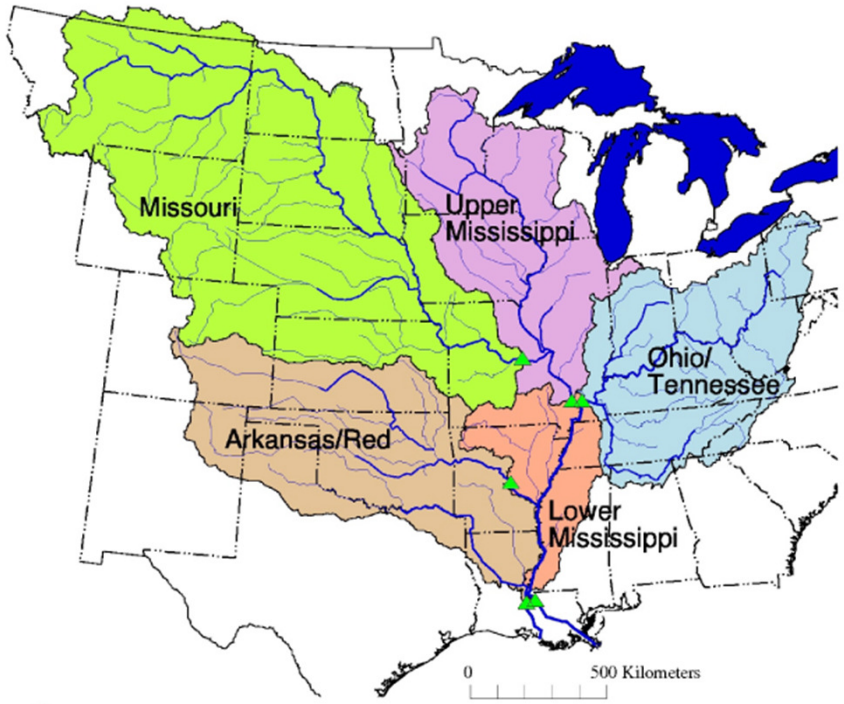
Using the Cloud to Model and Manage Large Watershed Systems

Jon Goodall
Assistant Professor
Dept of Civil and Environ. Engineering
University of South Carolina

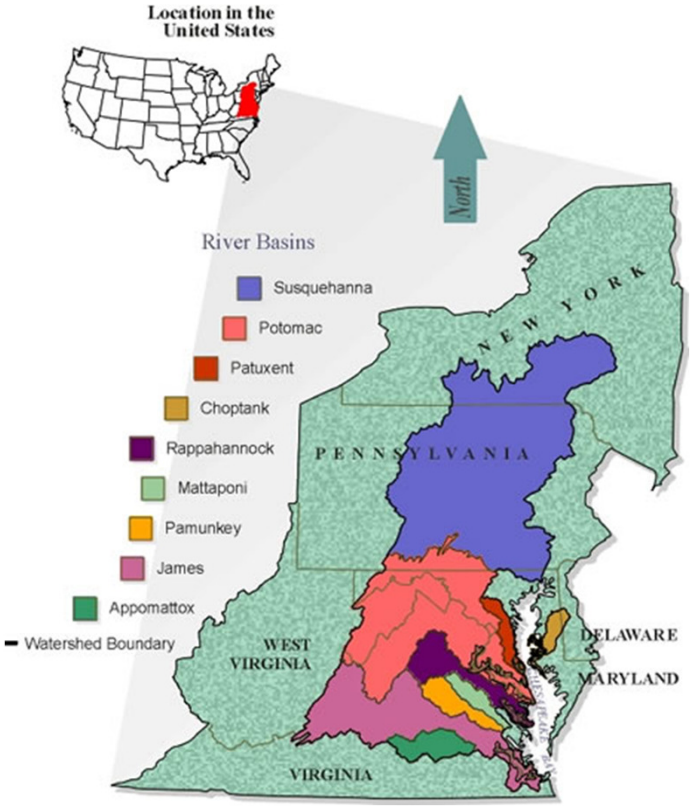
Marty Humphrey
Associate Professor
Dept of Computer Science
University of Virginia

Example Large Watersheds

Mississippi River Watershed

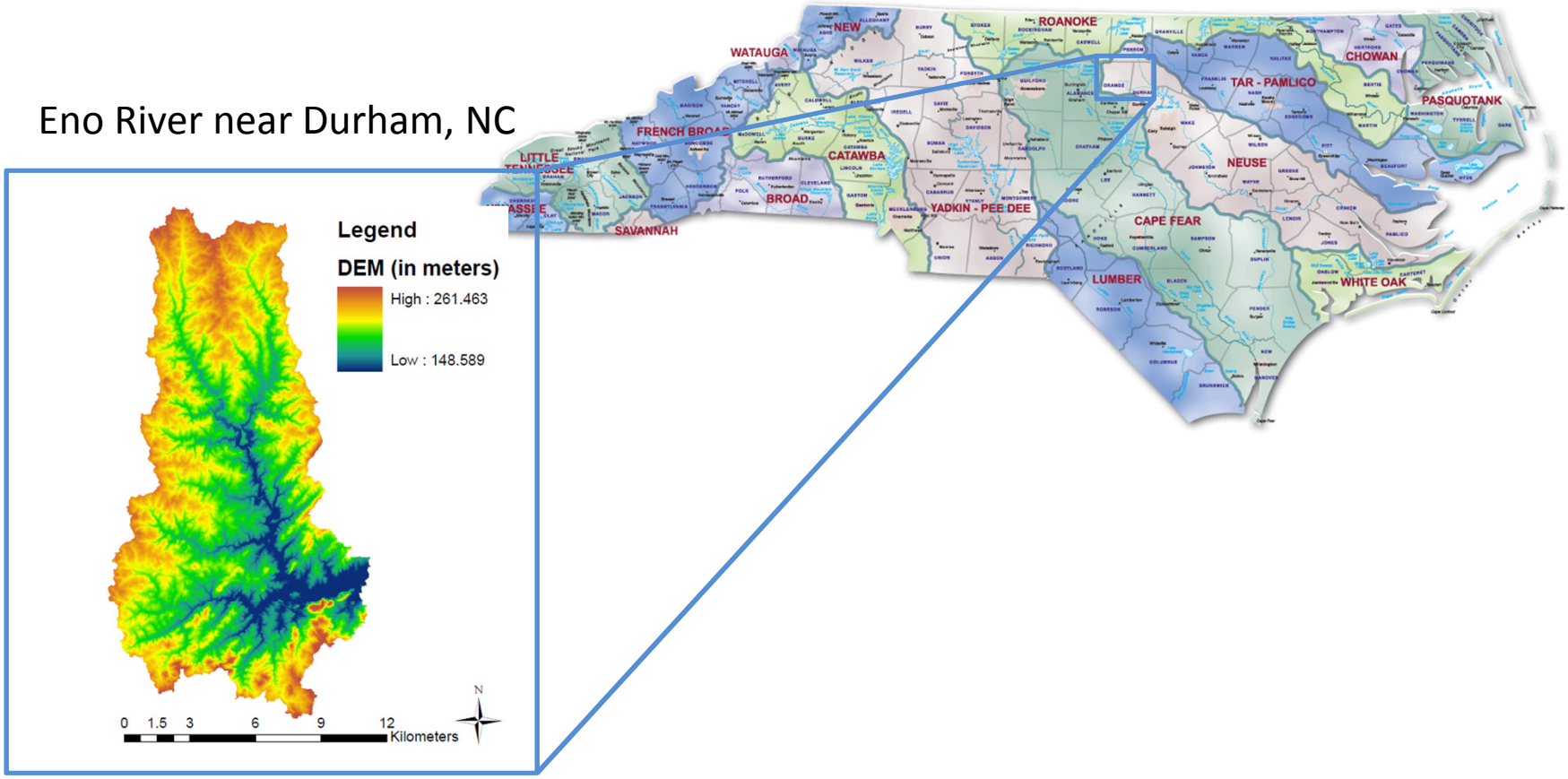


Chesapeake Bay Watershed



Typical Scale of Watershed Models

Eno River near Durham, NC



Motivation

- There are numerous scientific questions and policy challenges regarding water resources (both quantity and quality) that require understanding of large watershed systems.

- Watershed models used to answer these questions are typically limited to small systems due to computational demands.

- *Can we use the cloud to scale up our watershed models to large systems?*

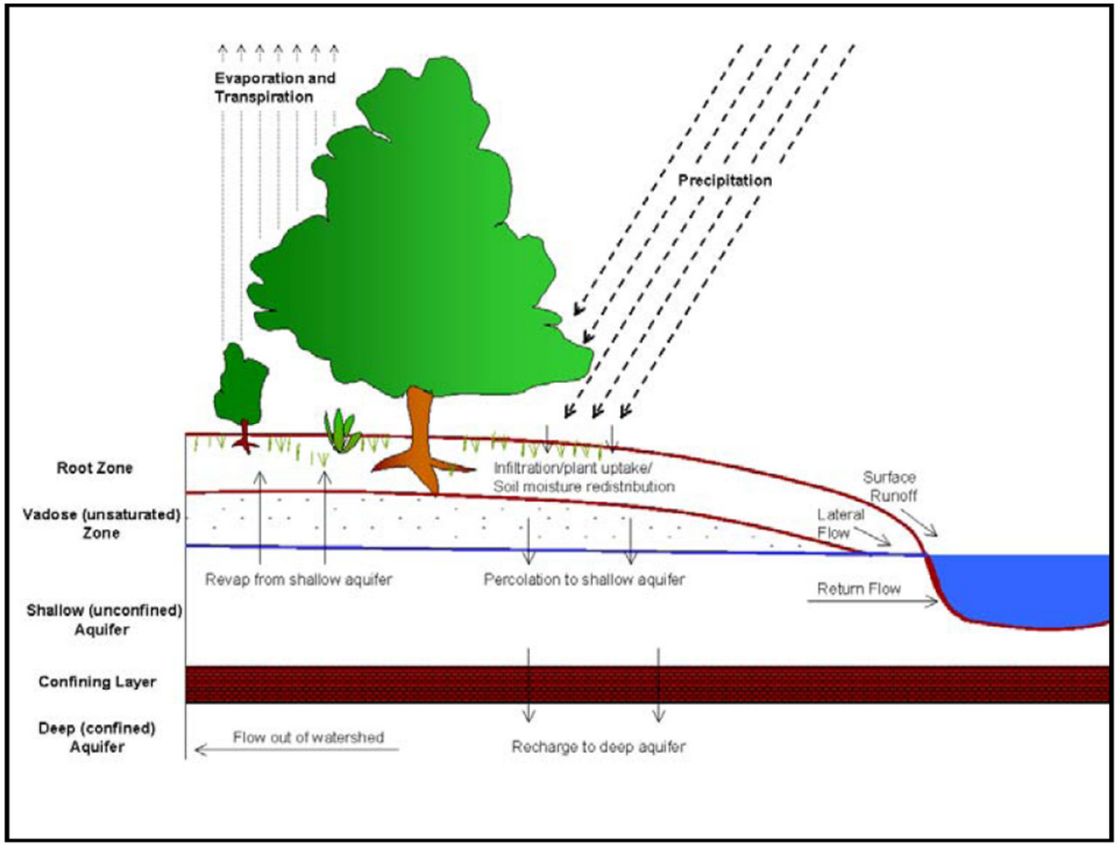
Scale

- Mississippi: 1,245,000 sq mi (3,220,000 km^2)
- Chesapeake: 64,000 sq mi (166,000 km^2)
- Eno (our case study watershed): 66 sq mi (171 km^2)

Eno to Chesapeake (~ 1,000 times)

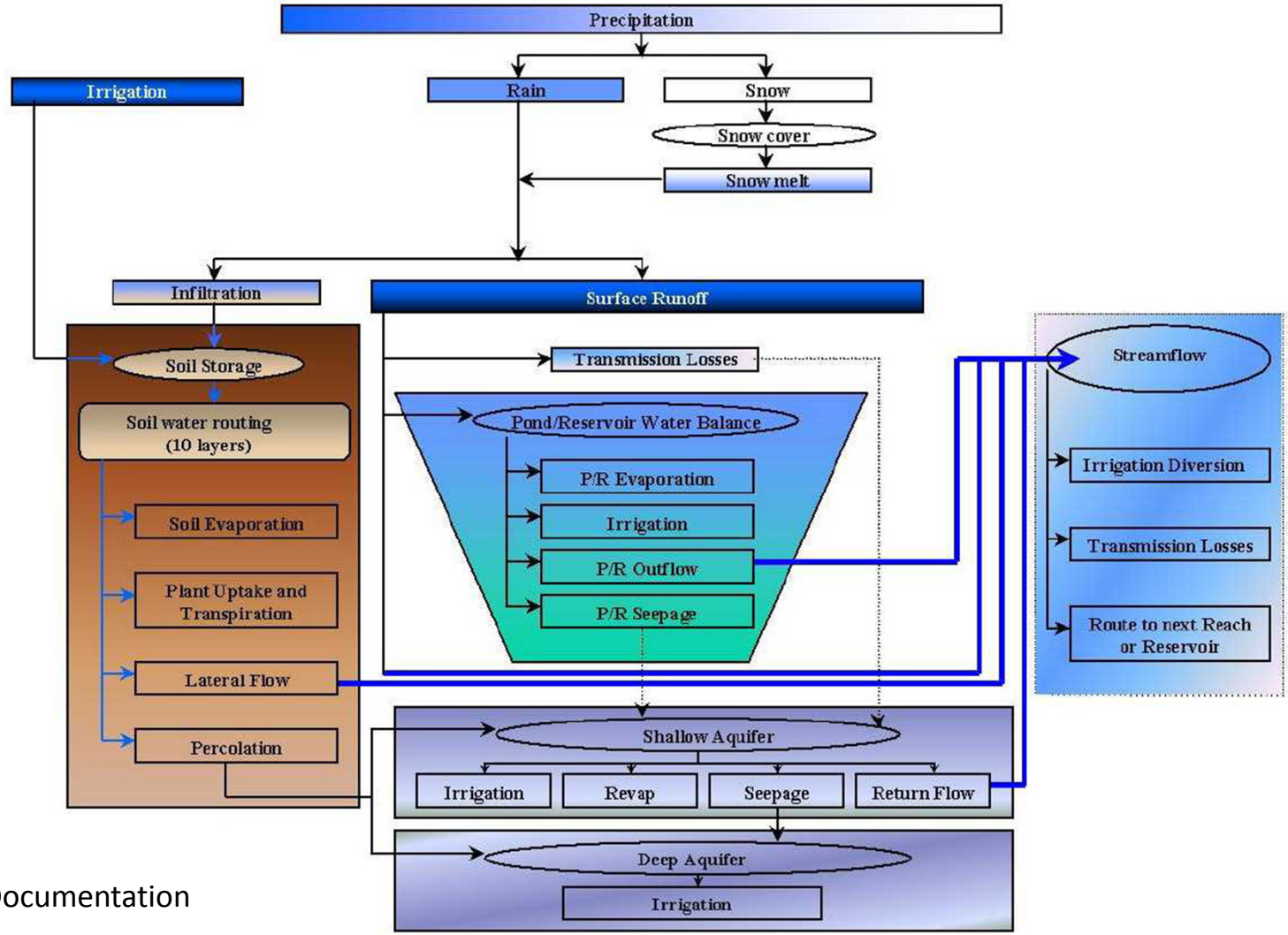
Eno to Mississippi (~ 20,000 times)

Background on Watershed Hydrology



Source: SWAT Theoretical Documentation

SWAT Model



Source: SWAT Theoretical Documentation

Challenges in Watershed Modeling

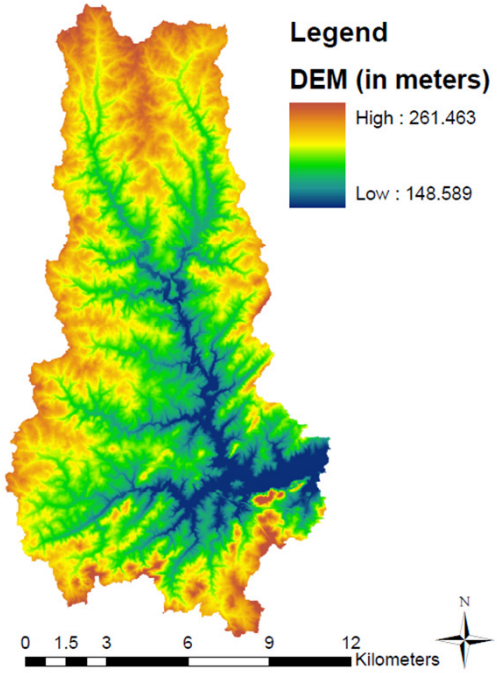
- Data Preparation
  - Data exists, but files are large and require preprocessing
- Model Calibration
  - Requires running the model multiple times with varying parameters
- Scale up Model to Large System (Chesapeake, Mississippi)
  - Impractical using current approaches

Example

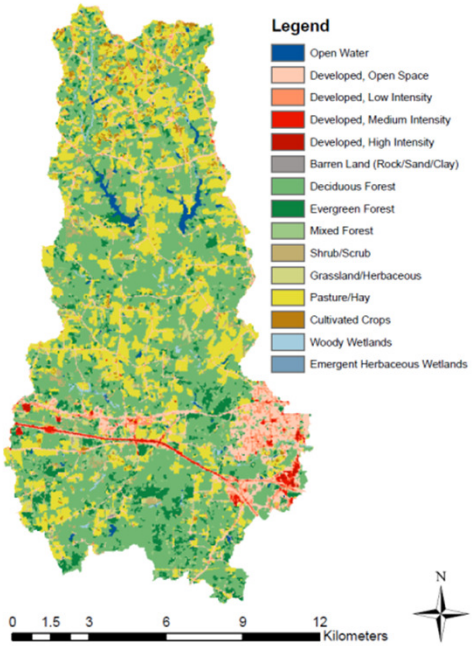
- Lucy would like to create a watershed model for the Eno River Basin for a scientific study on nutrient transport using the SWAT model

- Her first task is to perform a data “pre-processing” step where available datasets are used to create model input files.

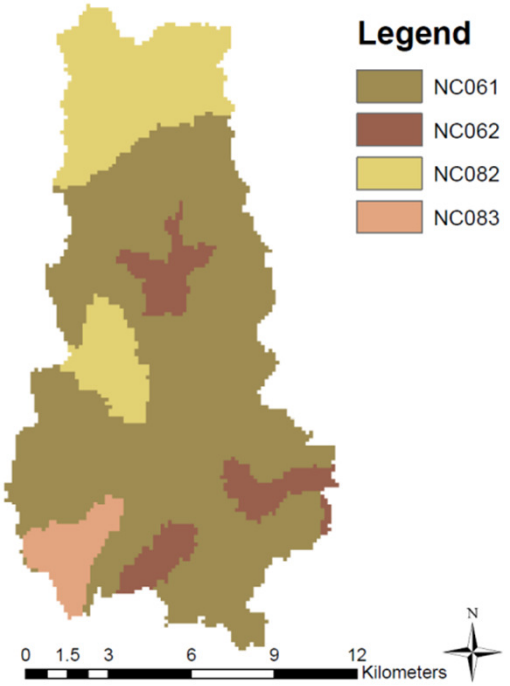
Data Preparation



Digital Elevation Models



Land Use Datasets

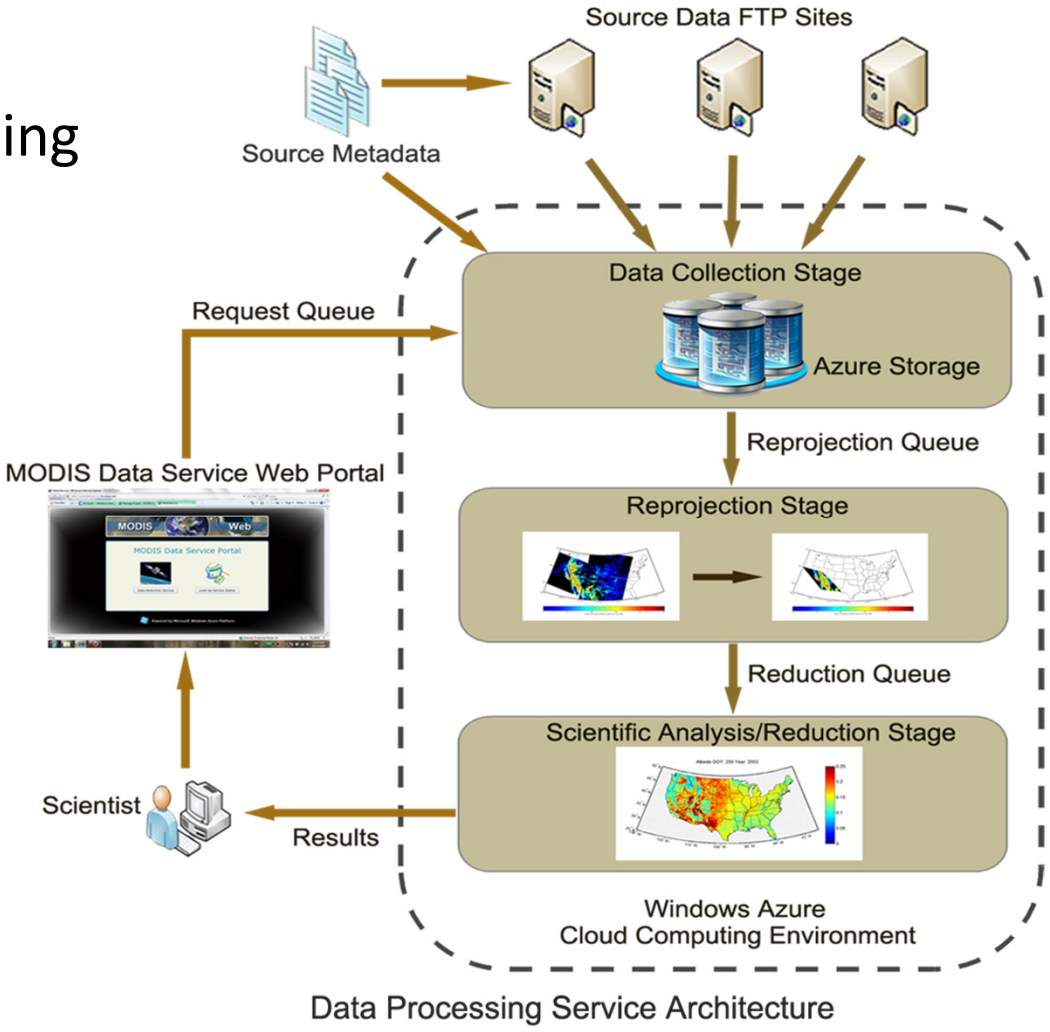


Digital Soil Surveys

Current Approach

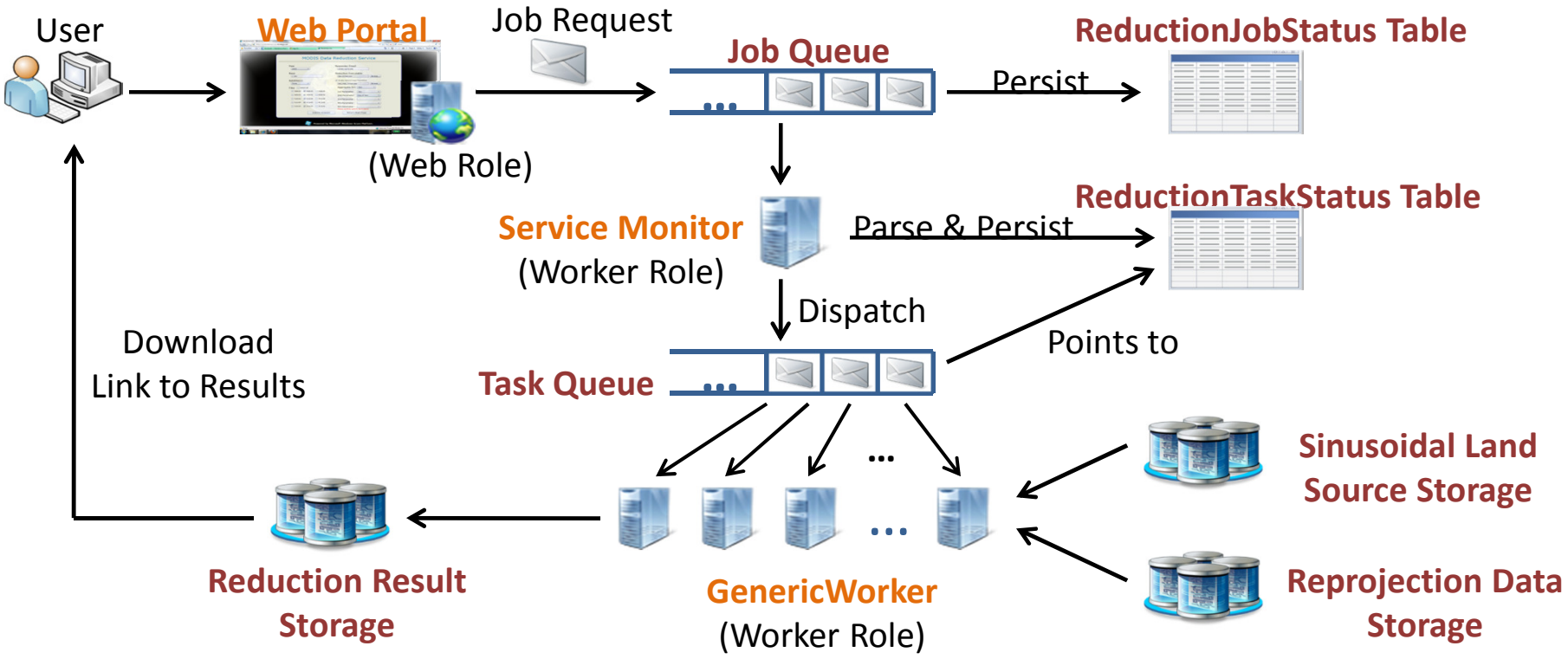
- Download, reproject, geoprocess data
- ArcSWAT is a GIS-based tool for this from USDA
- EPA BASINS is a tool for this
- Despite these tools, these steps remain tedious, time consuming, and error prone.

Proposed Approach using the Cloud



Data Processing Service Architecture

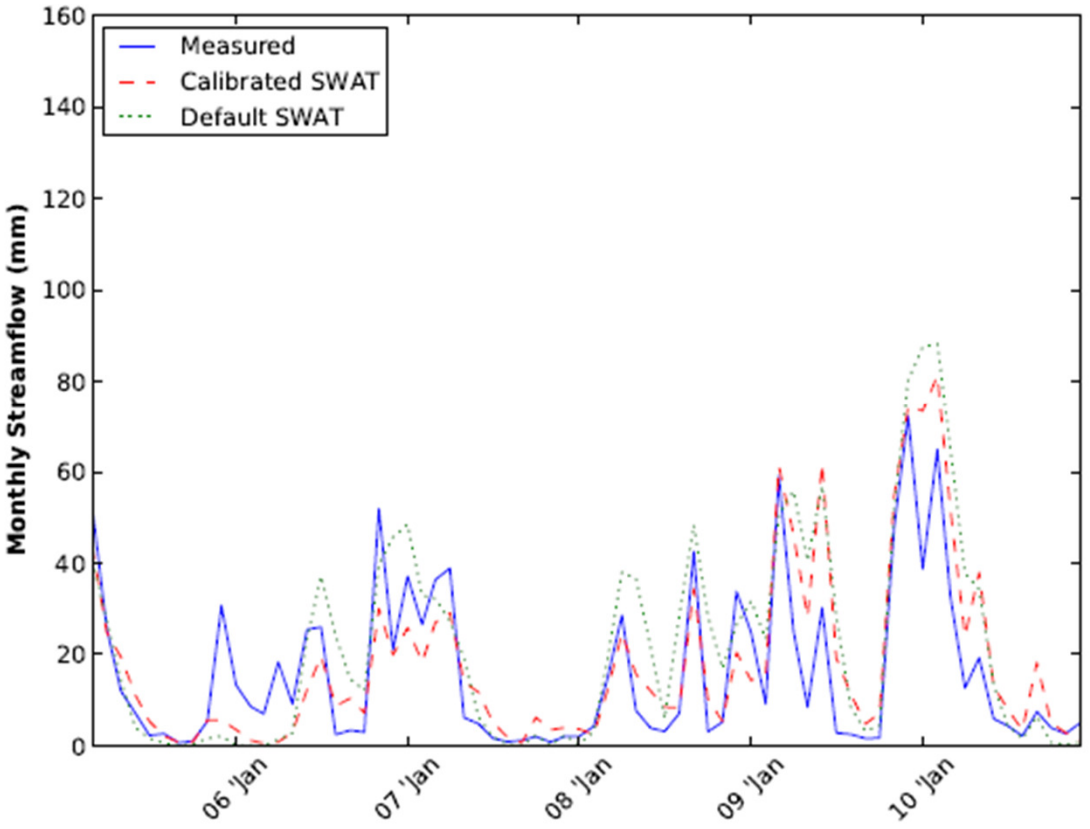
Internals of the Data Processing Service



Example (cont.)

- Lucy's next goal is to calibrate the model. She allows certain model parameters to vary within some limits in order to minimize the difference between observed and predicted streamflow.

Example Model Output

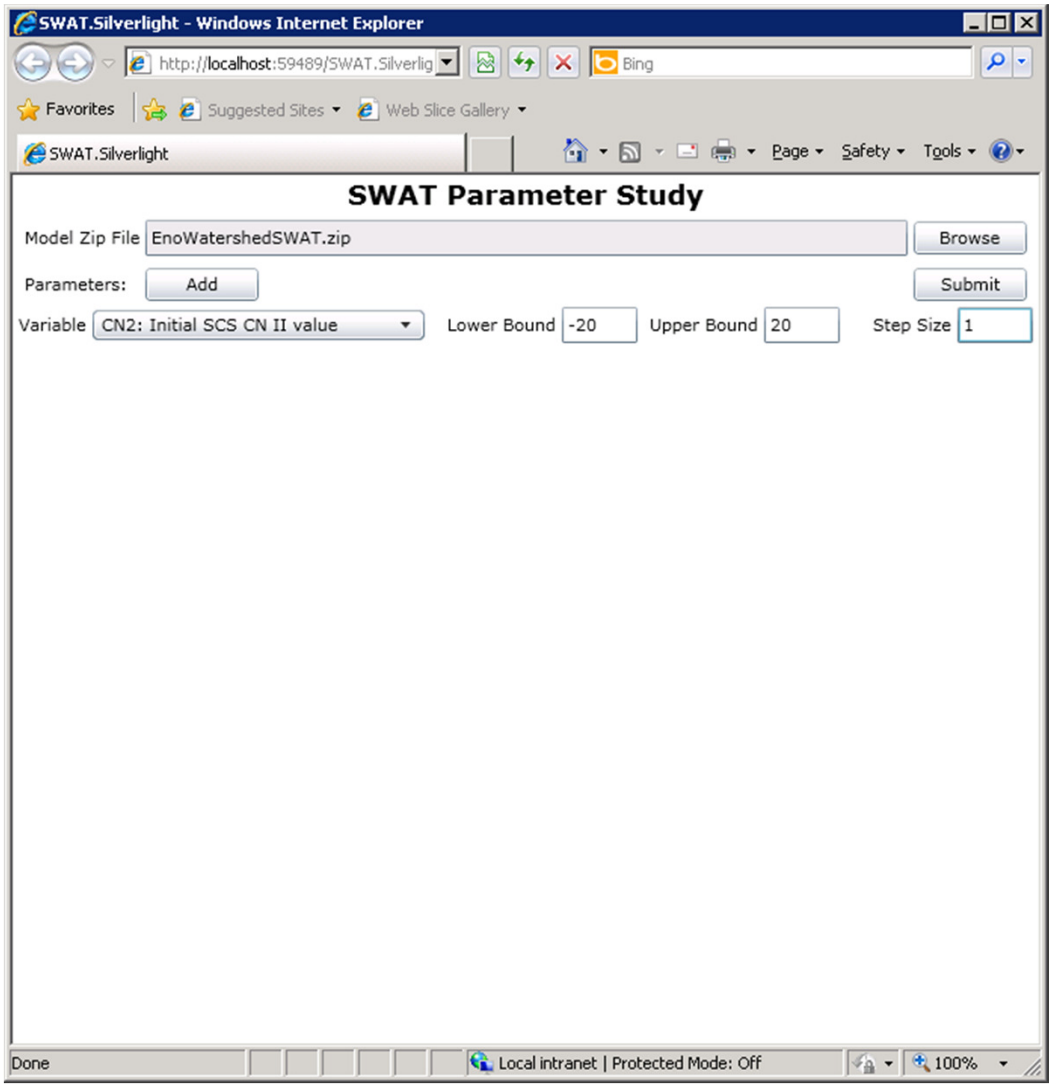


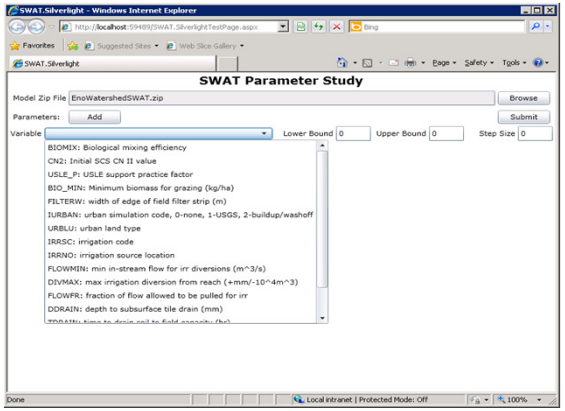
Current Approach

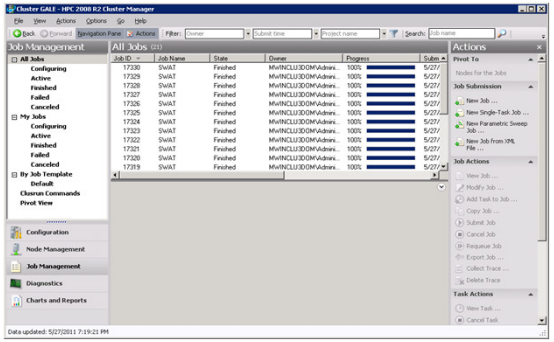
- Use calibration routines that are single threaded and run on the PC.

- Calibration takes over 1 day for the Eno watershed using current approaches

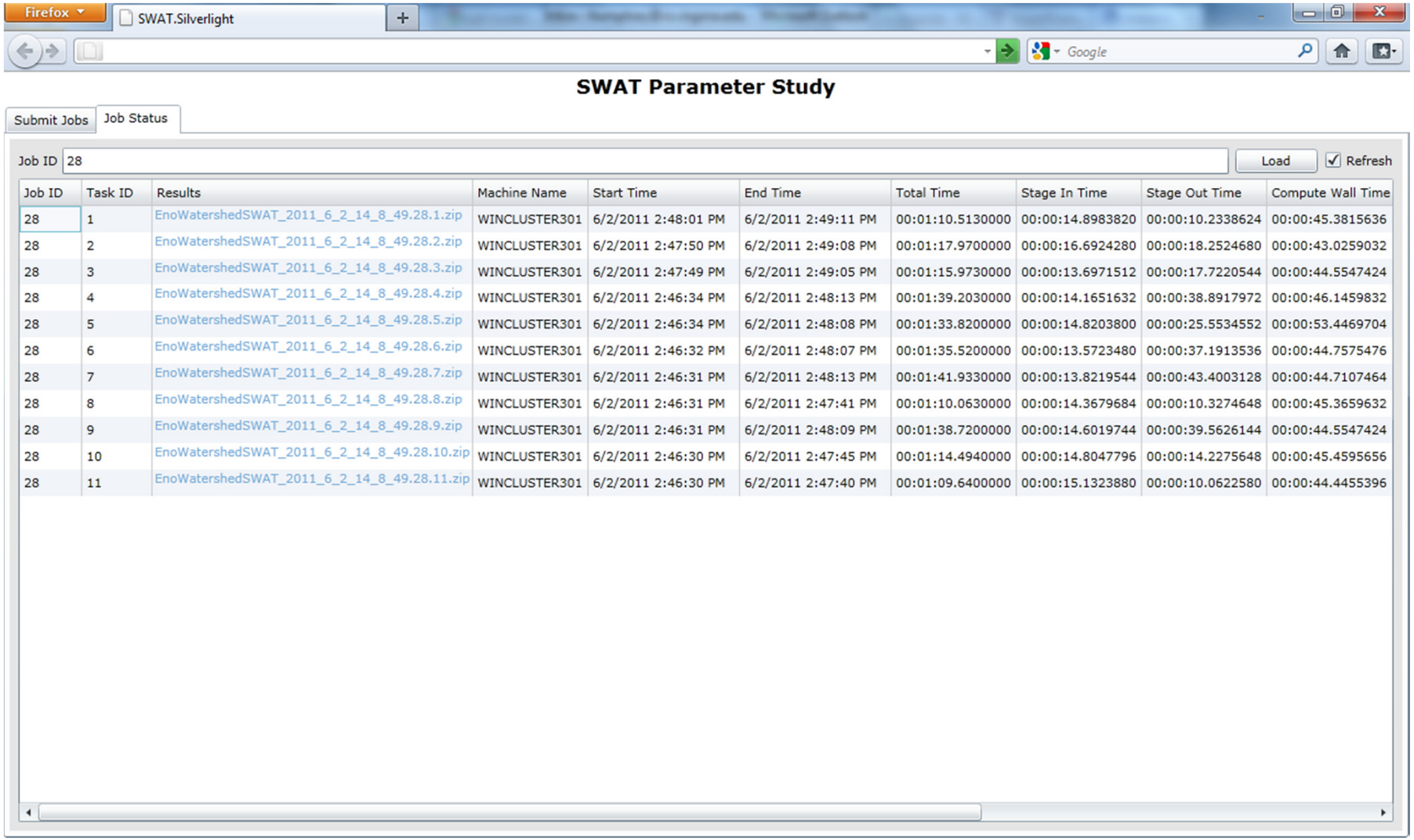
Proposed Approach using the Cloud




SWAT.Silverlight - Windows Internet Explorer
SWAT Parameter Study
Model Zip File
EnoWatershedSWAT.zip
Browse
Parameters:
Add
Submit
Variable
Lower Bound
Upper Bound
Step Size
BIOMIX: Biological mixing efficiency
CN2: Initial SCS CN II value
USLE_P: USLE support practice factor
BIO_MIN: Minimum biomass for grazing (kg/ha)
FILTERW: width of edge of field filter strip (m)
IURBAN: urban simulation code, 0-none, 1-USGS, 2-buildup/washoff
URBLU: urban land type
IRRSC: irrigation code
IRRNO: irrigation source location
FLOWMIN: min in-stream flow for irr diversions (m^3/s)
DIVMAX: max irrigation diversion from reach (+mm/-10^4m^3)
FLOWFR: fraction of flow allowed to be pulled for irr
DDRAIN: depth to subsurface tile drain (mm)


Cluster GALE - HPC 2008 R2 Cluster Manager
Job Management
All Jobs
Actions
Configuration
Node Management
Job Management
Diagnostics
Charts and Reports

HPC Cluster



Job ID	Task ID	Results	Machine Name	Start Time	End Time	Total Time	Stage In Time	Stage Out Time	Compute Wall Time
28	1	EnoWatershedSWAT_2011_6_2_14_8_49.28.1.zip	WINCLUSTER301	6/2/2011 2:48:01 PM	6/2/2011 2:49:11 PM	00:01:10.5130000	00:00:14.8983820	00:00:10.2338624	00:00:45.3815636
28	2	EnoWatershedSWAT_2011_6_2_14_8_49.28.2.zip	WINCLUSTER301	6/2/2011 2:47:50 PM	6/2/2011 2:49:08 PM	00:01:17.9700000	00:00:16.6924280	00:00:18.2524680	00:00:43.0259032
28	3	EnoWatershedSWAT_2011_6_2_14_8_49.28.3.zip	WINCLUSTER301	6/2/2011 2:47:49 PM	6/2/2011 2:49:05 PM	00:01:15.9730000	00:00:13.6971512	00:00:17.7220544	00:00:44.5547424
28	4	EnoWatershedSWAT_2011_6_2_14_8_49.28.4.zip	WINCLUSTER301	6/2/2011 2:46:34 PM	6/2/2011 2:48:13 PM	00:01:39.2030000	00:00:14.1651632	00:00:38.8917972	00:00:46.1459832
28	5	EnoWatershedSWAT_2011_6_2_14_8_49.28.5.zip	WINCLUSTER301	6/2/2011 2:46:34 PM	6/2/2011 2:48:08 PM	00:01:33.8200000	00:00:14.8203800	00:00:25.5534552	00:00:53.4469704
28	6	EnoWatershedSWAT_2011_6_2_14_8_49.28.6.zip	WINCLUSTER301	6/2/2011 2:46:32 PM	6/2/2011 2:48:07 PM	00:01:35.5200000	00:00:13.5723480	00:00:37.1913536	00:00:44.7575476
28	7	EnoWatershedSWAT_2011_6_2_14_8_49.28.7.zip	WINCLUSTER301	6/2/2011 2:46:31 PM	6/2/2011 2:48:13 PM	00:01:41.9330000	00:00:13.8219544	00:00:43.4003128	00:00:44.7107464
28	8	EnoWatershedSWAT_2011_6_2_14_8_49.28.8.zip	WINCLUSTER301	6/2/2011 2:46:31 PM	6/2/2011 2:47:41 PM	00:01:10.0630000	00:00:14.3679684	00:00:10.3274648	00:00:45.3659632
28	9	EnoWatershedSWAT_2011_6_2_14_8_49.28.9.zip	WINCLUSTER301	6/2/2011 2:46:31 PM	6/2/2011 2:48:09 PM	00:01:38.7200000	00:00:14.6019744	00:00:39.5626144	00:00:44.5547424
28	10	EnoWatershedSWAT_2011_6_2_14_8_49.28.10.zip	WINCLUSTER301	6/2/2011 2:46:30 PM	6/2/2011 2:47:45 PM	00:01:14.4940000	00:00:14.8047796	00:00:14.2275648	00:00:45.4595656
28	11	EnoWatershedSWAT_2011_6_2_14_8_49.28.11.zip	WINCLUSTER301	6/2/2011 2:46:30 PM	6/2/2011 2:47:40 PM	00:01:09.6400000	00:00:15.1323880	00:00:10.0622580	00:00:44.4455396

Early Results

	Stage-in	Compute	Total
Scientist laptop	0	55 sec	55 sec
Win2	5 sec	60 sec	65 sec
Azure (ex-large)	53 sec	32 sec	85 sec

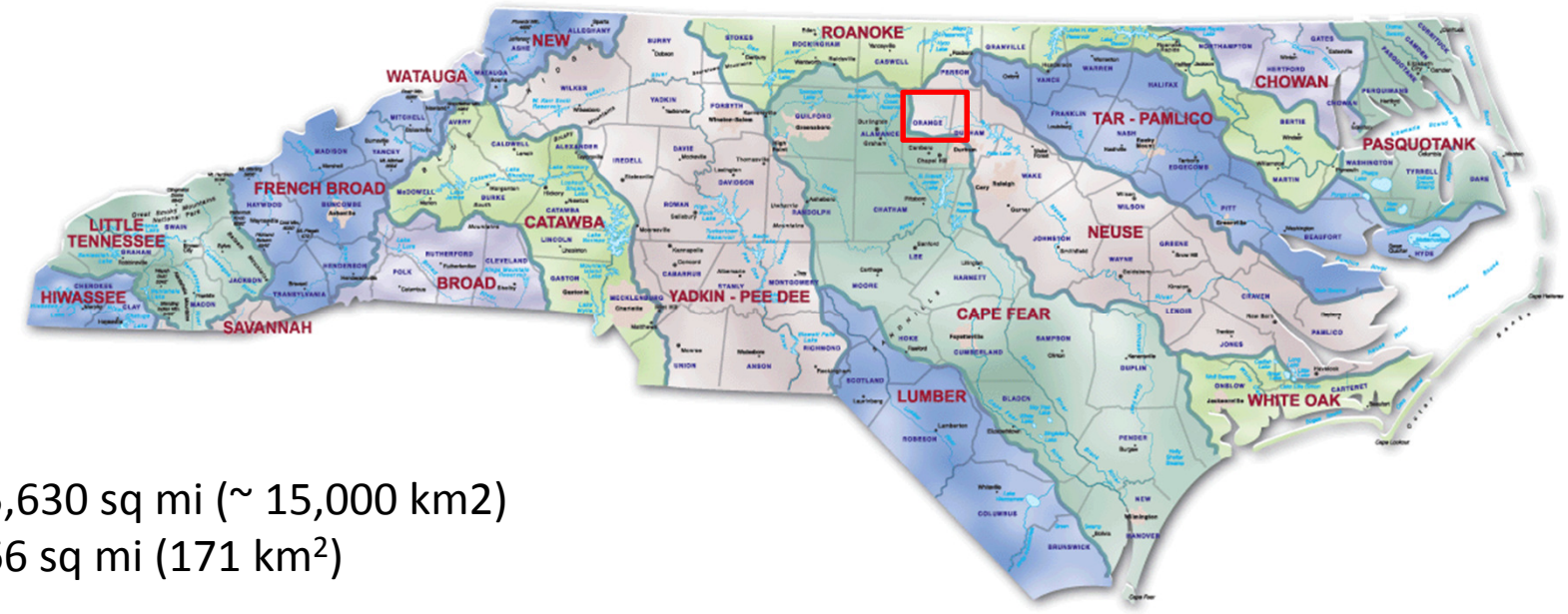
Details of the ENO SWAT Model

- Eno Watershed model includes 1,081 input and output files
- To update one of the parameters (CN2), 179 of these files will change
- Transferred to target node: 34.7 MB

Example (cont.)

- Lucy completes her study of the Eno, but now is interested in studying the larger Neuse River Basin (of which Eno is a part) to understand nutrient transport to the Tar-Pamlico Sound.

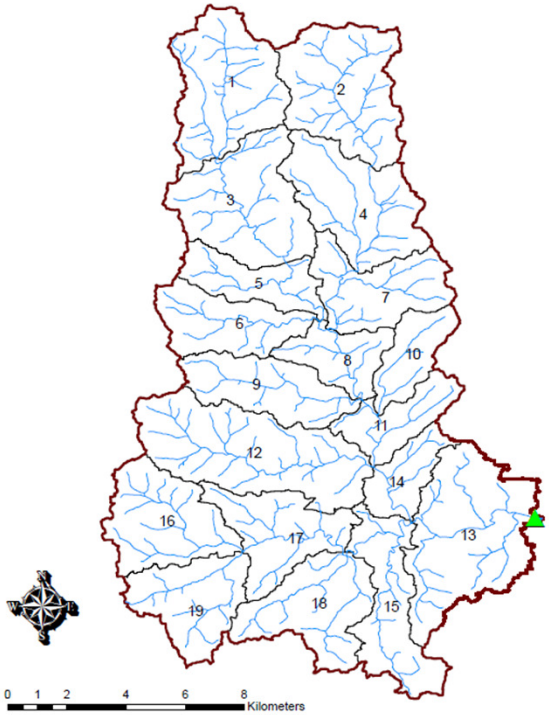
Neuse River Basin



5,630 sq mi (~ 15,000 km2)
66 sq mi (171 km^2)

Eno to Neuse = 85 times

Demonstration for Eno



- Modify model so that each subbasin is mapped to a different worker (Eno has 19 subbasins, Neuse might have 550)
- Outflow for each subbasin can then be reduced and used by a single VM to estimate streamflow
- Data preprocessing and calibration steps would be the same

Issues

- Azure only or cloudbursting?
- Data storage mechanism?
- Data sharing/reuse policy?
- Task granularity / coding?
- Task synchronization (e.g., MPI)?

Summary

- Goal: to transition compute intensive steps in watershed modeling to the Cloud
  - Data preparation and model calibration work is underway
  - Reengineering watershed model will be next step