STATE OF FLORIDA
DEPARTMENT OF ENVIRONMENTAL PROTECTION
TASK #1 OSTDS Database Development
FINAL REPORT
DEP AGREEMENT NO.:AT020

Florida State University, Earth, Ocean, & Atmospheric Science
Michael Core, Wei Mao, and Ming Ye
July 31, 2024

PROJECT TITLE: Continued Numerical Modeling for Investigation Onsite Sewage Treatment and Disposal Systems (OSTDS) Setback Distance

PROJECT LOCATION: All work for this project will be performed at the Florida State University's Main Campus in Tallahassee, Florida.

Project Overview

The Onsite Sewage Treatment and Disposal System (OSTDS) Database Project aims to consolidate and standardize environmental data related to water quality and waste management from various studies and reports shown in the Appendix in the Excel file titled OSTDS Database PlumeCount Metadata FieldNames DataType 20240723 (attached at the end of the document). The project involves georeferencing the monitoring well locations for 62 study sites, as the Florida Department of Environmental Protection (FDEP) provided, and organizing the data into a comprehensive database. This database includes information on analytes, environmental conditions, built environment variables, and tracer data, which aid in understanding the impact of OSTDS on water quality. The data is then visualized and analyzed using ArcGIS Pro and ArcGIS Online Experience Builder, facilitating data management and decision-making. The primary objective of the OSTDS Database is to provide a robust and reliable platform for storing, managing, and analyzing environmental data related to OSTDS. This database serves multiple purposes:

1. **Data Integration**: Integrating data from multiple sources into a standardized format to ensure consistency and accuracy.
2. **Geospatial Analysis**: Enabling spatial analysis and visualization of data to identify patterns and trends related to OSTDS and their impact on water quality.
3. **Decision Support**: Supporting environmental managers, policymakers, and researchers with accurate data to make informed decisions about water quality management and OSTDS regulations.
4. **Public Accessibility**: Providing access to comprehensive data for the public and stakeholders, promoting transparency and community engagement.

Accurate data collection and analysis are beneficial for understanding the environmental impact of OSTDS and making informed decisions about water quality management. Key reasons for the importance of precise data include informed decision-making, regulatory compliance, environmental protection, resource management, and scientific research. By consolidating data from various studies and reports, the OSTDS Database Project aims to enhance the understanding and management of OSTDS and their impact on water quality, contributing to the protection and sustainability of Florida's water resources.

Detailed Data Description and Record Counts

The following sections summarize and characterize the data within each category, including record counts from the provided tables and the OSTDS Database app.

1. Analytes data includes various measurements related to water quality, such as turbidity, total dissolved solids (TDS), total suspended solids (TSS), and conductivity. This dataset helps assess the chemical composition and potential pollutants in the water. The analytes data contains approximately 17,540 records.

   Example Fields:

   - NO3 N mg/L: Nitrate nitrogen in milligrams per liter ranged from 65 mg/L to 0.009 mg/L.
   - DO mg/L: Dissolved oxygen in milligrams per liter, crucial for aquatic life, ranging from 0.1 mg/L to 11.8 mg/L.
2. Environmental conditions data encompass parameters such as water table elevation, depth to groundwater, and soil moisture retention. This data is essential for understanding the ecological factors influencing water quality. The environmental conditions dataset comprises 4,560 records.

   Example Fields:

   - Water table elevation (ft): Water table elevation in feet ranged from 8.83 ft to 1.88 ft above the National Geodetic Vertical Datum.
   - K ft/day: Hydrologic conductivity in feet per day ranged from 0.45 ft/day to 9.26 ft/day.
3. Tracer studies provide insights into the movement of contaminants through the environment. The data includes initial tracer concentration, peak tracer concentration, and horizontal transport rate. Tracer data helps in tracking the spread and concentration of pollutants. This dataset contains 3,290 records.

   Example Fields:

   - Peak Tracer Concentration Fluorescein mg/L: Maximum concentration of fluorescein tracer in milligrams per liter, used to track water movement. The data ranged from 11.01 mg/L to 0.002 mg/L.

4. Built environment data includes information about households and infrastructure, such as the number of bedrooms, occupancy, and water use. This data is vital for understanding human activities' impact on water quality. The dataset for built/household variables contains 399 records.

   Example Fields:

- Number of bedrooms: The number of bedrooms in the household indicates potential occupancy and water use.
- Total Water Use gpd: Water is used in gallons per day, which helps estimate the household's impact on water quality and ranges from 1166.5 gpd to 173.2 gpd.

While constructing the OSTDS database, certain wells from specific reports could not be georeferenced. The reports (Brown et al., 2001; Hazen and Sawyer, 2011, 2013c, 2013b, 2013a) were not georeferenced due to missing figures and redaction from the public record by the Florida Department of Health (FDOH). Although the monitoring well locations were not georeferenced, the data remains in the CSV files. Additionally, several monitoring locations associated with piezometers, seepage meters, and monitoring wells lacked the geographic information necessary for georeferencing but are still included in the CSV file. These comprehensive datasets and methodologies used in their collection and organization ensure that the OSTDS database is a valuable resource for environmental research and management, providing insights into water quality and the impacts of human activities on the environment.

Data Collection and Georeferencing

The initial phase of creating the OSTDS database involved georeferencing reports from the FDEP. This step ensured the accurate mapping of monitoring well locations. The georeferencing process began with identifying all relevant FDEP reports containing data on monitoring wells, environmental conditions, and analyte values. Spatial information, such as latitude and longitude coordinates, addresses, or descriptions of monitoring locations, was then extracted from these reports. This information was digitized by plotting the locations on a map within ArcGIS Pro, followed by cross-checking against reference maps or additional data sources to ensure accuracy. When precise coordinates were unavailable, monitoring locations were estimated using parcel maps and historical aerial imagery to identify probable well locations based on property boundaries and report descriptions. Additionally, the proximity to known landmarks mentioned in the reports helped estimate locations and relative location methods were used to assess well locations between known points within a study area. After georeferencing, the observational data was organized in Excel to ensure consistency and ease of subsequent processing and analysis. The data was structured with clearly defined column headers, consistent data entry formats, and unique identifiers for each monitoring location to merge and join data tables (Esri, n.d.-b). For example, fields included the site name (e.g., "Jones"), location nomenclature (e.g., "J5"), sample depth (e.g., "9.27 ft"), date of observation (e.g., "1/31/1990, 7:00 PM"), analyte name (e.g., "Temp °C"), analyte value (e.g., "26.0"), and source literature citation (e.g., "Ayres Associates, 1993"). This structured approach enabled processing and analysis in ArcGIS Pro, creating maps, charts, and data tables for the OSTDS database.

The data preparation phase commenced with exporting the organized Excel sheets into CSV format, ensuring compatibility with ArcGIS Pro. Each Excel worksheet was exported individually, encompassing data such as site names, locations, sample depths, dates, analyte names, and unique well IDs. This methodical approach helped maintain data integrity and prevent any loss of information during the conversion process. Subsequently, the CSV files were imported into ArcGIS Pro, where they were treated as separate layers. This phase required attention to detail to ensure the accurate transfer of all attribute data. It was imperative to verify that the georeferenced monitoring well locations corresponded accurately to their expected positions on the map and that all attribute data, including analyte values and dates, were correctly associated with their respective sites and locations. This involved double-checking coordinates and attributing data to avoid discrepancies that could compromise subsequent analyses. The process set a foundation for geospatial analysis, ensuring the OSTDS database's reliability and accuracy for further processing and visualization within ArcGIS Pro.

Data Structuring in ArcGIS Pro

Data structuring in ArcGIS Pro for the OSTDS database project involved a series of steps to ensure the accuracy and usability of the geospatial data. Initially, the focus was on defining and organizing field descriptions for key attributes such as Site, Location, Sample Depth, Date, Analyte Name, and Well ID. Each field played a role in identifying and describing the data points, with the Well ID serving as a unique primary key constructed from the concatenation of Site, Location, and Sample Depth. Doubles were used instead of floats to maintain data integrity, especially in calculations and visualizations. This decision was driven by the need to prevent rounding errors that could arise from floating-point arithmetic, ensuring that all numerical values, particularly those related to environmental measurements, were represented with precision (Esri, n.d.-a). The structure and organization of the data were facilitated using the FDEP Win Waves search tool (FDEP, n.d.). Handling built environment variables required careful consideration of the unique characteristics of each site. Built environment data, which often was not directly tied to specific monitoring well locations, necessitated the creation of OSTDS locations and tap locations at residential sites. This approach ensured each site had a corresponding location to hold observations related to the built environment, such as water use per capita and other OSTDS-related variables. Furthermore, this workflow included transposing fields, calculating and standardizing units, and merging datasets while removing null values to prepare the data for spatial analysis and visualization.

To build the tables for the OSTDS database, each entity was separated, including environmental conditions, built environment variables, and tracer data, to ensure that each report data attribute was recorded correctly. This separation allowed for data management and facilitated the subsequent joining process. Each unit of measurement from the various reports was recorded and noted directly in the column names to maintain clarity and accuracy. The data preparation step involved calculating the units for each study and ensuring they were stated correctly. This process was carried out after transposing the fields, which required multiple iterations, one instance for each entity in the database. The data was then reviewed, and any necessary field calculations were performed to standardize the measurements across different studies. Finally, the Well ID was built by concatenating the site, location, and sample depth fields, ensuring each monitoring well location had a unique identifier. This Well ID was then used to join the tabular data with the spatial data, completing the data integration process. After joining, the joined data was exported, datasets were merged from multiple reports, and null values were carefully removed from the value fields to ensure the dataset's completeness and accuracy. This process resulted in an OSTDS database ready for further analysis and visualization.

Specific Arcade expressions were employed to standardize units and field names to ensure consistency and accuracy across the dataset. For instance, an expression was used to parse the Analyte Name field, removing underscores and capitalizing each word to create a clean, uniform presentation. The same methodology was applied to other fields, ensuring that all names followed a consistent format. For example, an Arcade expression was crafted to detect specific units within field names and then replace them with standardized units. This was needed for fields such as Analyte Name, where variations in units like "mg/L" or "µM" were common. By detecting and standardizing these units, the integrity and comparability of the data were maintained. A field named Source Literature was created within ArcGIS Pro to organize and reference the data. This field concisely encapsulates the literature citation for each dataset entry. This step helped to ensure that each data entry could be easily traced back to its source, providing transparency and verifiability to the dataset. By implementing these standardized practices, the OSTDS database achieves more data integrity and usability.

Subsequently, spatial joins were performed across multiple reports to consolidate the data from different studies into a single, unified dataset. This step is essential for enabling a comprehensive analysis of environmental conditions and their impact on water quality. The spatial join process involves overlaying different spatial datasets and transferring attributes from one layer to another based on their spatial relationships. This allows for the collection of data from various reports, ensuring a holistic view of the study area and facilitating detailed spatial analysis via the correlation between analyte concentrations and natural environment conditions. To ensure the integrity of the OSTDS database, a data cleanup and duplicate removal process was implemented. The initial step involved ensuring no null values existed in key fields, which are crucial for accurate data analysis and reporting. This was achieved by systematically selecting each entity, such as environmental conditions, and addressing null values individually. The data cleanup process maintained precision by focusing on one variable at a time. Field mapping and first-response methods were utilized to organize and refine the data. Field mapping involved assigning and standardizing field names and types, ensuring consistency across the dataset. The first response method allowed for quick identification and resolution of anomalies within the data, streamlining the cleanup process. Handling duplicate records was another aspect of the data cleanup process. Duplicate records can skew analysis results and reduce the reliability of the dataset. A systematic approach was adopted to identify and remove duplicate records. This involved using Summary Statistics and various tools within ArcGIS Pro to identify and eliminate redundant records caused by joining records in ArcGIS Pro (Esri, 2023). The process ensured that unique and accurate records remained, thereby enhancing the quality and reliability of the data.

Web Map and Popup Configuration

Configuring the web map and popups involved several steps to ensure a user-friendly and informative experience. First, new popups that required using Arcade expressions for HTML formatting were set up. These expressions allowed for dynamic content display and customization, enhancing the visual appeal and clarity of the information presented. The popups were designed to provide detailed information about each data point, with careful attention to formatting to ensure readability. In addition to the primary data fields, Google Street View links were integrated into the popups. This feature allows users to view the geographical context of each data point, providing a more immersive and informative experience (Esri, 2024). The integration was achieved through Arcade expressions that convert the point coordinates to latitude and longitude, forming a URL that opens Google Street View at the specific location. Overall, the configuration of the web map and popups aimed to provide a comprehensive, user-friendly interface that enhances the accessibility and usability of the OSTDS database. Creating and implementing web maps and hosted feature layers are necessary for developing the OSTDS database. Two primary ArcGIS Online hosted feature layers were created: OSTDS Database v1 202403 and OSTDS Database Charts and Graphs. These feature layers serve as the foundational data sources for the web maps, facilitating data visualization, analysis, and user interaction within the ArcGIS Online Experience Builder.

The OSTDS Database v1 202403 feature layer encompasses comprehensive datasets, including analyte data, environmental conditions, tracer data, and built environment variables. This feature layer ensures that all relevant data points are accessible and can be visualized through the web maps. The OSTDS Database Charts and Graphs feature layer is designed for statistical analysis and visualization. This layer supports the creation of the scatterplot, enabling users to explore correlations and trends within the data. The feature layer integrates seamlessly with the ArcGIS Online Experience Builder, providing dynamic updates to visualizations based on user-selected criteria (Esri, n.d.-d). Within the ArcGIS Online Experience Builder, these hosted feature layers power the interactive web maps and associated tools. Users can filter and select specific data points, generate charts and scatterplots, and access detailed information through popups. The experience is designed to provide a comprehensive and user-friendly interface for exploring the OSTDS database, making it a valuable resource for water quality research and analysis.

The ArcGIS Online Experience for the OSTDS project is designed to provide comprehensive tools for visualizing, analyzing, and interpreting water quality data. This experience is divided into three main components: Map and Table Visualization, Statistical Analysis Tools, and Scatter Plot Correlation Analysis. The Map and Table Visualization component is the starting point of the ArcGIS Online Experience. Users can explore georeferenced data on a map, which includes monitoring well point locations, environmental conditions, built environment details, and tracer data. The interactive map interface allows users to zoom in and out, select specific data points, and view detailed popups with relevant information. The associated tables provide a tabular data view, categorized into Analyte, Environment Condition, Built Environment, and Tracer Data. Users can filter, sort, and export data from these tables, facilitating easy access to specific subsets of the dataset for further analysis. The Statistical Analysis component offers a suite of tools for performing statistical analysis on the OSTDS data. Users can generate descriptive statistics such as minimum, maximum, mean, median, standard deviation, and record counts for nearly 300 unique variables. This feature helps users to understand the distribution and central tendencies of various water quality and environmental parameters. Additionally, histograms and charts are available to visualize the data distribution and identify potential anomalies or patterns. The Scatterplot Correlation Analysis component enables users to explore relationships between different variables. By selecting one variable from the analyte data and the environmental condition data, users can generate scatterplots to visualize potential correlations. The scatterplots are accompanied by R^2 values to quantify the strength of the correlation. This tool can be used to identify trends and relationships between environmental conditions and water quality parameters.

The interactive nature of the scatterplots allows users to hover over data points to see specific values and details, providing a deeper understanding of the underlying data. These three components provide a platform for analyzing OSTDS data, facilitating informed decision-making and research.

Conclusions and Future Work

The development of the OSTDS Database involved multiple steps to ensure accurate data management. The process began with georeferencing FDEP reports to locate and estimate the monitoring well locations on a parcel basis (Esri, n.d.-c). Subsequent organization of the data in Excel, followed by exporting to CSV files, provided a structured format for further processing in ArcGIS Pro. The final dataset was used to create two ArcGIS Online Hosted feature layers, serving as the data sources for the OSTDS Database web maps. These web maps provide an interface for visualizing and analyzing the data, including capabilities for statistical analysis and generating scatterplots to examine correlations between different variables within an ArcGIS Online Experience.

Future work for the OSTDS Database includes several enhancements to increase the dataset's comprehensiveness and the platform's usability. An area of focus will be incorporating more recent data and reports, ensuring the database remains up-to-date and reflects current environmental conditions. Optimizing the rendering speed of the map is essential to improving the user experience. This could be achieved by integrating Python with TypeScript, which would streamline data processing and enhance performance. Implementing an interpolation widget within the Esri Experience will provide insights by estimating analyte concentrations between monitoring wells. A raster estimation surface can be generated to visualize these calculated values, thus facilitating the identification of contamination trends and hotspots over a geographic area. Developing custom tools for data analysis is another priority. For example, a tool that allows users to calculate analyte attenuation within a buffered area of a selected monitoring well point feature could be beneficial. This tool would calculate the decrease in analyte concentration from the highest to the lowest measurement and record the spatial relationship between the monitoring wells, thus helping users understand the dispersion and degradation of contaminants in the environment. Standardizing data units is crucial for consistency and accuracy in environmental data reporting. Like the FDEP's WIN WAVES system, which reports both the original and a standardized value (FDEP, n.d.), the OSTDS Database could adopt this practice. Alternatively, an Arcade expression could convert units dynamically within the app or web map, allowing users to view data in their preferred units. Additionally, the database could benefit from an automated data extraction process. Training a GPT model to extract data from tables in PDF reports could significantly reduce manual data entry and improve accuracy. As an alternative, Python scripts can be utilized to extract tables from FDEP PDF reports, automating the data import process and ensuring timely updates to the database. Implementing these improvements will enhance the OSTDS Database's utility.

OSTDS_Database_Plu
meCount_Metadata_Fi

References

Brown, M. T., Annable, M. D., Delfino, J. J., Jawitz, J. W., Cohen, M., Hall, E., Harden, H., Chanton, J., & Burnett, W. (2001). *Determination of an appropriate onsite sewage system setback distance to seasonally inundated areas.* Florida Department of Health.

Esri. (n.d.-a). *ArcGIS field data types—ArcGIS Pro | Documentation*. Retrieved July 31, 2024, from https://pro.arcgis.com/en/pro-app/latest/help/data/geodatabases/overview/arcgis-field-data-types.htm

Esri. (n.d.-b). *Introduction to joins and relates—ArcGIS Pro | Documentation*. Retrieved July 31, 2024, from https://pro.arcgis.com/en/pro-app/latest/help/data/tables/joins-and-relates.htm

Esri. (n.d.-c). *Overview of georeferencing—ArcGIS Pro | Documentation*. Retrieved July 31, 2024, from https://pro.arcgis.com/en/pro-app/latest/help/data/imagery/overview-of-georeferencing.htm

Esri. (n.d.-d). *Spatial analysis in ArcGIS Pro—ArcGIS Pro | Documentation*. Retrieved July 31, 2024, from https://pro.arcgis.com/en/pro-app/latest/help/analysis/introduction/spatial-analysis-in-arcgis-pro.htm

Esri. (2023, May 9). *How To: Remove Duplicates from Joined Tables in ArcGIS Pro*. https://support.esri.com/en-us/knowledge-base/how-to-remove-duplicates-from-joined-tables-in-arcgis-p-000027634

Esri. (2024). *ArcGIS Experience Builder | ArcGIS Developers*. ArcGIS Experience Builder. https://developers.arcgis.com/experience-builder/

FDEP. (n.d.). *WAVES Search*. Retrieved July 31, 2024, from https://prodenv.dep.state.fl.us/DearWin/public/wavesSearchFilter?calledBy=menu#

Hazen and Sawyer. (2011). *C-HS1 Monitoring Data Summary Report No. 4 – Florida Onsite Sewage Nitrogen Reduction Strategies Study. Task C.25 Progress Report.* Florida Department of Health. https://depedms.dep.state.fl.us/Oculus/servlet/shell?command=view&[guid=95.38.1]&[profile=Research+Papers

Hazen and Sawyer. (2013a). *C-HS2 Monitoring Data Summary Report No. 4 – Florida Onsite Sewage Nitrogen Reduction Strategies Study. Task C.25 Progress Report.* https://depedms.dep.state.fl.us/Oculus/servlet/shell?command=view&[guid=95.35.1]&[profile=Research+Papers

Hazen and Sawyer. (2013b). *C-HS3 Monitoring Data Summary Report No. 4 – Florida Onsite Sewage Nitrogen Reduction Strategies Study. Task C.25 Progress Report.* https://depedms.dep.state.fl.us/Oculus/servlet/shell?command=view&[guid=95.36.1]&[profile=Research+Papers

Hazen and Sawyer. (2013c). *C-HS4 Monitoring Data Summary Report No. 4 – Florida Onsite Sewage Nitrogen Reduction Strategies Study.* Florida Department of Health.