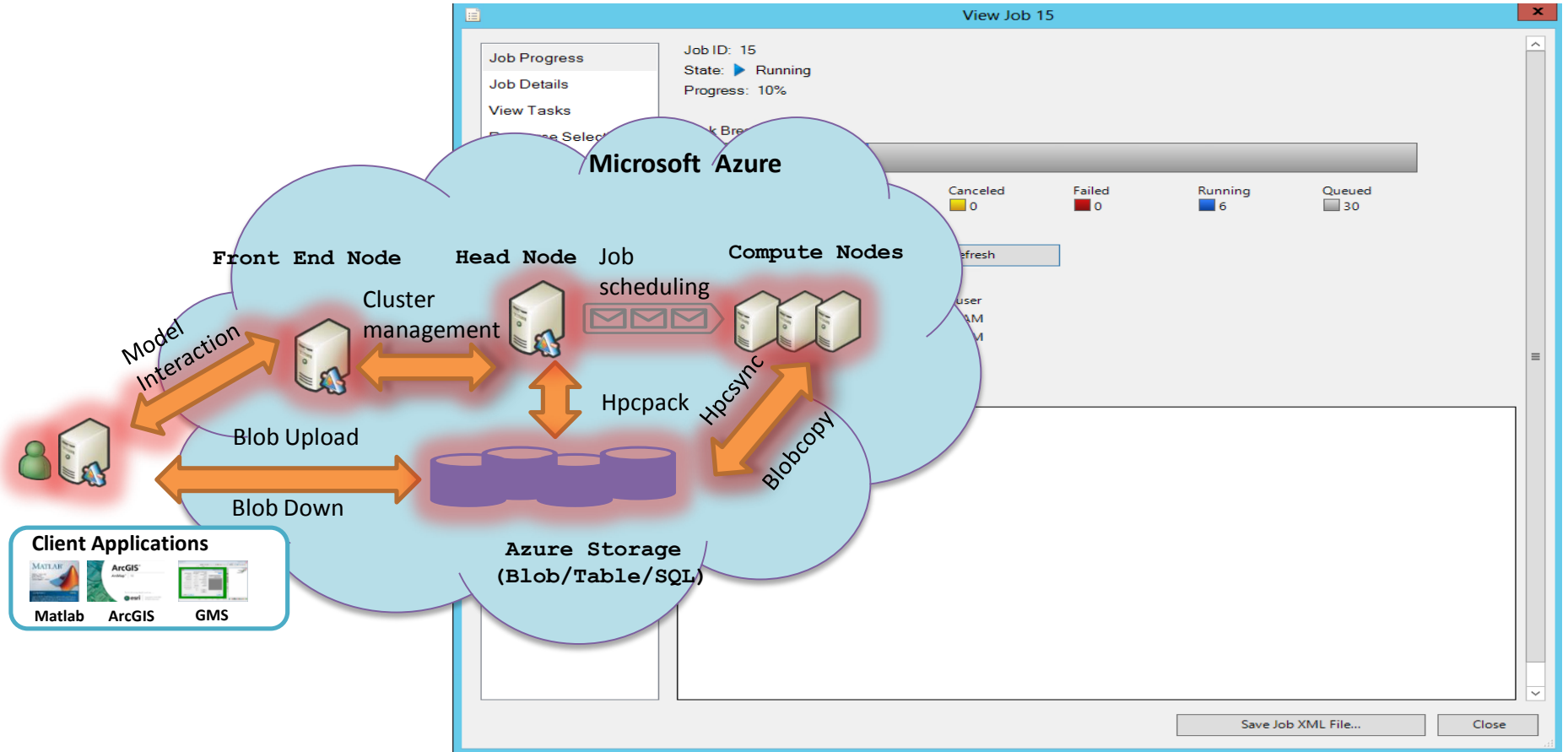
Ecohydrological model execution in HPC on Azure



Ecohydrological model execution in HPC on Azure



Ecohydrological model execution in HPC on Azure



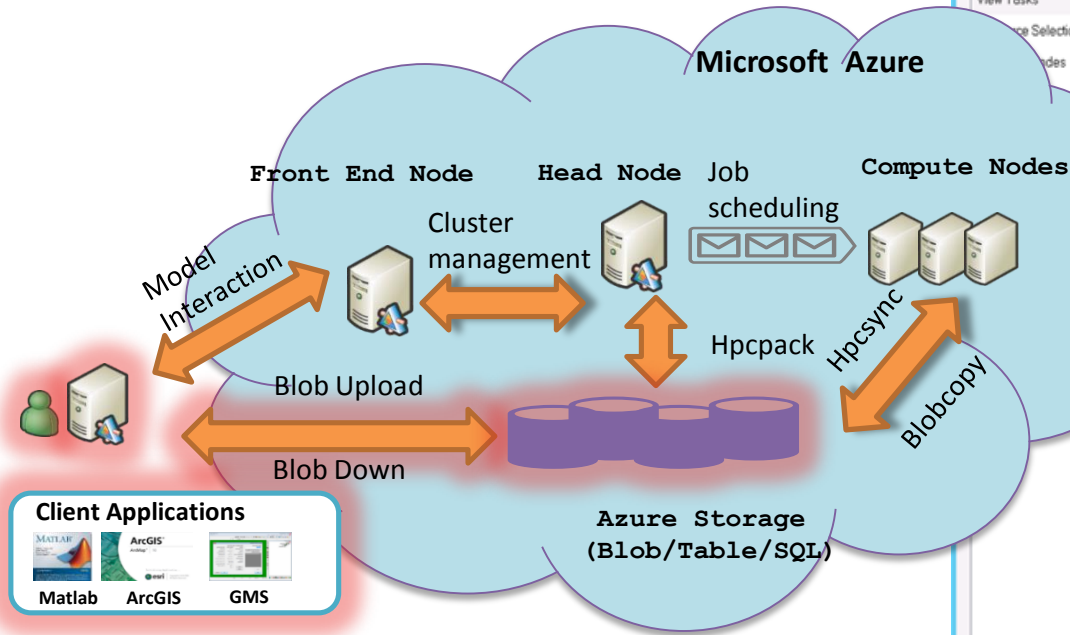
Ecohydrological model execution in HPC on Azure

Data Preparing & Uploading

Job Submitting

Status
Monitoring

Result Getting



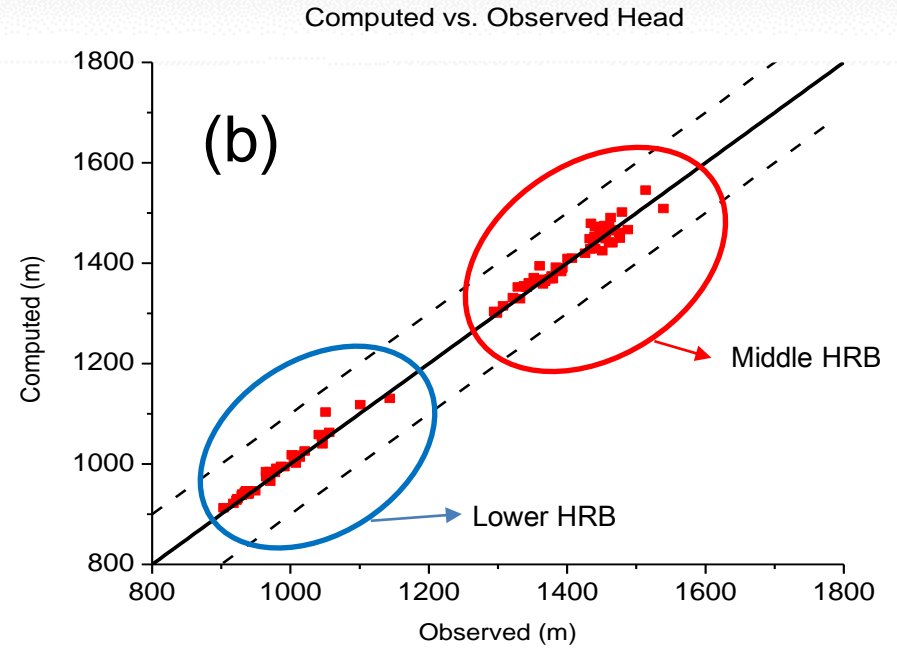
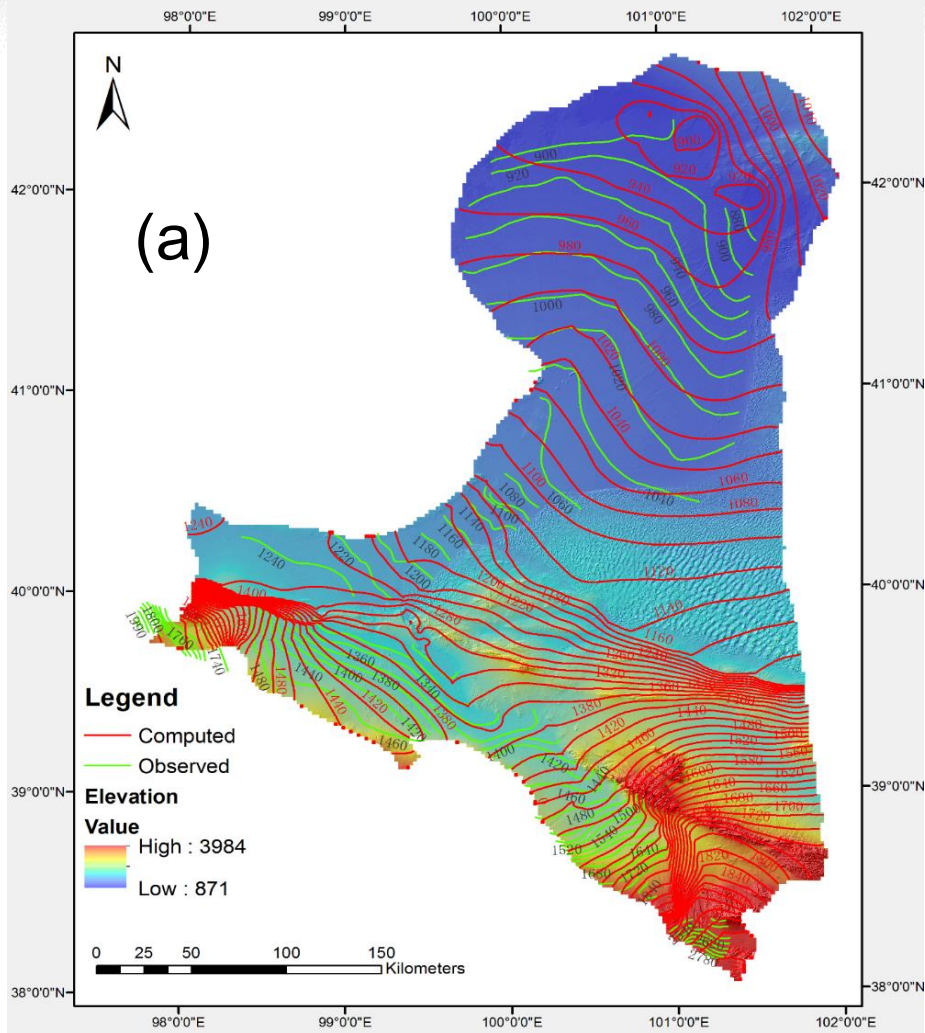
The screenshot shows the "View Job 6" interface. It includes a "Job Progress" section with a table of job details:

Task ID	State	Type	Task Name	Command Line
1.1 - 1.1	Finished	Parametric Sweep	Pest Sweep Task	run.bat *
1.1	Finished	Parametric Sweep	Pest Sweep Task	run.bat 1

Below the table is a "Refresh Task List" button. The interface also shows a "Details" section with a "Valid Exit Codes" button and a "Copy Output to Clipboard" button. The "Messages" section contains the following text:

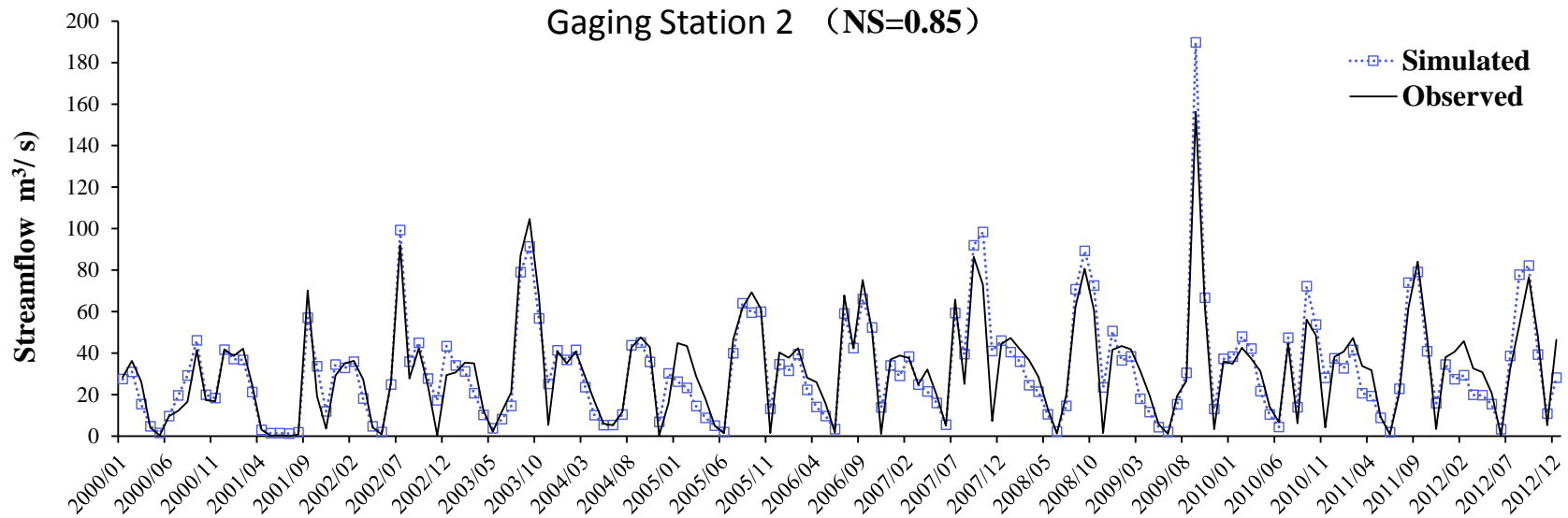
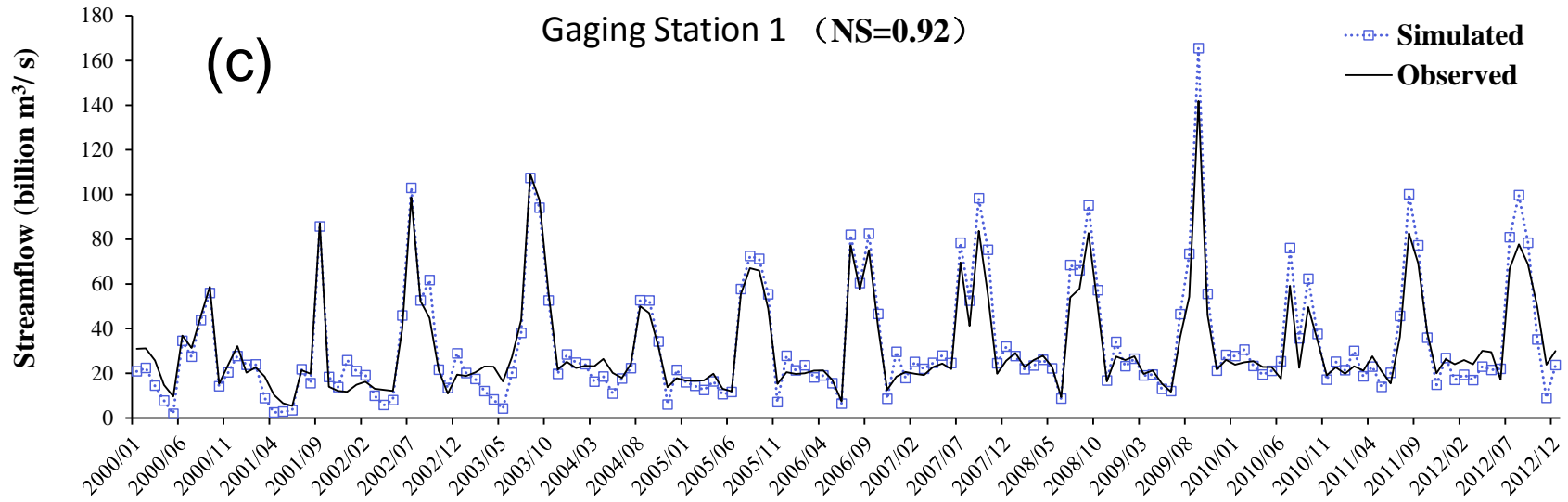
```
...ing model one last time with best parameters.....  
...nding run statistics .....
```

At the bottom, there are buttons for "Save Job XML File..." and "Close".

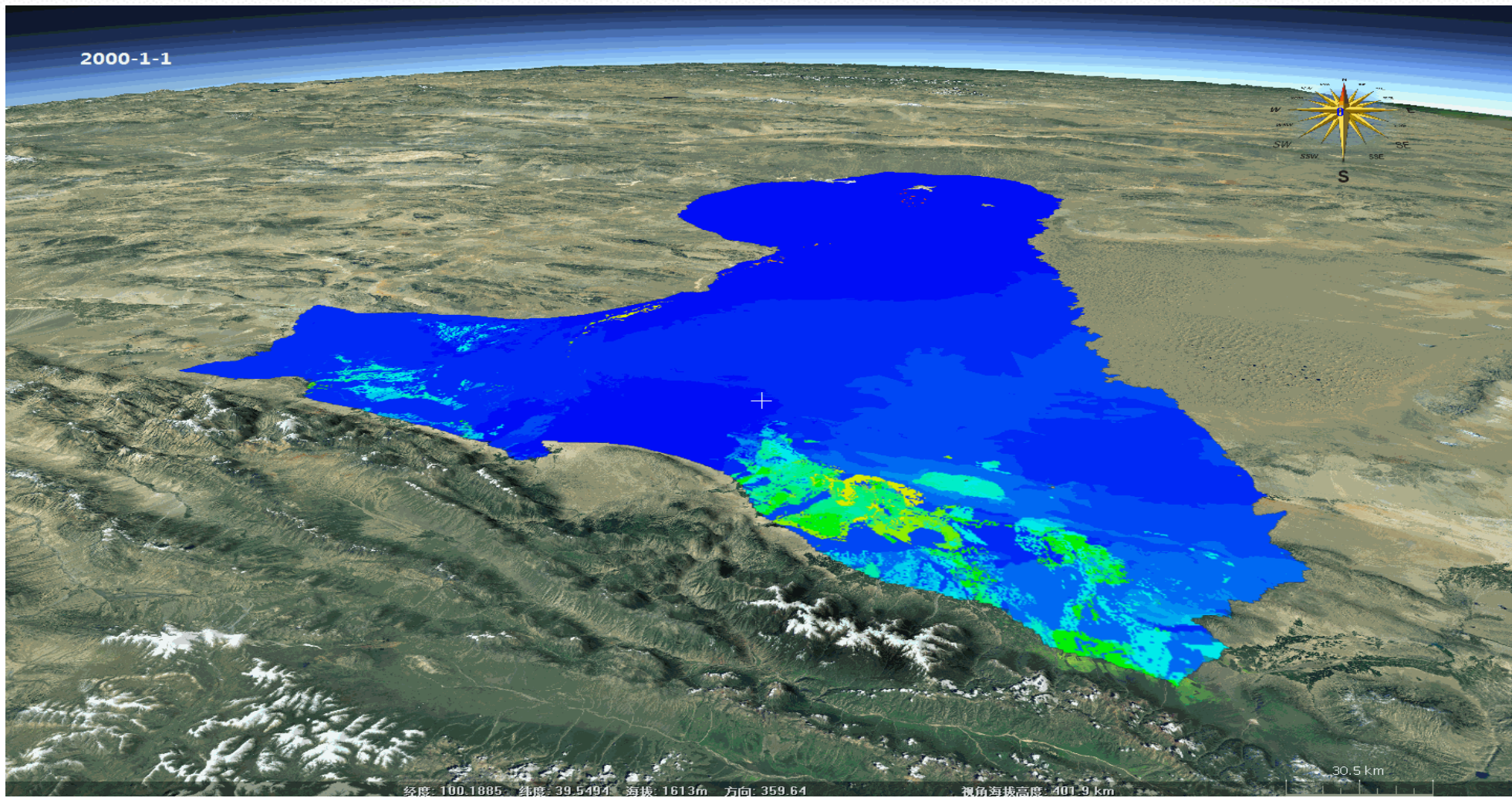


Model Calibration:

- Comparison between contour maps of computed and observed groundwater levels;
- Comparison between computed and observed heads at monitoring wells;
- Comparison of computed and observed streamflows and evapotranspiration



Simulated Evapotranspiration Dynamic Patterns (2000)



Element Model - Calibration

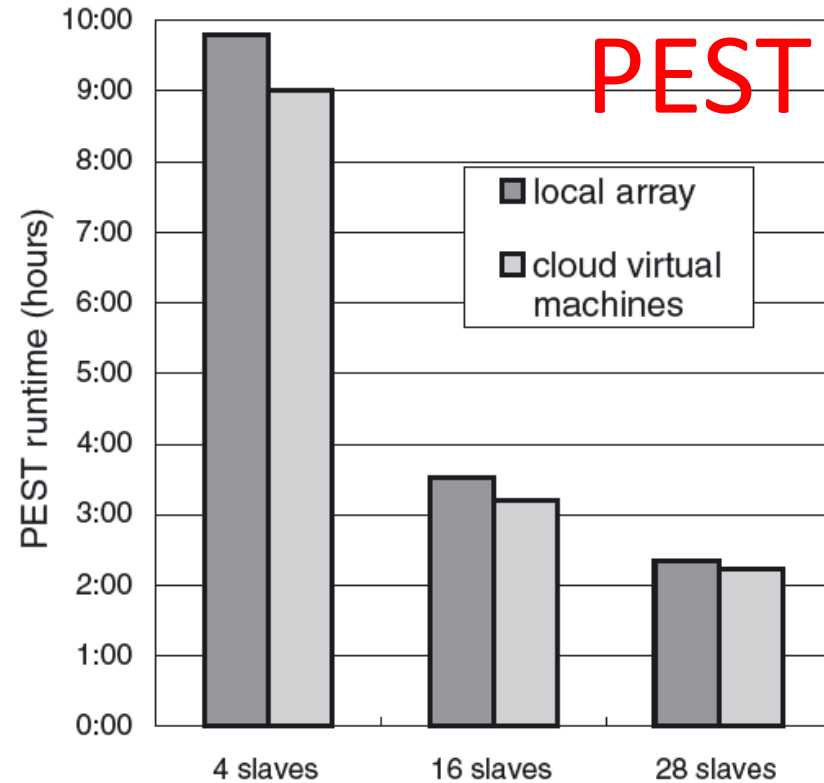
Computer	Time (s)
Q6700 Core 2 Quad (2.66 GHz)	85
GoGrid Cloud virtual machine	81
Xeon (3.0 GHz)	73
Q9650 Core 2 Quad (3.0 GHz)	71
i7 (3.33 GHz)	58

ground
water

Rapid Communication/

Using a Cloud to Replenish Parched Groundwater Modeling Efforts

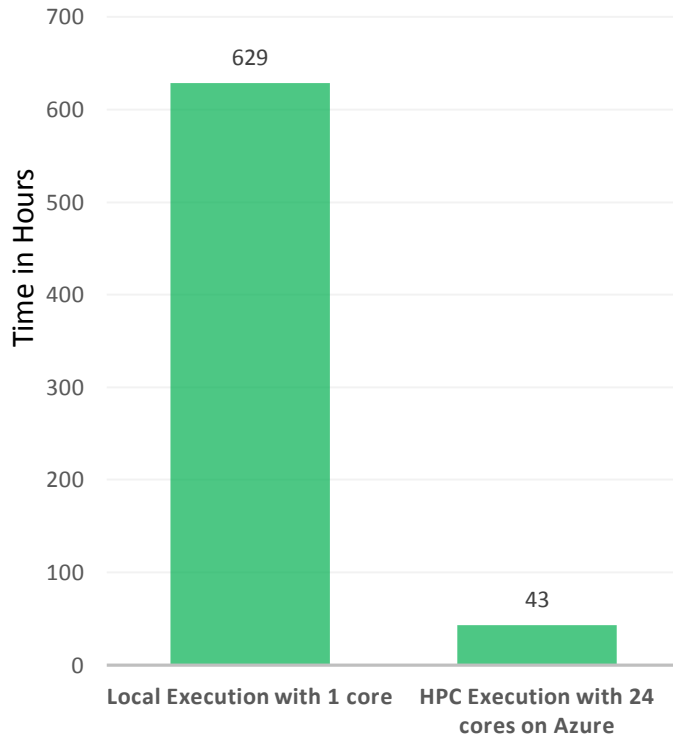
by Randall J. Hunt¹, Joseph Luchette², Willem A. Schreuder³, James O. Rumbaugh⁴, John Doherty^{5,6}, Matthew J. Tonkin⁷, and Douglas B. Rumbaugh⁴



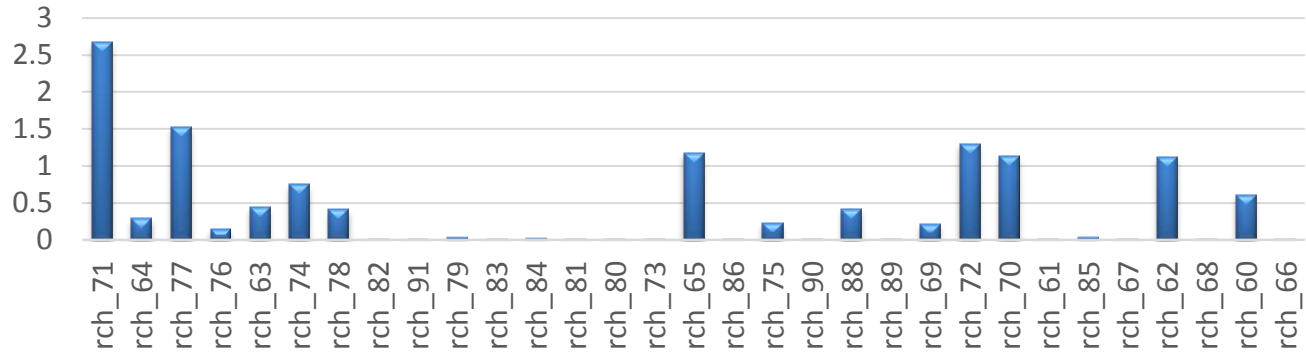
Comparison of parameter estimation runtimes obtained from a dedicated local desktop array and virtual machines run on the cloud

Results of sensitivity analysis on Cloud

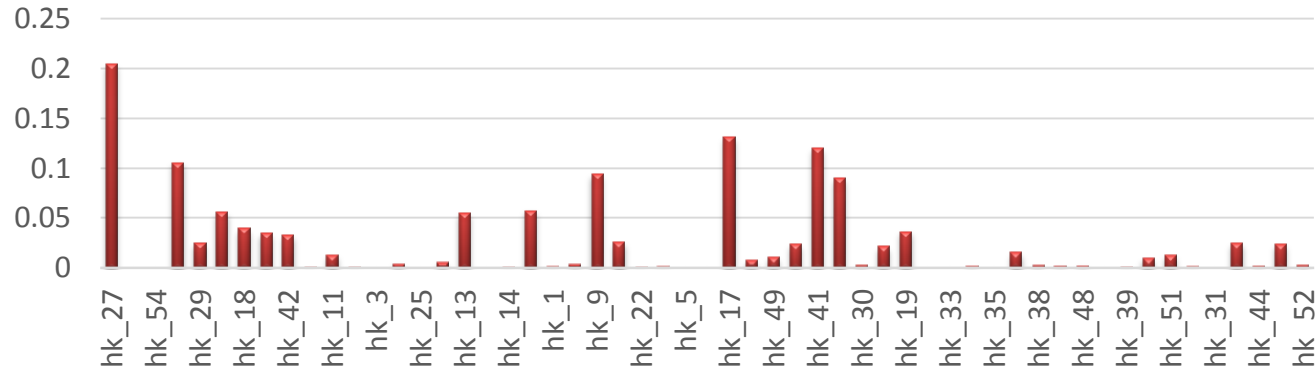
Computational performance of PEST for sensitivity analysis to HPC on Azure



Sensitivity_recharge

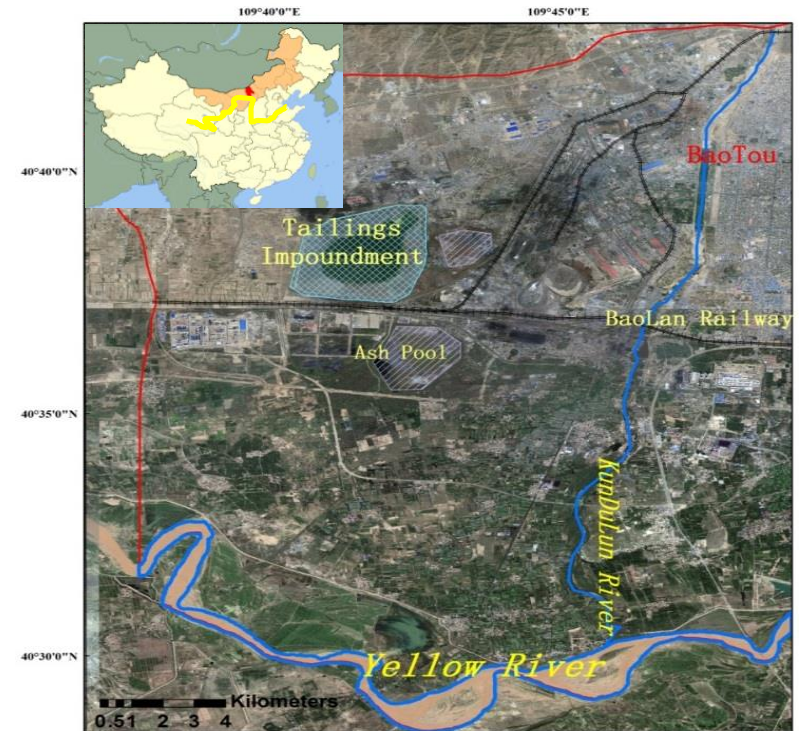


Sensitivity_conductivity

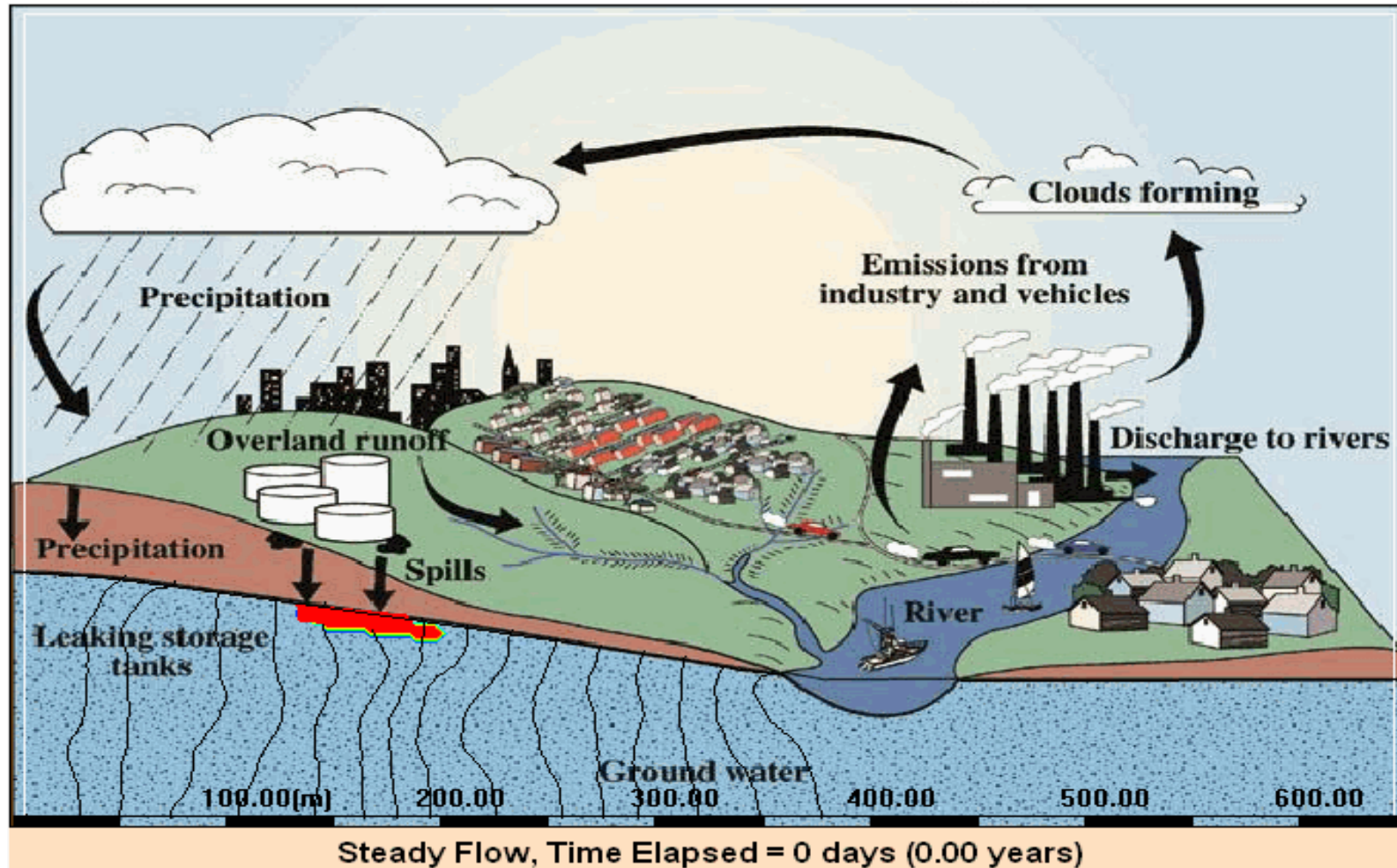


Case Study 2: Thorium Reactive Transport Modeling

- Baotou tailings pond, one of the largest tailings in China piled above ground surfaces.
- Most productive “secondary mines”, approximately 11 floors high!
- Of greatest concern is the potential for radioactive pollution of the Yellow River nearby which is the primary water source for 150 million people.



Modeling Contaminant Transport and Remediation



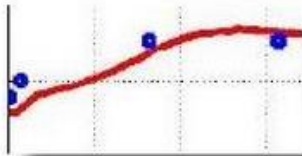


The Hydrogeology Group
The University of Alabama

Zheng 1990
Zheng and Wang 1999
Zheng 2010
Zheng et al. 2013

MT3DMS

A Modular 3-D Multi-Species Transport Model for Simulation of Advection, Dispersion and Chemical Reactions of Contaminants in Groundwater Systems

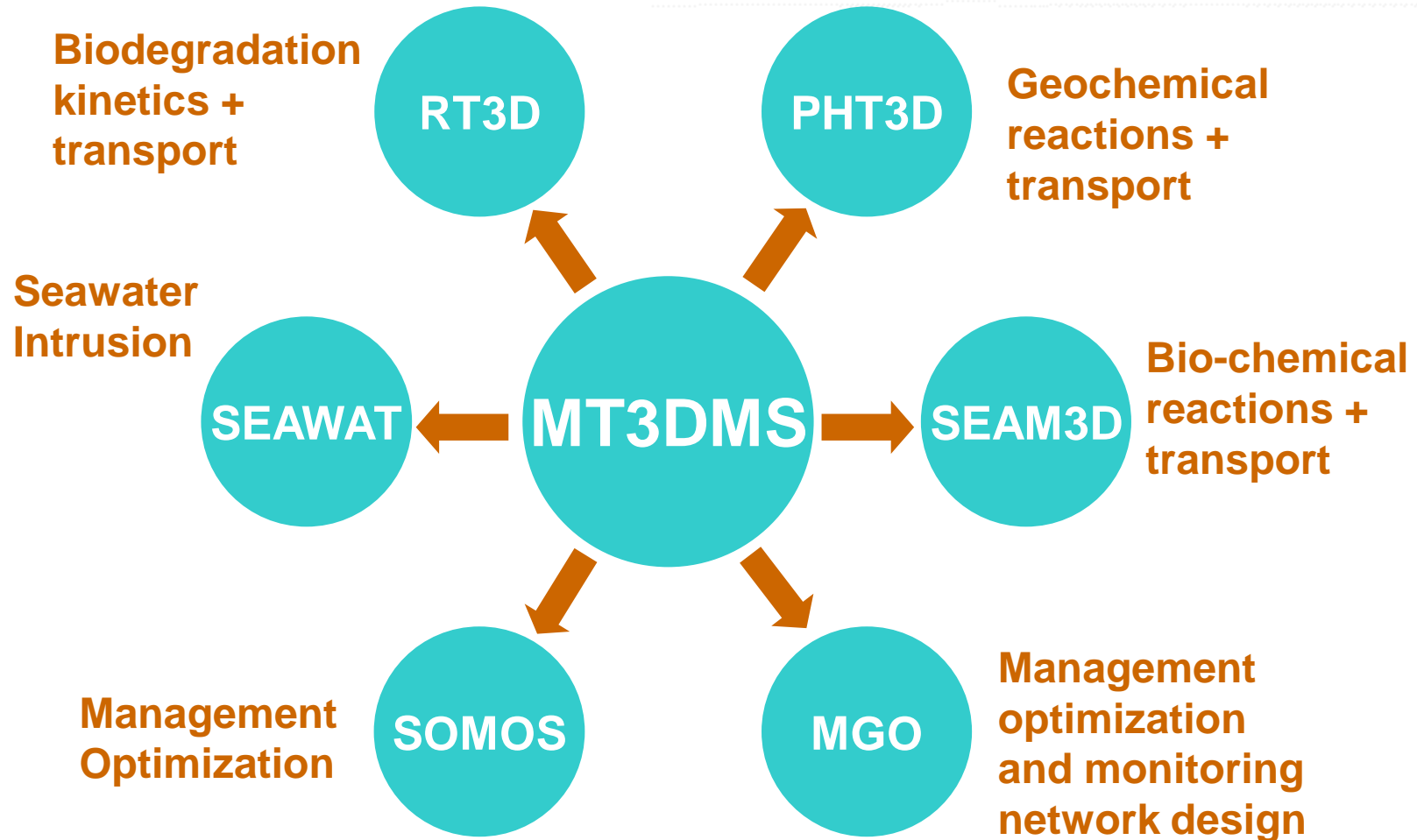


PHT3D

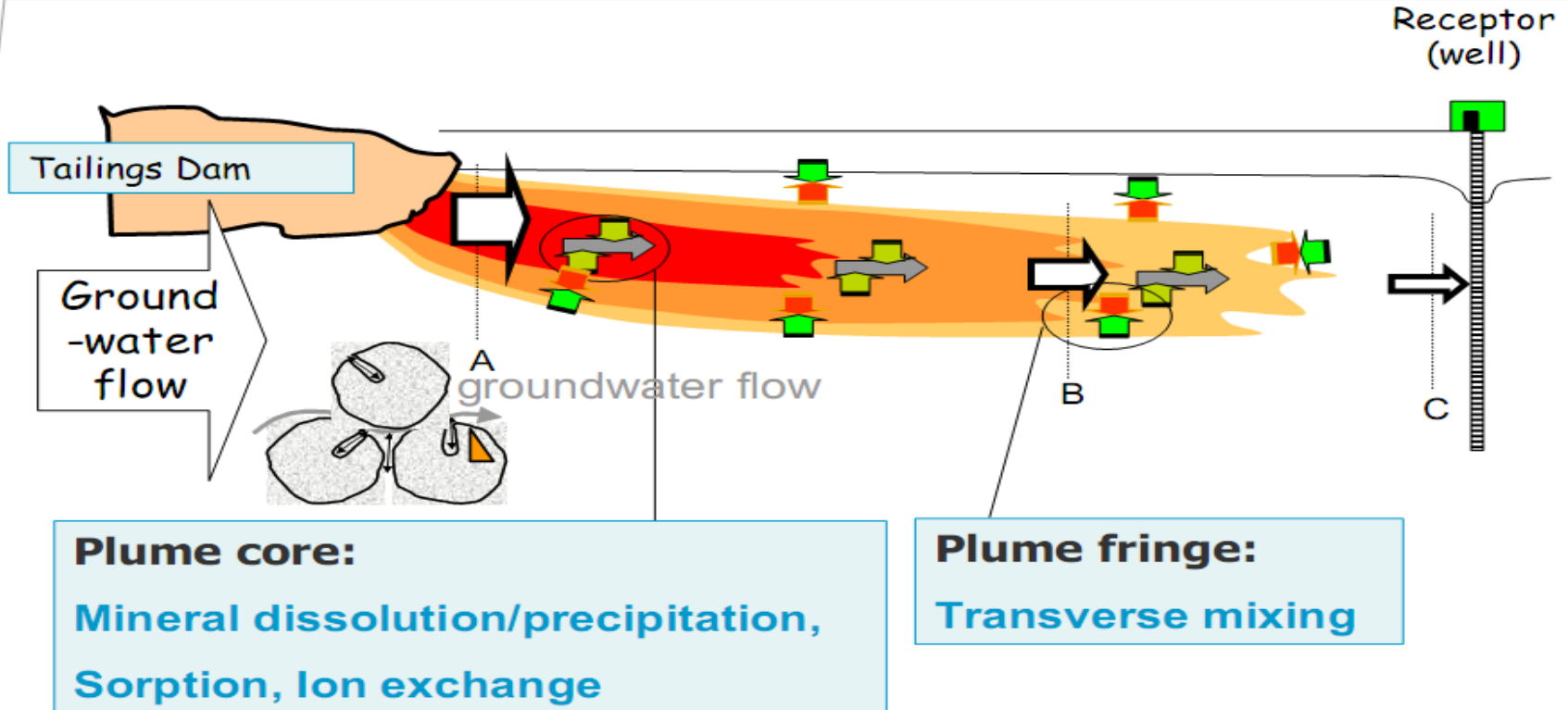


**A reactive multicomponent
transport model for saturated
porous media**

MT3DMS-Based Transport Modeling Tools



Conceptual model for acidic leachate attenuation



➡ Two major attenuation mechanisms:
Dilution and Reaction

(Henning Prommer, 2013)

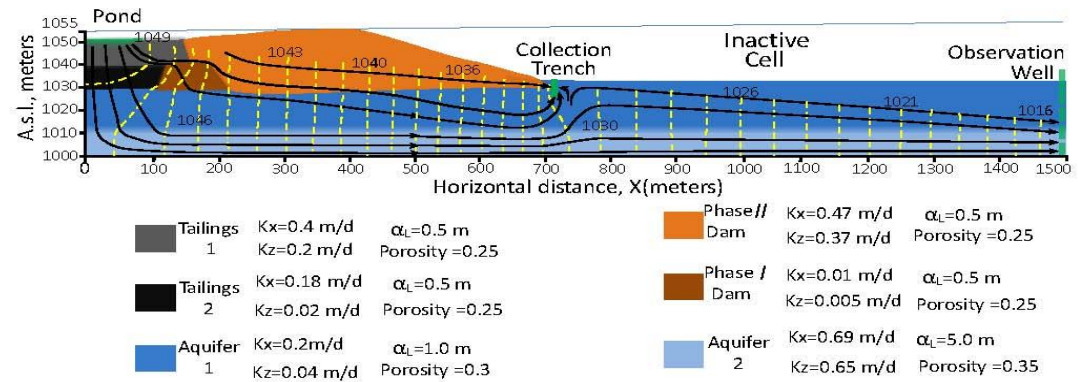
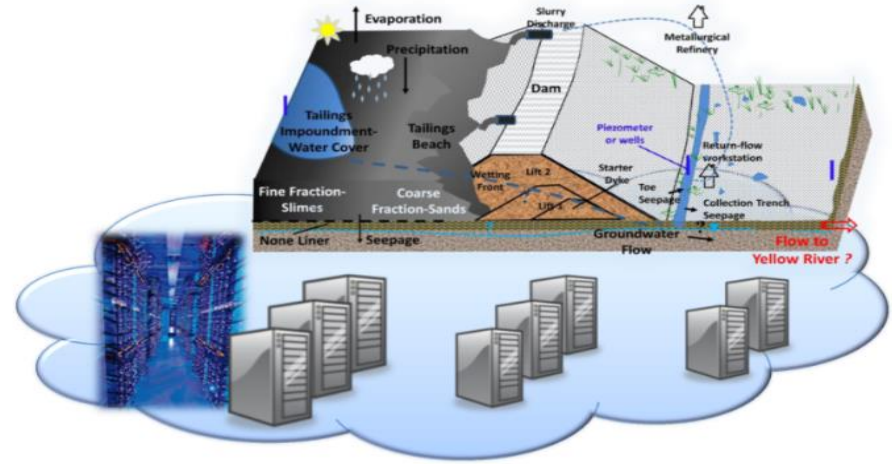
Numerical model

- Transported chemical undergoes surface complexation, with mineral dissolution/precipitation.

- Governing equation:

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial X_i} \left(D_{ij} \frac{\partial C}{\partial X_j} \right) - \frac{\partial}{\partial X_i} (v_i C) - R$$

- Single species transport model in MT3DMS with advection, multiple species reactive modeling in PHREEQC.



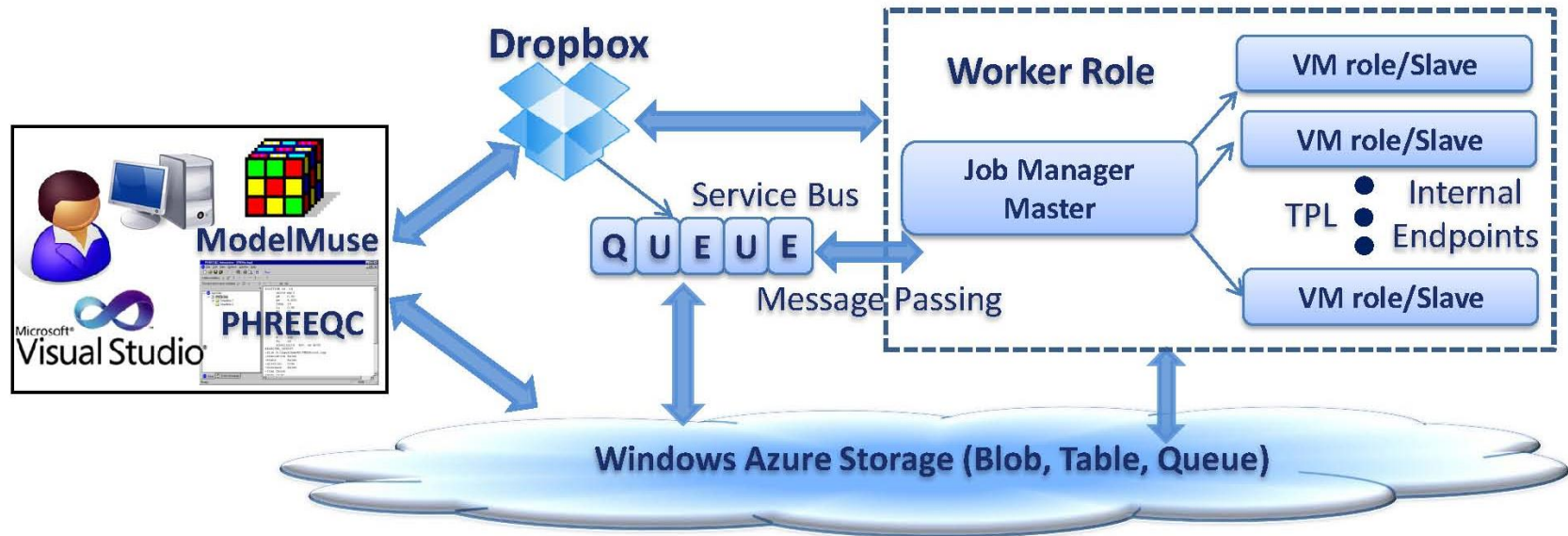
Architecture of Modeling Application on Azure

Job manager

- Communicate between Web portal and request queue.
- Coordinating computation workloads.
- Monitoring the execution status.

Task queue

- Communicate among Job manager, Slave VMs and Dropbox/GUIs.
- Releasing computation-input/-output queues according to bat files.



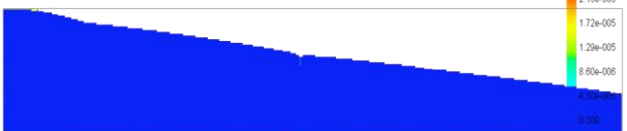
Simulation Results

Conservative transport plumes



Time = 10 (1-14)

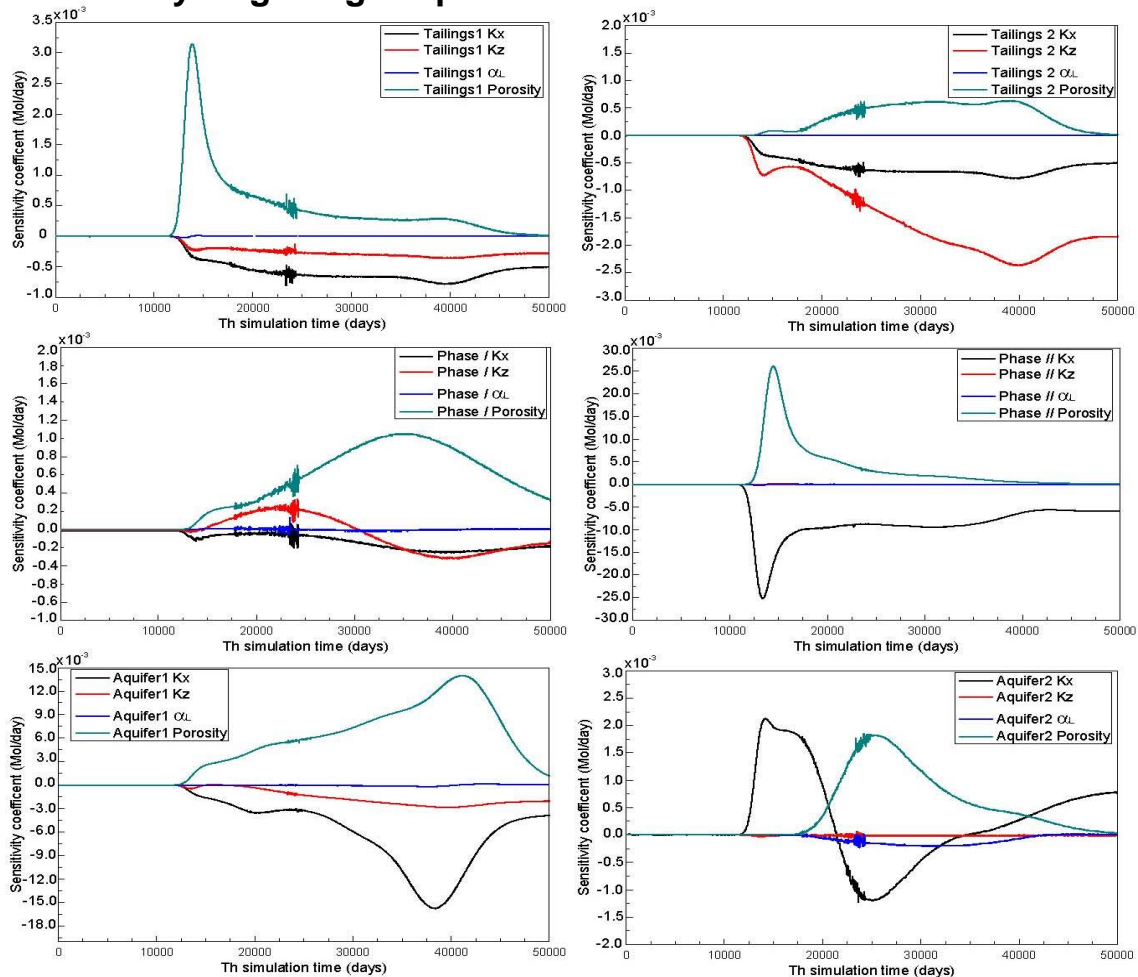
Thorium reactive transport plumes



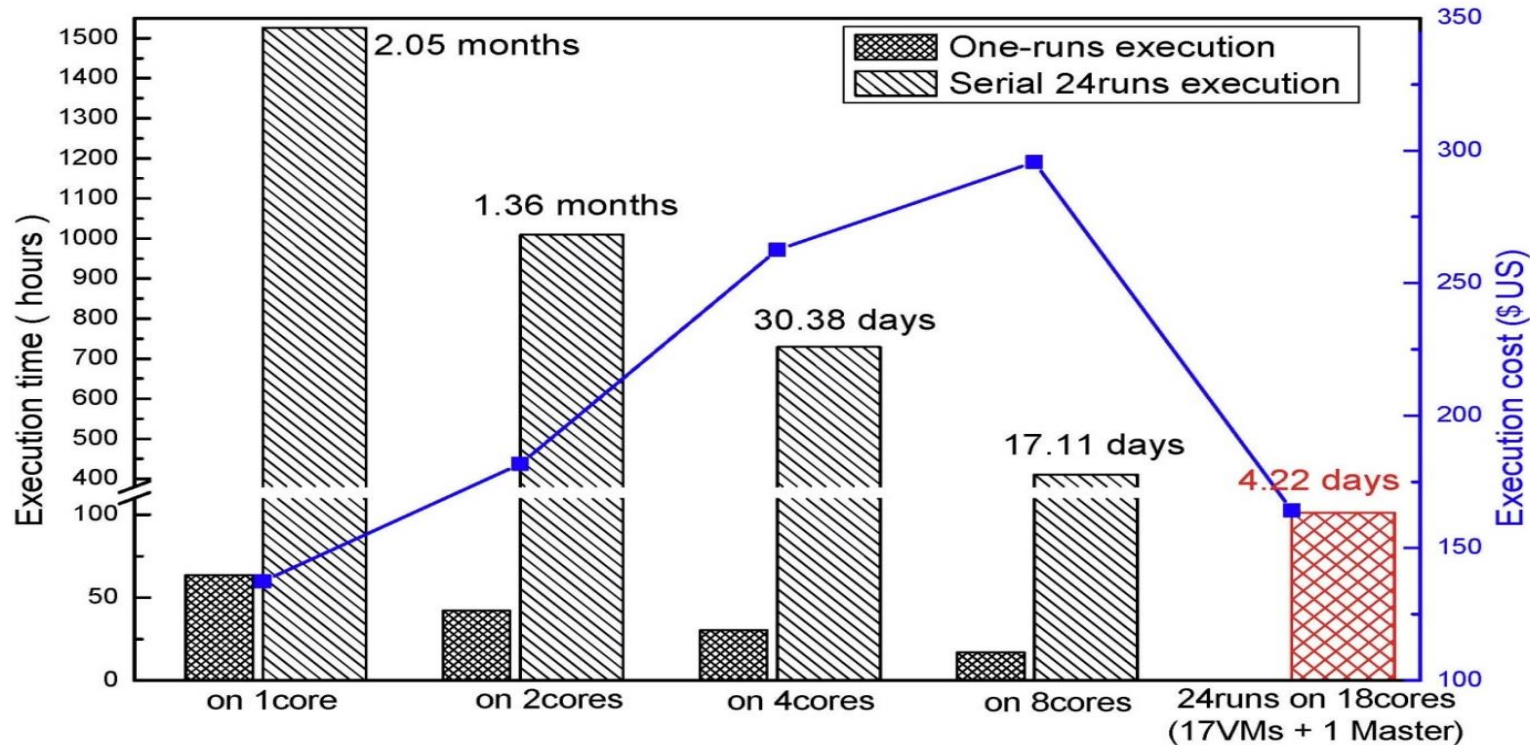
Time = 10 (1-14)

- FD, TVD, split-operator algorithm
- 72 layers, 745 columns

Sensitivities of thorium mass-fluxes into collection-trench to 24 hydrogeological parameters



Execution Time and Costs




Acknowledgements

Our cloud modeling research is supported by

 Microsoft Research Asia
(grant no. CNIC-MSR2013)



 National Natural Science Foundation
of China (grant no. 91225301)



PKU: Guangjun Zhang; Yingying Yao; Xiang Huang; Yong Tian

Microsoft Resxearch Asia: Lily Sun; Xin Ma

A scenic landscape featuring rolling green hills in the foreground, a field of vibrant yellow flowers in the middle ground, and distant mountains under a dramatic, cloudy sky. The text "Thank You!" is centered in the middle of the image.

Thank You!



Save the planet and return
your name badge before you
leave (on Tuesday)

