Towards a Cyberinfrastructure For Watershed Prediction and Data Assimilation

from Shared Discovery to Shared Decision Making

C. Duffy, L. Leonard, G. Bhatt, X. Yu, L. Giles, D. Wu. P. Ragavhan, Penn State Univ.

Numerical Watershed Prediction:
A Strategy

Issues & Questions

Simulation Framework

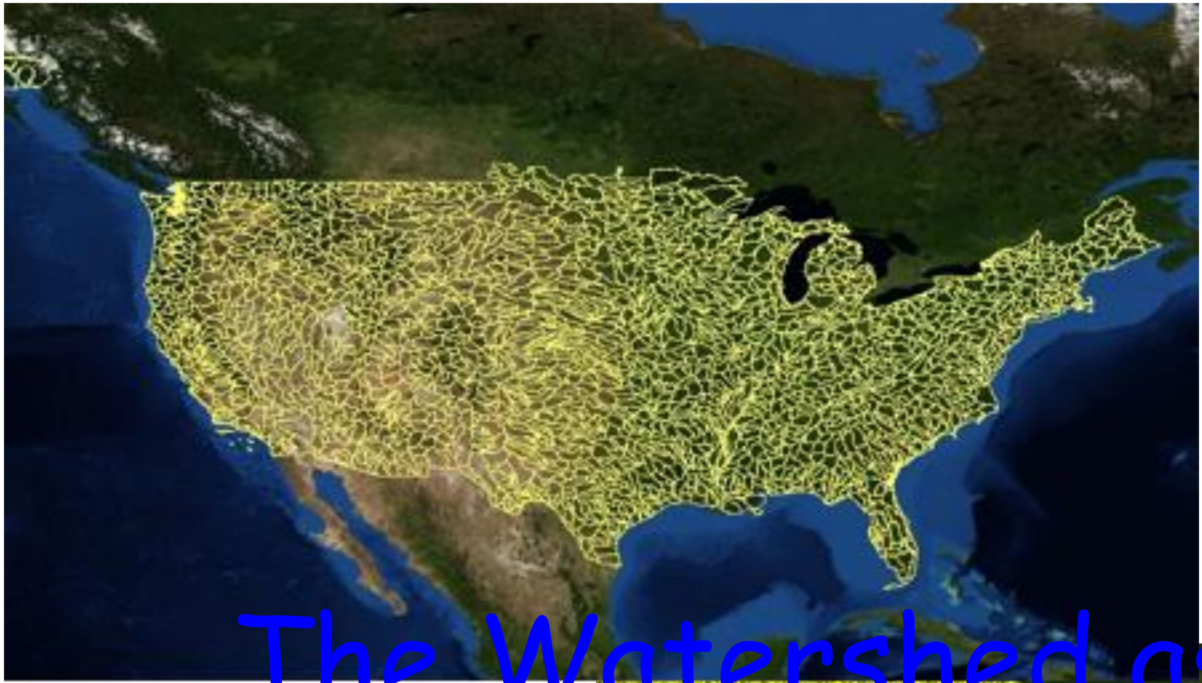
The Essential Terrestrial Variables - National Products?

Cyberinfrastructure for Data and Models-> Cloud Framework

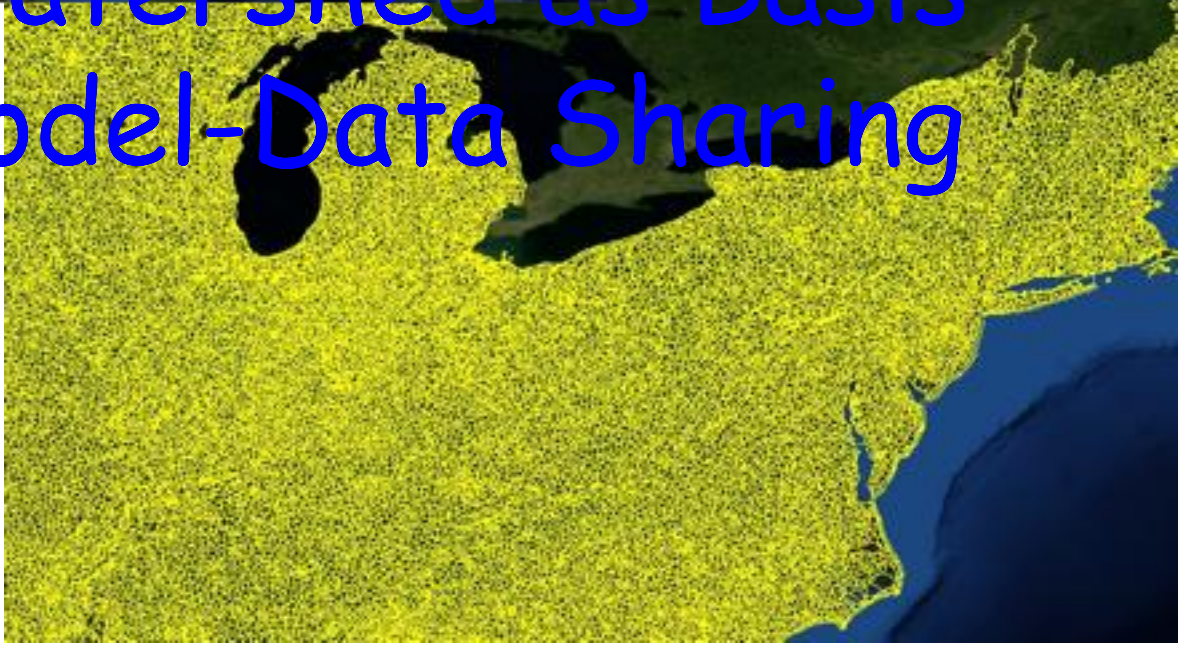
Reanalysis and the Role of Environmental Observatories

Platform for interagency collaboration & research support?

The Watershed as Basis For Model-Data Sharing



2,268 USGS
HUC 8
watersheds



103,444 USGS
HUC 12
watersheds

What are the Essential Terrestrial Variables?

Water-Energy-Solute-Vegetation-Management

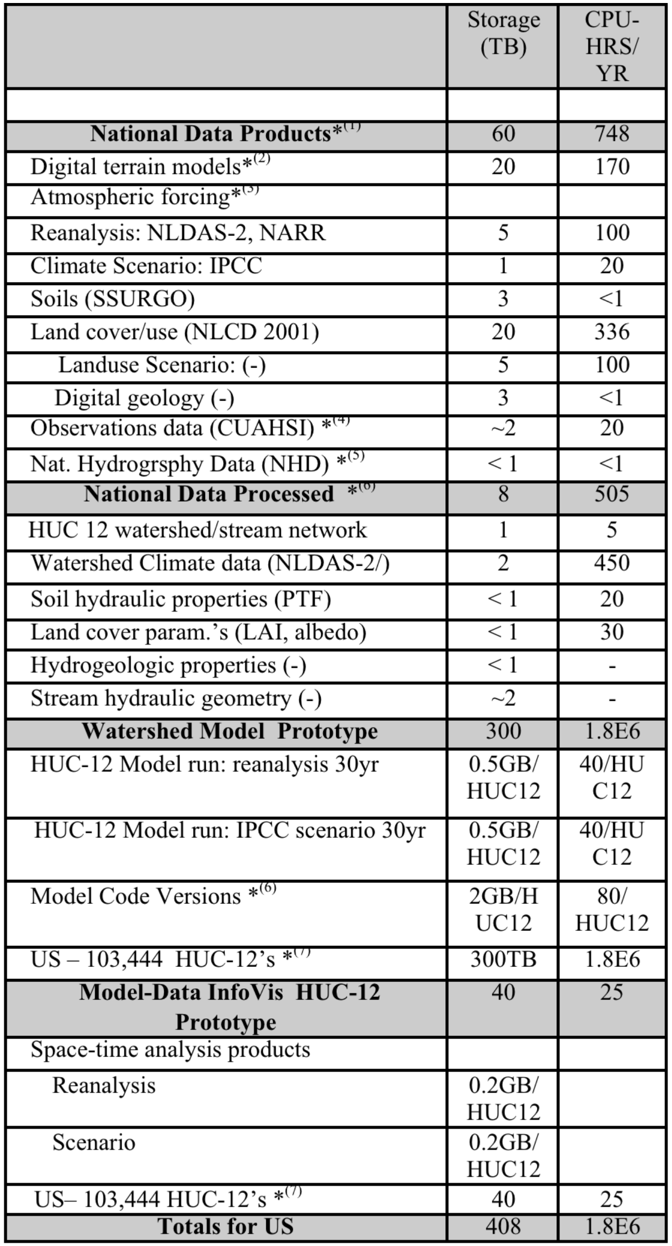
- Atmospheric Forcing (precipitation, snow cover, wind, relative humidity, temperature, net radiation, albedo, photosynthestic atmospheric radiation)
- Digital elevation models (30, 10, 3, 1m resolution)
- River/Stream discharge, stage, cross-section
- Soil (texture, C/N, organic, hydrologic & thermal properties)
- Groundwater (levels, extent, hydrogeologic properties)
- Land Cover (biomass/leaf area index, phenology,…….)
- Land Use (human infrastructure, demography, ecosystem disturbance, property & political boundaries)
- Environmental Tracers- stable isotopes
- Water Use and Water Transfers
- Lake/Reservoir/Diversion (levels, extent, discharge, operating rules)

Most data reside on federal servers ….many terabytes

A-Priori Data Sources

Feature/ Time Series	Property	Source
Soil	Porosity; Sand, Silt, Clay Fractions; Bulk Density	CONUS, SSURGO and STATSGO http://www.soilinfo.psu.edu/index.cgi?soil_data&conus http://datagateway.nrcs.usda.gov/NextPage.asp http://www.ncgc.nrcs.usda.gov/products/datasets/statsgo/
Geology	Bed Rock Depth; Horizontal and Vertical Hydraulic Conductivity	http://www.dcnr.state.pa.us/topogeo/, http://www.lias.psu.edu/emsl/guides/X.html
Land Cover	LAI	http://glcf.umiacs.umd.edu/data/landcover/data.shtml, http://ldas.gsfc.nasa.gov/LDAS8th/MAPPED.VEG/LDASmapveg.shtml;
	Manning's Roughtness	Hernandez et. al., 2000
River	Topology: From Node – To Node, Neighboring Elements;	Derived using PIHMgis (Bhatt et. al., 2008)
	Manning's Roughness;	Dingman (2002)
	Coefficient of Discharge	ModHms Manual (Panday and Huyakorn, 2004)
	Shape and Dimensions;	Derived from regression using depth, width and discharge data from http://nwis.waterdata.usgs.gov/usa/nwis/measurements
Forcing	Prec, Temp. RH, Wind, Rad.	National Land Data Assimilation System : NLDAS-2
DEM		http://seamless.usgs.gov/
Streamflow		http://nwis.waterdata.usgs.gov/nwis/sw
Groundwater		http://nwis.waterdata.usgs.gov/nwis/gw

Scale of Data and Model Storage And CPU-HRS/yr To Support a National Archive for Watershed Reanalysis 1979-2012 & IPCC Projections 2012-2065

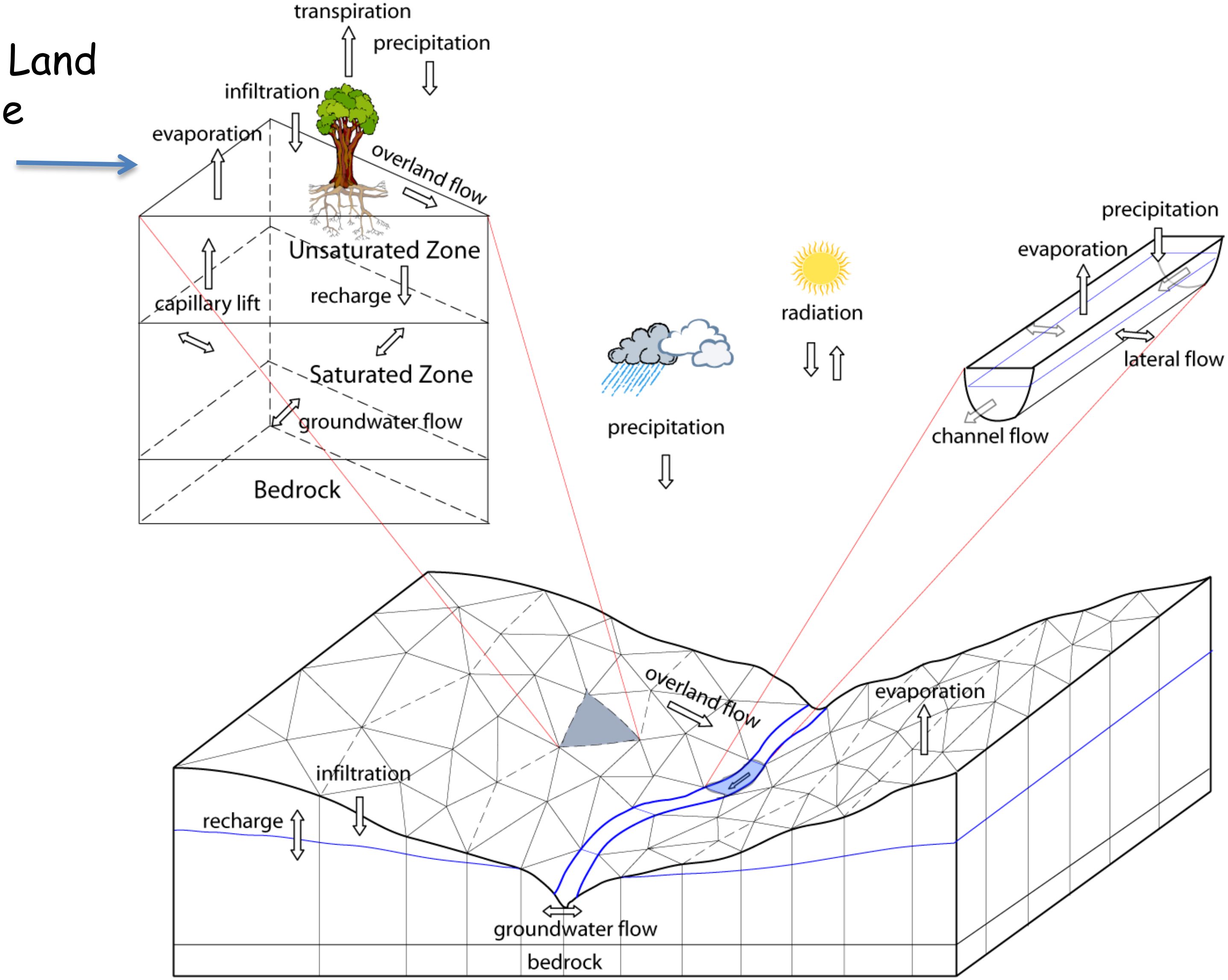


	Storage (TB)	CPU-HRS/ YR
National Data Products*[1]	60	748
Digital terrain models*[2]	20	170
Atmospheric forcing*[3]		
Reanalysis: NLDAS-2, NARR	5	100
Climate Scenario: IPCC	1	20
Soils (SSURGO)	3	<1
Land cover/use (NLCD 2001)	20	336
Landuse Scenario: (-)	5	100
Digital geology (-)	3	<1
Observations data (CUAHSI) *[4]	~2	20
Nat. Hydrogrsphy Data (NHD) *[5]	< 1	<1
National Data Processed *[6]	8	505
HUC 12 watershed/stream network	1	5
Watershed Climate data (NLDAS-2/)	2	450
Soil hydraulic properties (PTF)	< 1	20
Land cover param.'s (LAI, albedo)	< 1	30
Hydrogeologic properties (-)	< 1	-
Stream hydraulic geometry (-)	~2	-
Watershed Model Prototype	300	1.8E6
HUC-12 Model run: reanalysis 30yr	0.5GB/ HUC12	40/HU C12
HUC-12 Model run: IPCC scenario 30yr	0.5GB/ HUC12	40/HU C12
Model Code Versions *[6]	2GB/H UC12	80/ HUC12
US – 103,444 HUC-12's *[7]	300TB	1.8E6
Model-Data InfoVis HUC-12 Prototype	40	25
Space-time analysis products		
Reanalysis	0.2GB/ HUC12	
Scenario	0.2GB/ HUC12	
US– 103,444 HUC-12's *[7]	40	25
Totals for US	408	1.8E6

Estimated storage (TB) and cpu-hrs/yr for data processing, model runs and info-vis for the 103,444 watersheds in the coterminous US. Notes for Table 1: *[1] National data includes downloading and maintaining digital terrain, soils, hydrogeology, NLCD land cover/use, hydroclimatic data from NLDAS-2/NARR, and MODIS satellite data. *[2] Digital terrain includes 1m-lidar, 3m, 10m, 30m, 90m DEM. *[3] National data 1979-pres.: precipitation, net radiation, wind speed/direction, soil moisture, etc. on a 4km grid. *[4] National historical streamflow, soil moisture, weather station point gauging data.*[5] National hydrographic data for watersheds, streams, channel geometry in GIS formats. *[6] Processing of all National Data Products to generate input and model parameter database for all 103,444 watersheds, To be used to run the watershed model PIHM. *[7] US National coverage. There are 103,444 HUC12 watersheds in the continental US.

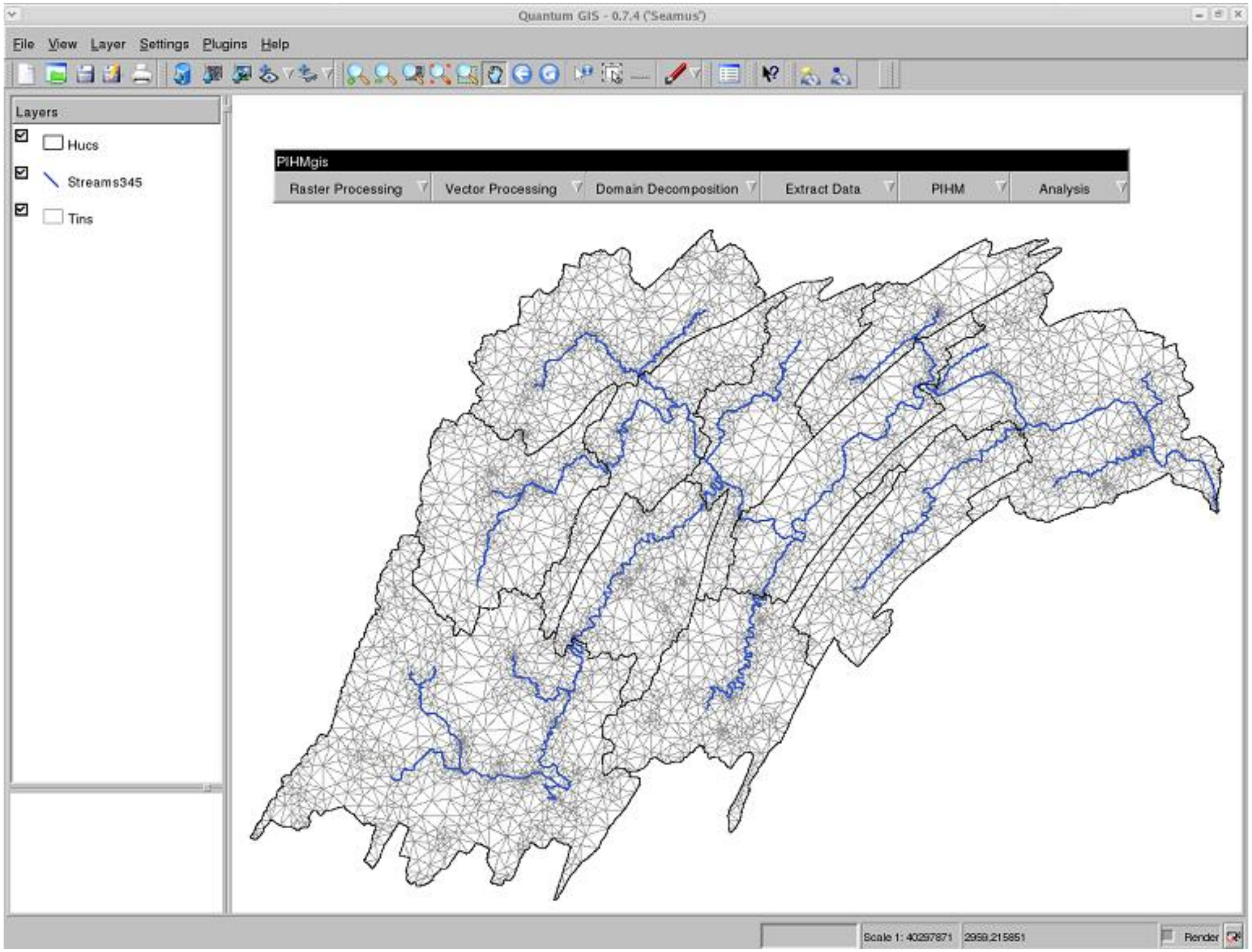
Penn State Integrated Hydrologic Model (PIHM)

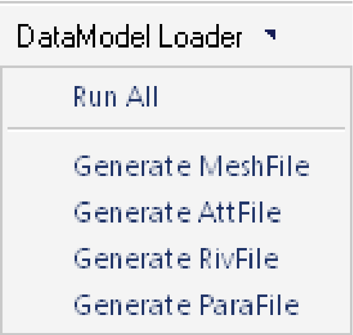
NOAH Land Surface Model

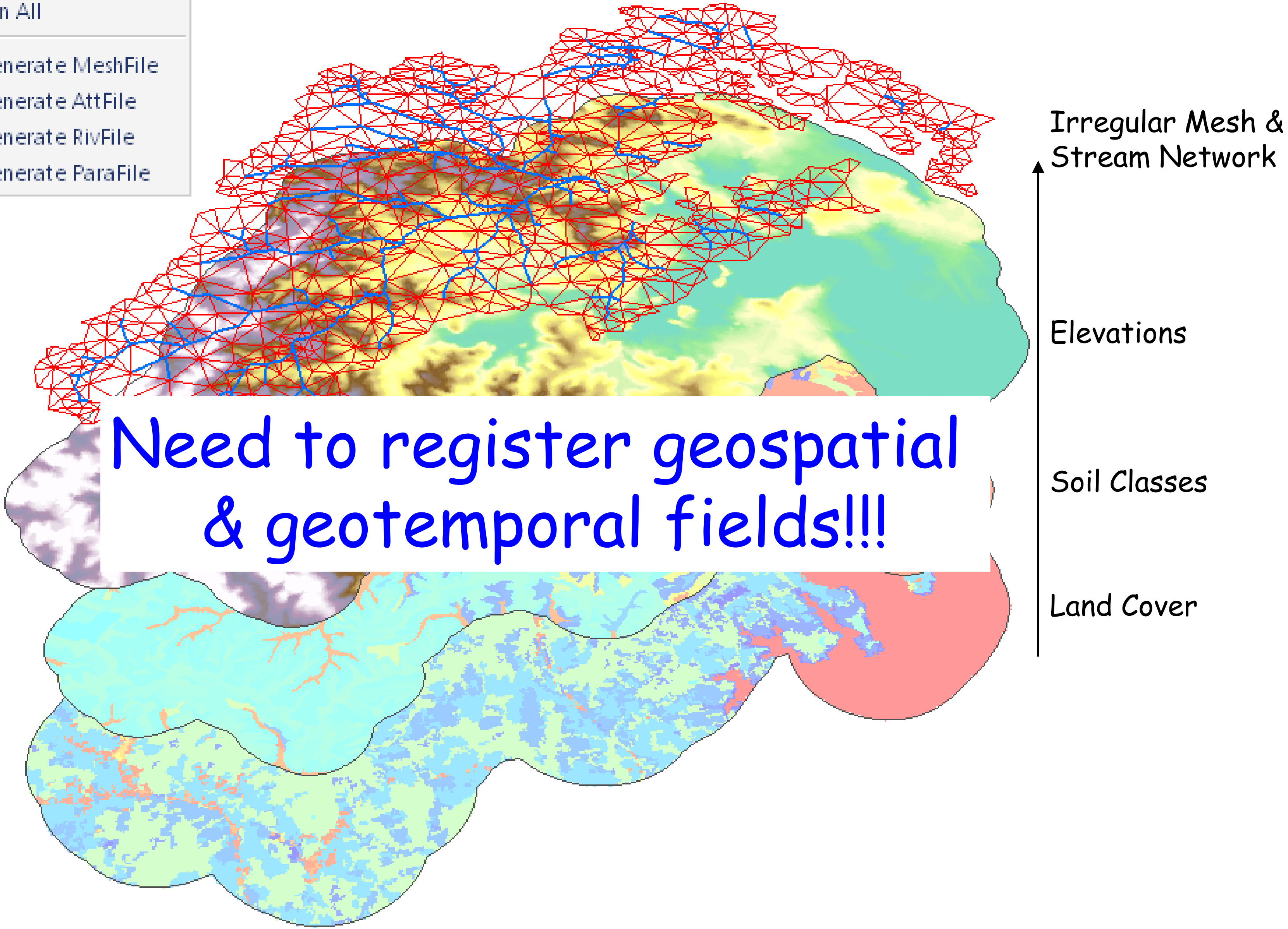


PIHM Qu and Duffy 2007. Kumar, Bhatt, Duffy 2009

PIHM GIS Desktop: Manipulating Geospatial Data and Models







Need to register geospatial & geotemporal fields!!!

Towards a New Cyberinfrastructure For Models and Data

Shared Discovery & Shared Decision Making

Workflow Concept

Scientific workflows are used to describe, compose, and execute ensembles of scientific computations on distributed resources.

The prescriptive and descriptive representations of workflows (called provenance) are useful for publishing, discovering, and reproducing computational results.

Scientific workflows within the geosciences range from pipelines used to create community data products to real-time processing of sensor data to individual researchers' unique computations (the "long tail")

Approach

Solicit input to develop an initial roadmap for defining scientific workflows in the geosciences

Provide examples of scientific workflows that:

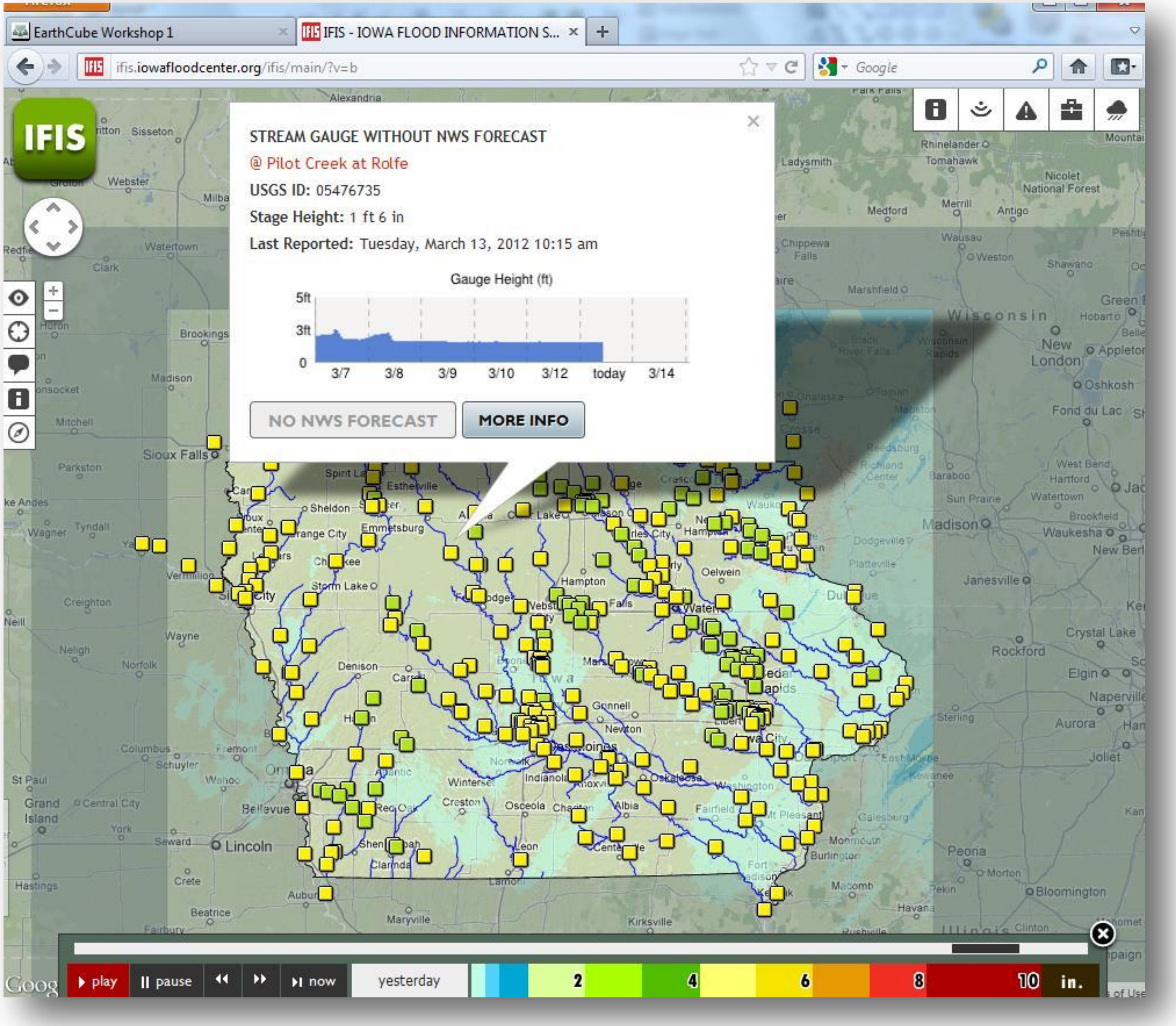
- Support automated and efficient data management
- Assures reliability for national and global data
- Offers a scalable solution …nation…..global
- Automatically captures provenance
- Accessible by to geoscience researchers, students, etc.
- Allows integration across geoscience disciplines

Possible Earth Science Workflow Paradigms

The Center Modeling Paradigm:

A group of people (a Center) design the workflow to run specific models, the results of executing the workflow are specialized data products that are then published to a community of users. An example of this is the Iowa Flood Center. A group of scientists develops the models and runs them, they integrate data from a specific set of sensors. Workflow execution engines can automate these processes. Once the workflows are run, the results are published so that all stakeholders in the Center can benefit.

Flood Information System



EarthCube Workshop 1

IFIS - IOWA FLOOD INFORMATION S...

ifis.iowafloodcenter.org/ifis/main/?v=b

STREAM GAUGE WITHOUT NWS FORECAST

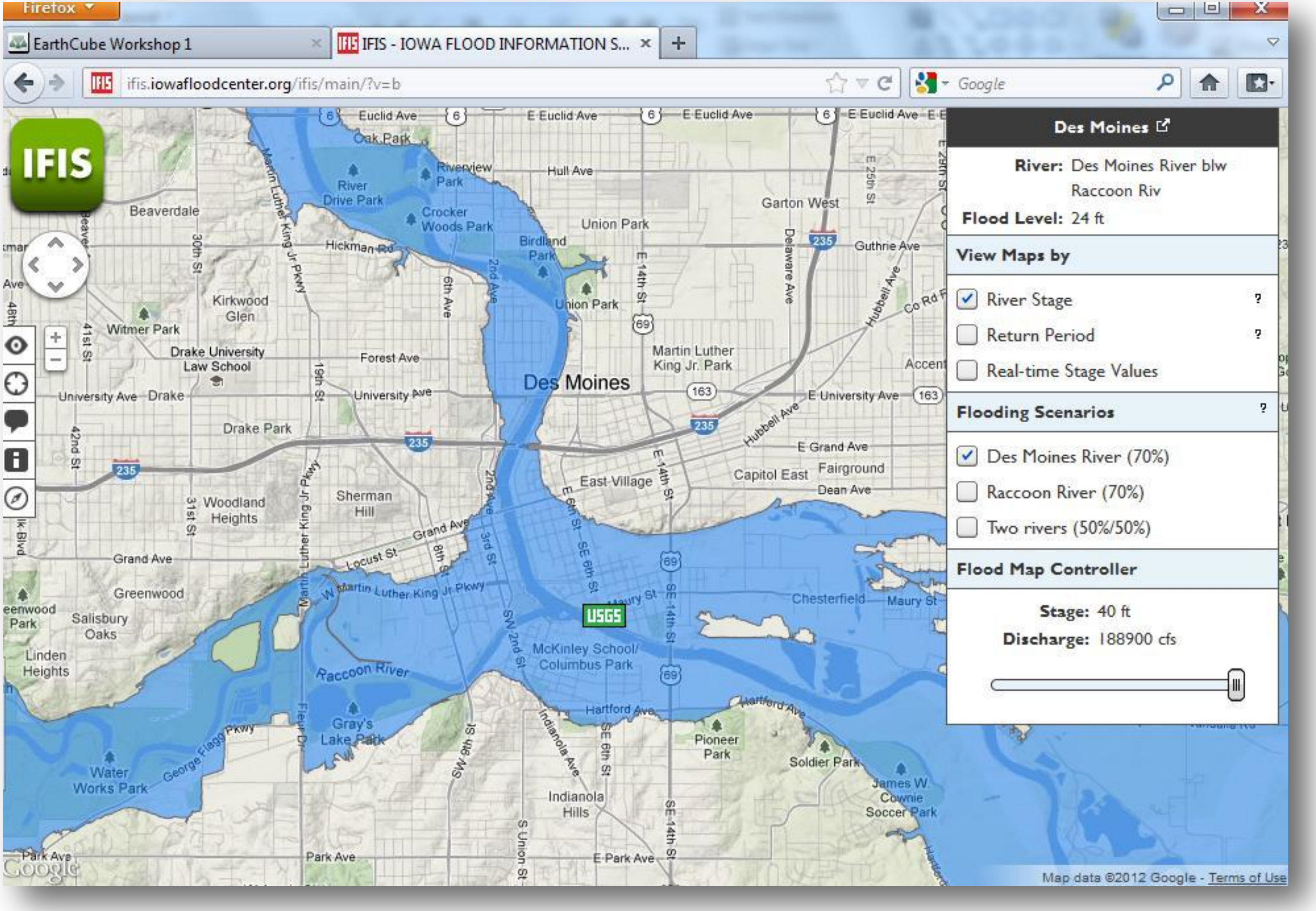
@ Pilot Creek at Rolfe

USGS ID: 05476735

Stage Height: 1 ft 6 in

Last Reported: Tuesday, March 13, 2012 10:15 am

Flood Information System



Des Moines

River: Des Moines River blw Raccoon Riv

Flood Level: 24 ft

View Maps by

☑ River Stage ?

☐ Return Period ?

☐ Real-time Stage Values

Flooding Scenarios ?

☑ Des Moines River (70%)

☐ Raccoon River (70%)

☐ Two rivers (50%/50%)

Flood Map Controller

Stage: 40 ft

Discharge: 188900 cfs

The Long Tail Paradigm:

Individual investigators run their own workflows. The investigators prepare the data themselves, integrate it with data from shared sources, and run their own models. An example of this is a river ecologist, interested in estimating water metabolism rates. She sets up her own local sensors, prepares the data, integrates it with data from government sites/services (about weather, other sensors in the area, etc), and then runs some models. Currently this investigator might just use spreadsheets to manage the data and model codes. In the future, she would like better support to do the routine data preparation, stream the data, pre-process the data automatically, and record all the steps. She might like to easily incorporate into her workflow the models developed by other people, run her workflow on other people's data, etc

Data Analysis: Water Metabolism

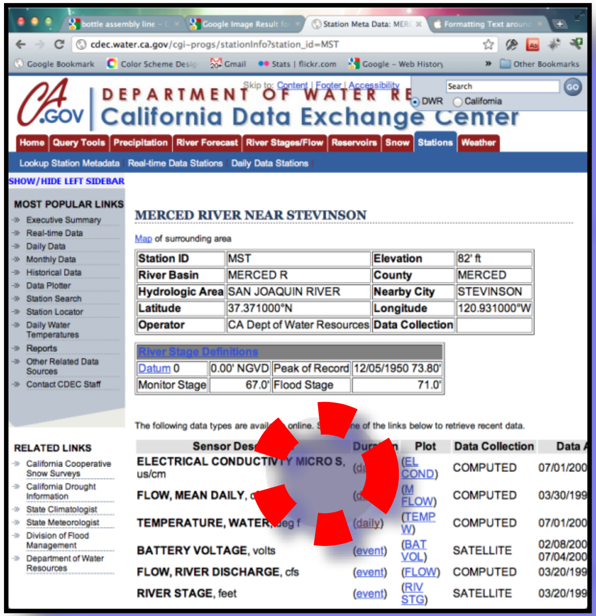
[Gil et al 11] [Villamizar et al 11]

Collaboration with Tom Harmon of UC Merced, C. Knoblock and P. Szekely of USC

California's Central Valley

Harmon's sensors

State data sources (other sensors)



MOST POPULAR LINKS

- Executive Summary
- Real-time Data
- Daily Data
- Monthly Data
- Historical Data
- Data Plotter
- Station Search
- Station Locator
- Daily Water Temperatures
- Reports
- Other Related Data Sources
- Contact CDEC Staff

RELATED LINKS

- California Cooperative Snow Surveys
- California Drought Information
- State Climatologist
- State Meteorologist
- Division of Flood Management
- Department of Water Resources

MERCED RIVER NEAR STEVINSON

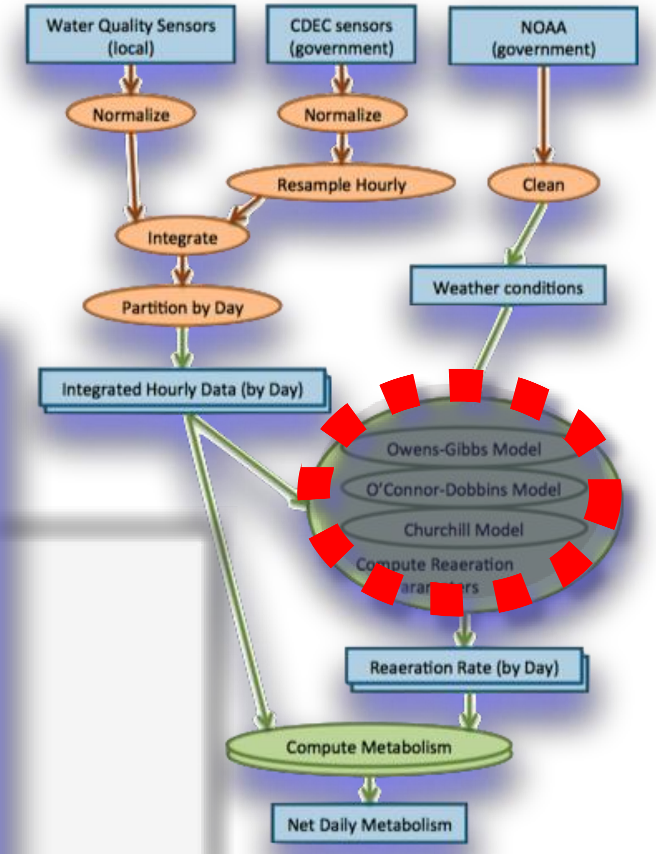
Map of surrounding area

Station ID	MST	Elevation	82' ft
River Basin	MERCED R	County	MERCED
Hydrologic Area	SAN JOAQUIN RIVER	Nearby City	STEVINSON
Latitude	37.371000°N	Longitude	120.931000°W
Operator	CA Dept of Water Resources	Data Collection	

River Stage Definitions			
Datum 0	0.00' NGVD	Peak of Record	12/05/1950 73.80'
Monitor Stage	67.0'	Flood Stage	71.0'

The following data types are avail online. S ne of the links below to retrieve recent data.

Sensor Des	Dur n	Plot	Data Collection	Data A
ELECTRICAL CONDUCTIV Y MICRO S, us/cm	(da	(EL COND)	COMPUTED	07/01/200
FLOW, MEAN DAILY, c	(d	(M FLOW)	COMPUTED	03/30/199
TEMPERATURE, WATER, deg f	(daily)	(TEMP W)	COMPUTED	07/01/200
BATTERY VOLTAGE, volts	(event)	(BAT VOL)	SATELLITE	02/08/200 07/04/200
FLOW, RIVER DISCHARGE, cfs	(event)	(FLOW)	COMPUTED	03/20/199
RIVER STAGE, feet	(event)	(RIV STG)	SATELLITE	03/20/199



The Big Data Paradigm:

Very large datasets with computationally intensive codes are routinely run, requiring high-end computing resources and scalable infrastructure. An example of this would be climate or weather models. The Big Data Paradigm includes the **Data Intensive Paradigm** where large computational models require very large datasets during runtime as well as pre- post-processing. An example of this is Ecological and Agro-Hydrologic models using vegetation, landuse, digital terrain, soils, weather and climate data as input to their models. This workflow requires the usual High Performance Computing (HPC) resources for modeling, but also requires HPC for large data handling during run-time.

The Collaborative Ocean Visualization Environment (COVE)

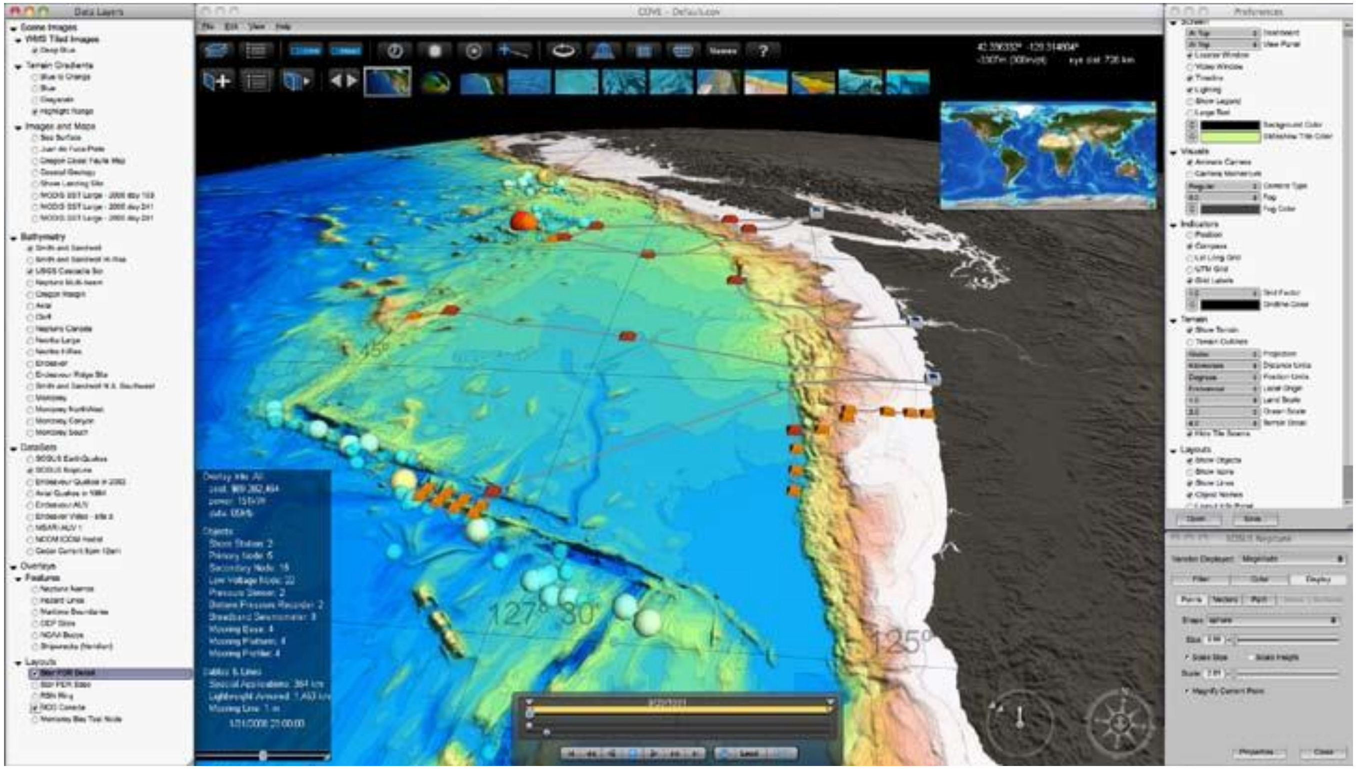


Figure 2: COVE displays a geo-positioned scientific data, seafloor terrain, images, and instrument layout with selectable layers on the left and rich visualization controls on the right.

The Metadata-Rich Paradigm:

A collaboration that wants to track metadata of all data used and produced by their analyses. The user in the Metadata-Rich Paradigm may be individuals or communities of geoscientists where data collection is a very intensive process and often uses very sophisticated laboratory analyses on each sample they collect. An example of this is geochemists that collect limited amounts of actual rock sample data for all rocks of the earth but processing and lab analyses lead to very large amounts of meta-data.

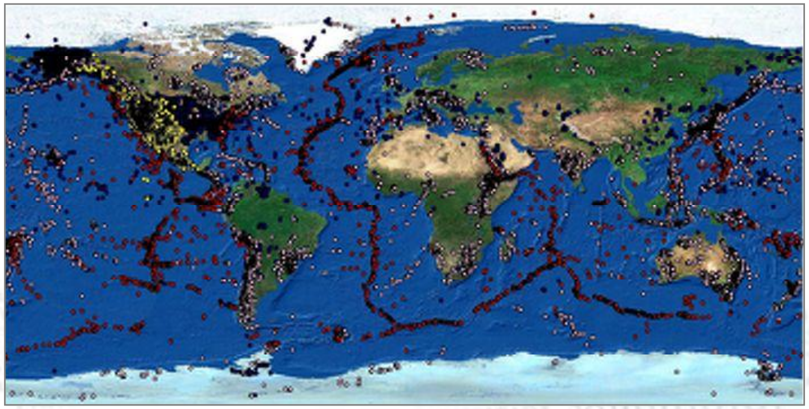
The Metadata-Rich Paradigm:

EarthChem

earthchem
ADVANCED DATA MANAGEMENT IN SOLID EARTH GEOCHEMISTRY

- Consortium founded in 2003
- Project funded by US NSF in 2005
  - To develop & operate the EarthChem Portal as a central discovery and access point for geochemical and geochronological data
  - To facilitate community-based growth of data holdings
  - To promote & implement standards for data management in Geochemistry

Access to >13 Million analytical values for >600,000 sample from GEOROC, NAVDAT, PetDB, USGS



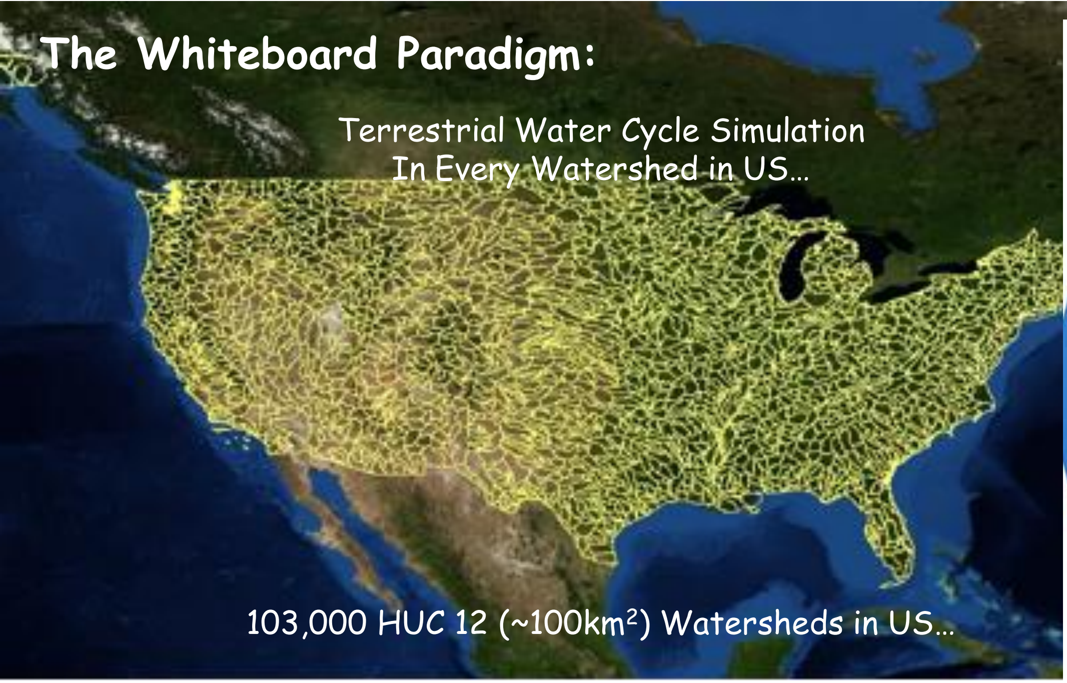
Geochemical Data

- Data are acquired by a plethora of analytical methods with procedures.
  - Tailored to scientific problems and customized to optimize data quality.
- Small data volumes.
- Large personal effort to generate data.
- Many data are unique and cannot be reproduced.

The Whiteboard Paradigm:

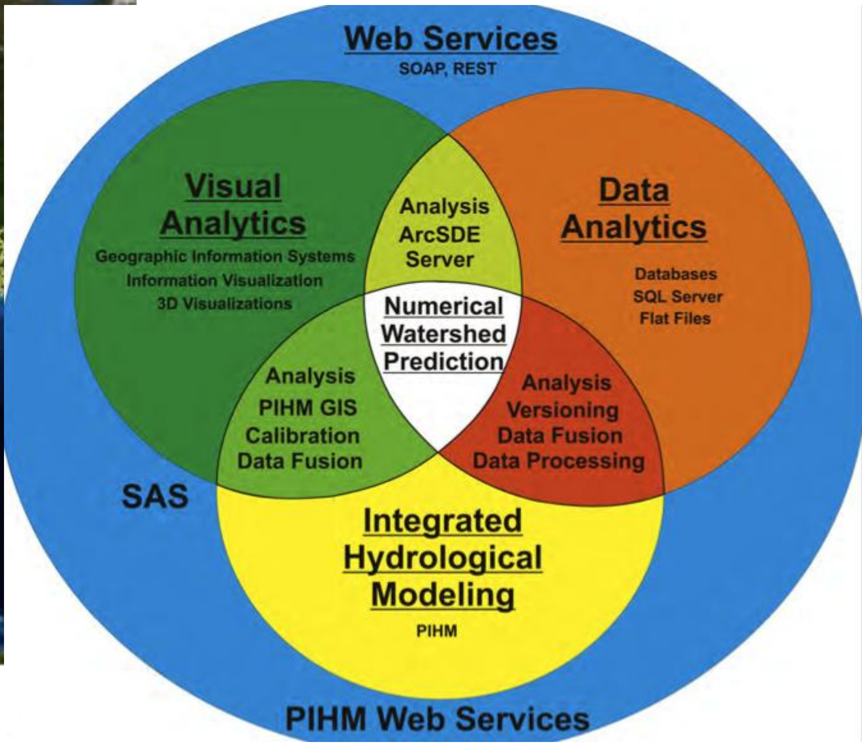
A group of investigators, decision makers in a collaboration on an interdisciplinary investigation drafts a sketch of what looks like a workflow, just to gain an understanding on what models they are agreeing to all run on their individual data. The team may have little understanding of each others needs at first but want to work together to solve some big science or multidisciplinary environmental question. An example of The Paradigm might be ecologists, hydrologists, coastal marine and social scientists working on the Chesapeake Bay watershed-estuary problems or possibly the Critical Zone Observatory program where geologists, hydrologists, geochemists, ecologists, weather, climate and soil scientists are all working together and want to use each others science in the most productive way. The important point here is the workflow needs a way to support team approach trying to resolve very hard holistic problems or trying to solve very specialized problem with a multidisciplinary worldview of data and models.

The Whiteboard Paradigm:



Terrestrial Water Cycle Simulation
In Every Watershed in US…

103,000 HUC 12 (~100km^2) Watersheds in US…



Web Services
SOAP, REST

Visual Analytics
Geographic Information Systems
Information Visualization
3D Visualizations

Data Analytics
Databases
SQL Server
Flat Files

Analysis
ArcSDE
Server

Numerical Watershed Prediction

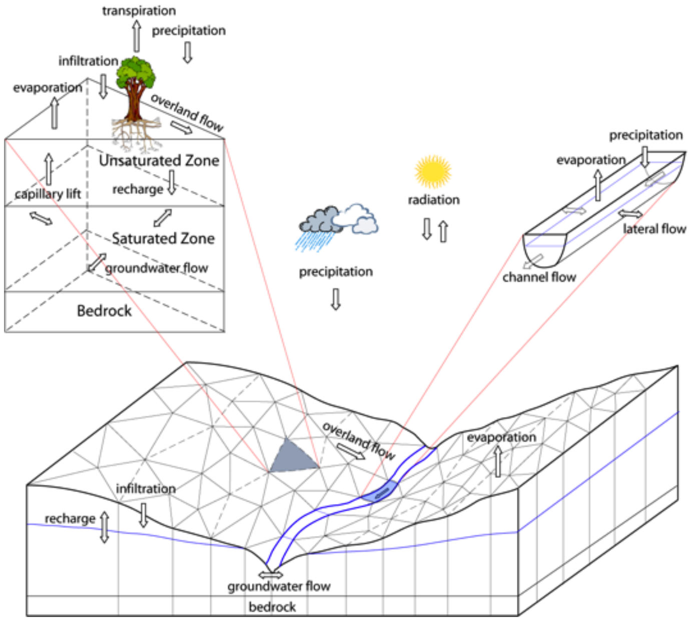
Analysis
PIHM GIS
Calibration
Data Fusion

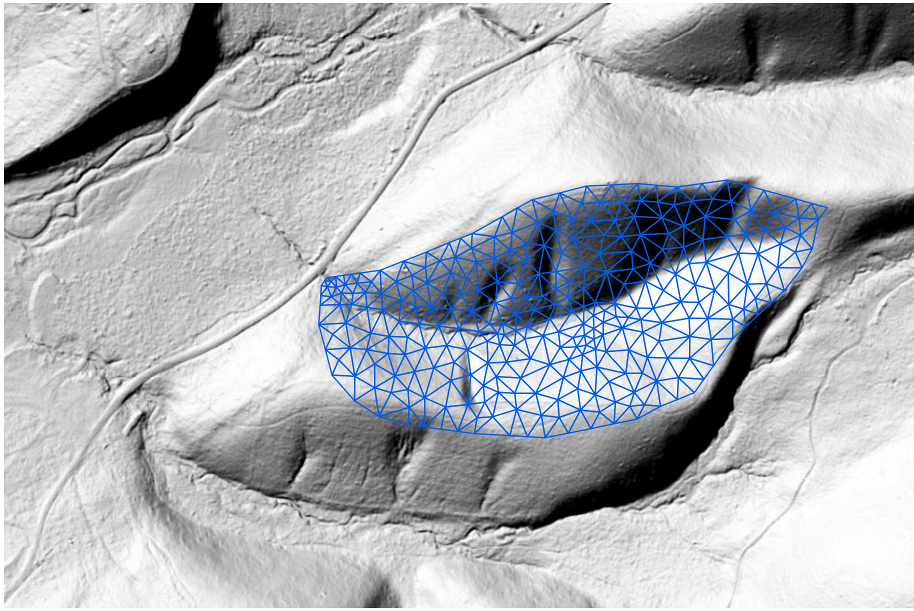
Analysis
Versioning
Data Fusion
Data Processing

SAS

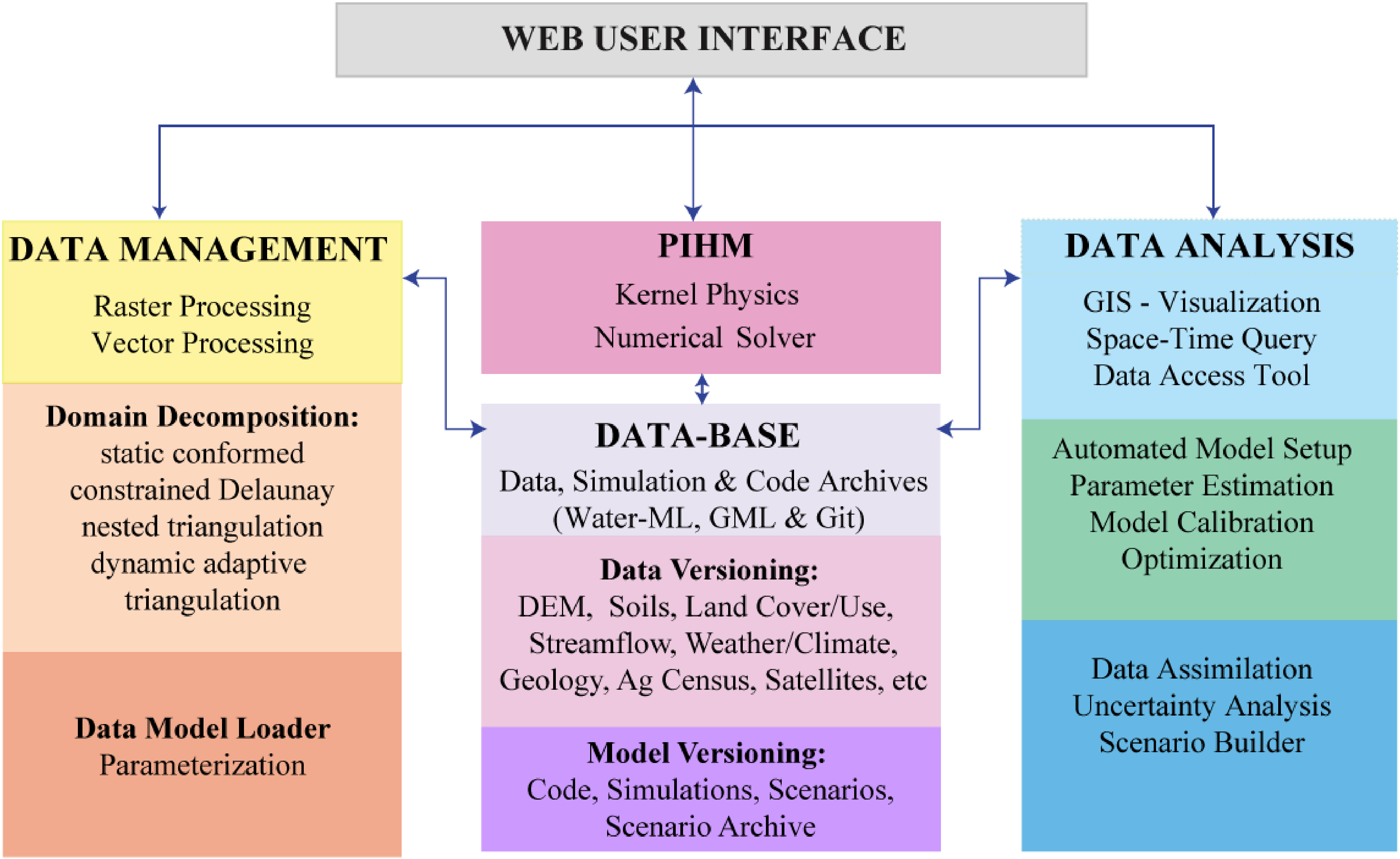
Integrated Hydrological Modeling
PIHM

PIHM Web Services





Model-Data Workflow



WEB USER INTERFACE

DATA MANAGEMENT

Raster Processing
Vector Processing

Domain Decomposition:
static conformed
constrained Delaunay
nested triangulation
dynamic adaptive
triangulation

Data Model Loader
Parameterization

PIHM

Kernel Physics
Numerical Solver

DATA-BASE

Data, Simulation & Code Archives
(Water-ML, GML & Git)

Data Versioning:
DEM, Soils, Land Cover/Use,
Streamflow, Weather/Climate,
Geology, Ag Census, Satellites, etc

Model Versioning:
Code, Simulations, Scenarios,
Scenario Archive

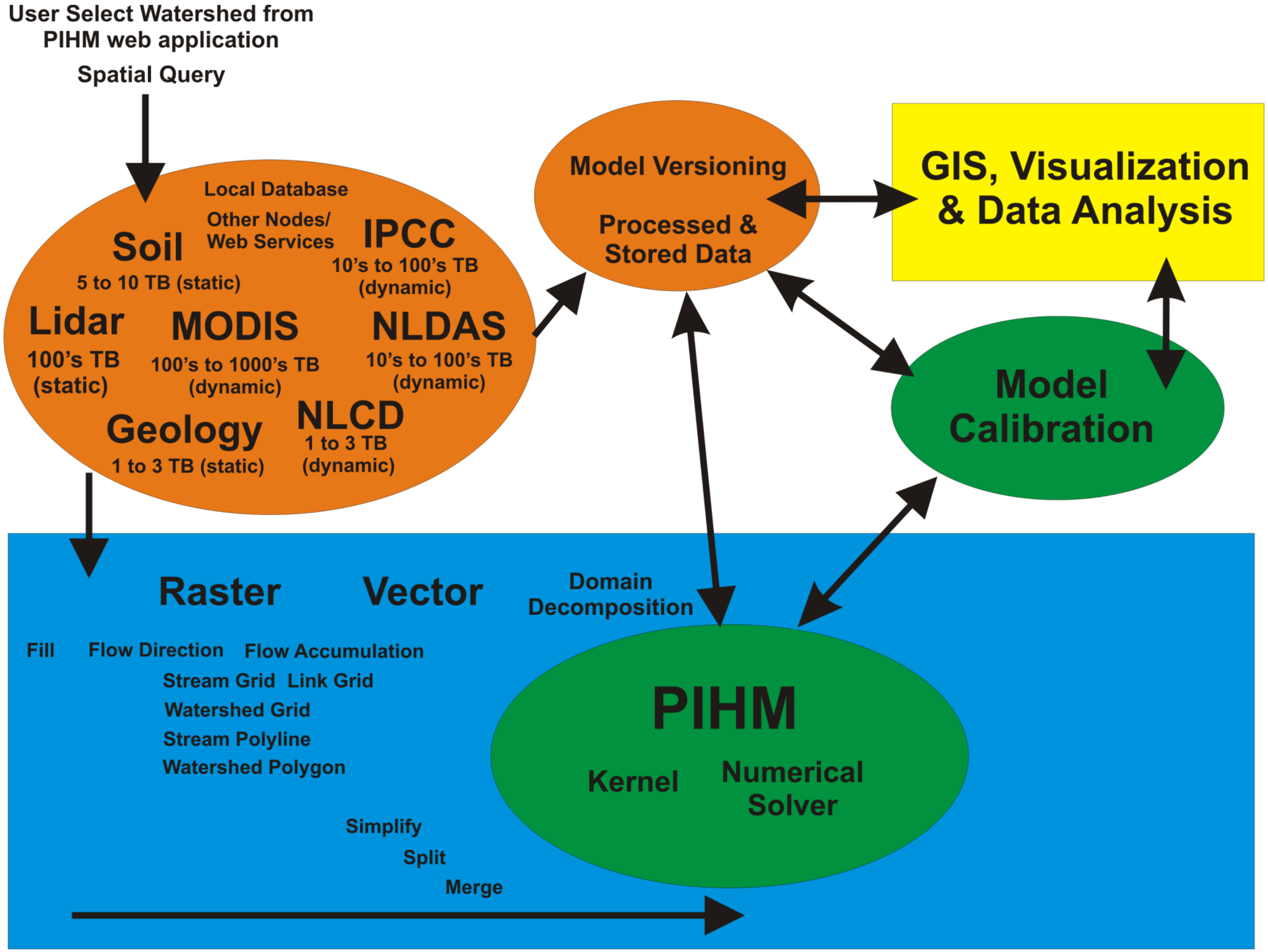
DATA ANALYSIS

GIS - Visualization
Space-Time Query
Data Access Tool

Automated Model Setup
Parameter Estimation
Model Calibration
Optimization

Data Assimilation
Uncertainty Analysis
Scenario Builder

Model-Data Integration Framework



User Select Watershed from PIHM web application

Spatial Query

Local Database

Other Nodes/ Web Services

Soil

5 to 10 TB (static)

IPCC

10's to 100's TB (dynamic)

Lidar

100's TB (static)

MODIS

100's to 1000's TB (dynamic)

NLDAS

10's to 100's TB (dynamic)

Geology

1 to 3 TB (static)

NLCD

1 to 3 TB (dynamic)

Model Versioning

Processed & Stored Data

GIS, Visualization & Data Analysis

Model Calibration

Raster

Vector

Domain Decomposition

Fill

Flow Direction

Flow Accumulation

Stream Grid

Link Grid

Watershed Grid

Stream Polyline

Watershed Polygon

Simplify

Split

Merge

PIHM

Kernel

Numerical Solver

PIHM USA Workflow

User Select Watershed from PIHM web application

Spatial Query

Soil

5 to 10 TB (static)

Lidar

100's TB (static)

MO

100's to 1

(dyna

Geolog

1 to 3 TB (sta

Loc

Oth

We

ization

alysis

n

Rast

Fill

Flow Direction

Stream

Watersh

Stream

Watersh

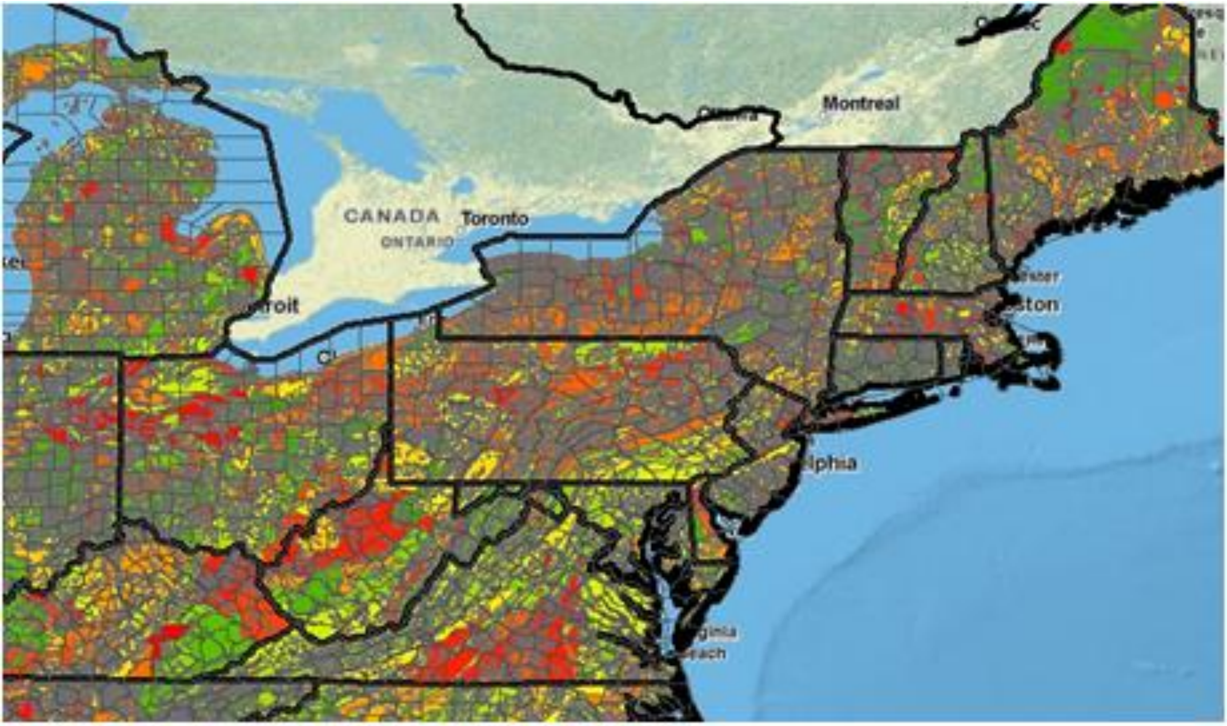
Kernel

Solver

Simplify

Split

Merge



http://soils.usda.gov/survey/geography/ssurgo/

http://soils.usda.gov/survey/geography/statsgo/

PIHM USA Workflow

User Select Watershed from
PIHM web ap

Spatial Qu

Soil

5 to 10 TB (s

Lidar

100's TB
(static)

M

100'

Geol

1 to 3 TB

Ra

Fill Flow Direct

Str

Wa

Str

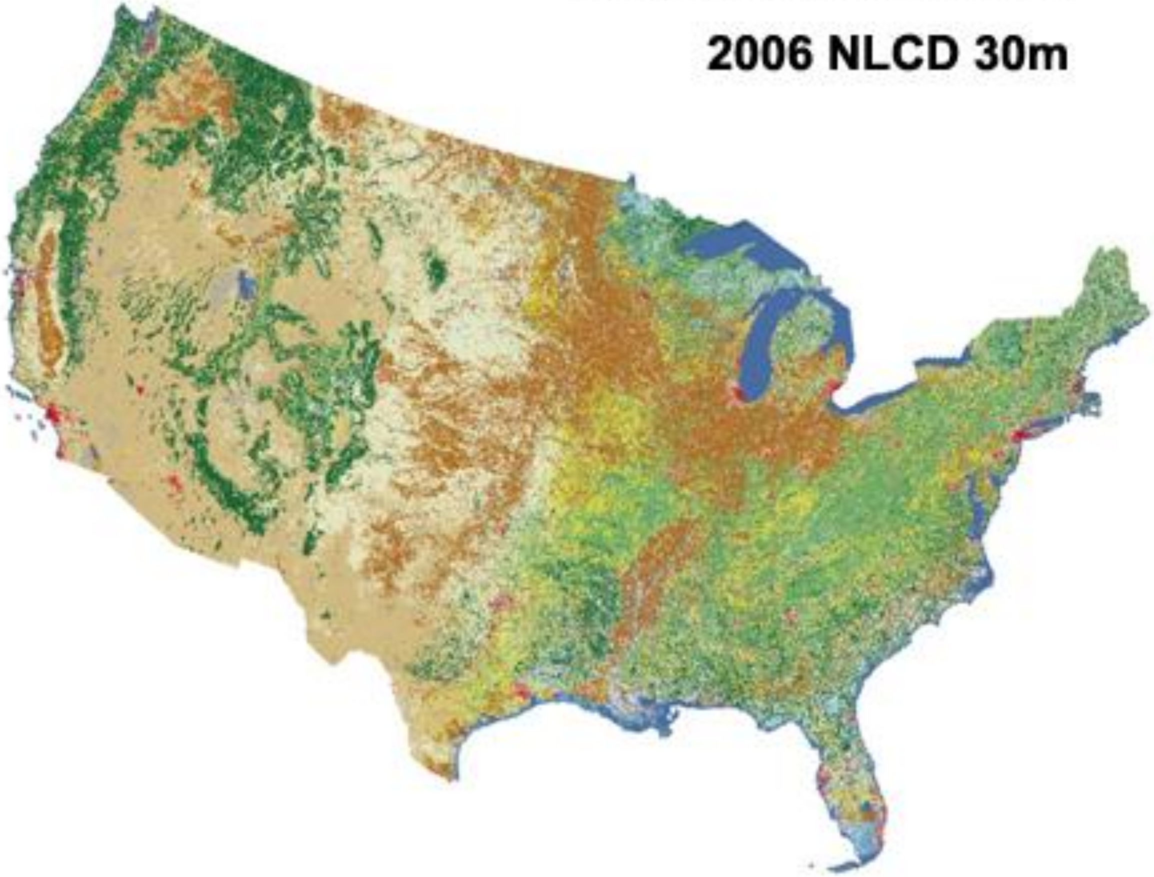
Wa

sualization
Analysis

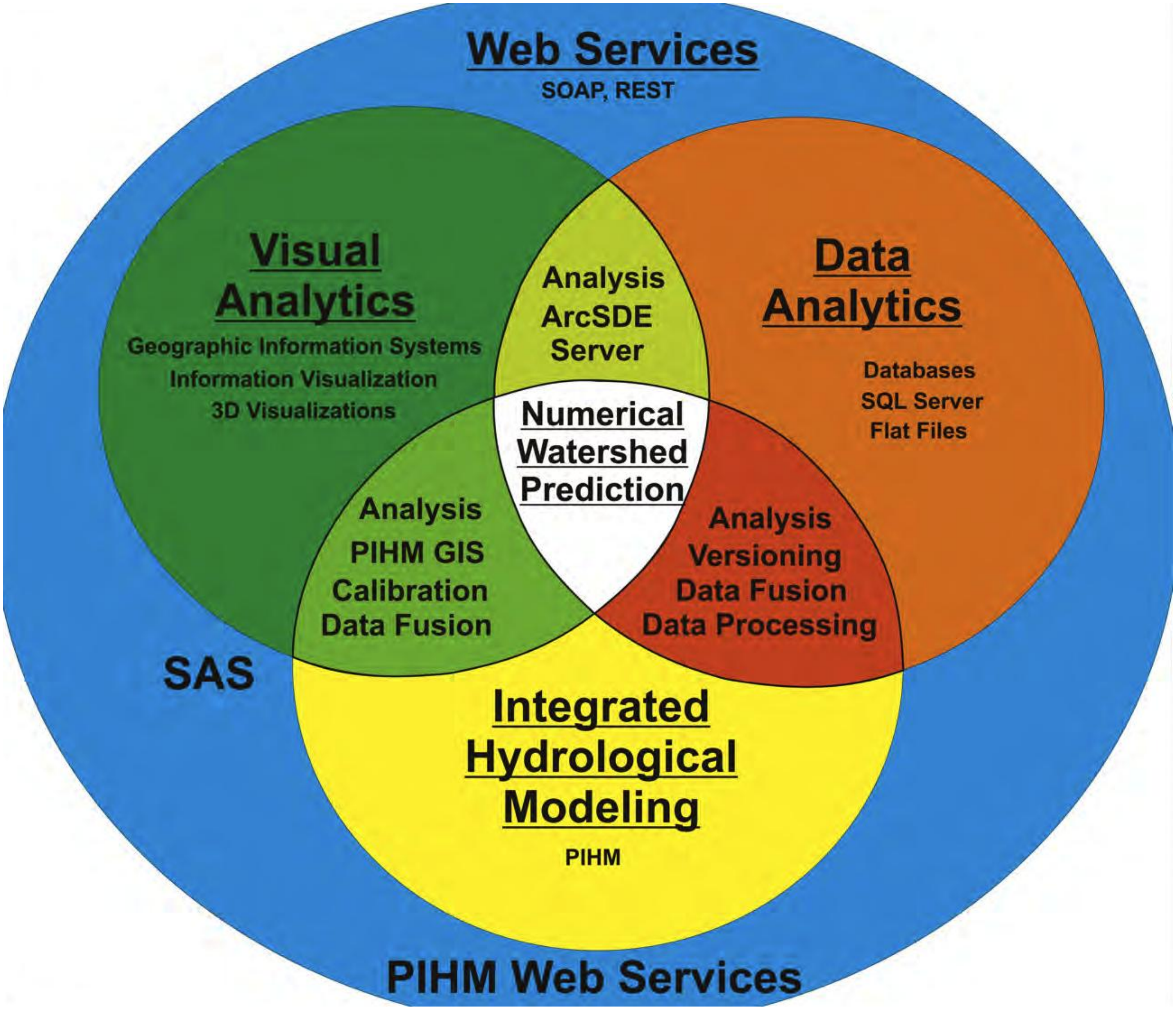
del
ation

14 GB Land cover File

2006 NLCD 30m



http://www.epa.gov/mrlc/nlcd-2006.html



Web Services

SOAP, REST

Visual Analytics

Geographic Information Systems

Information Visualization

3D Visualizations

Analysis

ArcSDE

Server

Data Analytics

Databases

SQL Server

Flat Files

Numerical Watershed Prediction

Analysis

PIHM GIS

Calibration

Data Fusion

Analysis

Versioning

Data Fusion

Data Processing

SAS

Integrated Hydrological Modeling

PIHM

PIHM Web Services

Cloud Everything As a Service

SaaS: Software as a Service

Web Services: Visual & Data Analytics, GIS API's, Modeling PIHM/SWAT/ETC.

PaaS: Platform as a Service

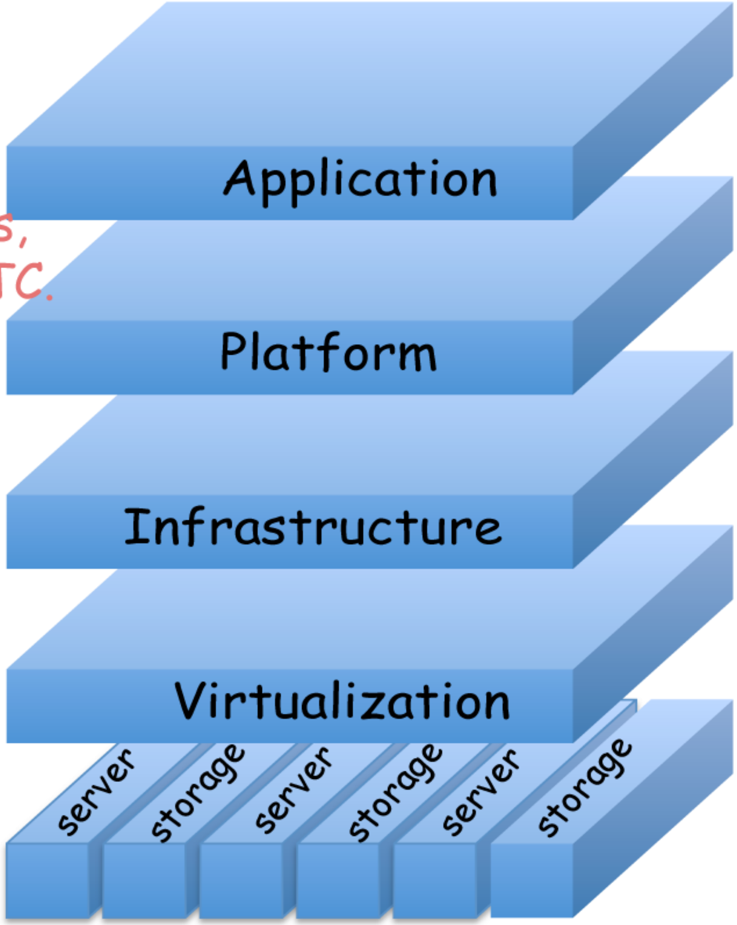
Linux, Windows, other

IaaS: Infrastructure as a Service

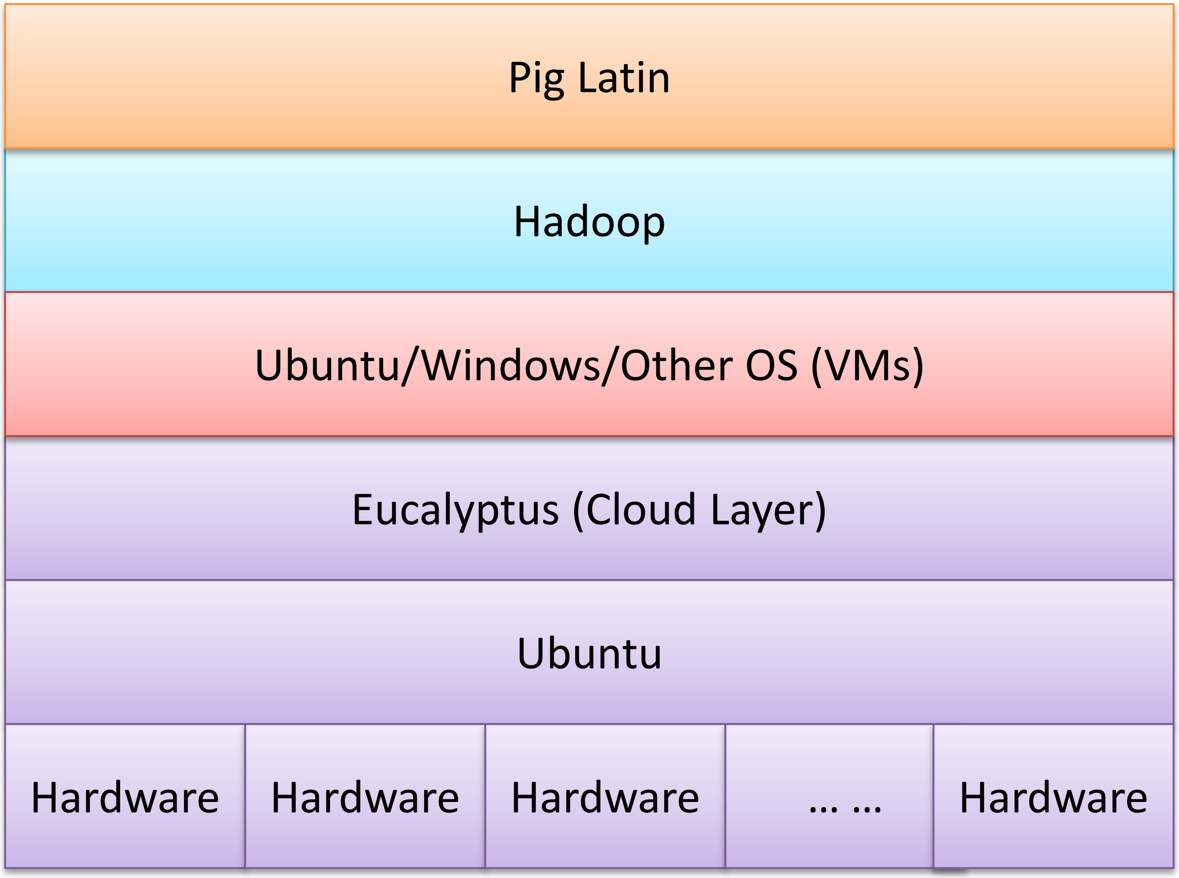
Eucalyptus Cloud layer

Virtual Machines

Hardware : cpu, disk, Memory & Bandwidth



Penn State Private Cloud Prototype



Pig Latin

Hadoop

Ubuntu/Windows/Other OS (VMs)

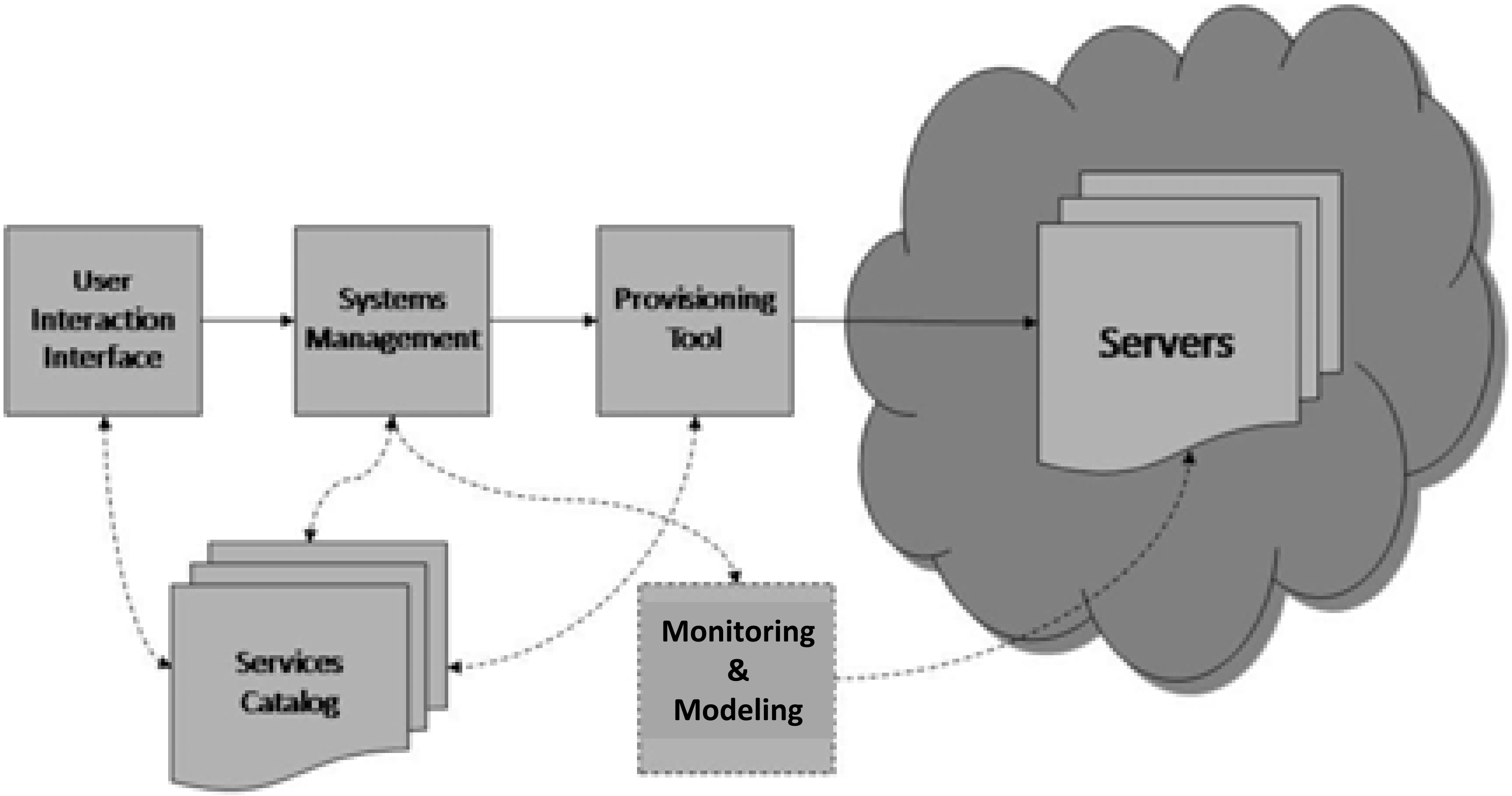
Eucalyptus (Cloud Layer)

Ubuntu

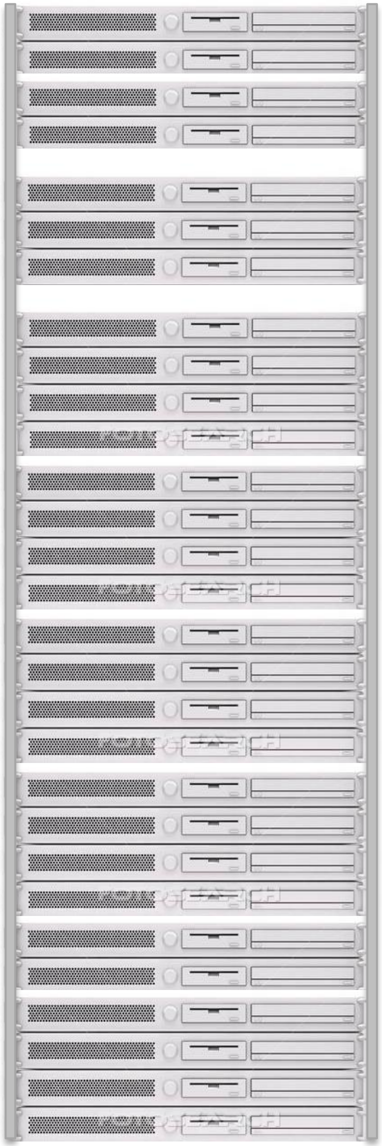
Hardware | Hardware | Hardware | … … | Hardware

D. Wu, S. Huang

Cloud Services Prototype



Micro-Cloud Hardware



2U OSX 10 TB Data Server ← CZO Server

2U OSX 10TB Disk Server ← CZO Server

3U 32 Core 20TB Server ← Cloud Services Layer

4U Linux Compute Server for PIHM ← PIHM Servers in Cloud

4U LINUX Compute Server for PIHM ← PIHM Servers in Cloud

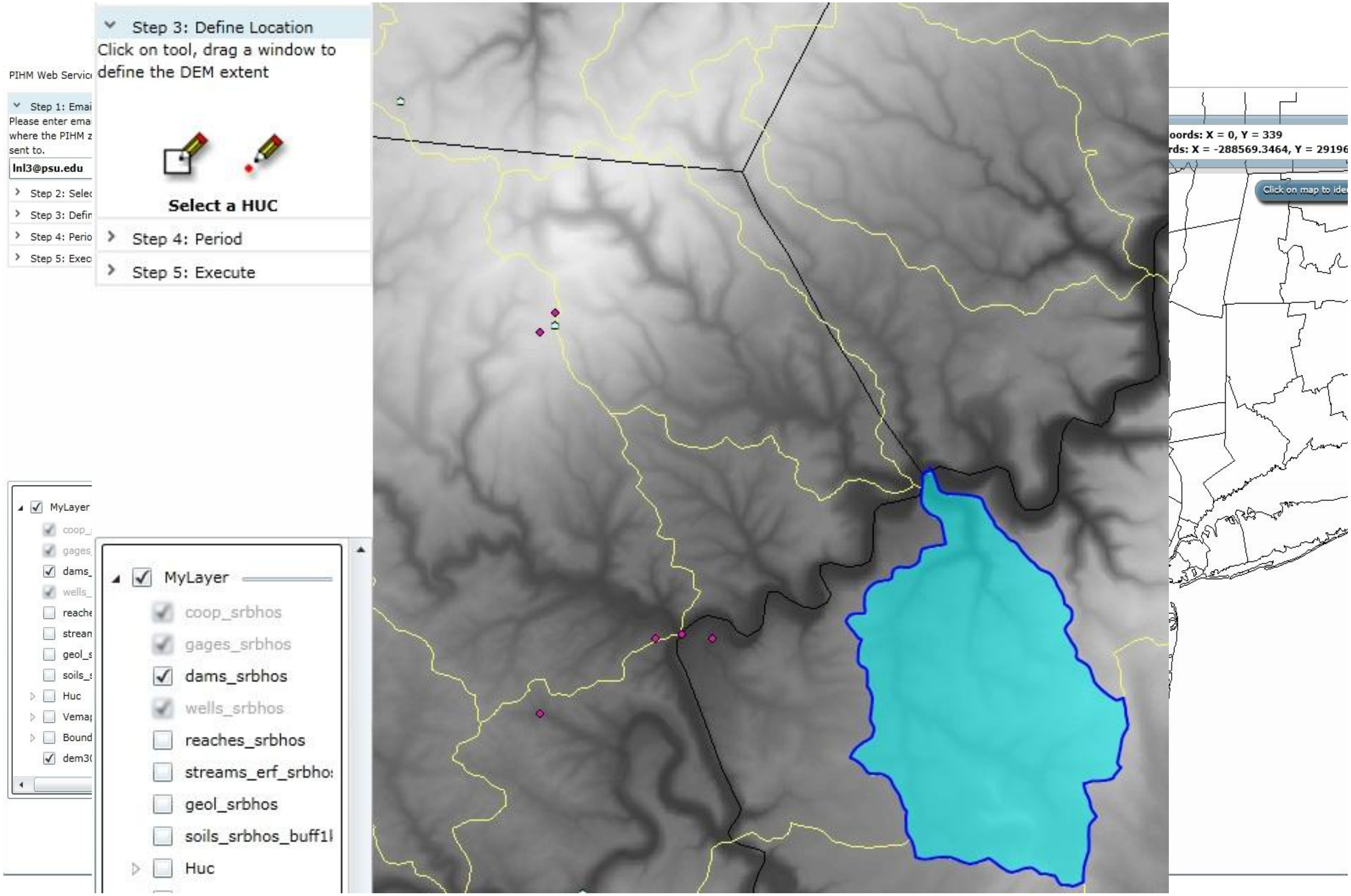
4U Linux Compute Server for PIHM ← PIHM Servers in Cloud

4U Windows Server 20 TB ← GIS Fail-Over

2U Web Server 20 TB ← Web App’s to outside world

4U Data Server 30TB ← Web App’s use this server

Web Services for data access & rapid Model prototyping



PIHM Web Servic

Step 1: Emai

Please enter emai where the PIHM z sent to.

lnl3@psu.edu

Step 2: Selec

Step 3: Defin

Step 4: Perio

Step 5: Exec

Step 3: Define Location

Click on tool, drag a window to define the DEM extent

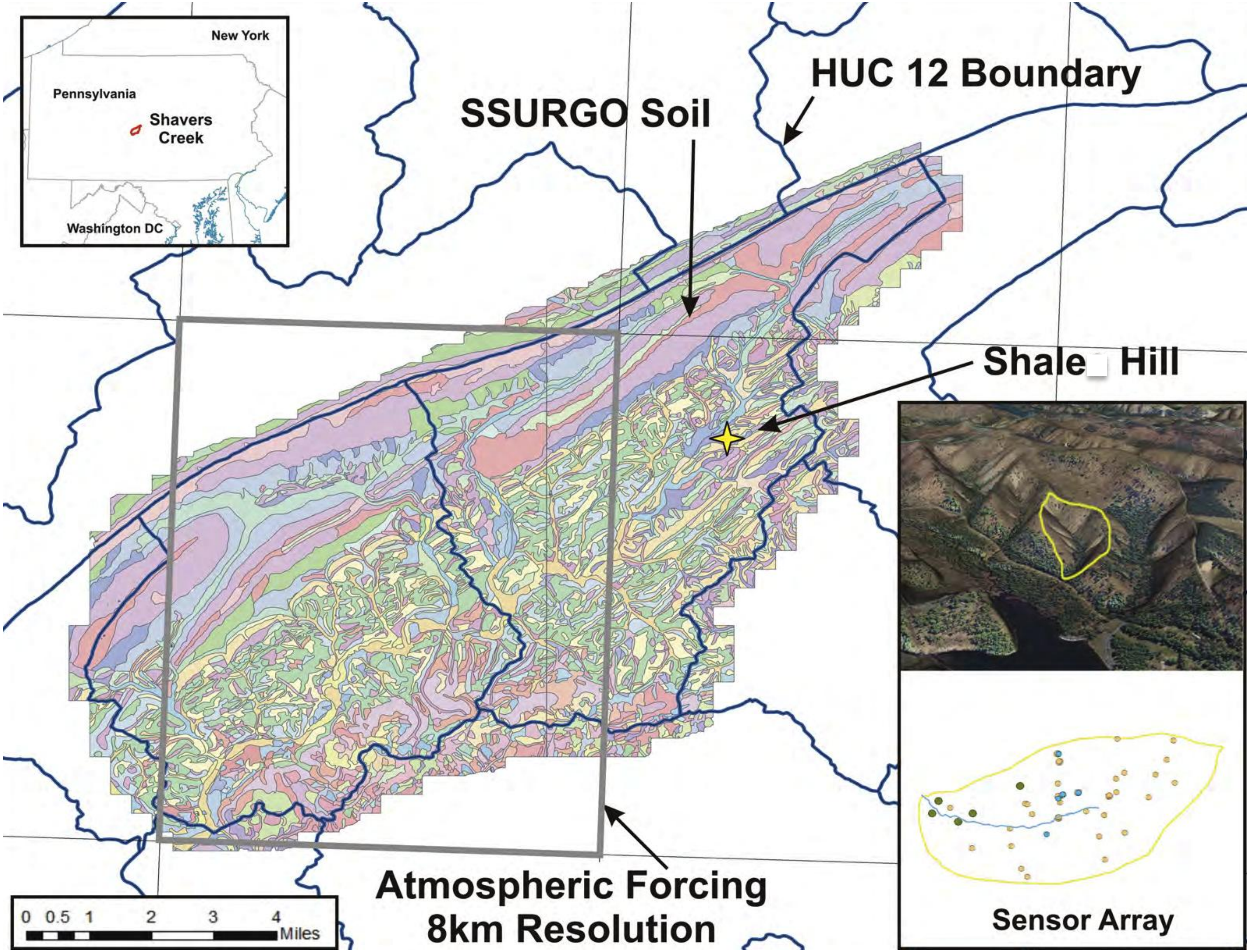
Select a HUC

Step 4: Period

Step 5: Execute

oords: X = 0, Y = 339

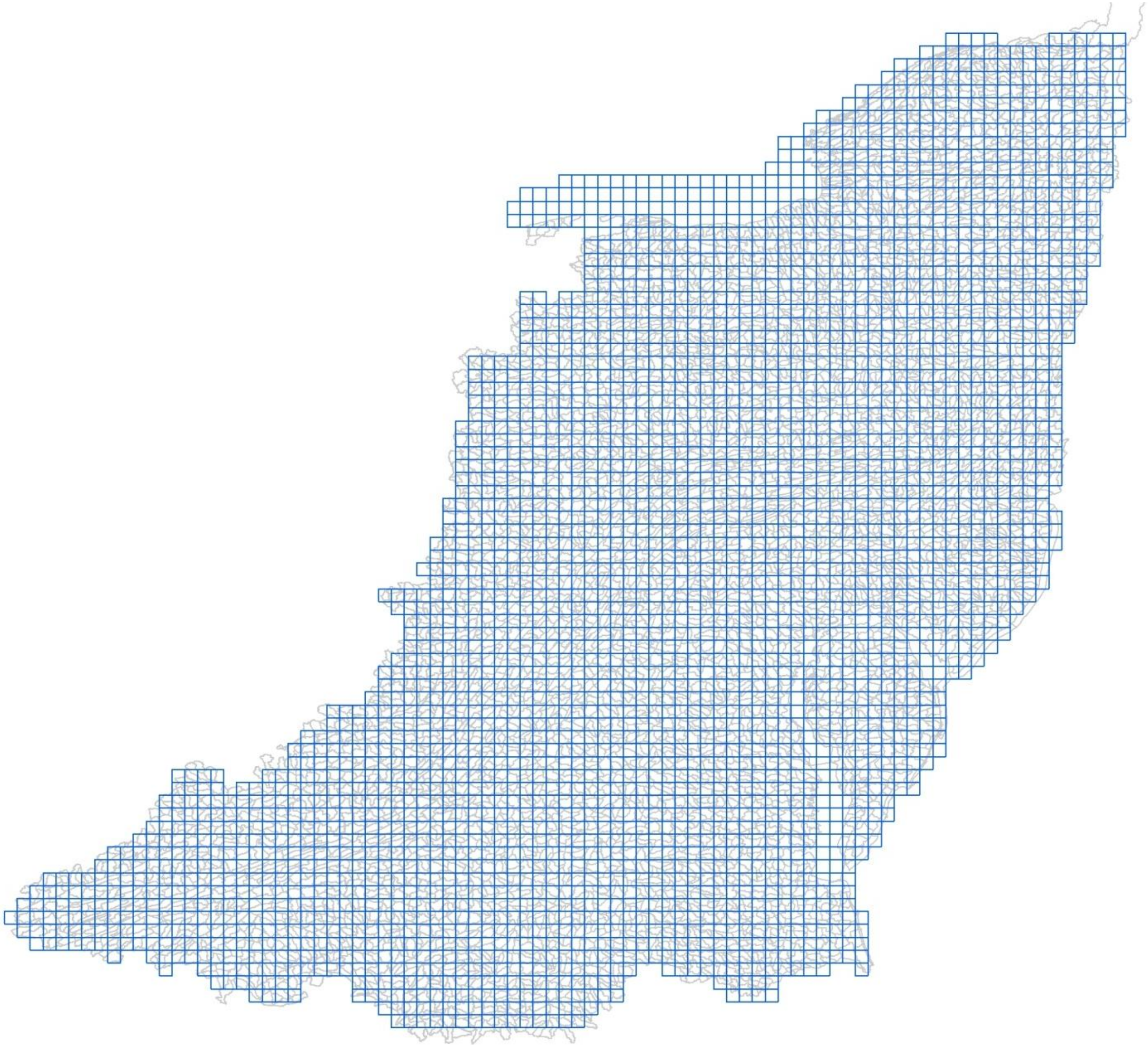
rds: X = -288569.3464, Y = 29196


New York
Pennsylvania
Shavers Creek
Washington DC
HUC 12 Boundary
SSURGO Soil
Shale Hill
Sensor Array
Atmospheric Forcing
8km Resolution
0 0.5 1 2 3 4
Miles

Upscaling Model-Data-Process: Chesapeake Bay

- Groundwater-Stream-Land-Surface model
- Fully coupled surface water, soil water, groundwater, and land surface components
- C-N-Sediments
- Vegetation Growth
- Environmental Tracers
- PIHM_HPC

Data :: Climate :: NLDAS II (1979 – present)



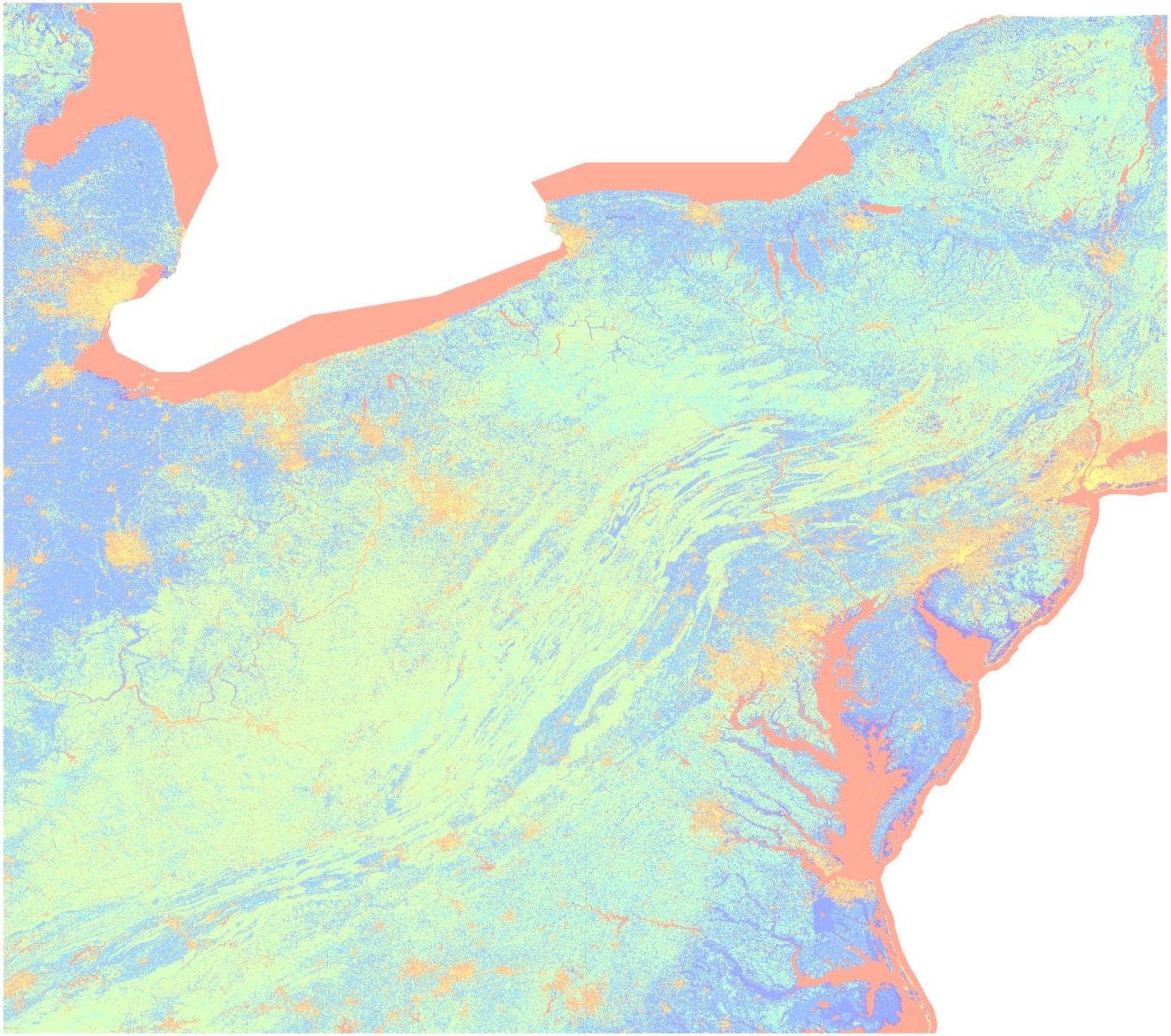
8km **Hourly** time-series

1. Precipitation
2. Temperature
3. Solar Radiation
4. Vapor Pressure
5. Relative Humidity
6. Wind Speed

We have developed tools that extracts forcing variables (from NLDAS-2 grib2 data) and formats all above mentioned variables according to PIHM data structure

~**3300** Climate Grids or Time-series Data for each climate variables!

Data :: Land Cover :: NLCD 2006 + Veg Parameters

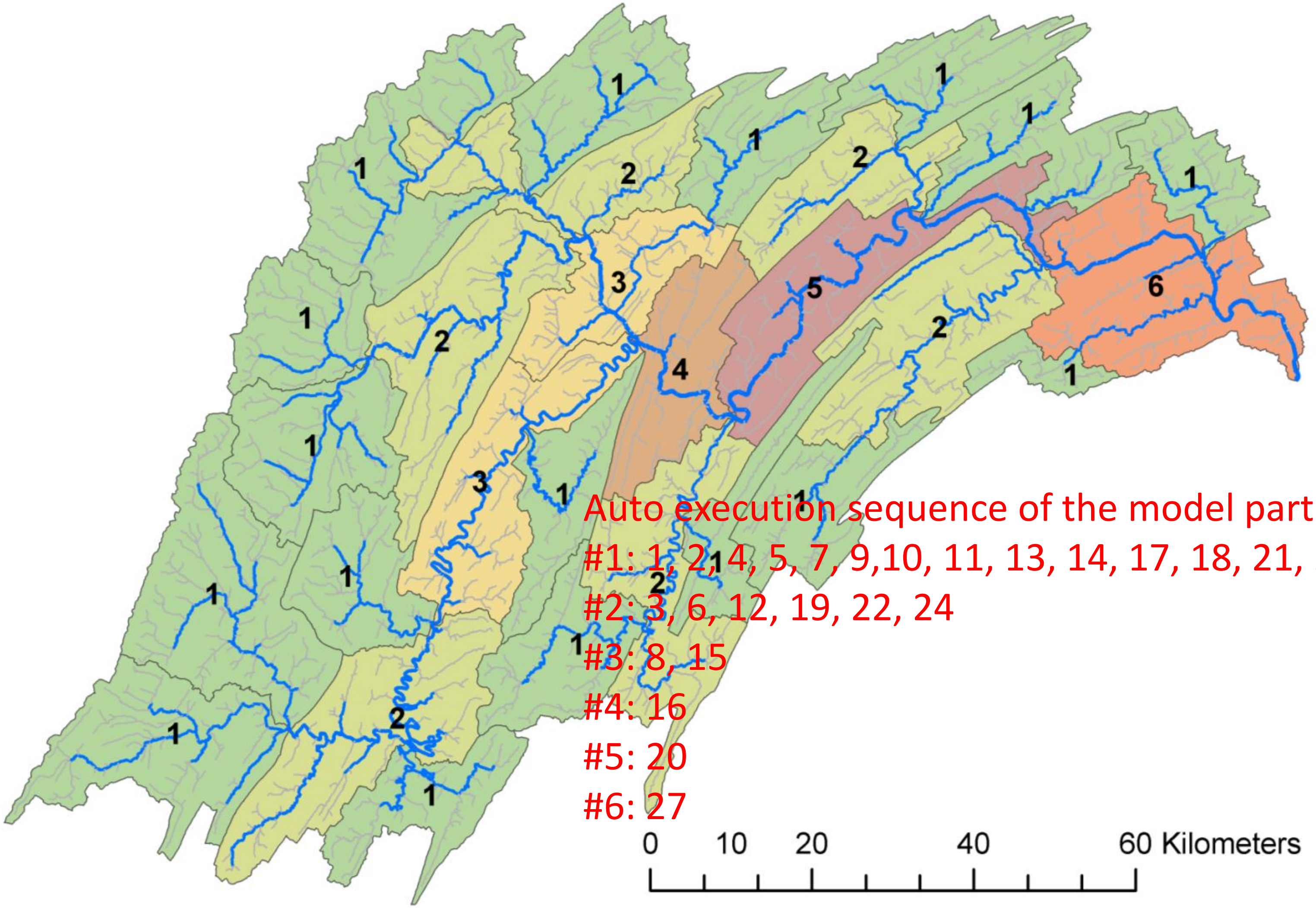


Vegetation Parameters

1. Leaf Area Index (TS)
2. Roughness Length (TS)
3. Min. Stomata Resist.
4. Albedo
5. Vegetation Fraction
6. Rooting Depth

Vegetation parameters are mapped to each LC-type using custom routines

HPC: Juniata River Domain Partitioning



Auto execution sequence of the model partitions would be:

#1: 1, 2, 4, 5, 7, 9,10, 11, 13, 14, 17, 18, 21, 23, 25, 26 28

#2: 3, 6, 12, 19, 22, 24

#3: 8, 15

#4: 16

#5: 20

#6: 27

Headwaters >> Downstream :: 6 Step Execution

2011 Seventh IEEE International Conference on e-Science Workshops

Watershed Reanalysis

Towards a National Strategy for Model-Data Integration

Christopher Duffy, Lorne Leonard, Gopal Bhatt,
Xuan Yu
Dept. of Civil & Environmental Engineering
Pennsylvania State University
University Park, PA, USA
cxd11@psu.edu

Lee Giles
College of Information Science and Technology
Pennsylvania State University
University Park, PA, USA
giles@ist.psu.edu

Abstract— **Reanalysis or retrospective analysis is the process of re-analyzing and assimilating climate and weather observations with the current modeling context. Reanalysis is an objective, quantitative method of synthesizing all sources of information (historical and real-time observations) within a unified framework. In this context, we propose a prototype for automated and virtualized web services software using national data products for climate reanalysis, soils, geology, terrain and land cover for the purpose of water resource simulation, prediction, data assimilation, calibration and archival. The prototype for model-data integration focuses on creating tools for fast data storage from selected national databases, as well as the computational resources necessary for a dynamic, distributed watershed prediction anywhere in the continental US. In the future implementation of virtualized services will benefit from the development of a cloud cyber infrastructure as the prototype evolves to data and model intensive computation for continental scale water resource predictions.**

Keywords-component; Climate Reanalysis, Data Analytics, Distributed Hydrologic Model, Numerical Watershed Prediction (NWP), PIHM, Software as a Service (SaaS), Visual Analytics, Web Services

I. INTRODUCTION

"*The challenge of water scarcity (or flooding) has to be confronted watershed by watershed according to the local physical and political conditions...*" (Stephen Solomon, Water and Civilization). There is a clear national need to provide researchers, educators, resource managers and the general public seamless and fast access to essential geo-spatial/geo-temporal data, physics-based numerical models, and data fusion tools that are necessary to understand, predict and manage the nations surface and groundwater resources. The evaluation of ecosystem and watershed services such as the detection and attribution of the impact of climatic change on floods and drought provides one of many examples of the pressing need for high resolution, spatially explicit resource assessments. At the present time, there is no unified cyber infrastructure for supporting watershed models, and the data resource itself (weather/climate reanalysis products, stream flow, groundwater, soils, land cover, satellite data products, etc.) resides on many federal servers with limited access. It is clear that fast and efficient access to the data during model development, analysis and simulation remains the challenge. It is also important to state that computation for terrestrial watershed modeling is a data-intensive process, requiring extensive data libraries for climate, soils, geology land-use and land-cover, etc. Once acquired, each of these data sources must be processed before it is useful for constructing the physical watershed model and successive versions of the data and model are desirable. Model-data versioning would also include retrospective simulations, real time forecasting as well as future scenarios for climate and landuse change simulations [7].

Predicting the spatial and temporal distribution of water on complex landscapes begins with a multi-physics model for water and energy that couples surface and subsurface flows, with a community model for land surface moisture and energy fluxes [12]. In our research we have designed and developed the Penn State Integrated Hydrologic Model (PIHM). The hydrological processes in PIHM are fully coupled on a spatially-distributed unstructured grid. The unstructured grid and domain decomposition allows the user to construct quality numerical grids that can be constrained to follow or preserve important features of the model domain (e.g. watershed boundaries, soil, geology, political boundaries, etc.). Once the model domain is formed, the process of acquiring and projecting the geo-spatial and geo-temporal data on the model grid is perhaps the most time consuming and difficult process in model development. The web-based strategy described here focuses on implementation of the cyber infrastructure and workflow facilitating model prototyping through rapid data access, model input generation, model-data archival and versioning, and visualization of the results. In principal, the strategy discussed here would enable World Wide Web users to have seamless access to all necessary data products and to be able to carry out a simulation from archived data for any watershed or HUC (Hydrologic Unit Code) in the United States. Figure 1 illustrates the 1248 HUC-8 product for the USA as an example.

II. SCALE OF COMPUTATION

Beyond the problem of access to national data, the scale of computation for both data processing and model computational represents a major hurdle. This predicament is especially true since our application promotes a national approach to watershed prediction. All data resources must be processed before they are useful as parameters, inputs, boundary and

Research funded by: NSF EAR 0725019 Critical Zone Observatory, NOAA NA10OAR4310166 Role of Groundwater in land surface processes and European Commission SoilTrEC Project.

Issues, Questions, Goals

Data Issues: What & Where are the Essential Terrestrial Variables?

Simulation Framework: What scales, processes, computation resources?

The Role of Testbeds to Scaling Up to Chesapeake bay

Unique Cyberinfrastructure Needs for Catchment Data and Models?

Towards a prototype for sharing geospatial/temporal data and models

Goal: State of the Art Data and Models -> improved decision making