

Utilization of GIS Technology to Support Development of Flow and Contaminant Fate and Transport Models at US DOE Sites – 15457

Angelique Lawrence, Georgio Tachiev
Applied Research Center, Florida International University

ABSTRACT

U.S. Department of Energy (DOE) sites have accumulated years of data and millions of spatial and temporal records related to the hydrological cycle, contaminant transport, and parameters of remediation technologies. In addition, numerical models which use field monitoring data to help understand flow and transport produce gigabytes of computed spatial and temporal data for each computation node. There is a need therefore for an advanced spatial data structure that can be used to address the management, processing, and analysis of spatial and temporal numerical modeling data derived from multiple sources. This can be accomplished using a geodatabase. The development of an ArcSDE geodatabase facilitates the centralized storage, backup, accessibility, organization, and management of model configuration and computed simulation data, putting it into a structured, coherent, and logical computer-supported system. GIS data processing and automation would enable faster and hence more complex analyses of field test data.

The Applied Research Center (ARC) at Florida International University (FIU) has provided analysis of flow and transport for several watersheds at DOE's Oak Ridge Reservation (ORR), including East Fork Poplar Creek (EFPC), Upper EFPC in the Y-12 National Security Complex area (Y-12 NSC) and White Oak Creek (WOC). Integrated surface and subsurface flow, fate and transport models were developed to provide analysis of contaminant patterns within each watershed. In addition, digital monitoring data available from the Oak Ridge Environmental Information System (OREIS) related to mercury (Hg) contamination and remediation within these watersheds was used for calibration and verification of the model. Experimental studies were also carried out which provided kinetic and equilibrium data about important parameters related to Hg transport, speciation and methylation/demethylation kinetics within the watershed. Geographic Information Systems (GIS) technology was employed to support the modeling work through storage and geoprocessing of spatial and temporal data required by the models and to produce hydrogeological maps for visualization. An ArcGIS geodatabase was developed for centralized storage and management of experimental and computed model data and its capabilities were extended over the years using tools such as ModelBuilder combined with Python scripting to automate repetitive tasks, perform statistical analyses and generate maps and reports. The hydrologic geodatabase model developed possesses a structure that enables linkage with scalable hydrologic modeling tools/applications to model hydrologic systems, and in this case, enables the testing of the potential impacts of various remediation scenarios on the ORR watersheds. The ArcSDE geodatabase can also be used to automate and simplify the process of calling stored GIS and timeseries data. This geodatabase can serve as a tool for contaminant flow and transport analyses which require large amounts of high-quality spatial and temporal data in order to ensure reliability and validity of modeling results. The existing geodatabase structure developed for the hydrological modeling at ORR was designed to be replicable for application at other DOE sites. An investigation of downloadable free/open source GIS software along with required security protocols to facilitate online querying of the database was also conducted to determine methods by which project-derived data can be more easily shared with other project stakeholders such as DOE personnel and site contractors.

INTRODUCTION

The Applied Research Center (ARC) at Florida International University (FIU) has supported the remediation efforts of the U.S. Department of Energy's Oak Ridge Reservation (ORR) in Tennessee through hydrological modeling of the fate and transport of inorganic and organic pollutants of concern with a focus on mercury (Hg). Experimental studies were also carried out which provided kinetic and equilibrium data about important parameters related to Hg transport, speciation and methylation/demethylation kinetics within the watershed. Integrated surface and subsurface flow, fate and transport models were developed for several watersheds including East Fork Poplar Creek (EFPC), Upper EFPC in the Y-12 National Security Complex area (Y-12 NSC) and White Oak Creek (WOC), to provide analysis of contaminant patterns within each watershed. More than a hundred simulations were completed for the purpose of calibrating the models, deriving model uncertainties, and for providing the analysis of remediation scenarios, resulting in gigabytes of computed spatial and temporal simulation data for each computation node. The accuracy and predictive forecasting ability of the hydrological models largely depend on the availability of timeseries data (daily/monthly/annual) as well as the period of time this data covers. Thus, the digital monitoring data related to mercury (Hg) contamination, hydrology and other relevant environmental parameters within these watersheds was used for calibration and verification of the models. This data, derived from sources such as the Oak Ridge Environmental Information System (OREIS), the U.S. Geological Survey (USGS), the Natural Resources Conservation Service (NRCS) STATSGO or SSURGO soil databases, and the U.S. Environmental Protection Agency (EPA) MRLC or NALC land cover databases, are being constantly updated, and as such it was necessary to periodically download and update the ORR geodatabase with more recent data. There was a need therefore for an advanced spatial data structure to address the management, processing, and analysis of spatial and temporal numerical modeling data derived from multiple sources. TABLE I below shows the model configuration files stored in the ORR Geodatabase. Geographic Information Systems (GIS) technology was employed to support the hydrological modeling work through storage and geoprocessing of spatial and temporal data required by the models and to produce hydrogeological maps for visualization, some of which are depicted in Fig. 1 and Fig. 2. An ArcSDE geodatabase was developed which facilitates centralized storage and management of experimental and computed model data, putting it into a structured, coherent, and logical computer-supported system. Its capabilities were extended over the years using tools such as ModelBuilder combined with Python scripting to automate repetitive tasks, perform statistical analyses and generate maps and reports. GIS data processing and automation enabled faster and hence more complex analyses of field test data.

TABLE I. Model Configuration Files Stored in the ORR Geodatabase

Spatial Data	Characteristics Represented
Admin_Features	EFPC, WOC, Y-12, Old Scrap Yard (OSY) Model domains (polygons)
	ORR Boundary
Admin_GRIDs	Model domains (GRIDs)
Conductivity_GRIDs	Hydraulic conductivity GRIDs
Contaminant_Conc_Features	Monitoring points (has associated timeseries attribute data)
	Plume Contours
Contaminant_Conc_GRIDs	Interpolated contaminant plumes (GRIDs)
DEMs	Clinch River, EFPC & WOC Watershed DEMs
Digital_Orthophotos	ORR DOQs (.bmp)
Drainage_GRIDs	Drainage Time Constant, Drainage Codes, Detention Storage (GRIDs)

GW_Features	Groundwater level contours
GW_GRIDs	Groundwater level GRIDs
Hydro_Features	Watersheds, subwatersheds, catchments, hydroareas (lakes/ponds) (polygons)
	Floodplain polygons
	Hydrography, Hydrodrainage, hydrostructures (polylines)
Impervious_GRIDs	Paved runoff coefficient (GRID)
Landcover_Landuse_Features	Landuse/Landcover polygons
Landcover_Landuse_GRIDs	Vegetation grid codes
Manning	Manning's coefficients (GRIDs)
Monitoring_Stations	USGS SW monitoring stations, outfalls, GW monitoring wells
Network_Features	Rivers, streams, reaches, cross sections, diversion ditch, utilities (polylines)
	Nodes (points)
Physical_Features	Buildings, obscured areas, natural outlines, man-made outlines (polygons)
	Margins, man-made structures (polylines)
Soils	Geology, soils (polygons)
Topo_Features	Elevation contours
Transport_Features	Roads, railroads, transportation structures (polylines)
Temporal Data	Characteristics Represented
Monthly_RF_TS	Monthly rainfall timeseries
Flow_Aug_TS	Flow augmentation timeseries
DHI_Timeseries	Flow rate/discharge timeseries

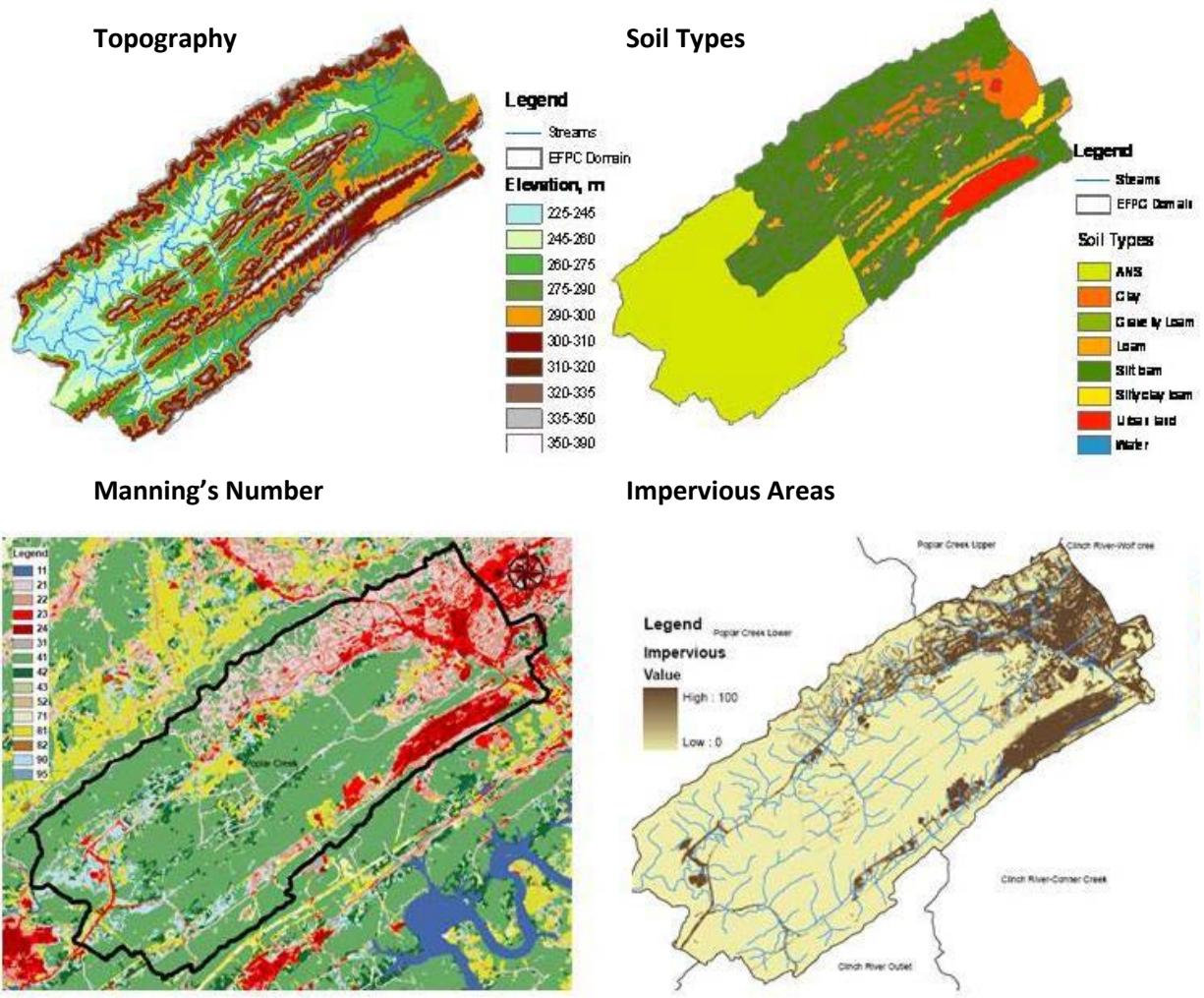


Fig. 1. Gridded input data files used for model development

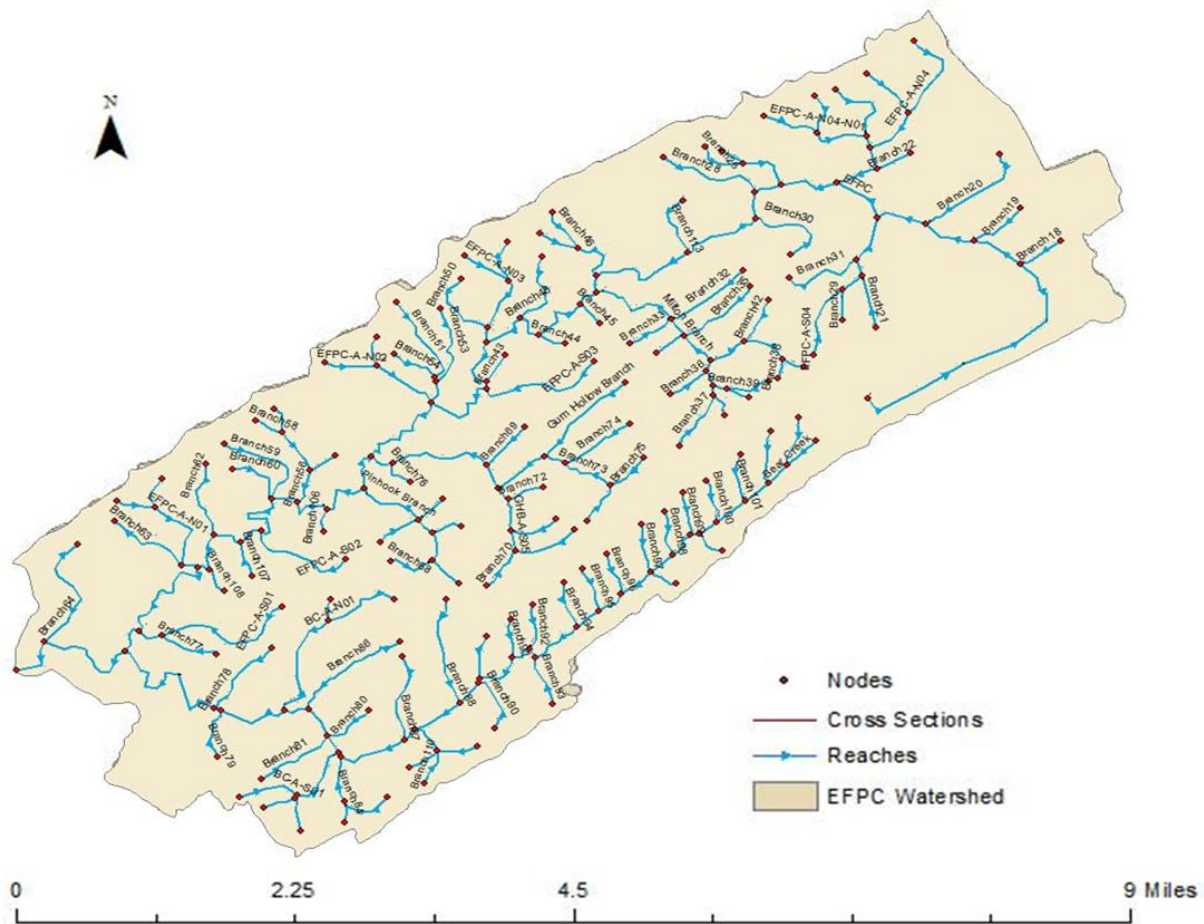


Fig. 2. GIS map showing EPFC network features used for model development

Details of the ORR geodatabase data and schema were generated using the ArcGIS Geodatabase Diagrammer utility for ArcGIS 10.2. ArcGIS Diagrammer is essentially a productivity tool used by GIS professionals to create, edit or analyze geodatabase schema. It generates diagrams and reports in the form of editable graphics within an interface similar to Microsoft Visual Studio. It serves as a visual editor which accepts XML workspace documents that are created from ESRI’s ArcMap or ArcCatalog. The reports generated depict the data structure of the ORR geodatabase and details of the features, rasters and tables that have been used or generated during hydrological model development, as well as any existing relationships and spatial references.

METHOD

Development of the Oak Ridge Reservation (ORR) Geodatabase

The ORR Geodatabase is a multiuser relational database management system (RDBMS) that was initially built upon a Microsoft SQL Server platform developed using Environmental Systems Research Institute (ESRI) ArcSDE technology (Fig. 3). The system was deployed on an advanced Windows server with the

latest technology and hardware and provides a user interface which facilitates data access, database connectivity, cryptography, web application development, numeric algorithms, and network communications. The system has since been updated to ArcGIS for Server to facilitate the sharing of GIS data and maps with project stakeholders as web services. The ArcGIS for Server system enables the hosting of GIS resources in-house and allows client applications, such as web mapping applications and mobile devices, to use and interact with these resources.

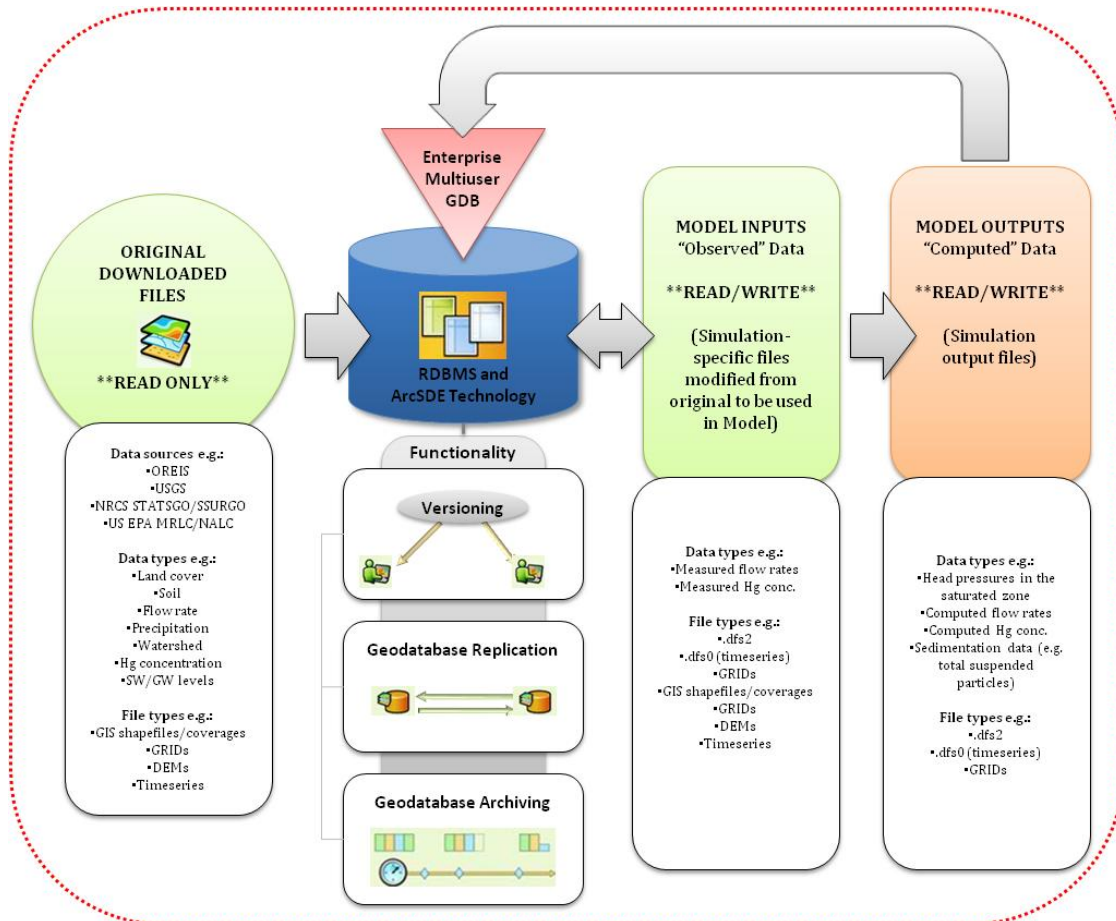


Fig. 3. The Oak Ridge Reservation Geodatabase system architecture

The ORR geodatabase is based on the ArcHydro and ArcGIS Base Map data models. The ArcHydro data model contains a set of accompanying tools designed to support water resources applications within the ArcGIS environment. These data models were also used as templates as there were many input data types in common with the ORR Geodatabase. Modifications were then made for project specific input parameters. The geodatabase developed possesses a structure that enables linkage with scalable hydrologic modeling tools/applications to model hydrologic systems, and in this case, enables the testing of the potential impacts of various remediation scenarios on the ORR watersheds. The ArcSDE geodatabase can also be used to automate and simplify the process of calling stored GIS and timeseries data. This geodatabase can serve as a tool for contaminant flow and transport analyses which require large amounts of high-quality spatial and temporal data in order to ensure reliability and validity of modeling results.

The existing geodatabase structure developed for the hydrological modeling work at ORR was structured to be replicable for application at other DOE sites and serves as a centralized data management system, providing access to data generated from simulations of contaminant fate and transport to all users and facilitating storage, concurrent editing and import/export of model configuration and output data that is specific to the hydrologic and transport models being used. The ORR geodatabase was configured for concurrent multi-user access and editing capability, adhering to the appropriate security and quality assurance protocols necessary to maintain data integrity. This process exerts control on the type of access all users have to the geodatabase and its datasets, and enables specification of user data management privileges. Connection to the geodatabase requires Windows-authenticated credentials, and permissions are assigned to users according to the existing roles contained in SQL Server which already have predefined sets of permissions; however, the option to create new roles and set associated permissions is also possible if necessary. Windows authentication for access to the geodatabase was selected for simple configuration utilizing a certificate-based security mechanism, which is more secure than an operating system (OS)-based authentication. Besides the FIU-ARC firewalls and the built-in software and hardware security protocols, the geodatabase resides on computers which must adhere to the FIU's University Technology Services (UTS) Security and IT Policies.

Geoprocessing Automation Using ArcGIS ModelBuilder and Python

The essence of this task was to provide modelers with ease of access to model data and pre-processing capability for files to be used as input into the hydrological models developed using MIKE SHE/11 (by DHI, Inc.). Post-processing of model results was also possible for analysis, use in reports and map production. Simple tasks such as retrieving data from the geodatabase; pre-processing the data; exporting for use in hydrological model development; subsequent import and post-processing of model data; data analysis; and production of graphs, maps and reports are repetitive but necessary. The use of Python scripts and ArcGIS ModelBuilder assists in automating these tasks which saves time and can facilitate batch processing of this data. One of the objectives of this task, therefore, was to also develop a reusable GIS tool, which can iterate over each set of the spatial input data parameters, perform geoprocessing actions and calculate statistical parameters. ArcGIS ModelBuilder can also be used to generate model workflow diagrams which are a great way of documenting and visually representing the geoprocessing tools and scripts being incorporated into the data model as development progresses.

FIU's work scope involved the development of customized Python scripts which required additional programming of built-in automated geoprocessing tools to call or retrieve data from the ORR geodatabase. A toolbox which combines built-in ArcGIS geoprocessing tools coupled with customized Python scripts was therefore developed and calibrated for use with the East Fork Poplar Creek (EFPC) model. This toolbox, however, is a scalable and reusable application that can be implemented for geodatabases that contain data relevant to other DOE sites. The tools and scripts developed automate the query and retrieval of timeseries data, including contaminant flow and transport parameters (e.g. mercury concentration, surface water and groundwater flow, discharge, groundwater level, etc.), from the existing ORR geodatabase. The ArcGIS data model iterates through selected features and exports the results in tabular format. The toolbox is calibrated to the project's location (i.e. the EFPC model domain) and has capabilities to:

1. Add GIS files to ArcMap and create layer files.
2. Select features within a specified area (e.g. the study domain) and zoom to selected features.
3. Clip/extract selected features and create new layer file of selected subset.
4. Export clipped feature in format to be used by MIKE SHE/11 model.

5. Export attributes of clipped feature in MS Excel or text format for statistical analysis and generation of graphs and reports.
6. Export map extent in various formats (e.g. JPEG, TIFF or PDF) for development of reports.
7. Interpolate timeseries data collected at various monitoring points, generate gridded surfaces, and finally create and export mapped results (Fig. 4).

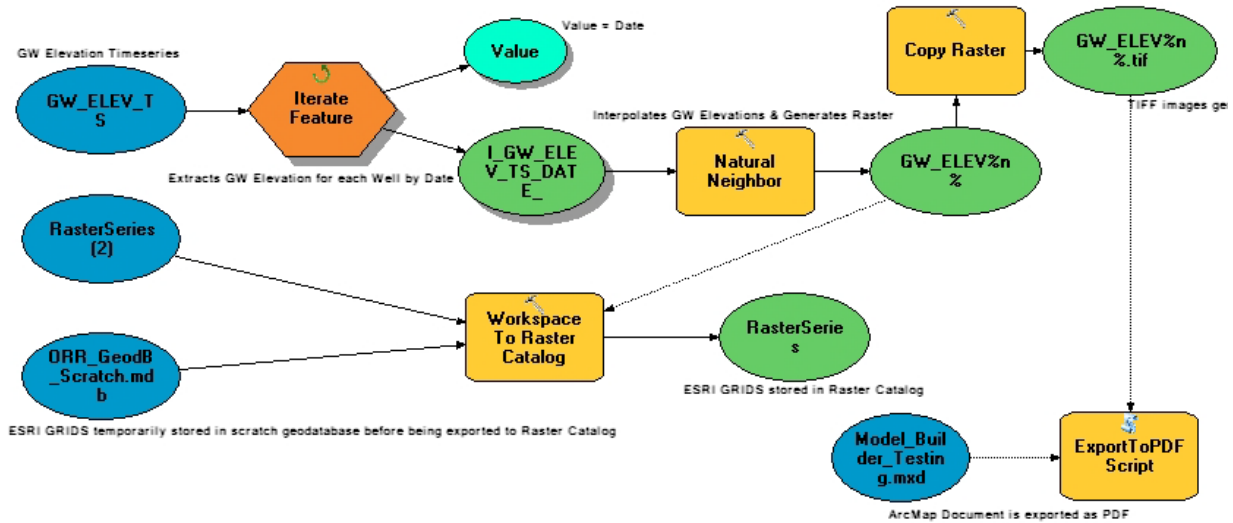


Fig. 4. Preliminary process workflow diagram created using ArcGIS ModelBuilder

Statistical Analysis Using ArcGIS ModelBuilder and Python

Additional customized Python scripts were also developed to further enhance the database capabilities to perform statistical analyses on model output data by implementing a library of scripts which can be coupled with other existing libraries used for mathematics, science, and engineering such as NumPy and SciPy. The goal was to create scripts to calculate model performance statistics for a subset of existing flow and contaminant monitoring stations. Data for the selected stations are available, in most cases, on an hourly basis and for some stations on a daily basis only. The following lists some of the parameters and their equations for which scripts have been developed.

1. *ME*: Mean error equal to the sum of the difference observed minus calculated divided by the total number of samples.
2. *MAE*: Mean absolute error is the sum of the absolute difference of observed minus calculated divided by the total number of samples.
3. *RMSE*: Root mean square error of the observed minus computed.
4. *STD*: Standard deviation of the residual.
5. *CoVar*: Covariance is defined as a measure of how much two variables change with respect to each other.

$$CoVar = \frac{\sum_{t=1}^T (h_o^t - \overline{h_o}) (h_m^t - \overline{h_m})}{T} \quad (\text{Eq. 1})$$

- T is number of timesteps
- h_o^t is observed stage at time t
- h_m^t is modeled stage at time t
- $\overline{h_o}$ is average observed stage
- $\overline{h_m}$ is average modeled stage

6. *Cor: Correlation* is defined as the covariance of the observed and modeled stage divided by the product of the standard deviations. The correlation and covariance describe the degree of similarity between the model and field data sets. The correlation is dimensionless and describes the degree to which the two data sets rise and fall in a similar manner.

$$Cor = \frac{\sum_{t=1}^T (h_o^t - \overline{h_o}) (h_m^t - \overline{h_m})}{\sqrt{\sum_{t=1}^T (h_o^t - \overline{h_o})^2 \sum_{t=1}^T (h_m^t - \overline{h_m})^2}} \quad (\text{Eq. 2})$$

7. *PEV: Percent explained variance* compares the variance of the residual to the variance of the measured data. This yields a percentage of the measured variability that is represented by the model.

$$PEV = 100\% * \left[1 - \left(\frac{\sum_{t=1}^T \left[(h_m^t - h_o^t) - \left(\frac{\sum_{t=1}^T (h_m^t - h_o^t)}{T} \right) \right]^2}{\sum_{t=1}^T (h_o^t - \overline{h_o})^2} \right) \right] \quad (\text{Eq. 3})$$

8. *NS: The Nash-Sutcliffe model efficiency coefficient* compares the residual squared to the measured variance. The Nash-Sutcliffe Coefficient yields a percentage of the error relative to the measured fluctuation. The *NS* is sensitive to mean differences in measured and computed, which can yield large negative values.

$$NS = 1 - \frac{\sum_{t=1}^T (h_m^t - h_o^t)^2}{\sum_{t=1}^T (h_o^t - \overline{h_o})^2} \quad (\text{Eq. 4})$$

Secured Internet Map Publishing of Hydrological Model Data

In order to facilitate the sharing of project-derived hydrological modeling results with project stakeholders such as DOE personnel and ORR site contractors, FIU researched the possibility of developing customized online mapping applications which utilize available software APIs. Special attention was paid to the security aspect of sharing project data via a web-based application and the potential cyber issues involved. Criteria for development had to take into consideration proper implementation of security protocols to ensure that data can be published via the Internet over a secured environment. In preparation for implementing secure protocols for data access, ArcGIS 10.2 was deployed on a secure server and research was conducted to determine the requirements for deploying hydrological model results from the server/PC onto an already existing platform that was developed by FIU for the U.S. Department of Energy. Users will be required to register with the customized ArcGIS platform developed by FIU for authentication and authorization to access data. This will secure the integrity of the site-specific data considered sensitive to cyber compromise.

The ArcGIS for Server architecture being implemented (Fig. 5) included three software technologies including: (1) ArcMap for creating, editing, and viewing the geospatial data and maps; (2) ArcGIS Server, which provided the platform for storing and sharing GIS resources with the user community; and (3) the Web Adaptor, which added an intermediate layer of protection between the end user viewing published maps and ArcGIS Server where the information is actually stored.

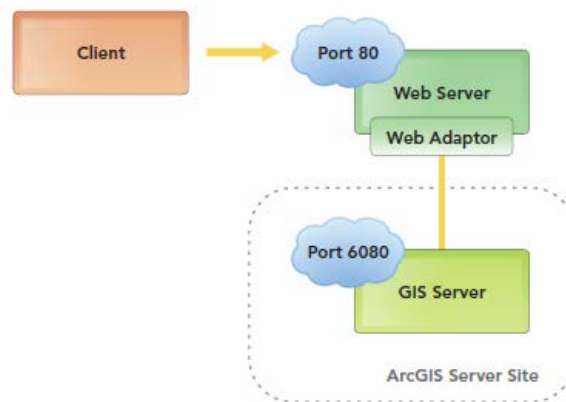


Fig. 5. ArcGIS for Server architecture

In order to display project data on the web, a map containing GIS-based hydrological data was created using ArcMap and then published using ArcGIS Server. Once published on the server, the map and associated data files become accessible as a GIS service or web service through port 80. However, to add an extra layer of protection and control to the web service, ArcGIS Web Adaptor was installed on an FIU-ARC GIS server, which ensures restricted access to the GIS service exclusively through the Web Adaptor.

With the Web Adaptor installed, a website was created that communicates with the web adaptor and consumes the GIS web service information, displaying it in an interactive map that users can view. The website uses ESRI's JavaScript API to create a map object from the web service created and using the map information, generates different map layers which are added to the map object to show all the different features in the map document.

A map Legend was also created to inform the user what the different geometries on the map represent. This functionality also uses JavaScript to retrieve the different map layer information from the web service and manipulates the web page to visually display this information to the user. Another feature added to the website is the ability to toggle different map layers on and off using checkboxes in the legend. Large maps can be crowded with many different layers and points of data; therefore, the layer toggle feature allows the user to view meaningful layers of information and omit irrelevant data crowding the map. Zooming and panning are also features implemented on the map allowing users to move and pan to any location within the map object.

Further development of the website then incorporated a layer selection feature. This feature allows the user to select a layer from the Legend which will also highlight and select the respective layer on the map. The Layer Selection feature works in conjunction with the popup feature. Once a user has selected a Layer, it will then be possible to click on a particular feature on the map from that layer and a popup will appear displaying the feature's attribute data.

CONCLUSION

GIS technology has proven useful in supporting the hydrological modeling work performed by FIU-ARC at the Oak Ridge Reservation (ORR) through development of a geodatabase which has provided an advanced spatial data structure for management, processing, and analysis of spatial and temporal numerical modeling data derived from multiple sources. ArcGIS ModelBuilder coupled with Python scripting enabled the automation of many of the repetitive geoprocessing tasks required for pre- and post-processing of hydrological modeling data. GIS data processing and automation facilitated faster and hence more complex analyses of field test data. The toolbox created is a scalable and reusable application that can be implemented at other DOE sites. Finally, a web-based GIS application can facilitate sharing of project derived data with stakeholders including DOE personnel and ORR site contractors.

REFERENCES

1. Ames, D.P., Horsburgh, J.S., Cao, Y., Kadlec, J., Whiteaker, T. and D. Valentine. (2012). HydroDesktop: Web Services-Based Software for Hydrologic Data Discovery, Download, Visualization, and Analysis. *Environmental Modelling & Software*. Vol 37, pp 146-156.
2. Ames, D.P., Michaelis, C.D., Anselmo, A., Chen, L. and H. Dunsford. (2008). MapWindow GIS. S. Shekhar, H. Xiong (Eds.) *Encyclopedia of GIS*. Springer, New York, pp. 633–634.
3. ArcGIS Online, ArcGIS Diagrammer User Guide, December 10, 2007. Retrieved from <http://maps.esri.com/Diagrammer/userguide.pdf>.
4. ArcGIS Resources, ArcGIS Help Library. Retrieved from http://resources.arcgis.com/en/help/main/10.2/index.html#/Welcome_to_the_ArcGIS_Help_Library/00qn0000001p000000/.
5. Armstrong, K., "Intermediate ModelBuilder," ESRI Southeast Regional User Group (SERUG) Conference Technical Workshop (2010).
6. Bradt, S. (2008). A comparison of open source or free GIS software packages. The New Hampshire Geospatial Extension Program, University of New Hampshire (UNH) Cooperative Extension.
7. Castle, E. (2003). *Geodatabases in Design: A Floodplain Analysis of Little Kitten Creek*. Thesis. Brigham Young University.
8. ESRI. (n.d.). ArcGIS Server .NET Help - Setting up a direct connection to SQL Server. Retrieved 2010, from ArcGIS Resource Center: <http://goo.gl/N4jppq>.

WM2015 Conference, March 15 – 19, 2015, Phoenix, Arizona, USA

9. ESRI. (n.d.). ArcGIS Server 10 Server Requirements. Retrieved 2010, from ArcGIS Resource Center: <http://goo.gl/hNMP5>.
10. ESRI. (2005). Raster Data in ArcSDE 9.1. An ESRI White Paper. Retrieved 2010, from ArcGIS Resource Center: <http://goo.gl/sRvOy>.
11. ESRI. (n.d.). Programming against ArcObjects. Introduction to COM. Retrieved 2010, from ArcGIS Resource Center: <http://goo.gl/qZjQR>.
12. ESRI. (n.d.). Server 10 .NET Install Guide - Installing ArcGIS Server for the Microsoft .NET Framework overview. Retrieved 2010, from ArcGIS Resource Center: <http://goo.gl/4Xy7W>.
13. Gogu, R. (2001). GIS-based hydrogeological databases and groundwater modelling. *Hydrogeology Journal*, 9:555–569. DOI 10.1007/s10040-001-0167-3.
14. Kaiser, D. and Bajwa, H., “Working with Temporal Data in ArcGIS”, ESRI International User Conference Technical Workshop (2010).
15. Law, D. (2010). Versioning 101: Essential information about ArcSDE geodatabases. *ArcUser Online* (Winter), 42-48.
16. Maidment, D. R. (2002). *Arc Hydro: GIS for Water Resources*. ESRI Press.
17. Miessau, R. R. (2007). Using ArcGIS to Support a Regional Hydrologic Model - Paper 2191. ESRI User Conference Proceedings. San Diego, CA: ESRI.
18. MIKE by DHI. (2011). *MIKE 11 GIS User's Guide*.
19. ModelBuilder Lab, Geoinformatics, Spring 2008, Purdue University Library.
20. Murrar, S., “ModelBuilder Advanced Techniques”, ESRI International User Conference Technical Workshop (2010).
21. NumPy Reference, Release 1.8.1, Written by the NumPy community, March 26, 2014.
22. NumPy User Guide, Release 1.8.1, Written by the NumPy community, March 26, 2014.
23. Olivera, F. V. (2006). ArcGIS-SWAT: A Geodata Model and GIS Interface for SWAT. *Journal of the American Water Resources Association*.
24. Robayo, O. and D. Maidment, 2005. Map to Map: Converting a NEXRAD Rainfall Map into a Flood Inundation Map. Center for Research in Water Resources, University of Texas at Austin, CRWR Online Report 05-1.
25. Rosenzweig, I. and B. Hodges, 2011. A Python Wrapper for Coupling Hydrodynamic and Oil Spill Models. Center for Research in Water Resources, University of Texas at Austin, CRWR Online Report 11-09. Submitted to Texas General Land Office Oil Spill Prevention & Response. FY 2011 Report under Contract No. 10-097-000-3928, October 31, 2011.
26. SciPy Reference Guide, Release 0.13.0, Written by the SciPy community, October 21, 2013.
27. Petersen, M. K., “Time-Series Contaminant Interpolation using ArcGIS and Spatial Analyst”, ESRI User Conference Proceedings (2006), Paper 1326.
28. Wikihow: “Creating time-series raster mosaics in ArcGIS 10 for Eye on Earth.”
29. Zimmer, R. (2010). GIS Matters: Something for Nothing. *The American Surveyor*, Vol 7 No 8.