Pacific Northwest
NATIONAL LABORATORY

Web-Based Data Access and Analysis Workbench to Support Environmental Restoration Decision-Making

March 31, 2021

Christian Johnson and **Patrick Royer**

Senior Development Engineer — Data Scientist

REMPLEX
CENTER FOR THE REMEDIATION OF COMPLEX SITES
@PNNL

U.S. DEPARTMENT OF ENERGY
BATTELLE

PNNL is operated by Battelle for the U.S. Department of Energy

PNNL-SA-161006

Pacific Northwest NATIONAL LABORATORY

Seminar Overview

Objective: Discuss web-based software for accessing and analyzing data in support of environmental decisions

Take-aways from today's seminar:

Challenge: Tools for environmental data access and analysis are required to meet needs for consistency, quality, communication, and decision-making.

Approach: Web-based software provides rapid access and analysis capabilities for disparate data from multiple sources.

Impact: Web-based tools effectively provide consistent analyses to support quality requirements and remedy decisions for complex sites.

REMPLEX CENTER FOR THE REMEDIATION OF COMPLEX SITES @PNNL

Outline of Discussion

- What is the nature of environmental data?
- Approaches to software tools for analysis
- Web-based applications using cloud computing infrastructure
- Analysis as support for decision making
- Web-based approaches for data visualization
- Web-based approaches for data analytics
- Conclusions, benefits, broader perspective

REMPLEX
CENTER FOR THE REMEDIATION OF COMPLEX SITES
@PNNL

Disparate Data Sources

- Data inherently comes in different forms and format
- Data access varies among organizations
- Common when multiple agency and organizational projects are participating
  - Overlapping "authoritative" data sources

REMPLEX
CENTER FOR THE REMEDIATION OF COMPLEX SITES
@PNNL

Environmental Data Types/Sources

- Electronic Tabulated data
  - Analytical chemistry
  - Well or waste site information
- Information from reports
  - Tables, figures
- Well log or geophysical data
- Sensor data
- Spatial data
  - Points, polygons
- 3D model information
- Satellite/remote sensing
  - Images, datasets

3.6.3 Proposed Input Parameters

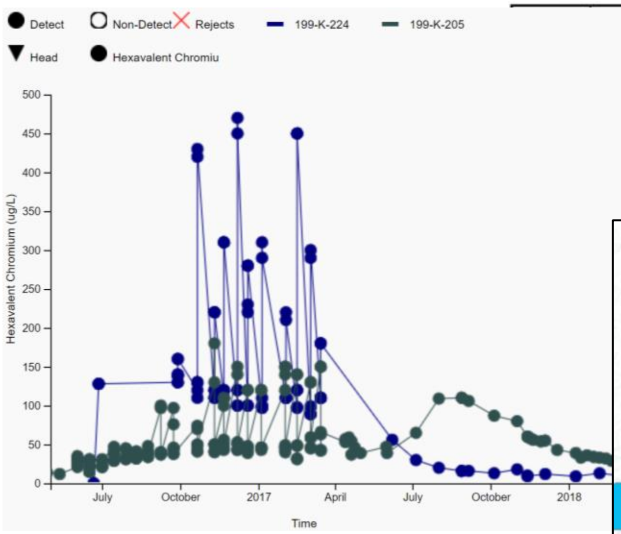
The nominal values in the governing concentration equation for the analytes used in the Composite Analysis are provided in Table 1. The values in Table 1 have been modified to have units of Ci/m³. The modification is performed by multiplying the original data in pCi/L by 10^{-9}.

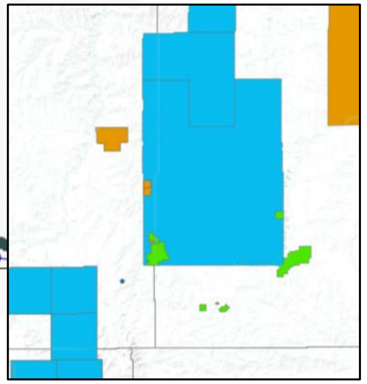
The stochastic distributions associated with the nonzero coefficients defined in Table 1 are defined using the following rules:

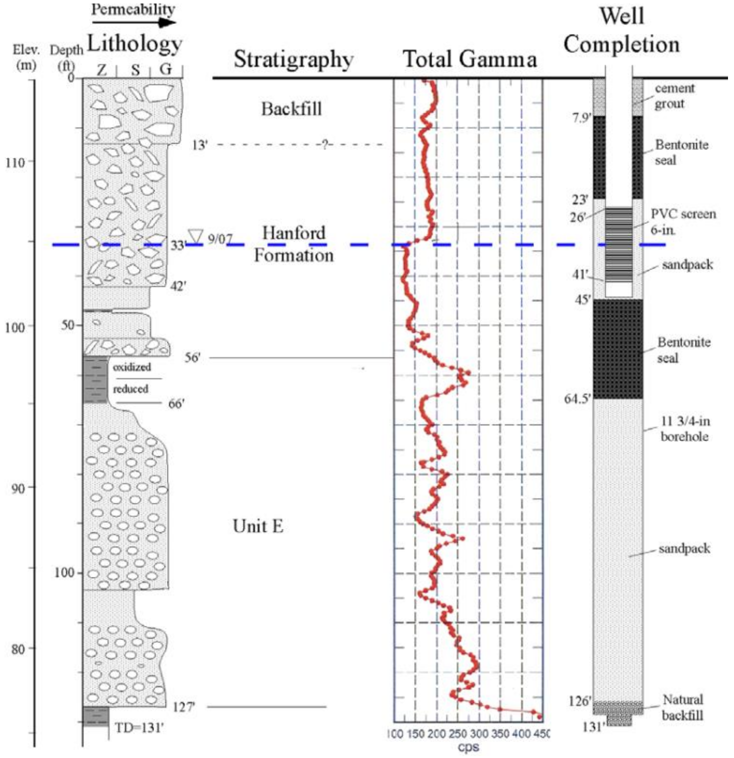
- **Variable C_B.** The triangular distribution will be used for all values of C_B. The distribution will be symmetric about the midpoint, and the half-range will be 50% of the mid-point. The variable tag will be CB.

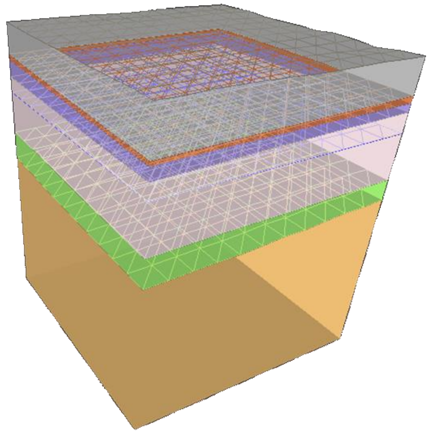
Table 1. Nominal Coefficient Values for Background Concentrations in the Columbia River

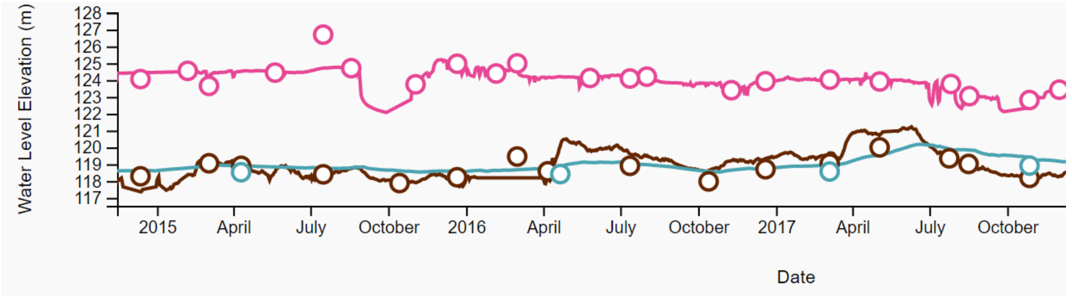
λ_B	Fallout?
0	No
0	No
0.223	Yes
0	Very small
0.0562	Yes
4.41×10^{-4}	Very small
0	Very small
0	No
Not modeled	No
0	Very small
0.0241	Yes











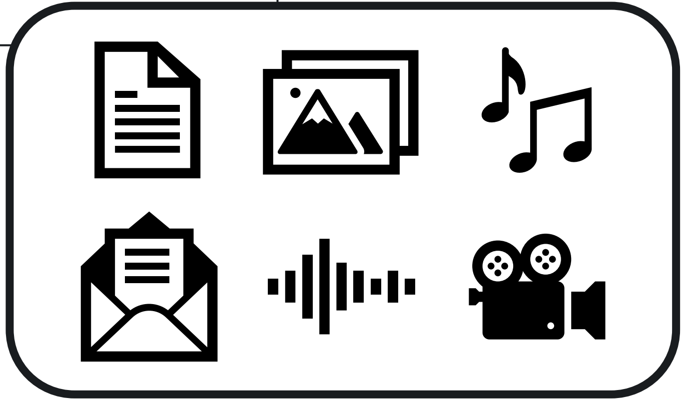
Nature of Data

- Impacts storage, access, and manipulation
- Structured data
  - Data fields in a record
  - Tabular data
  - Typically stored in a relational database
- Semi-structured
  - Non-tabular
  - Uses markers to separate semantic elements and enforce record/field hierarchies
  - e.g., JSON
- Unstructured
  - No pre-defined model/organization
  - e.g., text documents, audio, sensor data, etc.

Date	AvgVa	StdDe	MaxVa	MinVa	NumVa
12/18/2019	101.1904	10.34665	124.566	82.68952	90
12/18/2019	47.39977	0.98256	48.8643	44.81337	90
12/18/2019	28.8651	27.91818	69.87847	0	90
12/18/2019	407.5266	1.755548	411.603	403.2118	85
12/18/2019	404.7679	1.518011	408.2755	399.7396	86
12/18/2019	163.6203	0.416473	164.3519	162.2685	78
12/18/2019	166.1039	0.432044	166.8981	165.0463	87
12/18/2019	100.6407	0.102121	100.8391	100.4051	72
12/18/2019	61.09903	0.126984	61.34259	60.65538	79
12/18/2019	89.84978	0.101705	90.09694	89.62674	78

```
{
    "scenarioTitle": "Inga Site 3",
    "analysisConfigurationAlgorithms": [
    "aquiferPorosity": {
        "title": "Aquifer Porosity",
        "value": 0.3
    },
    "aquiferThickness": {
        "title": "Aquifer Thickness",
        "value": 100,
        "units": "ft"
    },
```

JSON (JavaScript Object Notation)



REMPLEX
CENTER FOR THE REMEDIATION OF COMPLEX SITES
@PNNL

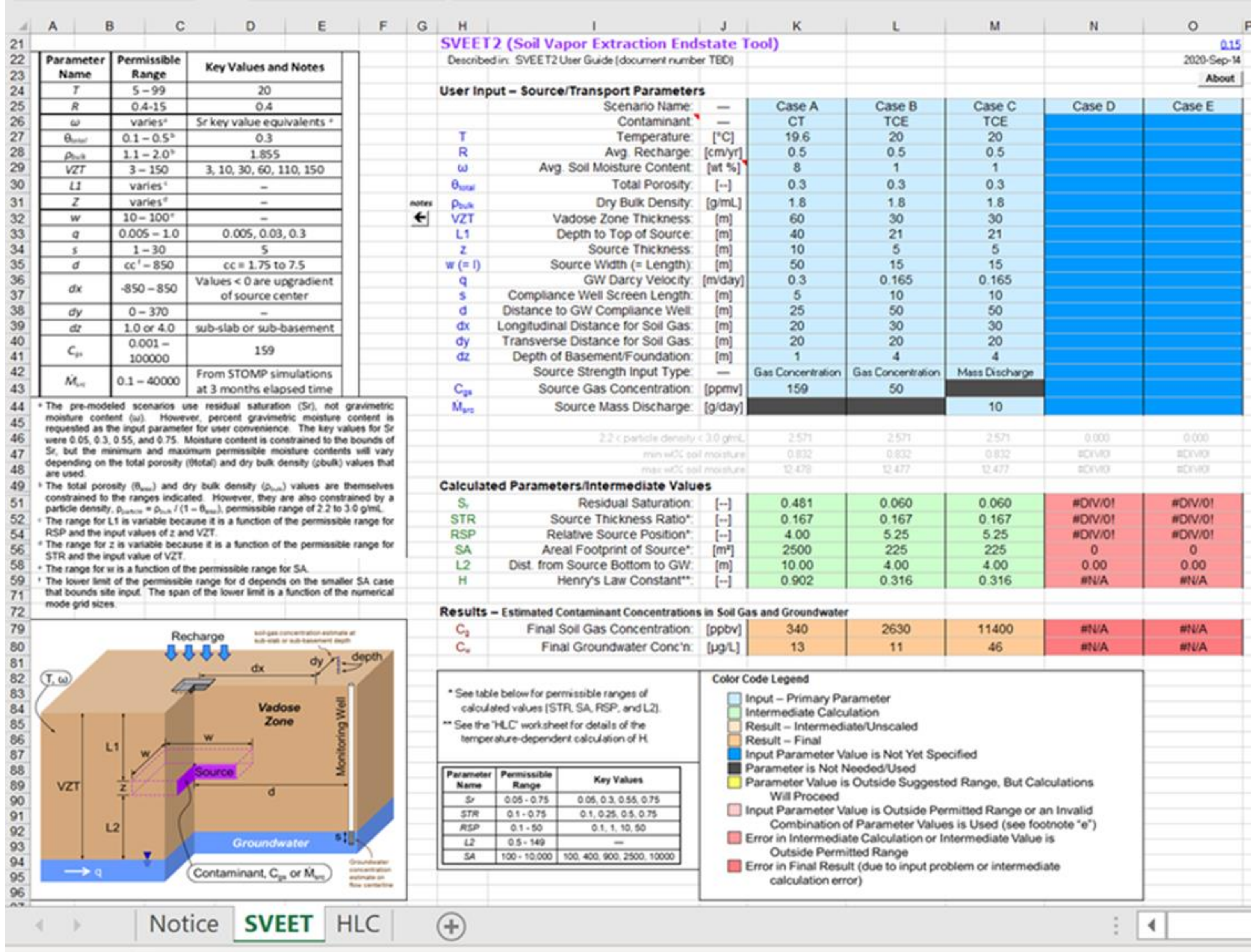
Quality Assurance – Data and Software

- Data provenance
  - How, where, when, why, and by whom was the data was produced
  - Key information for trust, credibility, and reproducibility of data and results using that data
- NQA-1
  - ASME Nuclear Quality Assurance standard
  - Structured program of procedures
    - Evaluating, reviewing, and documenting data and calculations
- Key quality concepts:
  - Traceability
  - Reviewing the work
  - Software requirements and testing

REMPLEX
CENTER FOR THE REMEDIATION OF COMPLEX SITES
@PNNL

Traditional Approaches to Software Tools

- Traditional
  - Compiled executable for desktop
  - Microsoft Excel-based tool
- Issues
  - Installation on desktop/mobile platforms
  - Users need to have specific software (Excel)
  - Potential version issues
  - Less flexible data access
  - Less flexible geospatial functionality
  - Indirect sharing of results



(Truex et al., 2013; Johnson et al., 2021a)

REMPLEX CENTER FOR THE REMEDIATION OF COMPLEX SITES @PNNL

Pacific Northwest
NATIONAL LABORATORY

PNNL Custom Development vs Commercial-off-the-Shelf (COTS) Platforms

- Geospatial based
  - Esri ArcGIS Online
    - Intuitive, programming knowledge not required
    - Free and subscription-based hosted applications
  - Kepler.gl
    - Open-source geospatial analysis tool for large data sets
- Analytical – with some geospatial
  - Power BI
  - Tableau

REMPLEX
CENTER FOR THE REMEDIATION
OF COMPLEX SITES
@PNNL

Use the Approach that Fits the Need

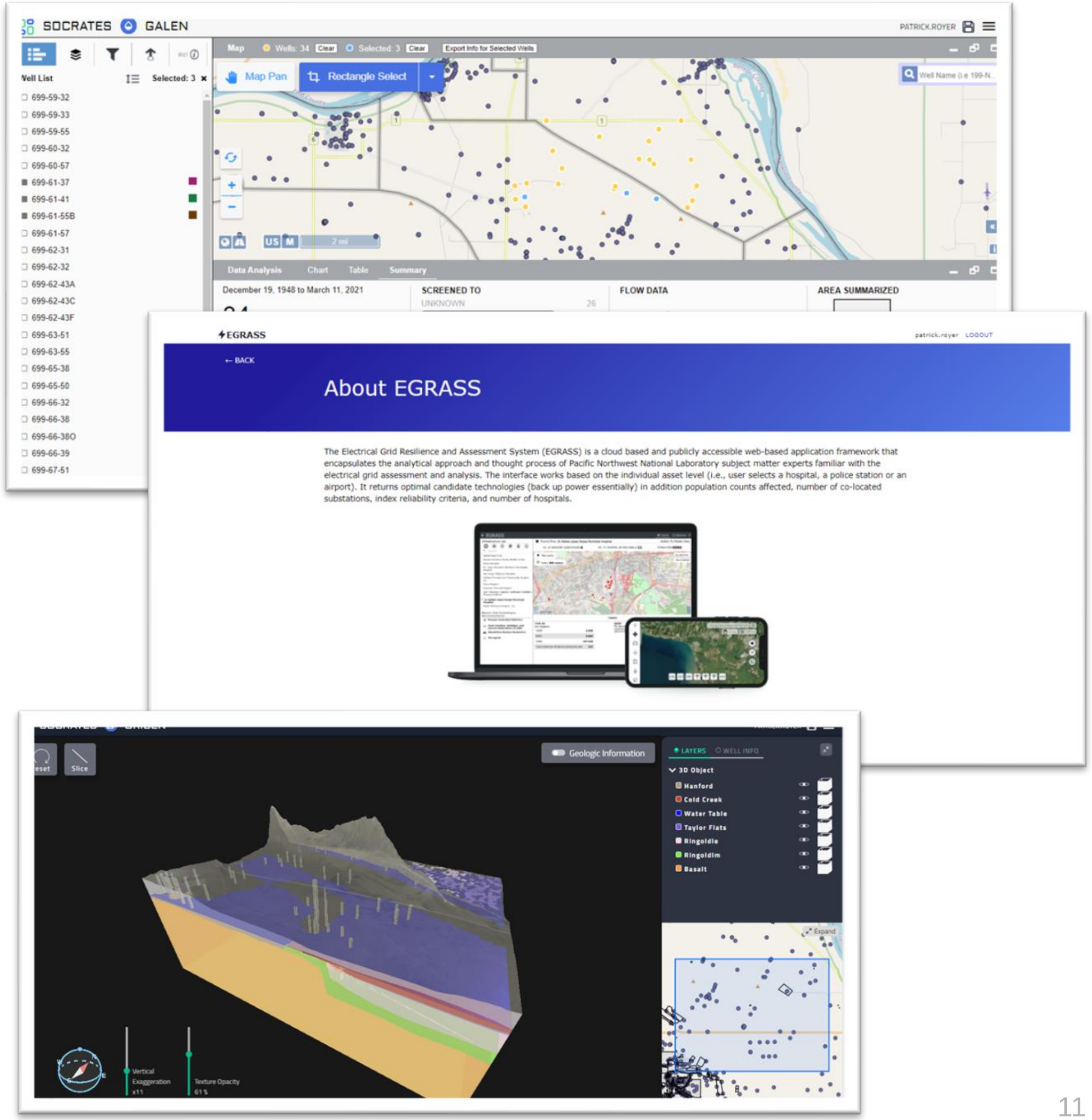
- Spreadsheet tools are suitable when
  - Users have the spreadsheet software
  - Quick development is needed
  - Familiar interface is useful
  - External data and geospatial needs are minimal
- Use exploratory data analysis when
  - Very little is known about the data
  - Objectives are vague – mainly about data capture and discovery
- Power user type tools used when
  - Analytics are independent and fragmented
  - Expertise in the tool already exists
- But…web-based applications have advantages

Pacific Northwest NATIONAL LABORATORY

REMPLEX CENTER FOR THE REMEDIATION OF COMPLEX SITES @PNNL

Web-based Applications that Leverage a Modern Cloud Approach

- Domain knowledge coupled with computational and development expertise
- Custom web applications with emphasis on cloud-based architecture
- Managed cloud services
  - AZURE, AWS, and Google Cloud
    - Government, Gov Cloud
  - Support from internal cloud team to guide application developers
- Modular and extensible platform translated to different domains
  - Environmental management
  - Electrical grid – emergency and response
  - Chemicals of mass destruction throughout Asia

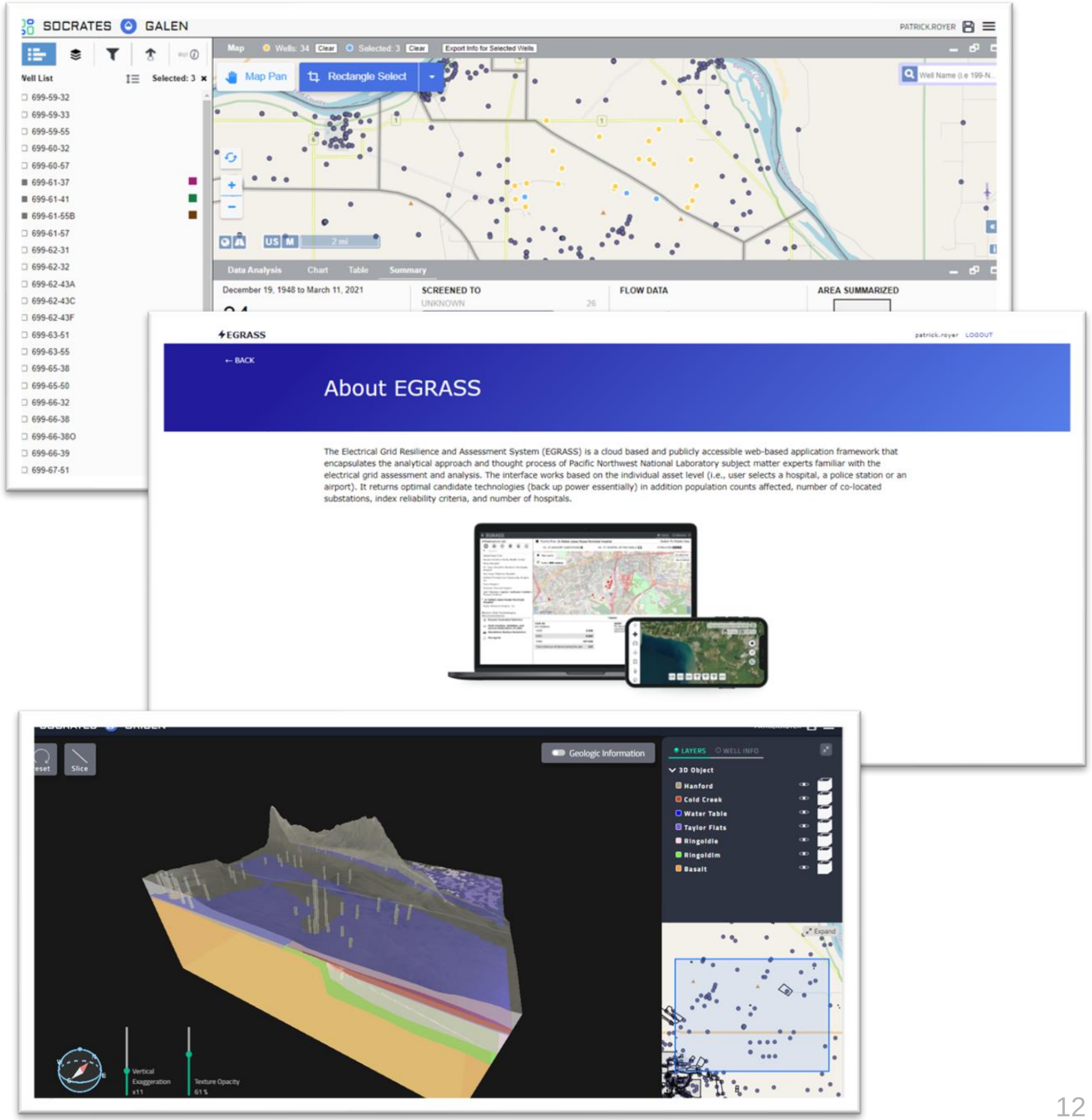


REMPLEX
CENTER FOR THE REMEDIATION OF COMPLEX SITES
@PNNL

Pacific Northwest
NATIONAL LABORATORY

Web-based Applications Leveraging Modern Cloud Approach

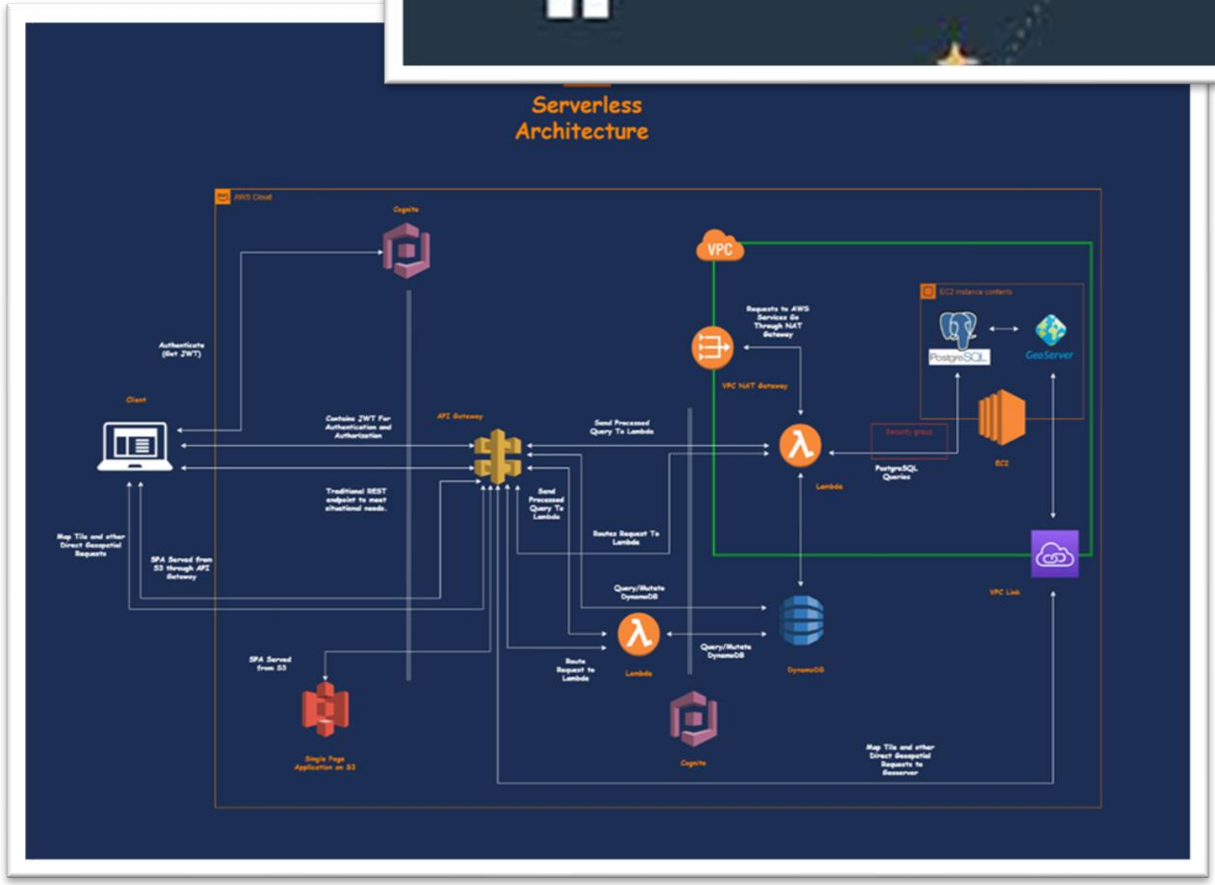
- Targeted ***vertical*** applications
  - Meet enterprise-scale analytic objectives
  - Extensive user experience (UX) and user elicitation outreach
  - Tightly coupled with regulatory requirements
  - Transferable to different cloud accounts
- Build and deploy modern, focused analytic and data services
  - Analysis as a Service (AaaS)
  - Data as a Service (DaaS)
- Role based access and profile management



REMPLEX
CENTER FOR THE REMEDIATION OF COMPLEX SITES
@PNNL

Migration to Cloud? Shifting to Cloud Architecture with emphasis on Serverless

- Serverless computing
  - Machine resources allocated on-demand opposed to virtual machine paradigm
  - Cost-effective
  - Maintenance assumed by cloud provider and customer
  - Elastic scalability
  - Improved productivity
- When to use virtual machines
- Implications for migrating from an on-premise server to the cloud
  - Legacy to modern –“lift and shift”



REMPLEX
CENTER FOR THE REMEDIATION OF COMPLEX SITES
@PNNL

Amazon Web Services (AWS) Components

- S3 – Simple Storage Solutions
  - File storage "bucket"; shared access, like FTP
- EC2 – Elastic Cloud Computing
  - Virtual machine; can be used for automated processing
- Lambda – serverless compute service; automated processing
- RDS – Relational Database Service
  - e.g., SQL Server
- DynamoDB unstructured/noSQL key-value data
- Cloudwatch – monitoring services & applications
- Cognito – authentication and user management

aws

Amazon S3

Amazon Lambda

Amazon EC2

Amazon RDS

Amazon DynamoDB

Amazon CloudWatch

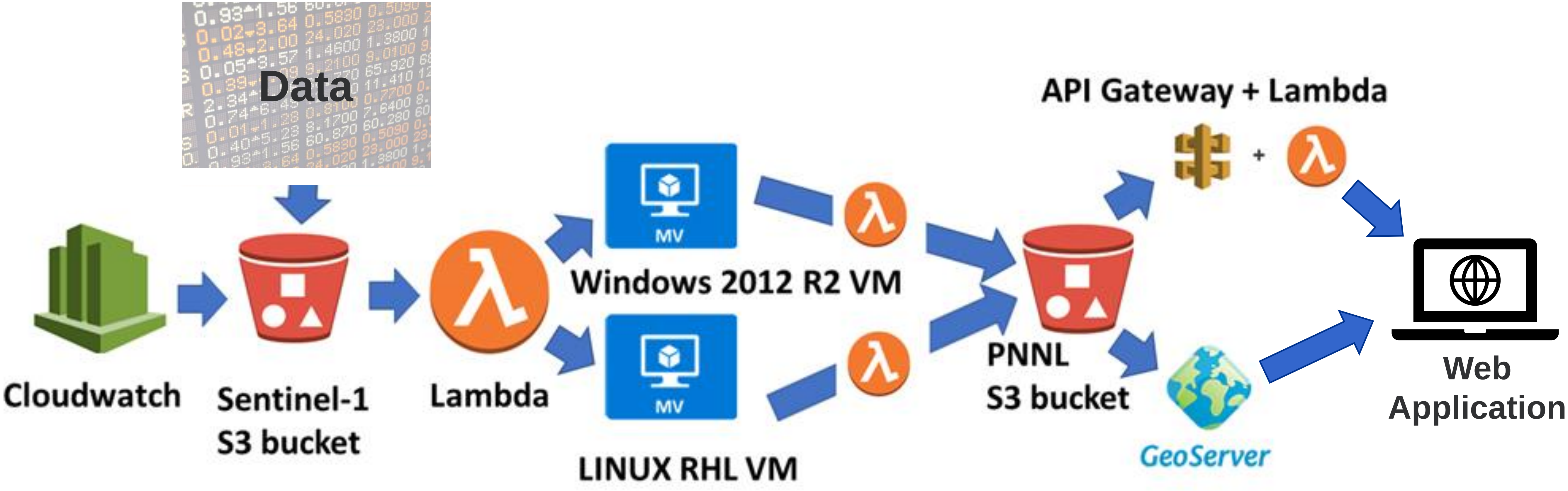
Amazon Cognito

Pacific Northwest NATIONAL LABORATORY

REMPLEX CENTER FOR THE REMEDIATION OF COMPLEX SITES @PNNL

Automated Workflow Example

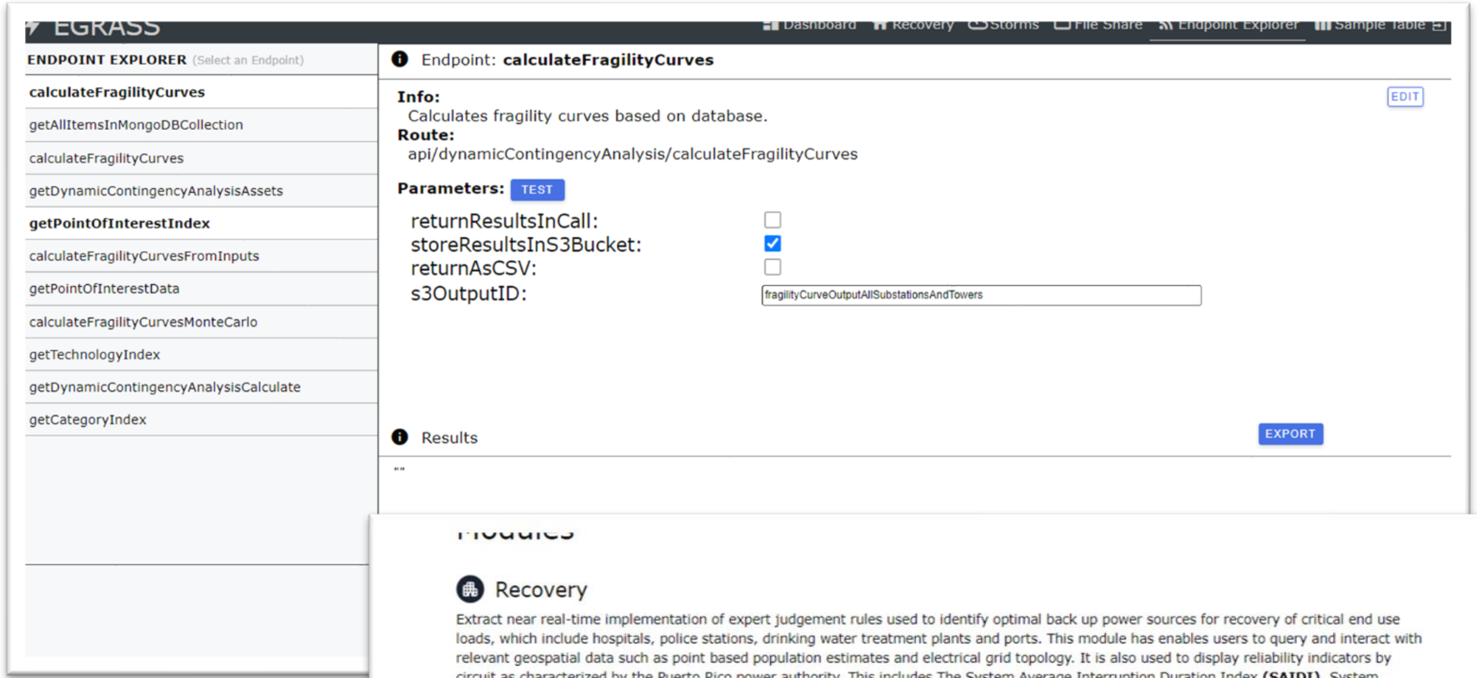
- AWS and cloud architecture for automating complex, multistep workflow

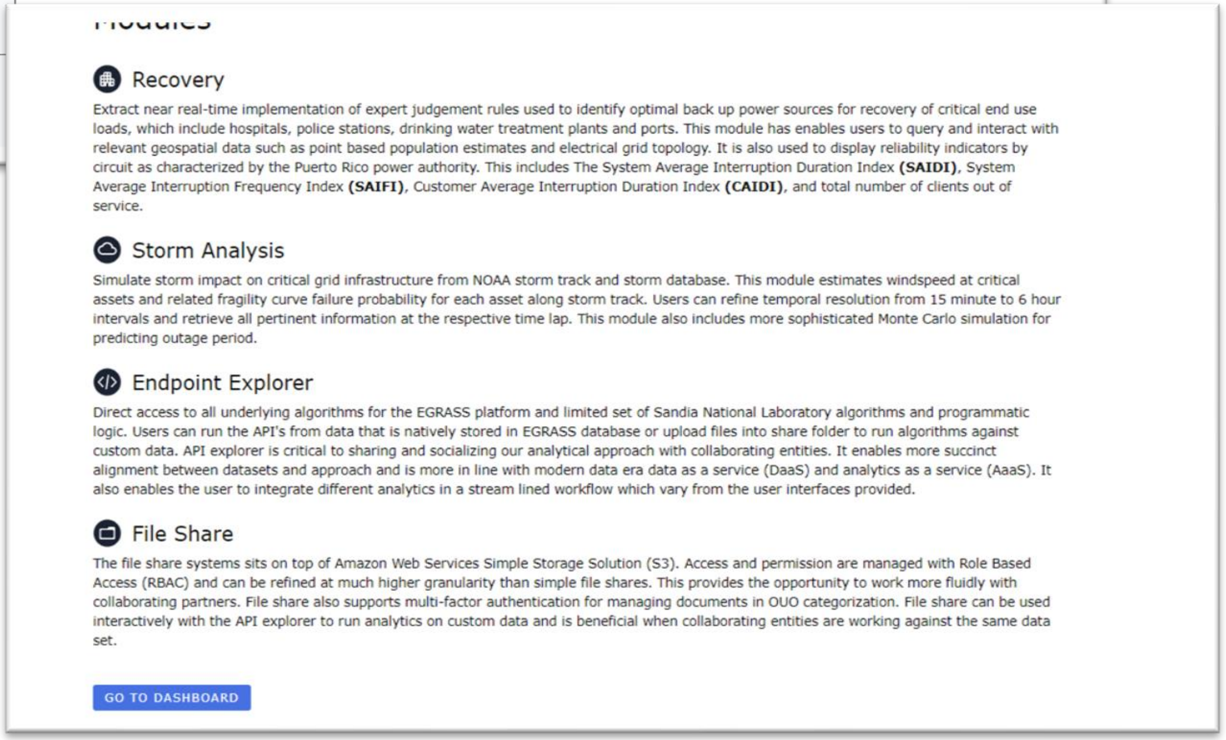


REMPLEX
CENTER FOR THE REMEDIATION OF COMPLEX SITES
@PNNL

Exposing the Application Programming Interface (API) Directly for Power Users

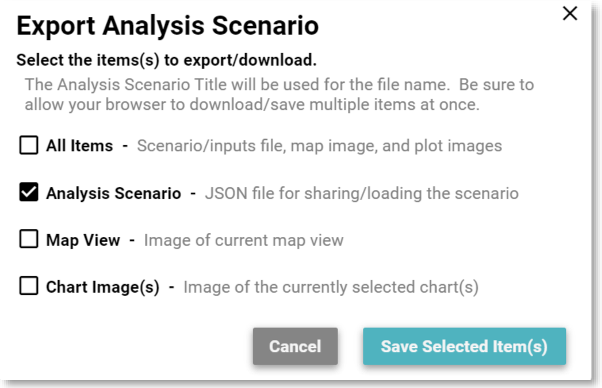
- Combining role-based access with varying data solutions and direct API access
- Data integration and assimilation
- Custom, input-driven API
- Facilitates collaboration between organizations

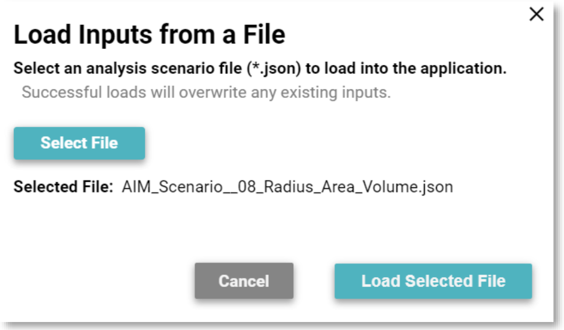


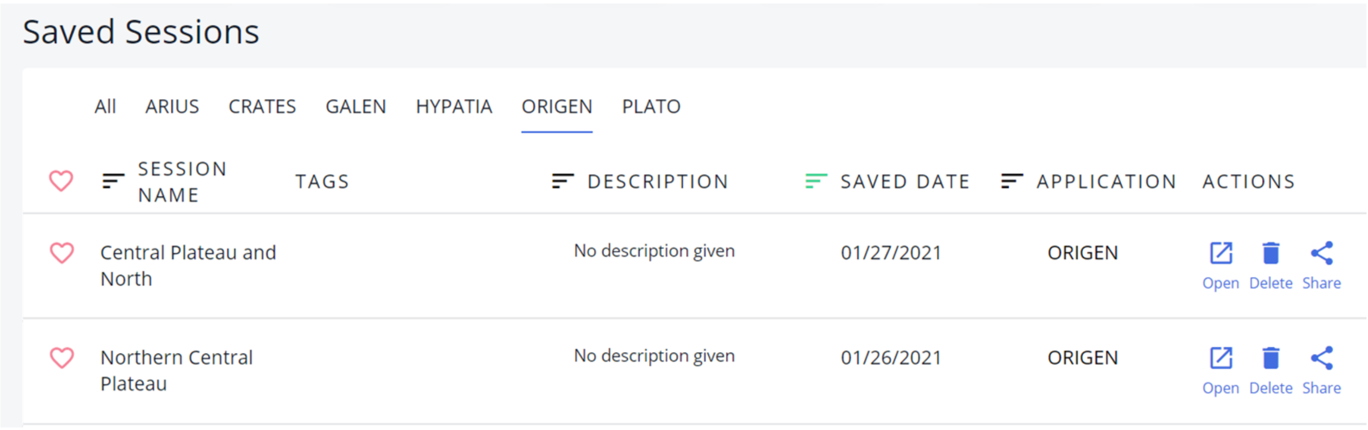


Profile Management

- User login allows user-specific customization
- Control user access based on role, organization, etc.
  - Data access
  - Functionality
  - Pre-defined styles
- Define and maintain user-specific options
  - Favorite contaminants, waste sites, etc.
  - Colors, styles, analysis options, etc.
- Save scenarios
  - Record of work, reproducibility
  - For sharing with colleagues







REMPLEX
CENTER FOR THE REMEDIATION OF COMPLEX SITES
@PNNL

Pacific Northwest NATIONAL LABORATORY

Analytics for Environmental Decision Making

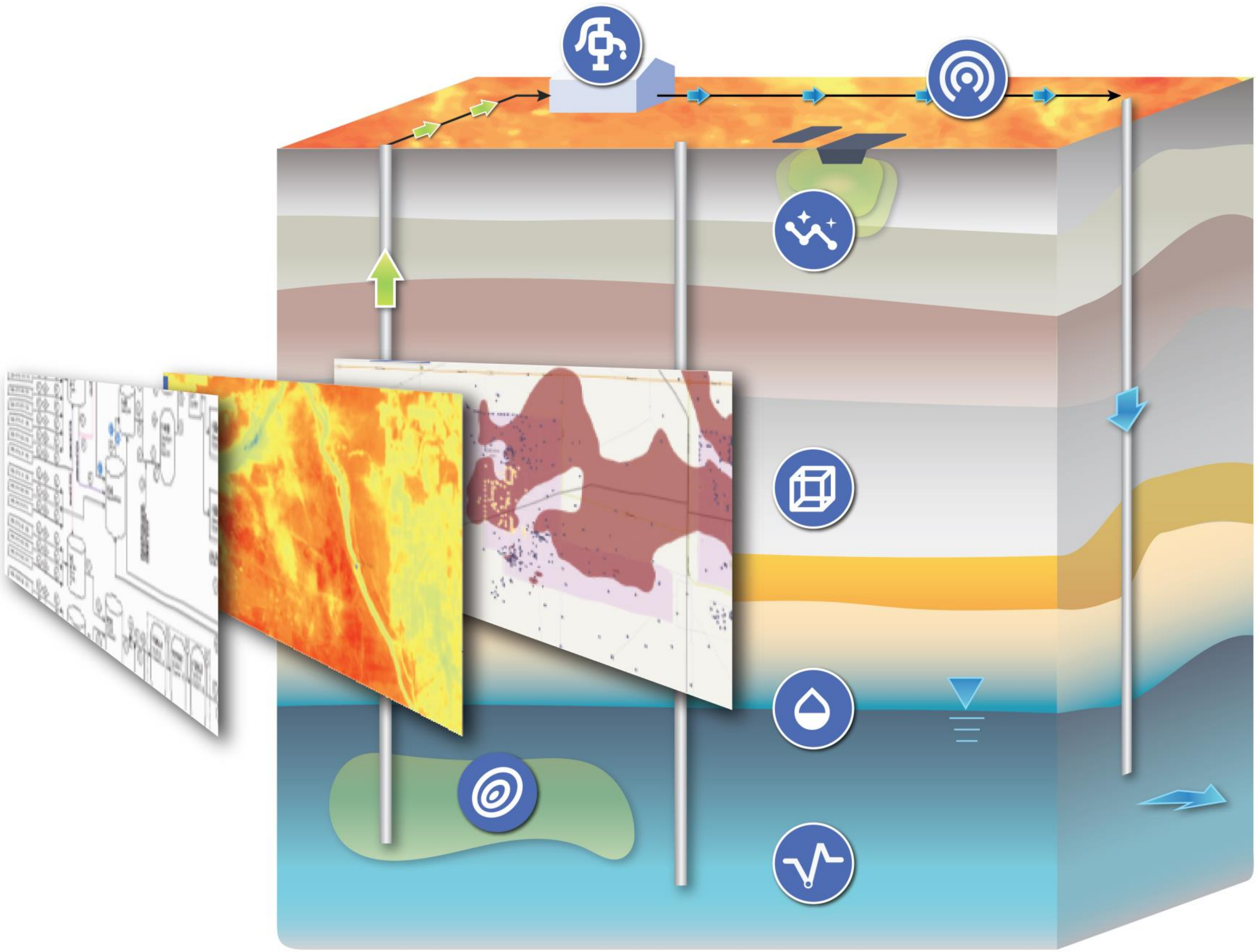
- Web-based applications provide
  - Availability, convenience, rapid assessment
  - Consistency and reproducibility in analyses
  - Inter-operability of tools and data
  - Vehicle for communication with technical, management, and regulatory personnel
    - Sharing, reporting
  - Basis for supporting environmental decisions
    - Feeds into approaches such as adaptive management of complex sites
- Examples of tools to support decision making
  - Analysis based on published guidance (e.g., EPA) and standard statistical methods
  - Data access and visualization
  - Rapid data analytics

REMPLEX CENTER FOR THE REMEDIATION OF COMPLEX SITES @PNNL

EPA = U.S. Environmental Protection Agency

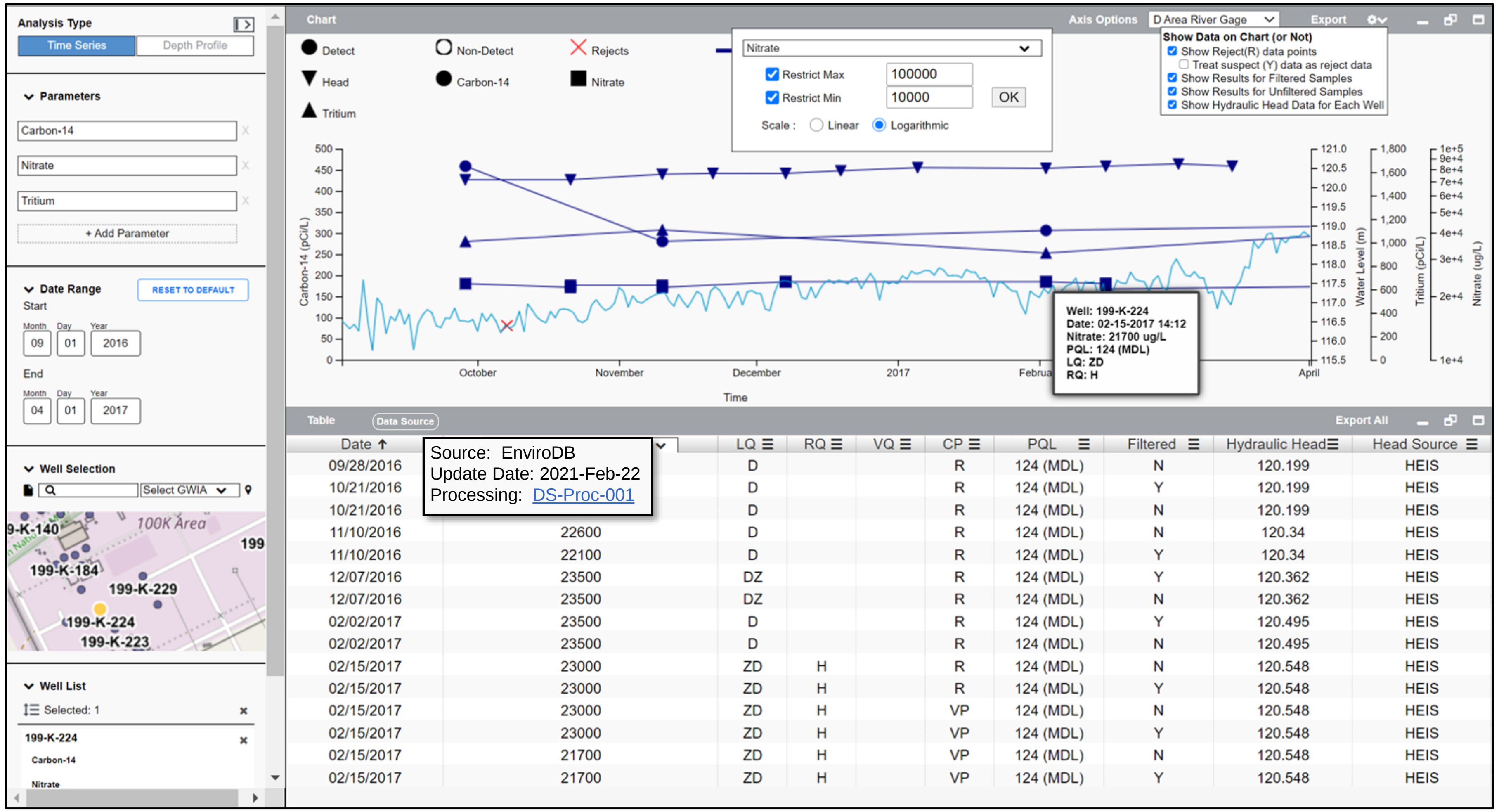
Web-based Tools for Environmental Decision-Making

- Environmental aspects
  - Groundwater
  - Wells
  - Waste sites
  - Remedies
  - Remote sensing
- Supports
  - Exit strategies/endpoints
  - Optimization
  - Adaptive remedy decisions
  - Decisions related to protecting human health and the environment



Access and Visualization of Groundwater Data

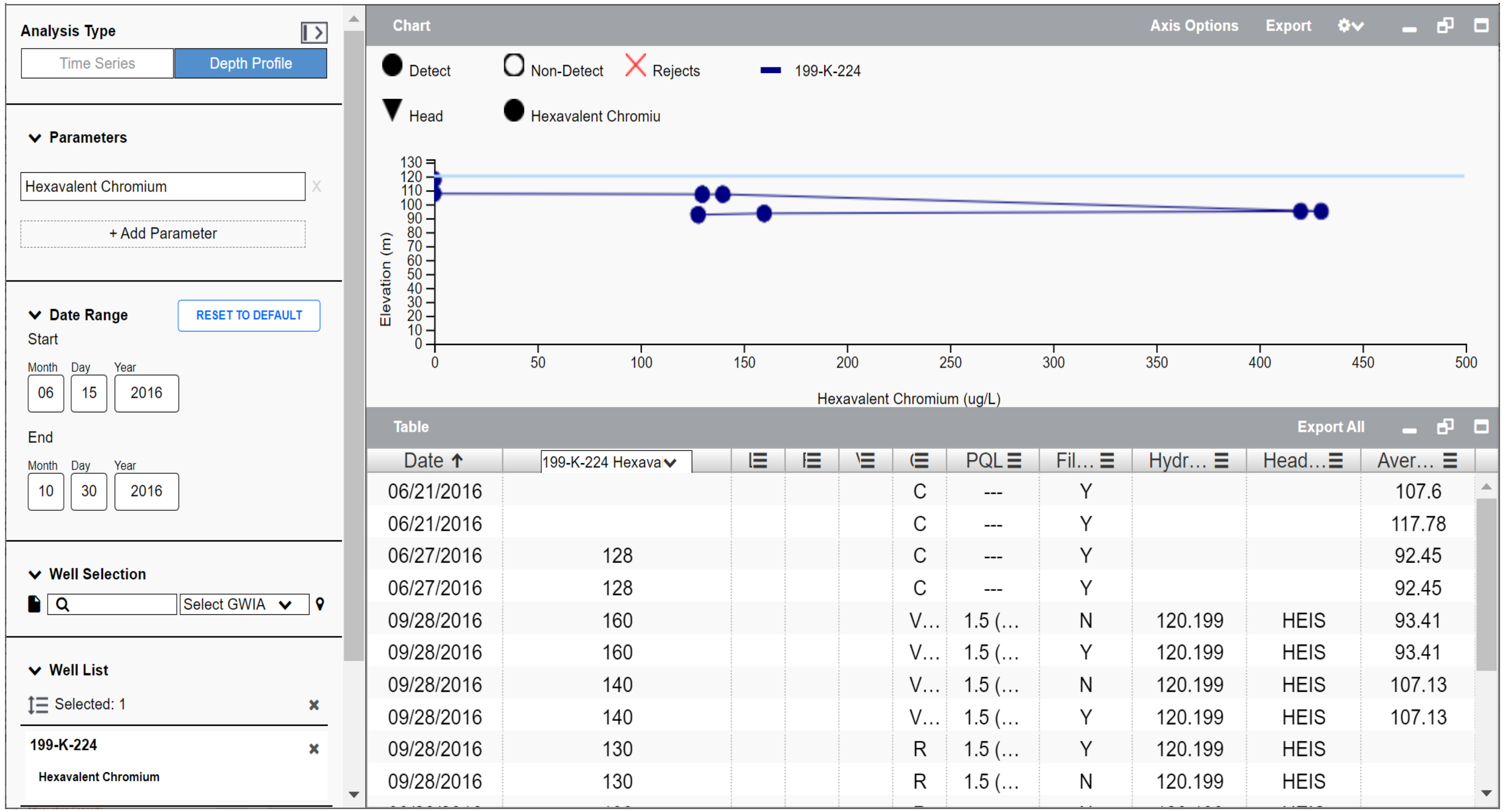
- Time series and depth profile visualizations
- Show data in chart and table
  - Multiple measured parameters
  - Specified time frame
  - Filtering based on attributes
- Water table/river levels
- Export figures/data for reporting
- Used in context of remedy and guidance (e.g., EPA, 2013)



REMPLEX
CENTER FOR THE REMEDIATION OF COMPLEX SITES
@PNNL

Access and Visualization of Groundwater Data

- Time series and depth profile visualizations
- Show data in chart and table
  - Multiple measured parameters
  - Specified time frame
  - Filtering based on attributes
- Water table/river levels
- Export figures/data for reporting
- Used in context of remedy and guidance (e.g., EPA, 2013)

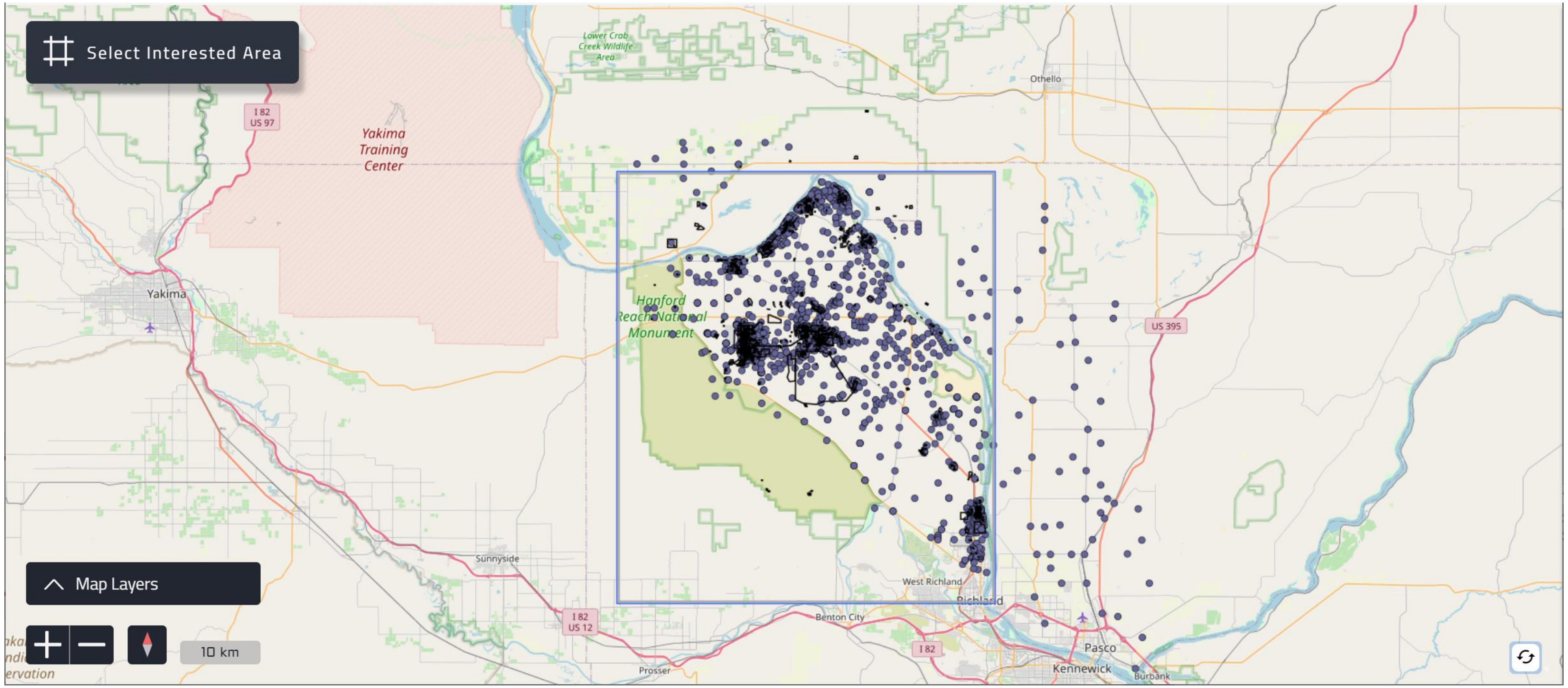


Depth profile information – can help target optimization or adaptive remedy approach

REMPLEX
CENTER FOR THE REMEDIATION OF COMPLEX SITES
@PNNL

Access and Visualization of 3D Geology

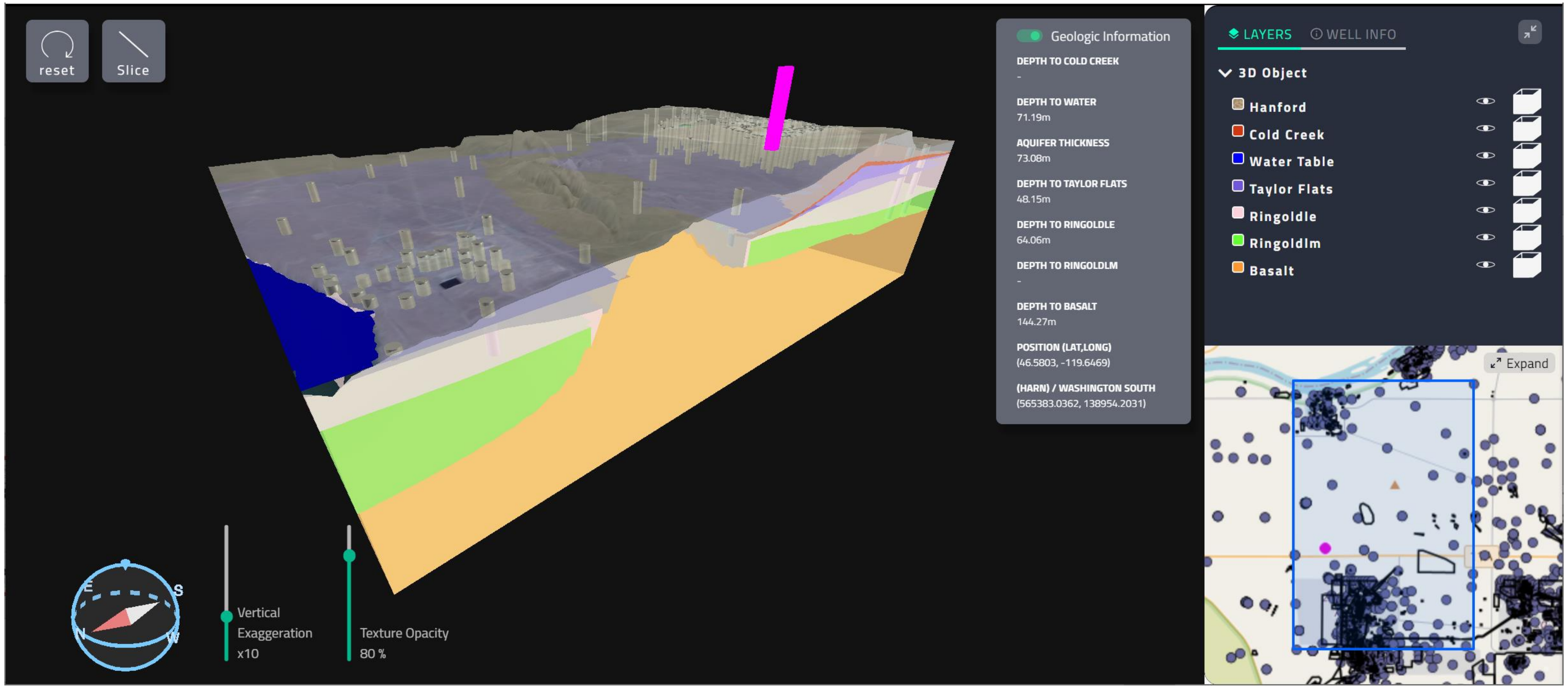
- Hydrogeological units
- Point elevations
- Cross section viewer
- Well construction information
- Link to other tools
  - Provides relevant context



REMPLEX
CENTER FOR THE REMEDIATION OF COMPLEX SITES
@PNNL

Access and Visualization of 3D Geology

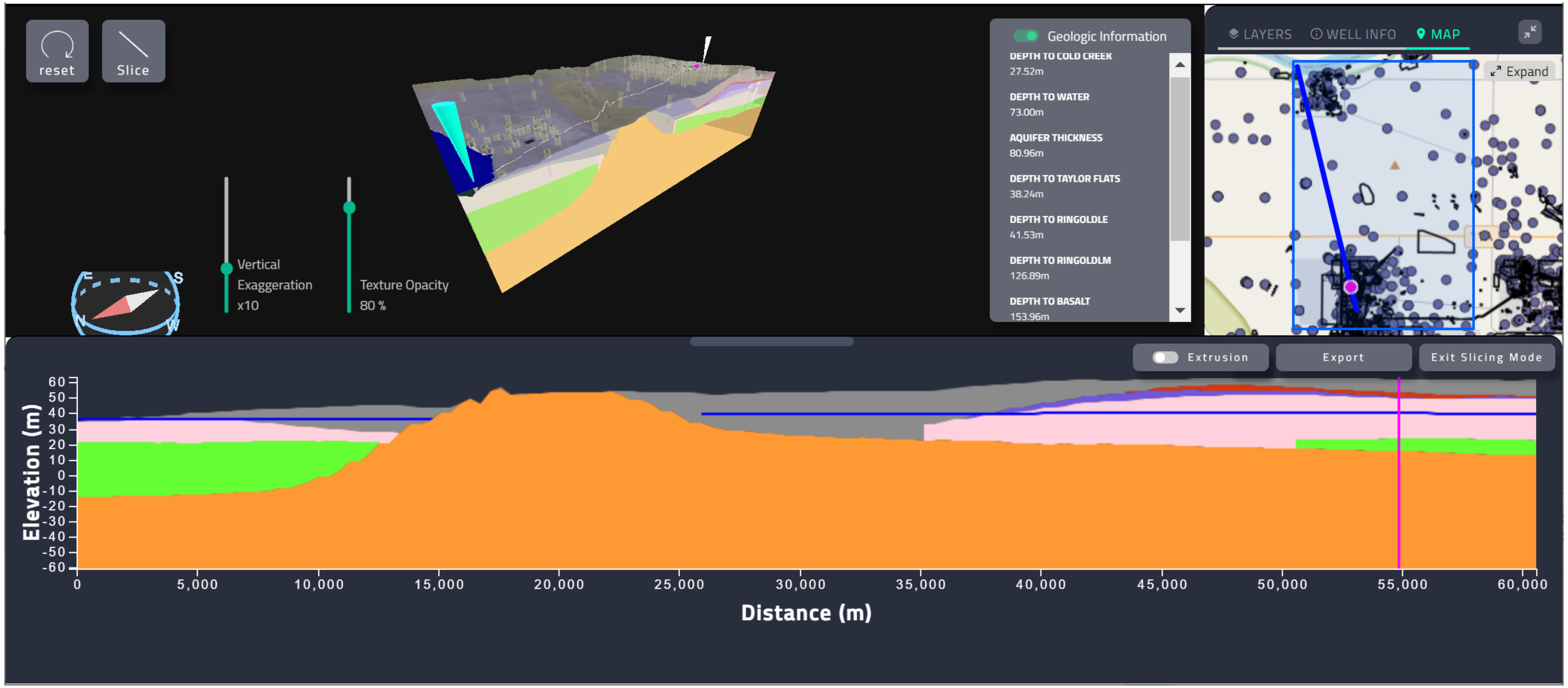
- Hydrogeological units
- Point elevations
- Cross section viewer
- Well construction information
- Link to other tools
  - Provides relevant context



Pacific Northwest NATIONAL LABORATORY

Access and Visualization of 3D Geology

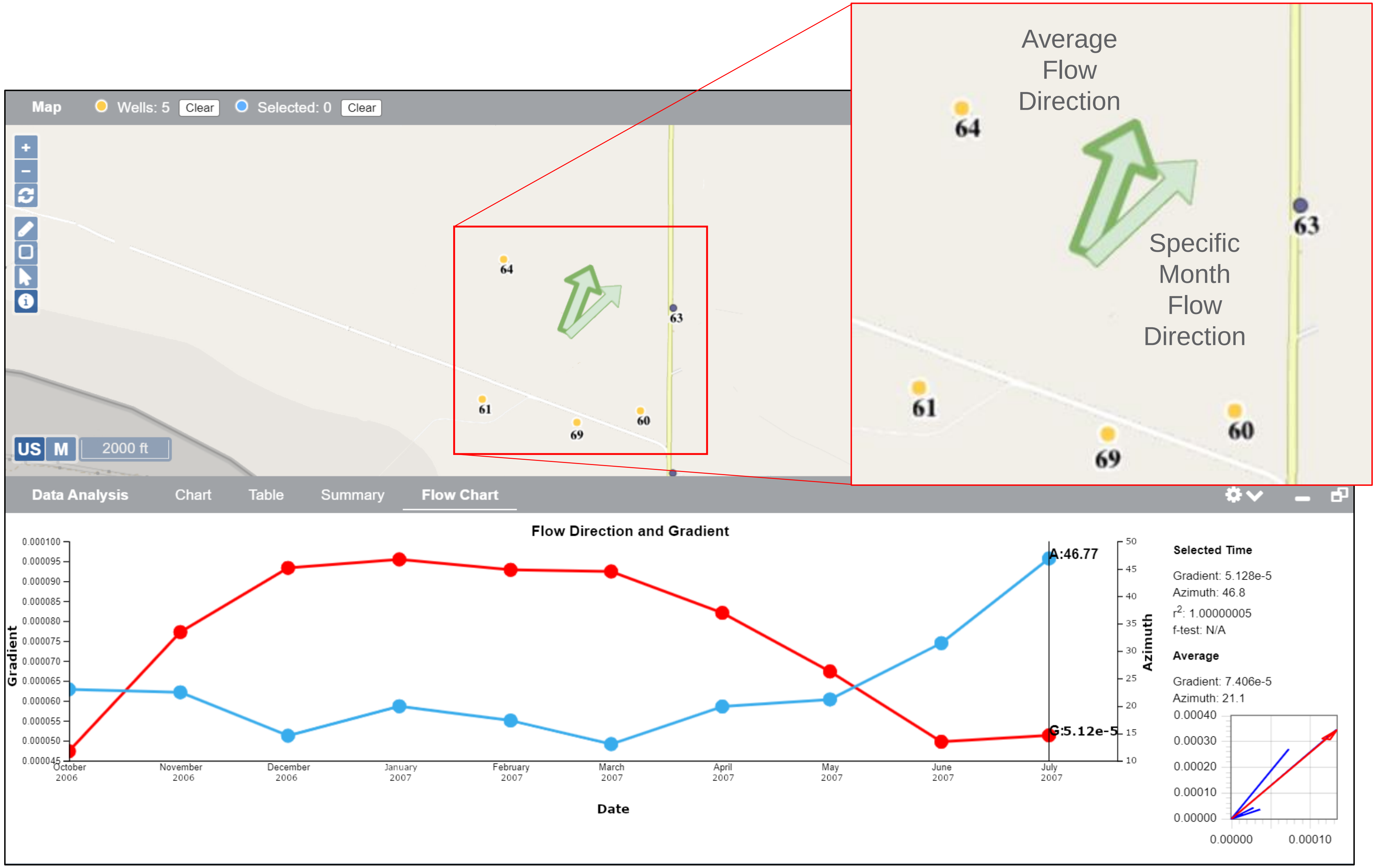
- Hydrogeological units
- Point elevations
- Cross section viewer
- Well construction information
- Link to other tools
  - Provides relevant context



REMPLEX CENTER FOR THE REMEDIATION OF COMPLEX SITES @PNNL

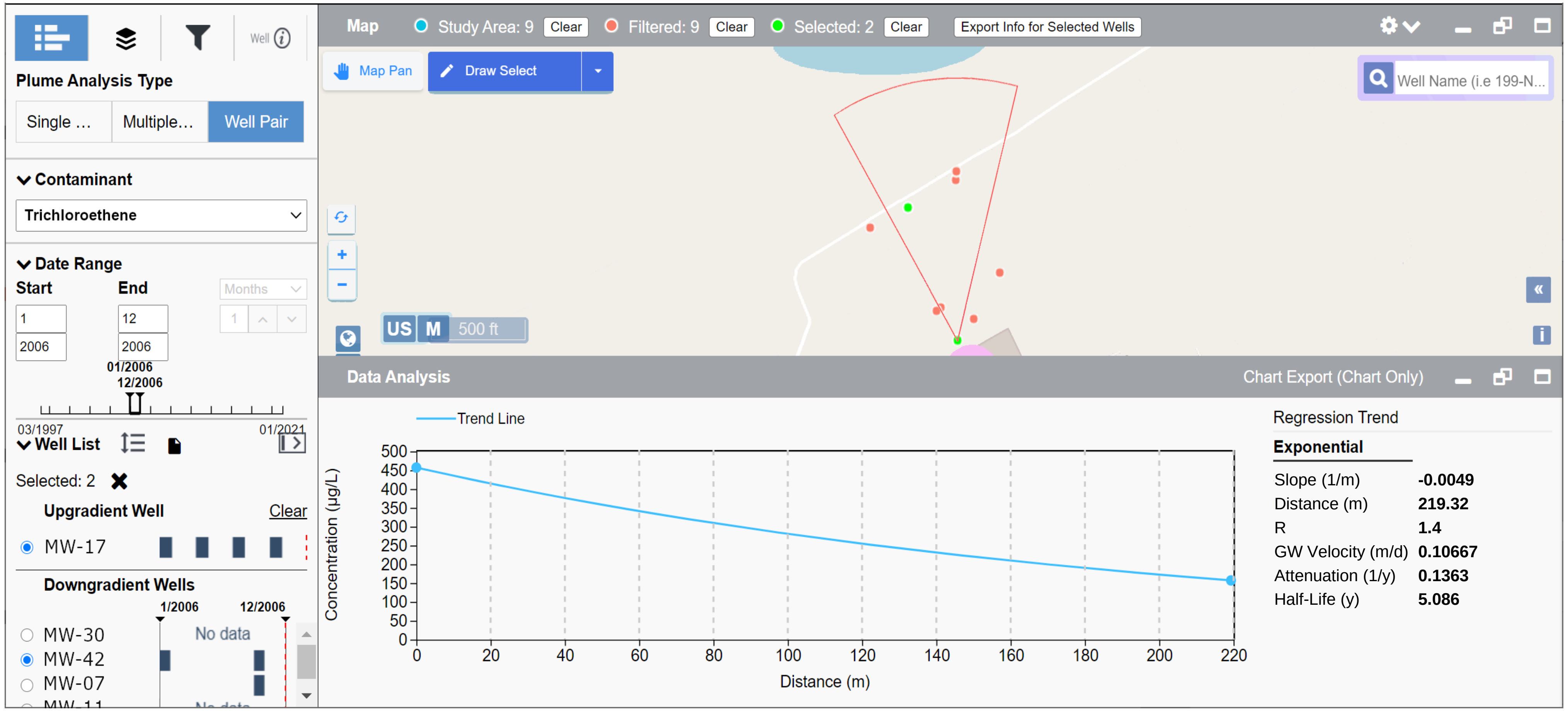
Analysis of Water Level and Flow Direction

- Groundwater elevation data
  - Key for plume understanding
- Where is contamination going?
- What seasonal effects exist?
- How fast is water and contaminant migrating?



Attenuation Capacity Analysis

- Contaminant attenuation capacity?
  - Attenuation zone between source and compliance point
  - Distance/time and attenuation
  - Under natural conditions
- Threshold source concentration
  - Or max. mass flux
- Based on guidance
  - EPA, 2002; Truex et al. 2015

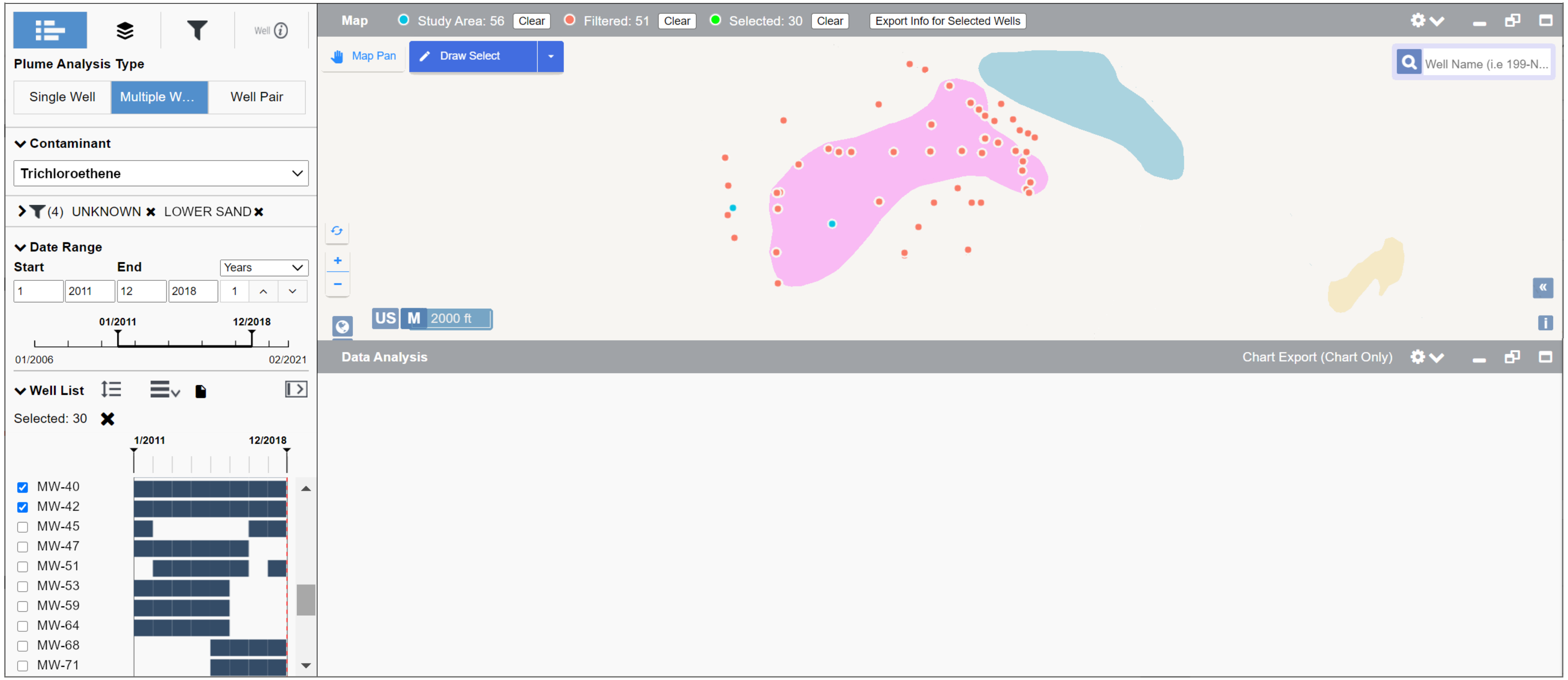


Compliance point is about twice the distance, say 400 m
Compliance point target concentration = 5 µg/L
Thus, attenuation capacity → source concentration must be 35.5 µg/L to meet target

REMPLEX
CENTER FOR THE REMEDIATION OF COMPLEX SITES
@PNNL

Plume Analysis

- Plume behavior over time
- Temporally consistent well set
- Trend/ attenuation
- Is remedy working?
  - Optimize? Done?
- Based on guidance
  - EPA, 2002; Truex et al. 2015

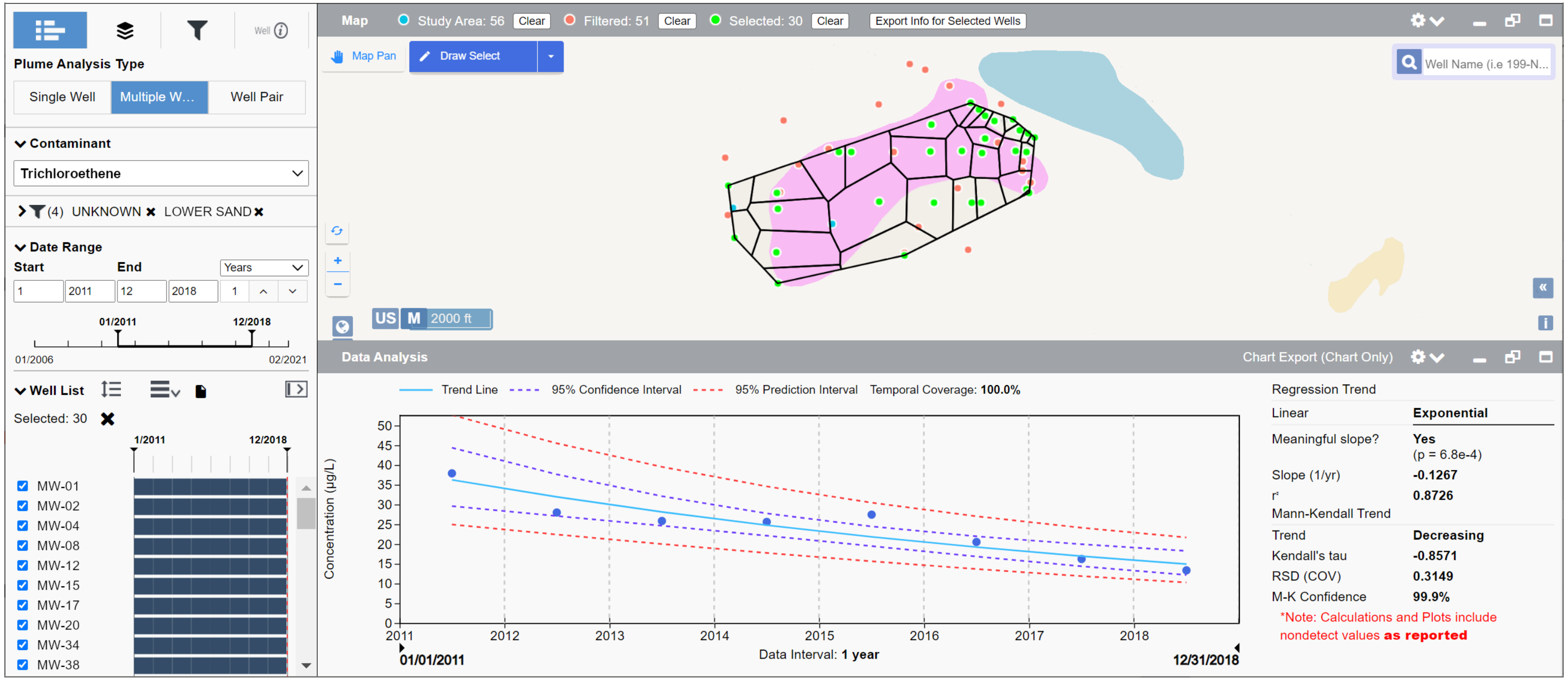


Wells sampled at different times (different temporal patterns)

REMPLEX
CENTER FOR THE REMEDIATION OF COMPLEX SITES
@PNNL

Plume Analysis

- Plume behavior over time
- Temporally consistent well set
- Trend/ attenuation
- Is remedy working?
  - Optimize? Done?
- Based on guidance
  - EPA, 2002; Truex et al. 2015

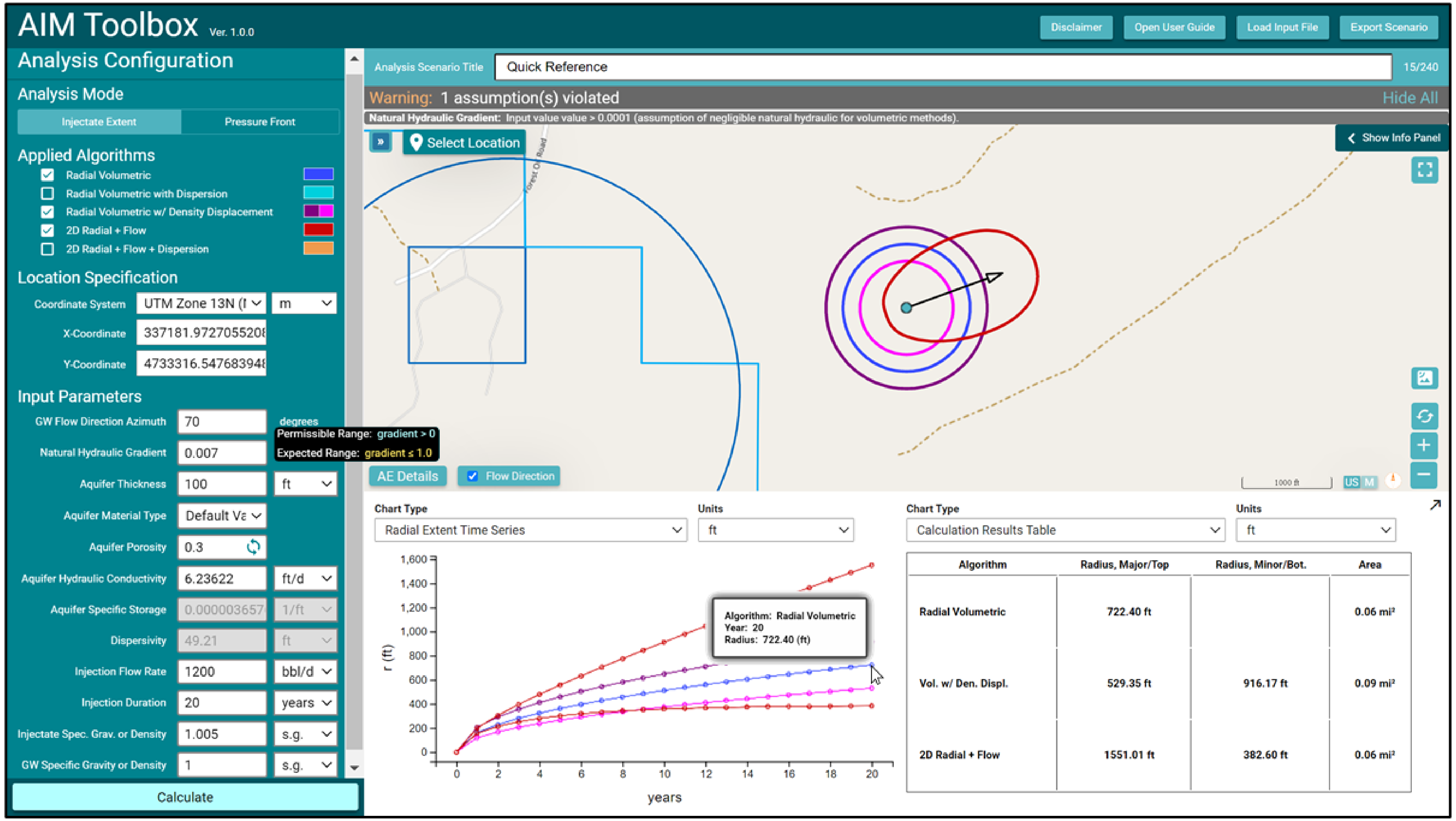


Pacific Northwest NATIONAL LABORATORY

REMPLEX CENTER FOR THE REMEDIATION OF COMPLEX SITES @PNNL

Aquifer Injection Evaluation

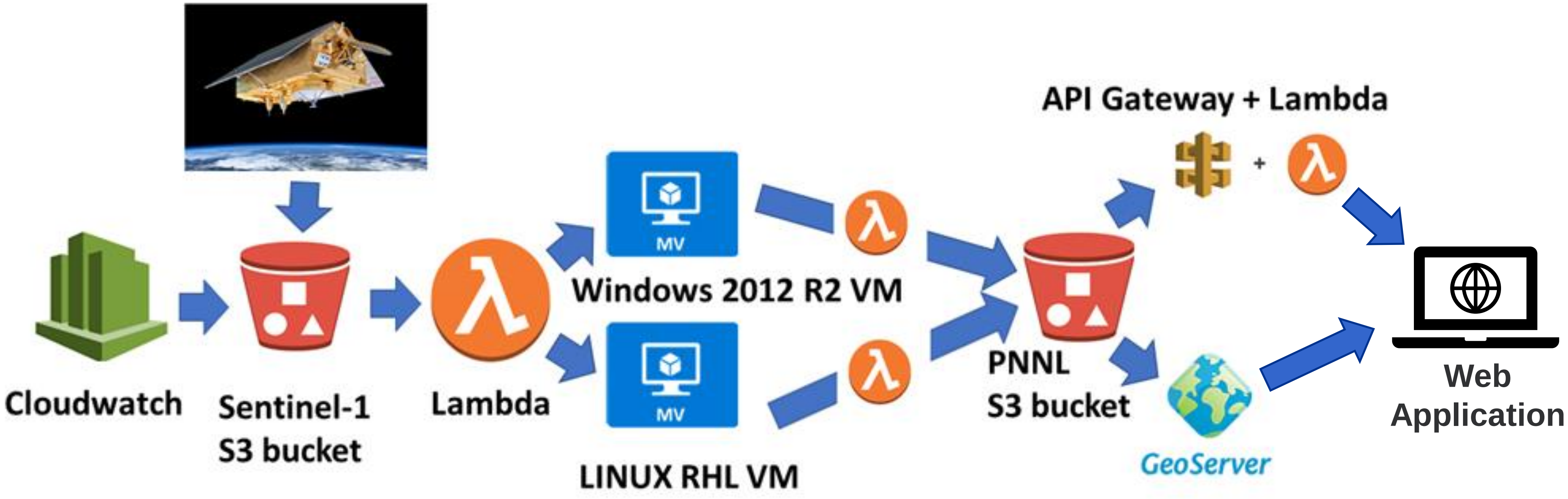
- Injection of produced water (typically a brine)
- What is the radius or area of impact?
  - Displacement
  - Dispersion
  - Groundwater flow
  - Density differences
- Used for assessing aquifer exemption applications



Aquifer Injection Modeling Toolbox – set of analytical models (Johnson et al., 2021b)

Remote Sensing Analyses

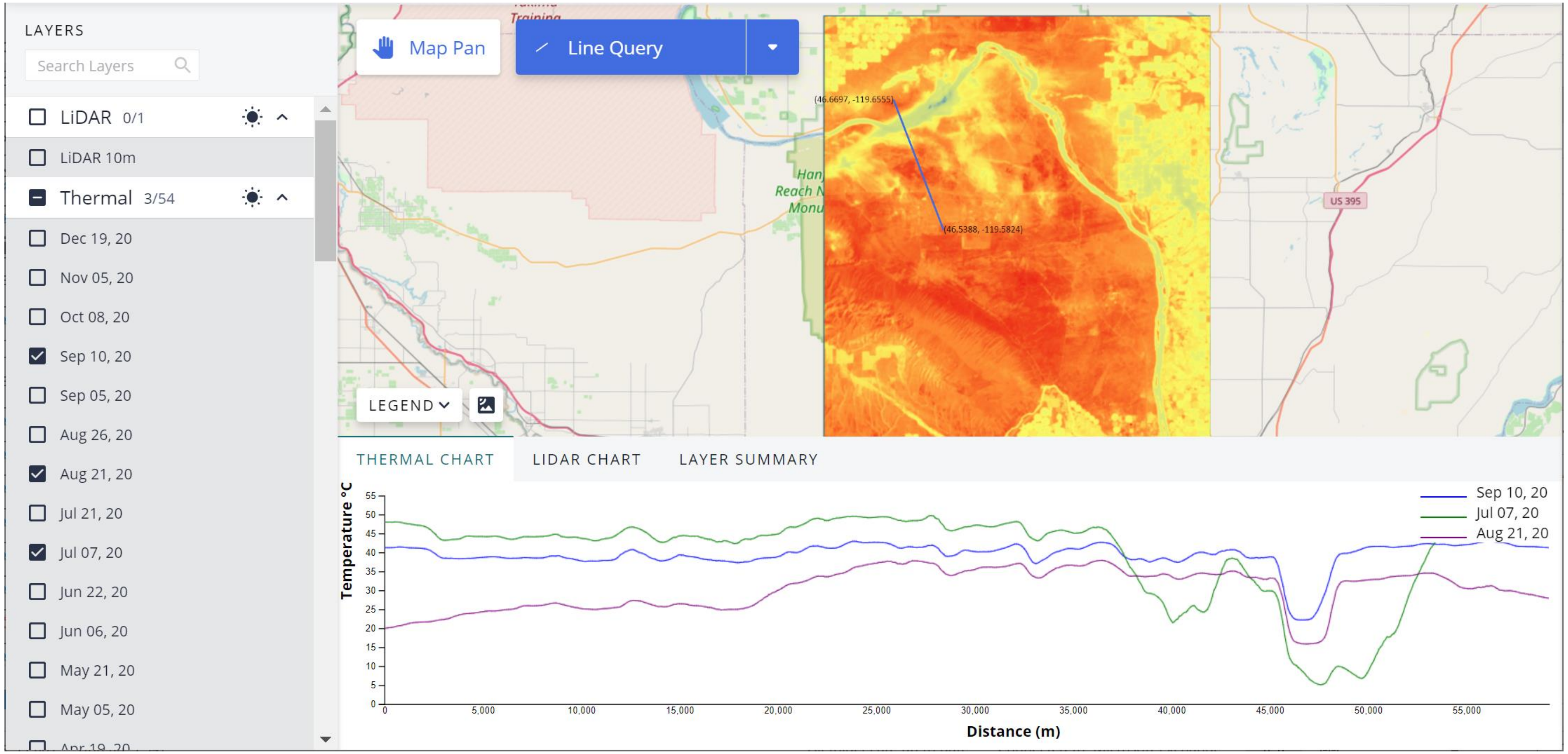
- Satellite, airborne data
  - Automated processing
- Thermal data to assess groundwater flow into river
- Subsidence detection
  - Automated alerts



REMPLEX
CENTER FOR THE REMEDIATION OF COMPLEX SITES
@PNNL

Remote Sensing Analyses

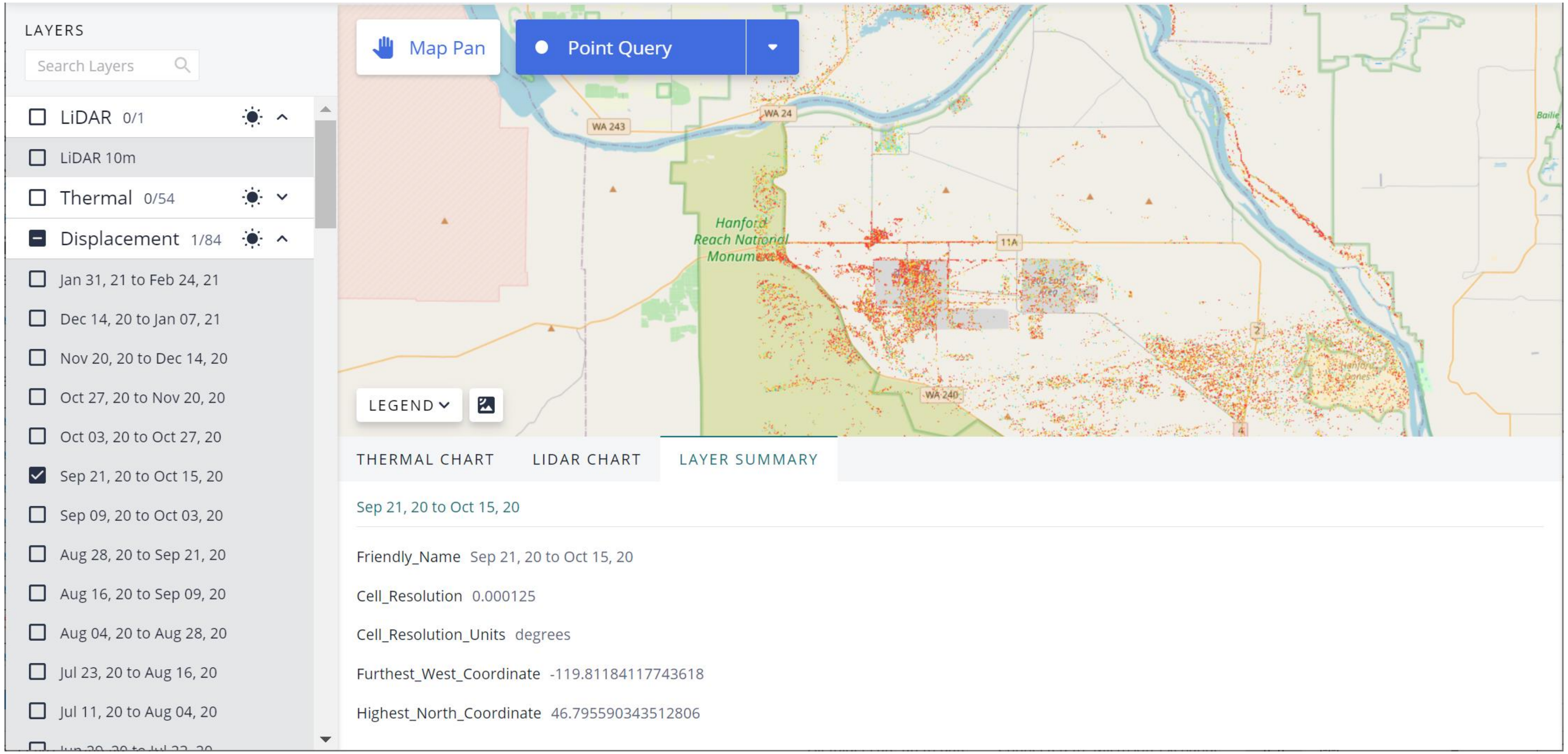
- Satellite, airborne data
  - Automated processing
- Thermal data to assess groundwater flow into river
- Subsidence detection
  - Automated alerts



REMPLEX
CENTER FOR THE REMEDIATION OF COMPLEX SITES
@PNNL

Remote Sensing Analyses

- Satellite, airborne data
  - Automated processing
- Thermal data to assess groundwater flow into river
- Subsidence detection
  - Automated alerts



Pacific Northwest NATIONAL LABORATORY

REMPLEX CENTER FOR THE REMEDIATION OF COMPLEX SITES @PNNL

Conclusions and Benefits

- Environmental professionals and managers need tools
  - Access to environmental/remedy data
  - Consistent and quality-approved approaches for analysis
  - Communication with other technical, management, and regulatory personnel
  - Supporting remedial decisions
- Web-based applications for subsurface remediation and long-term management
  - Minimize complexity associated with large data sets and analytics
  - Maximize consistency and productivity
  - Promote technical communication through sharing, reproducibility, and availability
  - Provide robust and reliable performance
  - Provide a technical basis for environmental decision-making

Pacific Northwest NATIONAL LABORATORY

Impact and Broader Application

- Data management resilience/adaptability
  - Large volumes of data (e.g., satellite, sensor, climate)
  - Variety of analytics
  - User experience – adapting to changing human interaction patterns
- Cloud computing/serverless architecture
  - Robust and rapid analytics – from analytical methods to complicated machine learning
  - Big data processing
  - Integrating disparate data types, relational and unstructured data
- Web-based tools can support decision making in a range of contexts
  - Water and other natural resources
  - Climate resilience
  - Disaster response

REMPLEX CENTER FOR THE REMEDIATION OF COMPLEX SITES @PNNL

References

- EPA. 2002. Calculation and Use of First-Order Rate Constants for Monitored Natural Attenuation Studies. EPA/540/S-02/500, U.S. Environmental Protection Agency, National Risk Management Research Laboratory, Cincinnati, Ohio.
- EPA. 2013. Guidance for Evaluating Completion of Groundwater Restoration Remedial Actions. OSWER 9355.0-129, U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response, Washington, D.C.
- Johnson, C.D., K.A. Muller, M.J. Truex, G. Tartakovsky, D.J. Becker, C M. Harms, and J. Popovic. 2021a. *SVEET2 Final Report*. ESTCP Project ER-201731, U.S. Department of Defense, Environmental Security Technology Certification Program.
- Johnson, C.D., T. Franklin, V. Molina, J.Q. Wassing, P.K. Tran, M.S. Kump, I. Demirkanli, C.M. Yonkofski, F. Zhang. 2021b. *Aquifer Injection Modeling (AIM) Toolbox User Guide*. PNNL-31087, Pacific Northwest National Laboratory, Richland, Washington.
- Truex, M.J., D.J. Becker, M.A. Simon, M. Oostrom, A.K. Rice, and C.D. Johnson. 2013. *Soil Vapor Extraction System Optimization, Transition, and Closure Guidance*. PNNL-21843, Pacific Northwest National Laboratory, Richland, Washington.
- Truex M.J., C.D. Johnson, D. Becker, M.H. Lee, and M.J. Nimmons. 2015. Performance Assessment for Pump-and-Treat Closure or Transition. PNNL-24696, Pacific Northwest National Laboratory, Richland, Washington.

Pacific Northwest NATIONAL LABORATORY

REMPLEX CENTER FOR THE REMEDIATION OF COMPLEX SITES @PNNL

Pacific Northwest
NATIONAL LABORATORY

Center for the Remediation of Complex Sites

Technical Leadership

Independent technical resource with proven track record of supporting deployment of advanced technologies and alternative strategies

Multi-institutional Collaborations

Integration and leveraging across federal and private partnerships to facilitate solution development

Solution Development

Leverage existing capabilities spanning all TRLs to provide solutions in adaptive remediation and long-term stewardship that enable risk-based remediation

Pacific Northwest
NATIONAL LABORATORY

Thank you

Christian Johnson
cd_johnson@pnnl.gov

Patrick Royer
Patrick.Royer@pnnl.gov

remplex@pnnl.gov
www.pnnl.gov/projects/remplex

REMPLEX
CENTER FOR THE REMEDIATION OF COMPLEX SITES
@PNNL