



A Design Document: Custom Analytical Tools and Automation for the Industrial Hygienist

September 2019

P Royer
M Triplett

B Simpson
K Rohlifing

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, makes **any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.** Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY

operated by

BATTELLE

for the

UNITED STATES DEPARTMENT OF ENERGY

under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service,
U.S. Department of Commerce, 5285 Port Royal Rd., Springfield, VA 22161
ph: (800) 553-6847
fax: (703) 605-6900
email: orders@ntis.fedworld.gov
online ordering: <http://www.ntis.gov/ordering.htm>



This document was printed on recycled paper.

(9/2003)

A Design Document: Custom Analytical Tools and Automation for the Industrial Hygienist

September 2019

P Royer
M Triplett

B Simpson
K Rohlfing

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

Executive Summary

This report outlines an approach for developing a modular and scalable web-based application suite, Internal Data Access and Visualization (I-DAV), that would be developed to meet the specific analytic requirements for Industrial Hygienist (IH) analysts and tank vapors Subject Matter Expert's (SME) at the Hanford site.

Through interviews and requirements elicitation with IH analyst, and on-site in-depth review of workflow patterns and expected outcomes, PNNL has developed a framework for proposed collection of interdependent goal-oriented software applications. Commercial off the shelf applications, which require only minimal custom code, were deployed this year as a mechanism of rapid prototyping and providing proof of concept. This report describes these efforts and accomplishments to date, as well as concepts for future applications and general enhancement of the DAV and core concepts for the I-DAV.

Additional content was integrated into this report as of September 2019. Updated content was added to reflect the inclusion of I-DAV concepts and ideas, and refined versions of original concepts set forth as the work commenced and continued through Q1 to Q3 in FY19. Except for analysis related to mobile Proton-Transfer-Reaction Mass Spectrometry (PTRMS) and Chemical Mixtures Model (CMM), all designs appended after Q3 reflect the final I-DAV application. Furthermore, this report is limited in the amount of design content relevant to the I-DAV because the current development site contains of this information and is available to the sponsor at <https://tank-vapors-test.azurewebsites.net/>

Acronyms and Abbreviations

AOP	Abnormal Operating Procedure
AWZ	Application Web Zone
DAV	Data Access and Visualization
DMZ	Demilitarized Zone
DRI	Direct Read Instruments
EA	Exposure Assessment
FY	fiscal year
HCN	Health Code Number
HI	Hazard Index
HLAN	Hanford Local Area Network
I-DAV	Internal Data Access and Visualization IH
IH	Industrial Hygienist
MS	Microsoft
MSA	Mission Support Alliance
OEL	Operational Exposure Limit
SDDS	Surveillance Data Display System
SME	Subject Matter Expert
SWIHD	Site Wide Industrial Hygiene Database
TVDVAV	Tank Vapor Data Access and Visualization Tools
UX/UI	User Experience/User Design
VMDS	Vapors Monitoring and Detection System
WRPS	Washington River Protection Solutions

Contents

Executive Summary	iii
Acronyms and Abbreviations	v
1.0 Introduction	1.1
1.1 Modular Approach—Analyst Tools and Support	1.1
1.2 Enterprise Application for Exposure Assessment—Development and Deployment	1.2
2.0 Background.....	2.1
3.0 Identifying Knowledge Gaps and Opportunities for Improving Knowledge Transfer.....	3.1
3.1 Identifying Organizational “Pain Points” in Data Analytics: Migrating from Manual to Automation.....	3.1
3.2 Opportunities with Sensor Integration	3.2
4.0 Data Acquisition, Synchronization, Publishing and Path Forward	4.1
5.0 I-DAV Application Concepts: Developing Tools to Support Internal Tank Vapor Data Analysis Functions	5.1
5.1 Chemical Mixtures Hazard Index Assessment.....	5.2
5.2 AOP-15 Event Chemical Concentration and Site Assessment.....	5.2
5.3 Waste Disturbing Even Impact Analyzer.....	5.3
5.4 APGEMS (or atmospheric plume dispersion model) Web Interface	5.4
6.0 Exposure Risk Calculator: An Implementation for Automated Analytics	6.1
6.1 Exposure Risk Calculator.....	6.1
6.2 Exposure Assessment Team Rapid Prototype and Tools Ready for Use.....	6.3
6.2.1 Outlier and Error Detection, Identification and Visualization	6.3
6.2.2 Exposure Risk Category Visualization.....	6.4
6.2.3 COPC and COC Quick Screening Tool	6.5
6.2.4 Shapiro-Wilk Batch Goodness of Fit & Test for Normality Tools	6.6
7.0 Updated Module Concepts March 2019.....	7.1
7.1 Automation constraints	7.1
8.0 Analysis of PTRMS data	8.1
9.0 Conclusion.....	9.1
Appendix A – Excel-based Workflow for Exposure Risk Rating Calculator.....	A.1
Appendix B – Attached as Separate Document	B.1
Appendix C Chemical Mixtures Model Method.....	C.1
Appendix D IDAV EA Results Views –and Multiple Detailed Results (Download only).....	D.1

Figures

Figure 1. DAV Explore depicting VMDS data with concentration and meteorological data.....	1.2
Figure 2. DAV Explore depicting VMDS data with concentration and meteorological data.....	4.1
Figure 3. Data Conditioning and IDAV Exposure Assessment Workflow.....	5.1
Figure 4. IDAV Data Flow and Application Interfaces	5.2
Figure 5. Chemical Mixtures Model Rendering	5.3
Figure 6. Mock up design for Waste Disturbing Event Assessment.....	5.4
Figure 7. Notional concept of migrating away from many different Excel work books to a unified application which replicates the same analytical sequence.	6.1
Figure 8. Step-wise sequence for pre-processing and analyzing data for the Risk Exposure Calculator.....	6.2
Figure 9. Exposure Assessment Workflow.....	6.2
Figure 10. Error Detection and Outlier Analysis Visualization.....	6.3
Figure 11. Field Exposure Assessment Data Selection and Visualization.....	6.5
Figure 12. Exposure Risk Category Determination.....	6.5
Figure 13. COPC and COC Quick Screening Tool.....	6.6
Figure 14. Boxplot and ANOVA analysis of ammonia data compared across sample activity (F =22.46, P-value < 0.01).....	8.2
Figure 15. Boxplot and ANOVA analysis of furan data compared across sample activity (F =1.24, p-value=0.28).....	8.3
Figure 16. Scatter correlation matrix with select chemical values (log transformed).....	8.5
Figure 17. Distribution of correlation coefficients for Wind speed, wind direction, and barometric pressure against all chemicals.....	8.5
Figure 18. Log-transformed scatter plots and best fit model, depicted as a distance from nearest tank farm for 100 m (red), 500 m (blue), and 5000 m (green) distances. Circular buffers around tank farms indicate buffer distance from tank farm centroid.	8.6
Figure 19. 10 minute moving average concentration behavior of selected leading indicators or contaminants of potential concern in 200 East.....	8.7
Figure 20. 10 minute moving average concentration behavior of selected leading indicators or contaminants of potential concern at an off-site onion processing plant upwind from Hanford. Note the behavior of NDMA in the 10 am to 11 am window.	8.7

Tables

1 Databases, Interfaces, and Update Frequency for proposed I-DAV application suite.....	4.2
---	-----

1.0 Introduction

This report outlines an approach for developing a modular and scalable web-based application suite, Internal Data Access and Visualization (I-DAV), that would be developed to meet the specific analytic requirements for Industrial Hygienist (IH) analysts and tank vapors Subject Matter Expert's (SME) at the Hanford site. I-DAV is not intended to replace mission-critical work that analysts have training and expertise to perform. Rather it would be used to automate much of the mundane, time-consuming, and error-prone preprocessing and data manipulation that is required prior to evaluation. Additionally, the I-DAV application would streamline data access and data manipulation, leading to a more consistent and unified approach for characterizing data and reporting data-driven findings. It would be configured with a multi-tiered content management approach, allowing application administration to grant and revoke permissions for components within I-DAV.

The proposed application, or suite of tools, would offer significant time and resource savings for routine reporting, data processing and data preparation. Furthermore, computational algorithms, statistics and other necessary quantitative functions would be merged into the I-DAV application framework. The proposed design allows for streamlining workflow and improving data quality for the analysts, fostering a more consistent analytic approach, with the ability to add new capabilities in a modular fashion as they mature.

1.1 Modular Approach—Analyst Tools and Support

Through the process of developing the DAV in fiscal year (FY)17 and ongoing interviews and elicitation with Washington River Protection Solutions (WRPS) analysts in FY18, Pacific Northwest National Laboratory (PNNL) was able to systematically distill step-wise workflows for various analytical endeavors used by the Exposure Assessment (EA) process and IH team. Although the end goal is to develop a cohesive, production caliber web-based application, the interview and elicitation process resulted in the development of various intermediate products which also proved to be highly useful for analysis and quality assurance purposes. These intermediate products include:

- Automated end-to-end workflow captured by the development of highly portable and maintainable programs written in R and Python scripting languages. These scripts are written in a framework optimized for collaboration, offering ease of maintenance and strong reproducibility across computing platforms using Jupyter Notebooks and Databricks. Sample outcomes of this product are attached (Appendix A and Appendix B; Appendix B attached as separate document).
- Tableau data access and visualization. This included maintaining a Tableau server instance whereby PNNL developed and deployed data visualization software, performed rapid prototyping of different dashboards or analytical reports, and made the outcome accessible to WRPS. In addition, on premise workshops, tutorials, and collaborative hands on data exploration were also included.

1.2 Enterprise Application for Exposure Assessment—Development and Deployment

In order to facilitate rapid prototyping during the interview process, PNNL configured and deployed an instance of Tableau for server and access to Tableau products was made accessible to WRPS partners by allowing connection from Hanford Local Area Network (HLAN) hosted machines to application portal on premise at PNNL (<http://vaporsih02.pnl.gov/>). PNNL met with WRPS directly on site and has provided overview and site tutorials to facilitate immediate collaboration.

The I-DAV folds the outcome of all existing products, and outcome of requirements elicitation and interviews into a single containerized application, which would be linked to streaming data sources (database) throughout the product lifecycle. This implementation would mitigate a wide range of potential issues related to data conditioning, inconsistency of interpretation and latency related to data acquisition and ingestion. Furthermore, the I-DAV would increase user accessibility and lower lifecycle cost, as a web-based application does not require licensing, specific hardware, programming knowledge, or software other than a web browser. It would also allow for managed, purposeful growth as new analytical and visualization needs were identified and approved.

I-DAV would be accessible from the DAV main page <https://www.tankvaporsexplorer.com>, but access would be restricted to registered users. Augmenting the DAV with project-side analyst tools – as opposed to developing a completely separate application (or using a suite of applications that do not naturally work together) – would add benefit in that users would maintain their entire workflow, start to finish, in the same application. The I-DAV would build on the 2017 DAV application (Figure 1) which provides direct access to data via a public viewer for selected SWIHD data sets. Internal users can currently access Vapors Monitoring and Detection System (VMDS) and meteorological data. The DAV was not developed specifically as an analytic tool, but it incorporates essential concepts and methods an analyst might need in terms of efficiently and effectively accessing and visualizing data. In addition, the development of the DAV led to a successful collaborative effort for sharing data across PNNL and HLAN network fire walls, and demonstrated compliance with rigorous security standards and requirements; extending the platform and infrastructure already in place, rather than creating a new application is a more practical use of project resources.

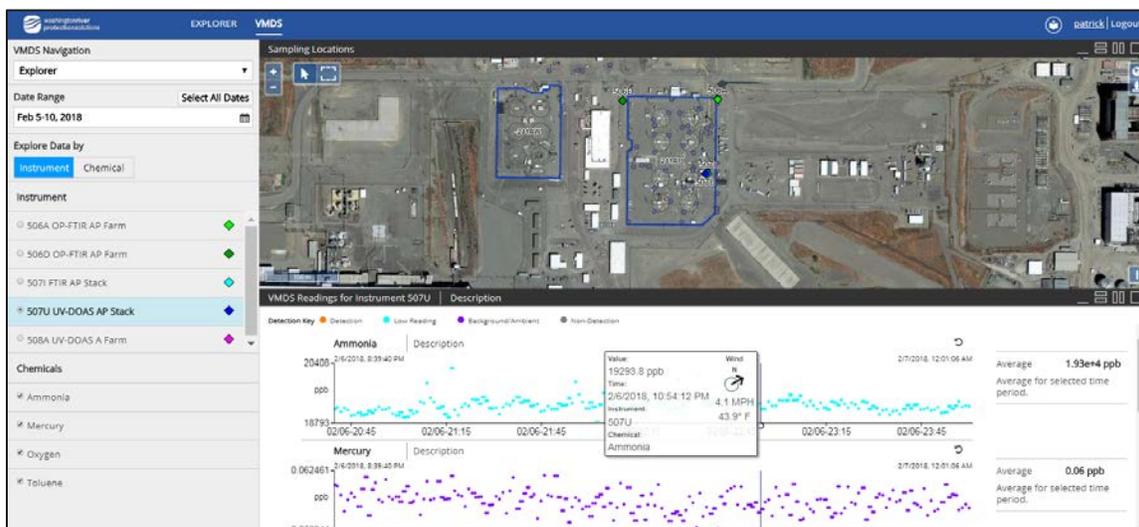


Figure 1. DAV Explore depicting VMDS data with concentration and meteorological data

The objective of this preliminary phase was to design tool concepts based on specific IH analyst requirements. In addition to proposed data conditioning workflows, or developmental design concepts, several discrete products related to the work of the WRPS Exposure Assessment team are proposed for production in FY19. The visualization and analysis tools are the outcome of a collaborative effort between PNNL and WRPS Exposure Assessment team that began in April 2018.

These analytic products have been developed using Tableau data visualization software¹ and Databricks² high performance analytic platform. Tableau is a commercial off the shelf software package used for rapid prototyping, data visualization, and reporting using production databases or user selected data extracts. Databricks is a data capture and discovery analytic platform that leverages Hadoop³ style, scalable big data analytics. In Databricks framework data is accessed and manipulated in common scripting languages, including SQL, HIVE, Python and R. The analytic process and workflow can be shared in a common notebook, which can then be published and shared to socialize and to review the technical coding approach. Databricks allows multiple scripting languages to be used in concert, leveraging the strengths of each in one environment. Using these development tools helps to quickly define the production scale workflow and illustrate how end products will function. The speed at which feedback can be incorporated into prototypes can accommodate user needs in a structured and timely fashion. PNNL will share these initial products, and has already provided assistance and support in using these new tools, resources permitting, throughout FY18. Rapid prototype tools include the following (described in more detail in Section 6.0):

1. Outlier and error detection, identification and visualization (*Tableau*)
2. Exposure Risk Category visualization (*Tableau*)
3. COPC and COC Quick Screening Tool (*Tableau*)
4. Shapiro-Wilk Batch Goodness of Fit & Test for Normality Tool (*Databricks: R, Python*)

Experience gained in 2018 with using these initial tools guided the design and development of the I-DAV application through 2019. Potential components of the I-DAV have been identified and are described in Section 6.0.

¹ <https://www.tableau.com/>

² <https://databricks.com/>

³ <http://hadoop.apache.org/>

2.0 Background

A common challenge for data analytics in large and collaborative ecosystems is the inherent fragmented, and often disparate nature of data repositories. Data stewardship and management practices, and variation of these practices between different organizations, makes this challenge even more difficult to overcome in applying analytic solutions. Furthermore, the sequence for objective-driven analysis is often bound to personal workstations. Workflows are not always documented and socialized, and become unsustainable at large scales and/or under rapidly changing data conditions or staff changes.

These limitations in technical knowledge transfer may limit the efficacy, overall integrity, and timeliness of analytic products as a result of loss of organizational resilience and data processing repeatability. Although at the local organizational level there are often “tools” in place to *download* and *retrieve* data, these efforts often represent a “snapshot” of the data at the time it was downloaded.

The results along with all of the associated data conditioning (potentially incorporating any new rules that were adopted since the last analysis) must be refreshed at each update of the database and potentially propagate through the workflow, as design choices, enabling assumptions, and/or domain knowledge changes. Operating procedures and regulatory thresholds often change over time. Furthermore, because individual analysts often develop workflow processes at a local level exclusively, using applications such as Microsoft (MS) Excel (and introducing VBA or other scripting languages), they are not easily maintained and distributed to other analysts. Even in applications that seem widely used and well supported, such as MS Excel and VBA scripts, the underlying libraries often change. Version updates to these software often render older versions incompatible with present users without sustained attention and change control.

3.0 Identifying Knowledge Gaps and Opportunities for Improving Knowledge Transfer

3.1 Identifying Organizational “Pain Points” in Data Analytics: Migrating from Manual to Automation

Analyzing data on a personal workstation that has been downloaded from a shared repository or a database with client-side access is a convention that generally worked for many years. This mode of working may be challenging to overcome for the simple reason that this approach has become the accepted routine for many analysts. The ongoing evolution of sensor technology, sensor integration, and the need to manage “big data” has made the shift away from the conventional personal workstation approach even more imperative. Applying analytics at Hanford is an exceptionally large and complex undertaking, given the wide range of resource access, characterization of disparate data sets, and level of network and data security. For the following reasons, we believe that migration to a shared, and where possible, automated workflow would offer tremendous benefit in time/cost savings, improved analysis quality, and faster cycle time response for WRPS:

- **Technology and Knowledge transfer and single point of failure:** Ultimately, the knowledge gained and the process for arriving at a defensible conclusion must be shared in a manner that work can continue and be replicated/sustained without the author of analytical workflow, if necessary. Single point of failure from a pure data analytics perspective refers to the notion that a process cannot be repeated and/or sustained. Repeatability is the underpinning of experimental sciences; sustainability is a prerequisite for project performance.
- **Data snapshot in time vs ongoing (streaming) analytics:** There is often good reason for analysis of data using a snapshot in time, such as concentration data coincident with an Abnormal Operating Procedure (AOP) event or to identify potential trends or patterns over long periods of performance. However, for maintaining situational awareness during a job or process, it is more practical and efficient to run analytics in “real time”, or as close to “real-time” as possible, so the analyses can be used to maximum beneficial effect, taking action in the moment, as necessary. In this manner a “snapshot” can still be retrieved as part of the larger data stream that is collected. The speed at which the analyses happens however, raises a legitimate concern that data has not been sufficiently pedigreed, scrubbed, and/or filtered before analytics are performed. While this concern is warranted, we believe that, with time and dedicated effort, a large majority of data pre-processing can be automated in order to provide value in near-real time for decision-making. In cases, where the engineer analyst still has valid reason to be concerned about the results of the automation, an “alert” system can be established that points the scientist/engineer to potential off-normal or out-of-specification conditions.
- **Client-side libraries and custom code version control:** A common frustration in a workstation environment is that an individual may invest a great deal of time, thought, and energy into a value-added custom algorithm, only to learn that the same function will not run on a different computer or architecture or scale as the input data features scale. This is perhaps one of the most important incentives for migrating work to an enterprise (server-client) paradigm. Web applications are becoming increasingly utilized and preferred over native applications because of this ‘works anywhere’ feature. Applications that appear “native” often have a prominent web based component, such as Windows Office 365 that is familiar and maintains the same capabilities as a desktop version, but is always up to date with respect to configuration control.

- **Transcription Errors and Error Propagation:** A majority of current analytic work is derived from working with spreadsheets, with much of the pre-processing involving tedious aspects of data formatting and data table population, such as copying and pasting, referencing other tables, filtering columns, and other repetitive functions. In processing large data sets, some of these errors can be very difficult to recognize as they occur and may be propagated as others become involved with the same data and workflow.

3.2 Opportunities with Sensor Integration

Through the course of interaction, meetings, and one-on-one interviews with both WRPS and PNNL analysts, it has become clear that there are sensor data available that would be valuable to analyze, but these data are not being preserved and maintained in any systematic fashion (at least at the time of this writing). Consequently, there is no mechanism to retrieve this information for analysis at the same granularity that data is typically captured. A series of wirelessly integrated direct reading instruments (DRIs) are being tested for use at the Hanford Site. These are mobile – or fixed, opportunistic sampling sensors that collect airborne species concentration data. These data are not being transferred electronically, but are being quantitatively summarized and manually entered by field technicians. Much of the data collected from these instruments are in fact “reading zero”, “NULLS”, or “non-detect”, and are thought to be irrelevant (or merely noise). Valid concentration measurements are being summarized (i.e., average, daily maximum) as part of the field campaign and are then transcribed manually, and entered into a database.

Capturing and archiving these data streams at the highest granularity possible offers several potential advantages, requiring few or no incremental resources. With big data management, “cold” and “hot” storage are terms and techniques applied to data warehousing to account for data that have relatively low perceived value and will not be retrieved often (*cold storage*), and data that are highly relevant and that will be actively used in making decisions (*hot storage*). There is significant value in archiving data even if it is in a test or pilot phase. For instance, machine learning techniques typically need large quantities of data to calibrate and test performance before fielding a production version algorithm. PNNL has been involved in many large scale sensor integration projects, and could help directly with this effort.

Similarly, there is a pressing need to know the exact location where vapor sample data are acquired in order to properly apply atmospheric dispersion models. However, the descriptive information in the database is not geotagged. Currently, there are options for geo-encoding in many sensory devices, and there may be additional opportunities with existing, and to-be deployed, equipment to improve fidelity and resolution of current analytic techniques. This data application presents another area of expertise at PNNL and the inclusion of geolocation would benefit most analytical processes. Of particular importance is the quantifying the location of the species concentration data relative to its potential (or actual) emission source(s).

4.0 Data Acquisition, Synchronization, Publishing and Path Forward

There are a variety of approaches for gaining access to off premise (outside of PNNL) databases and maintaining a synchronized connection. PNNL has been working with site partners for many years to develop strategic opportunities in this domain. Direct access, such as a linked server or similar technical approach, almost always requires participation of the Mission Support Alliance (MSA) contractor in order to add, update, or change firewall rules.

In the near future MSA will no longer permit direct internet-based access to Hanford Site databases. All data traffic will need to be routed between servers via an extranet service provided by both parties. At present this is referred to as the Demilitarized Zone (DMZ) from the MSA side, and the Application Web Zone (AWZ) from the PNNL side. While some of these connections cannot be put into place with the current network security stance, PNNL has implemented other mechanisms for connecting to source data (Figure 2).

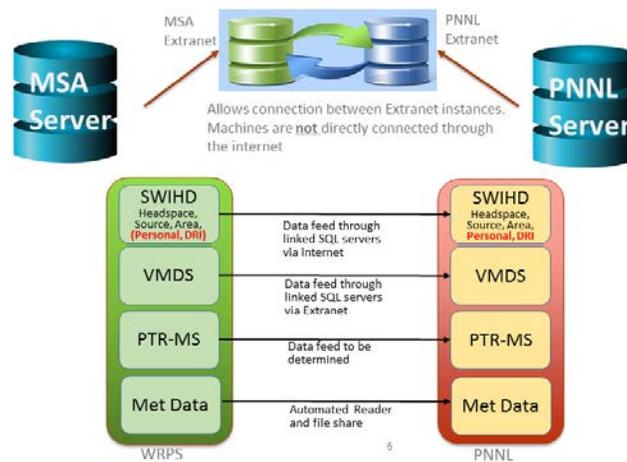


Figure 2. DAV Explore depicting VMDS data with concentration and meteorological data

More recently, MSA has expressed willingness to work directly with Microsoft Azure and Amazon AWS Cloud services, provided the instance meets FedRAMP Moderate security designation¹. This new avenue for data hosting represents an excellent opportunity for PNNL and other site contractors to leverage other cloud services and software as a platform for a wide variety of data needs. Using a FedRAMP approved cloud service is the desired path forward, once the process is out of development and into production. Access to current data is described in Table 1.

¹ <https://www.fedramp.gov/understanding-baselines-and-impact-levels/>

Table 1. Databases, Interfaces, and Update Frequency for proposed I-DAV application suite

Data Source	Connection Type	Description	Update Frequency
VMDS	SQL Database – linked server in extranet	Stack Monitors and open path monitors from AP area	Live synchronization, records updated from WRPS daily
SWIHD ¹	SQL Database – linked server in internet	Site wide Industrial Hygiene database	Live synchronization, new records inserted from WRPS ~4-6 months
SDDS	Excel Spreadsheet downloaded from SDDS database	Download from SDDS database, includes tank transfer data	Snapshot upon request
MET DATA	Shared FTP	Hanford Meteorological data	15 minute intervals updated daily

¹ Currently PNNL has access to a limited subset of SWIHD headspace, source and area data that were filtered to meet the requirements of the public-facing DAV application. I-DAV users would need access to an unfiltered set of SWIHD data and additional data sets.

5.0 I-DAV Application Concepts: Developing Tools to Support Internal Tank Vapor Data Analysis Functions

This section describes potential tools that could be incorporated into the I-DAV application. These tools address recurring data analysis functions and have been identified through user elicitation interviews (Figure 3). An additional set of tools related to the immediate Exposure Assessment work flow is described in Section 7.0 along with some near-term (early FY-19) deployment options for those tools (Figure 4).

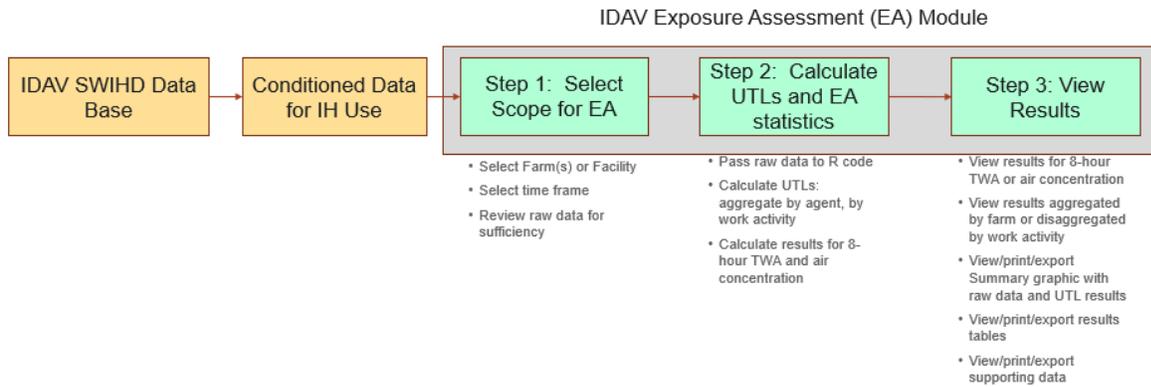


Figure 3. Data Conditioning and IDAV Exposure Assessment Workflow

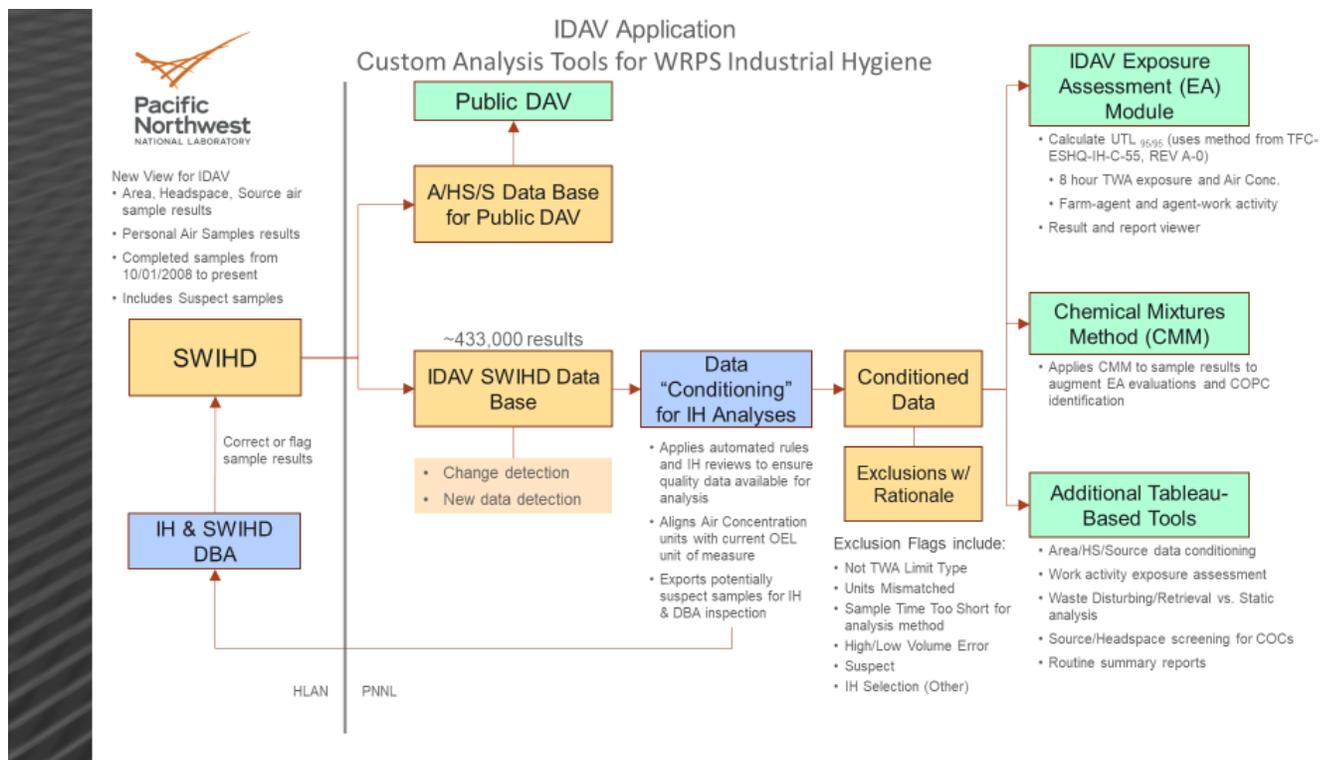


Figure 4. IDAV Data Flow and Application Interfaces

5.1 Chemical Mixtures Hazard Index Assessment

A chemical mixture methodology relevant to Hanford chemical mixture cases is presented in Figure 5. This is part of a 2-step screening method. First, the Hazard Index (HI) is calculated for each chemical in an airborne mixture at the same receptor location. Then the sum of all chemical HI's is developed to represent the cumulative HI. For mixtures with sum greater than 1.0, the cumulative target organ effects are evaluated by binning the same or similar toxic health effects using Health Code Numbers (HCNs). The most appropriate implementation of the chemical mixture hazard index model is to incorporate coincident chemical measurements, although this is extremely challenging given the nature of Site Wide Industrial Hygiene Database (SWIHD) data collection, absence of geolocation information, and the difficulty in obtaining VMDS data. The proposed implementation of this tool would allow users to select a set of sample data from any area location with continuous integrated measurements of concentration data (i.e., VMDS, SWIHD), and run HI computations on the fly.

5.2 AOP-15 Event Chemical Concentration and Site Assessment

AOP-15 event tool would be designed to simulate the conditions present during an AOP-15 event. This would include, for example, meteorological conditions, atmospheric plume dispersion simulations, chemical concentration data, and relevant site data as recorded in operational notes. The user can then select an atmospheric dispersion model and run a simulation with inputs from conditions present during the time of the AOP-15 event. The synthesis of this information can be used to form a more complete picture of what occurred during the time of the AOP-15 event.

The proposed AOP-15 Module would enable users to select historical AOP-15 data as part of the analytical process. This module would be integrated with an Atmospheric Plume Dispersion modeling tool, allowing users to select model of choice for simulating plumes from sources in the vicinity of the AOP-15.

Personal Samples:

Target Organ System Effect (TOSE) and Specific Target Organ Effect (STOE) Hazard Index (HI) Sums for all Surveys
 Farm = AX FARM, AX FARM, AY FARM, AX FARM, AZ FARM

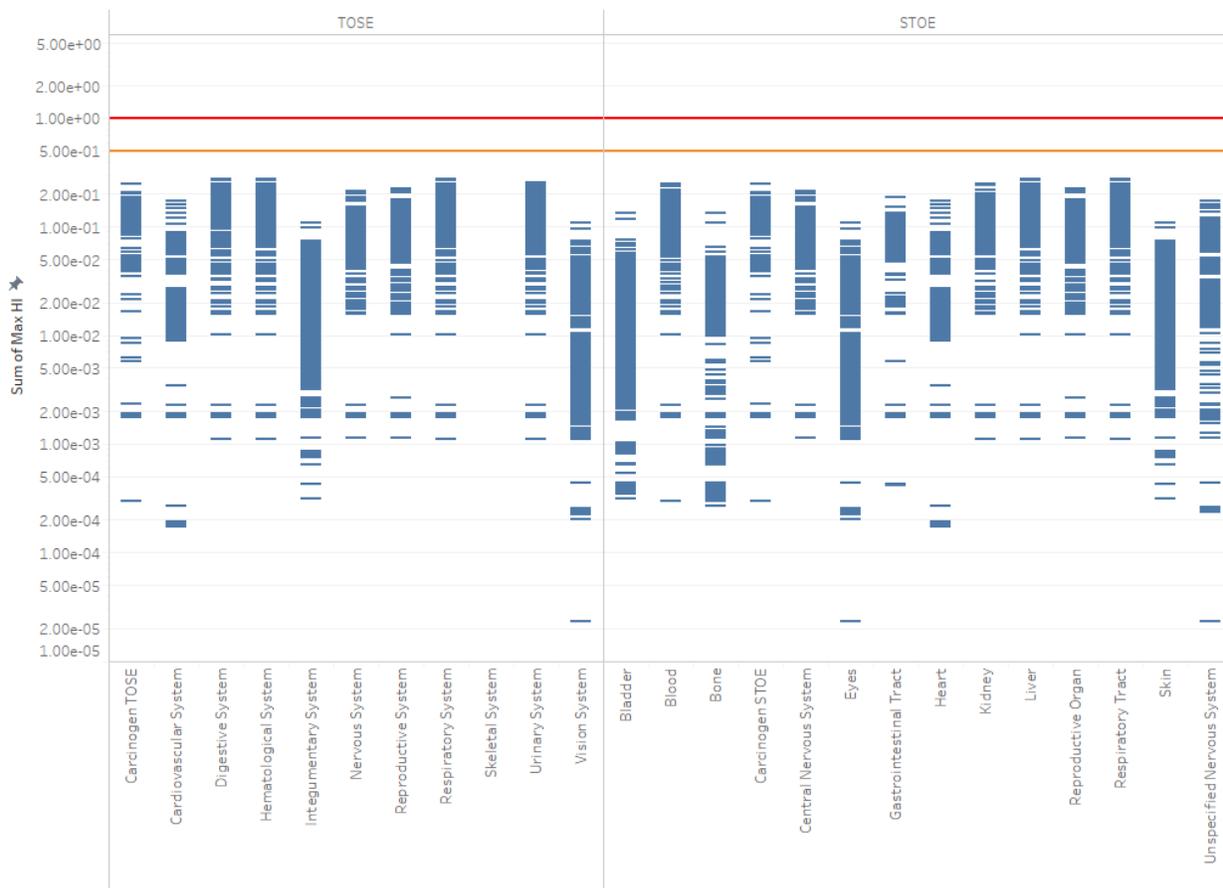


Figure 5. Chemical Mixtures Model Rendering

5.3 Waste Disturbing Event Impact Analyzer

Waste Disturbing Event analyzer would be used to characterize the impact of waste disturbing events on vapor concentration data (headspace/area/exhauster). The assessment would pull from vapor data sources, SWIHD, VMDS, and potentially PTRMS. In addition, analyses in this module would pull data from relevant tank data bases, such as SDDS and Tank Waste Information Network System (TWINS).

Through a process of group and one-on-one interviews with WRPS staff, PNNL was able to document the step-wise process in systematically identifying various data sources in order to replicate WRPS approach in an orderly and repeatable fashion.

The approach involves pairing coincident sample data from SWIHD that are taken during a waste disturbing event. Importantly, the start and stop time from waste disturbing event requires a time stamp, which is beyond what is available in TWINS directly. This start/stop time from waste disturbing event is determined qualitatively by looking at surface level information and pumping and making a technical judgement as to when the pumping/surface levels appear to change significantly. The mid-point during the sample period is ascertained from the start and end time. The total change in waste level is characterized at that point (Figure 6). The data referred to in this process can be queried directly from these databases (TWINS, SWIDS, SDDS – TMAC). While PNNL does have direct access to TWINS and SWIHD, direct access to SDDS is not currently in place, and would need to be accommodated in FY19 to fully automate this process.

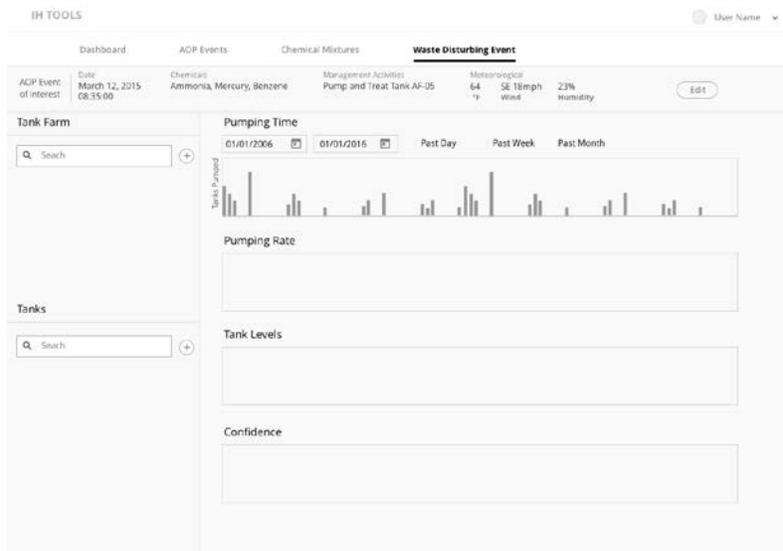


Figure 6. Mock up design for Waste Disturbing Event Assessment

Supplementary data would be used to better characterize transfer volume and from tank to tank and pumping rates during transfer. This information can be used to characterize potential relationships between transfer activities and vapor concentration data in various places. More specifically, these analyses could be used to determine if there is a causal relationship between waste disturbing events and elevated chemical vapor concentrations.

5.4 APGEMS (or atmospheric plume dispersion model) Web Interface

This feature would not necessarily be limited to the APGEMS model, but would potentially offer the user a list of models to choose from for running simulation. It has been observed that APGEMS lacks utility in some of the near-field calculations but generally conveys relevant information (plume direction, distance) at longer distances. Other models may better suffice for near-field estimates; these alternatives would need to be identified by WRPS. Irrespective of the model selection, simulations would pull from concentration source data, instantaneous (real-time monitoring results) if available, and generate plumes “on the fly”.

6.0 Exposure Risk Calculator: An Implementation for Automated Analytics

6.1 Exposure Risk Calculator

PNNL interviewed and worked with various analysts at WRPS in order to better characterize overall needs and requirements. PNNL endeavored to capture the logic and thought processes involved, and identify routine data processing aspects that could be improved by automation (Figure 7, Figure 8, Appendix A). The EA team has a workflow in place that is largely dependent on MS Excel spreadsheets and VBA scripts to characterize and quantify chemical concentration/tank farm paired data in terms of potential exposure.

Using this approach, the EA staff download personal monitoring data from the SWIHD site, manually preprocess all data, and then copy this data into a MS Excel workbook.

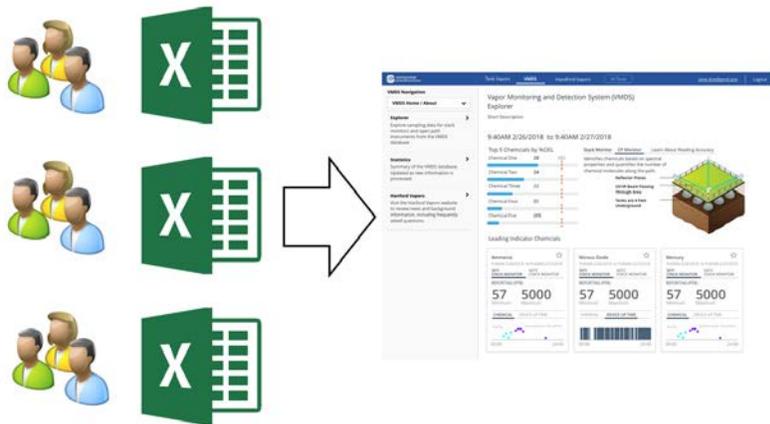


Figure 7. Notional concept of migrating away from many different Excel work books to a unified application which replicates the same analytical sequence.

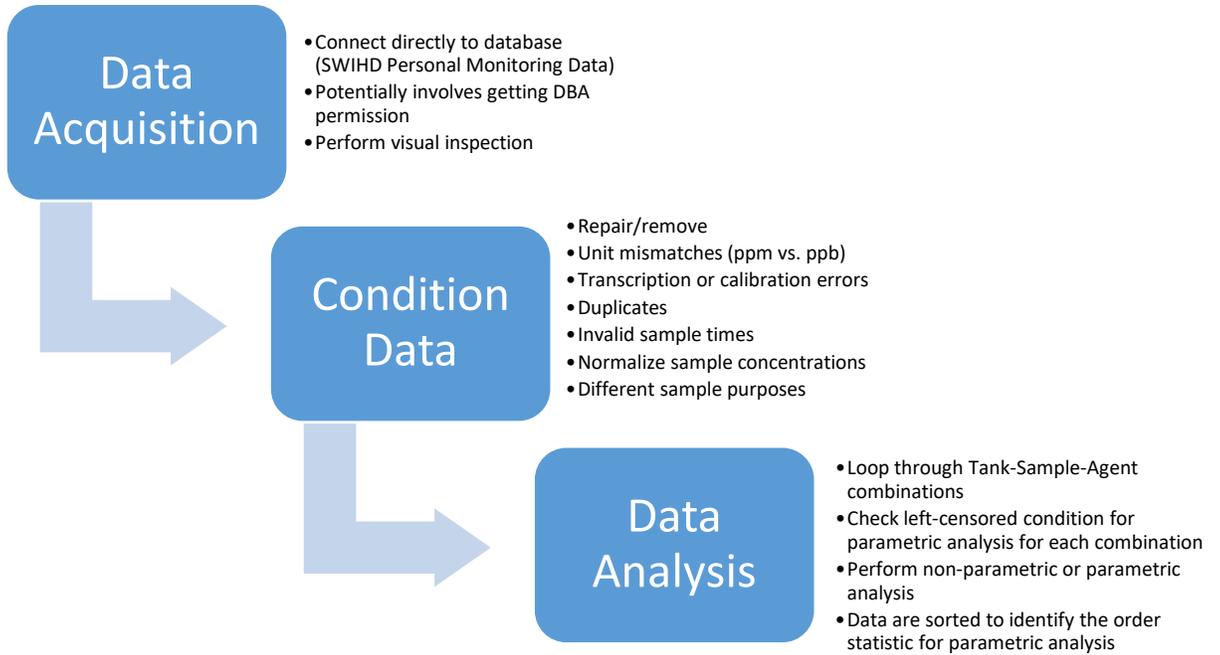


Figure 8. Step-wise sequence for pre-processing and analyzing data for the Risk Exposure Calculator

PNNL has described this process using logical blocks of workflow as shown in Figure 9. Exposure Assessment Workflow. PNNL has been able to replicate this process and obtain equivalent results.

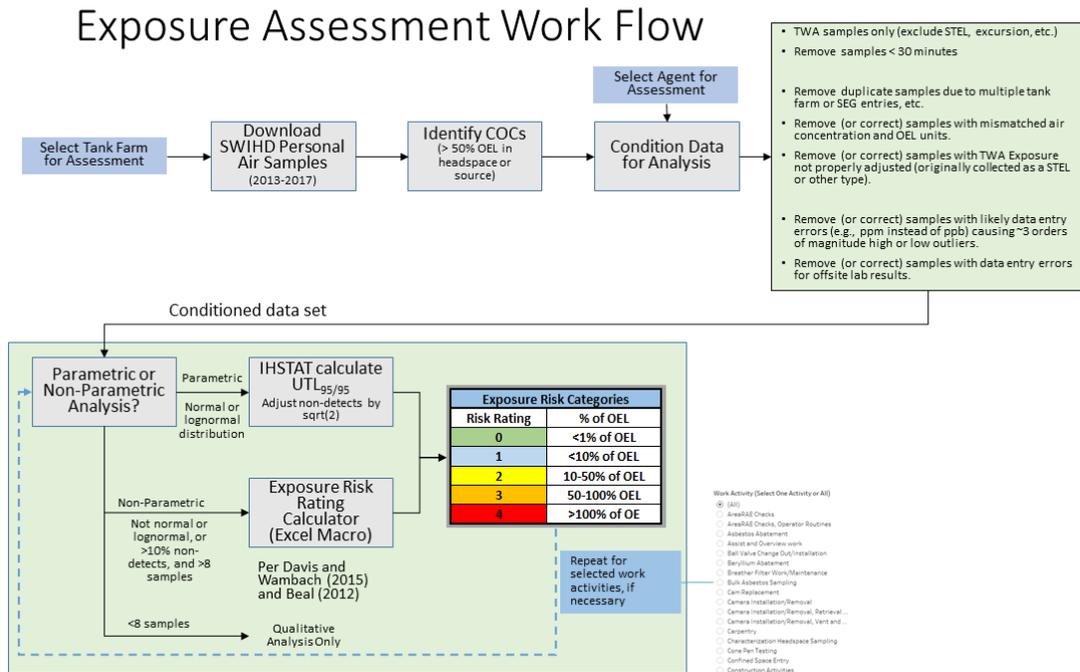


Figure 9. Exposure Assessment Workflow

6.2 Exposure Assessment Team Rapid Prototype and Tools Ready for Use

A set of tools was developed for this project and is ready for use to support ongoing Exposure Assessments. Personal air monitoring data from 2013 through 2017 was downloaded from SWIHD and then filtered to be consistent with the data sets being used for the Exposure Category determination, e.g., 8 hour TWA samples, sample times at least 30 minutes, matching air concentration and OEL units of measure, etc. These tools were developed using commercial software Tableau® and Databricks®, and are published in a manner that they can be utilized by WRPS Exposure Assessment staff.

6.2.1 Outlier and Error Detection, Identification and Visualization

Visual data analysis is a convenient and fast approach for isolating outliers in data sets. These tools rapidly identify errors pertaining to sample volume (e.g., mL instead of L) but could be adapted for other types of outliers (e.g., air concentration units) as a function of the various filters available for use. Data is depicted as the percent OEL (along the y-axis) as a function of total sample volume (along the x-axis; Figure 10).

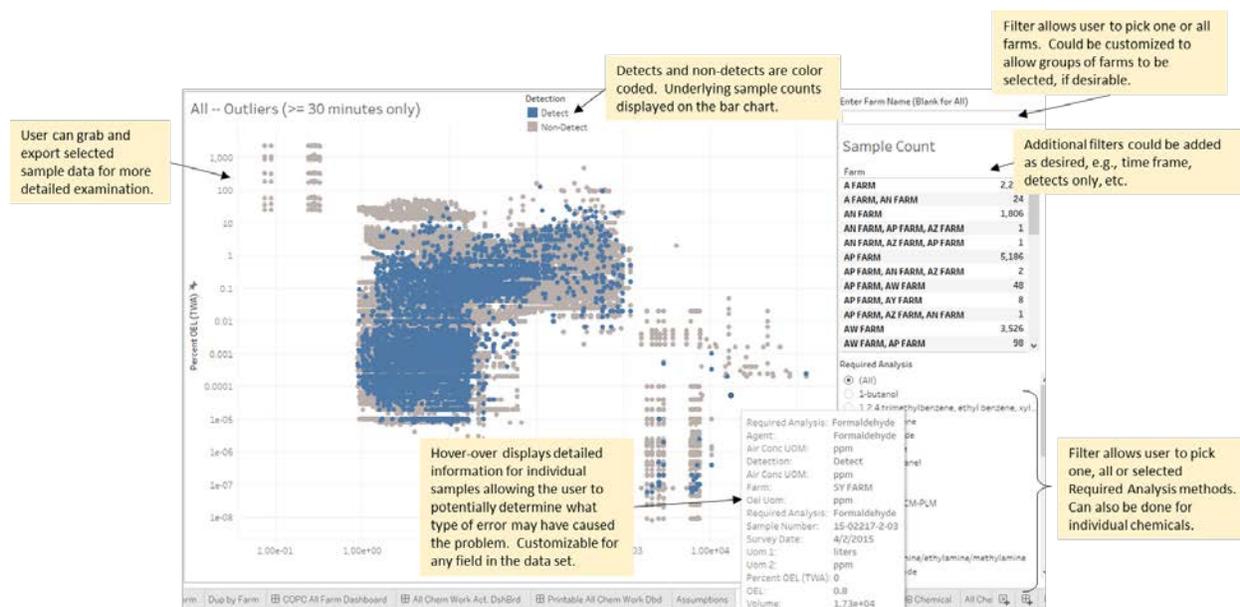


Figure 10. Error Detection and Outlier Analysis Visualization

Outliers can be quickly identified along lower and upper extent of the volume sampled (Figure 9). These are samples that appear to have incorrect volumes (e.g., mL instead of L or vice versa) leading to percent OEL that is three orders of magnitude too high or too low.

By default, the data is shown for all tank farm/chemical combinations, but users can filter by tank farm(s) and by analysis method (e.g., VOC, nitrosamines, furans, etc.). Other filters can be added to rapidly identify subsets of potential outliers. Users can quickly select and export outlier data for consideration as data that should be marked “suspect” in SWIHD, or corrected.

This application is implemented in Tableau® and can be accessed through Tableau Server using a Tableau Explorer license¹. If Tableau Explorer is not an option, IH staff can work directly with PNNL staff to generate Excel files containing all suspected outlier samples for additional analysis and generate visualizations as part of the reporting strategy.

6.2.2 Exposure Risk Category Visualization

This tool provides visual access to all data used to support the Exposure Risk Category determination (Figure 11, Figure 12):

- Complements Risk Category determination by providing intuitive visual confirmation of all underlying data used in the formal calculation.
- Allows rapid selection by farm and even work activity.
- Connected to full SWIHD data set for Personal Air Samples.
- Supports levels of aggregation, e.g., all or any work activity, all or any tank farm, etc.
 - Uses the same pre-conditioned data set as used by the Exposure Risk Rating spreadsheet (TWA samples only, 30+ minute samples, no duplicates, etc.).
 - Allows users to quickly see those chemicals that have no values above 1% of the OEL and that may not need to be carried through the formal statistical analysis.
 - Additional feature – it is possible to add the capability to create files identifying all samples that were screened out and not used for an Exposure Assessment, e.g., due to mismatched units, sample times less than 30 minutes, etc. and examine them for trends or anomalies.

This application is implemented in Tableau® and can be accessed through Tableau Server using an Explorer license. If Tableau Explorer is not an option, IH staff can work directly with PNNL staff to generate visualizations that will complement the Exposure Assessment process.

¹ Requires individual license for ~\$420/year, per user.

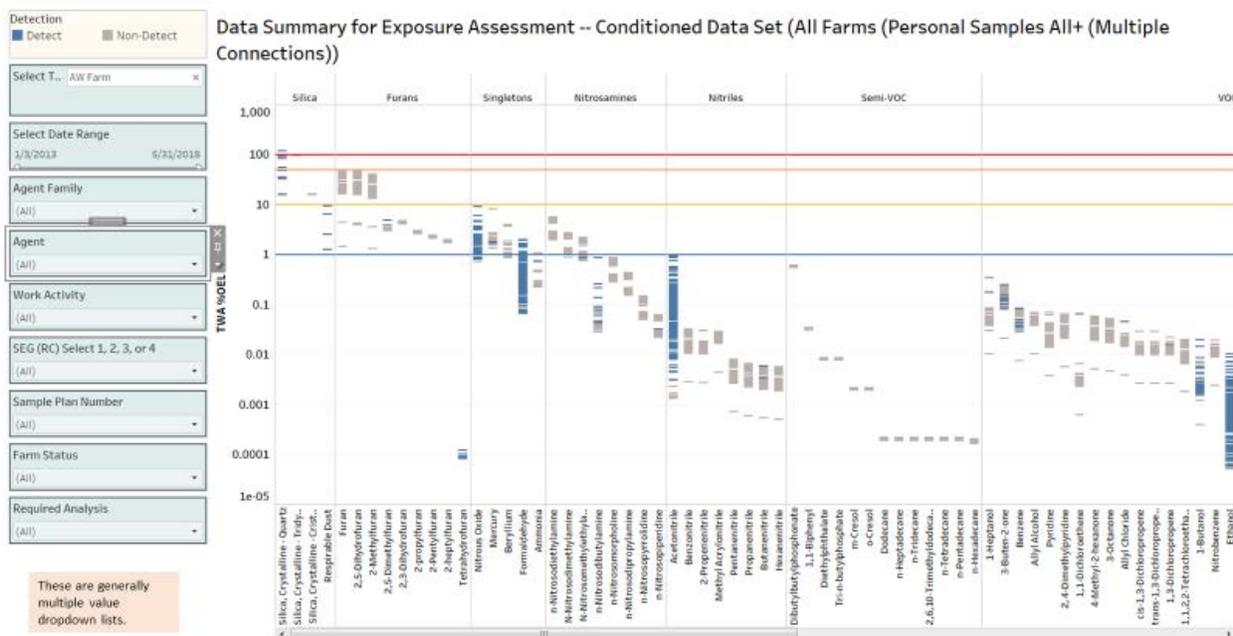


Figure 11. Field Exposure Assessment Data Selection and Visualization

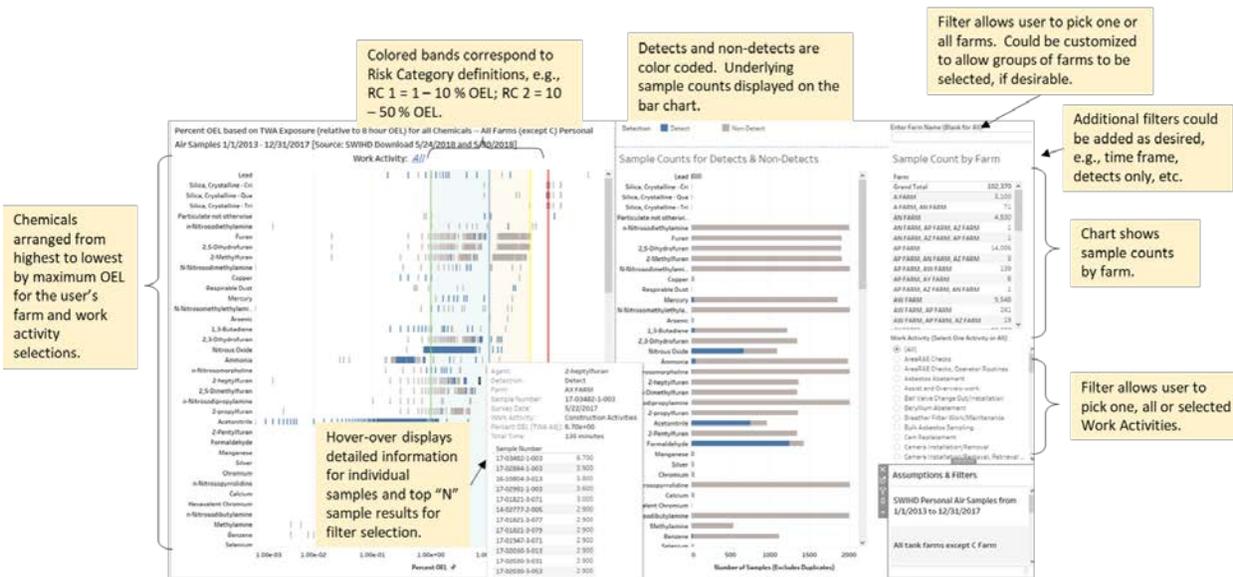


Figure 12. Exposure Risk Category Determination

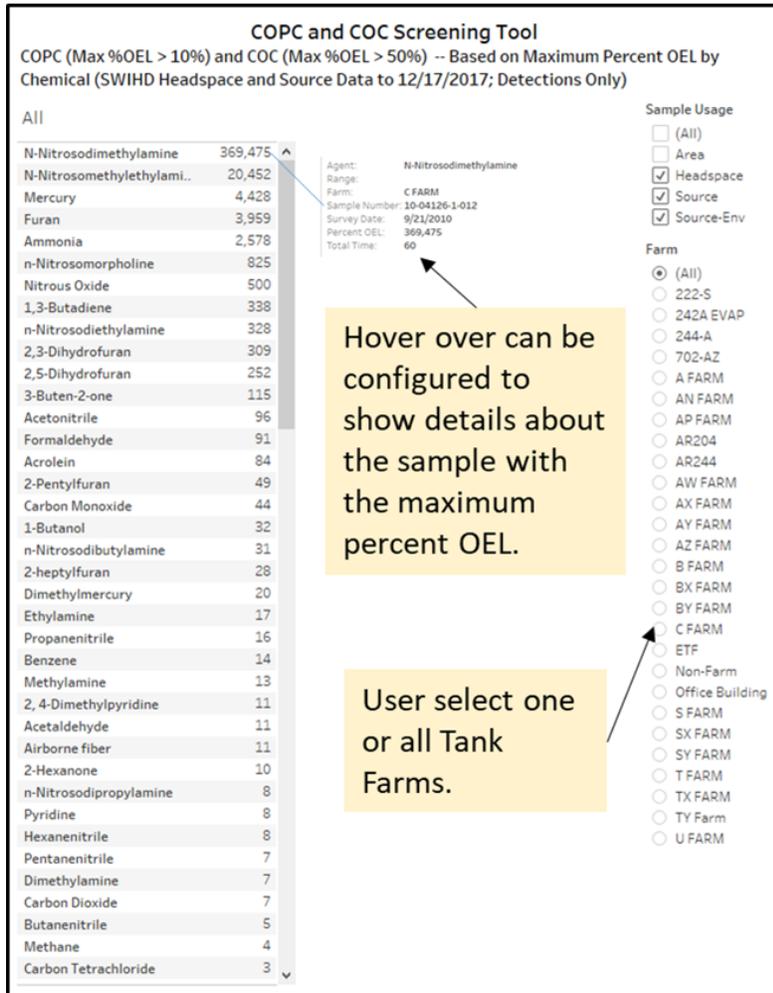
6.2.3 COPC and COC Quick Screening Tool

This tool provides rapid identification of COPCs and COCs based on SWIHD headspace and source data:

- Connects to all SWIHD headspace and source data
- Allows user to quickly identify chemicals that have been detected in headspace and source samples above 50% OEL (COCs) and above 10% OEL (COPCs).

- Provides filter by tank farm, or other fields that IH staff might requests.

This application is implemented in Tableau® and can be accessed through Tableau Server using an Explorer license. If Tableau and Tableau Explorer are not options, IH staff can work directly with PNNL staff to generate COC/COPC lists by tank farm or other selections. Figure 13 provides an overview of the COPCS and COC Screening Tool.



Additional filters can be added as needed. For example, time ranges could be set to restrict dates as desired, e.g., only last 5 years.

Hover over can be configured to show details about the sample with the maximum percent OEL.

User select one or all Tank Farms.

Figure 13. COPC and COC Quick Screening Tool

6.2.4 Shapiro-Wilk Batch Goodness of Fit & Test for Normality Tools

At present, WRPS is manually filtering data for each tank farm/chemical combination and characterizing each data set using Shapiro-Wilk goodness of fit test for normality. The results of this are used to determine with the remaining methods should include parametric or non-parametric approach to completing the analysis. PNNL has automated this process for every tank farm/chemical combination and can run it “on the fly” against the database directly. PNNL also recommends incorporating a modest change identification feature to allow users to understand what additions or deletions from the database have occurred since the last refresh.

7.0 Updated Module Concepts March 2019

The WRPS and PNNL I-DAV working group began meeting bi-weekly at the beginning of FY 2019. The purpose of these meetings was to evaluate and refine module concepts proposed in this original document, and to advance new ideas for implementation of prototype designs that offered the greatest benefit for WRPS analysts and management. During these meetings PNNL and WRPS worked collectively to identify organizational pain points, or bottle necks, with respect to work throughput and associated data pipeline. In addition, applications were optimized to highlight and streamline analytic workflow and automate this process to the greatest extent possible. Similarly, concepts that were proposed that were not deemed to be highly relevant to mission objectives were de-emphasized in ongoing prototyping efforts. Overall scope was adjusted to meet evolving project requirements.

In this light, module concepts related to exposure assessment process, and chemical mixtures model were determined to offer the greatest return, and analyses related to AOP events, waste disturbing activities, and atmospheric plume dispersion modeling were the least beneficial. Updated design concepts relevant to exposure assessment and chemical mixtures model are shown in Appendix C. Importantly, design concepts shown here are separate from the I-DAV software design, which is submitted as a separate deliverable. In addition to module concepts show in Appendix C, there are roughly 30 other Tableau workbooks available to WRPS staff on our PNNL Tableau server instance <https://vaporsih02.pnl.gov/>

7.1 Automation constraints

Many of the tools developed during the rapid prototype phase were designed to facilitate data capture and discovery, and to assist in the identification of outliers or spurious measurements that were necessary to exclude before further analysis. Logical query statements were helpful in identifying spurious data and errors; however, manual review of suspect data was still necessary in order to ensure that data was correctly identified and removed. Given this feature, complete end-to-end automation, directly from the database to the application, was not feasible even though a large portion of the work was automatable. The exposure assessment, operates on the notion that a scrubbed and pedigreed data set, abstracted from SWIHD is available.

8.0 Analysis of PTRMS data

At the beginning of FY19 there were a series of questions set forth by WRPS IH analyst in which they felt the mobile PTRMS data would be helpful. The PTRMS data is large in terms of overall magnitude and repeat frequency, and as such an approach was required to migrate this data to a larger data format and analyze it in a meaningful way. This was carried out using MS SQL database and Jupyter notebooks, and Tableau workbooks. Jupyter notebooks can be used in a fashion like Tableau and Microsoft Power BI for developing prototype concepts for web-based applications. The program logic used in Jupyter can be easily ported to a web type architecture, and if using AWS Lambda and AWS Gateway API, python can be used directly with minimal edits to base code. For this reason, it is an ideal tool for socializing concepts for integration into the IDAV. Variance in PTRMS concentration data from leading indicators was characterized accounting for the effect of sample activity. Sample activity refers to the purpose for sampling; for example, routine analysis, background study, AOP event. This analysis was carried out using python programming language in the Jupyter notebook platform, which provides an interactive experience, enabling collaboration, rapid sharing, and visualizations developed in common scripting languages. Jupyter notebooks support python, R, Julia and base SQL languages, and has emerged as a dominant platform for scientists and data analysts. For this exercise, PTRMS data, between the dates of July 6, 2018 and May 5, 2019 (roughly 1.7M samples; Figure 1) was loaded into a Microsoft SQL instance and data was accessed henceforth with python SQL alchemy library. Storing data in a SQL database, and using it as a primary access point offers performance benefit due to the retrieval speed of accessing data from SQL database as opposed to a flat file (i.e., .CSV file). In addition, archiving data in a single point portal mitigates data fragmentation.

Concentration data from leading indicator was compared across different sample activities to assess the effect of sample type on concentration data. 1000 stratified random samples were selected (by both numeric values and sample activity), data was log transformed before statistics were estimated, the outcome was depicted in boxplots (Figure 14. Boxplot and ANOVA analysis of ammonia data compared across sample activity ($F = 22.46$, $P\text{-value} < 0.01$) Figure 15. Boxplot and ANOVA analysis of furan data compared across sample activity ($F = 1.24$, $p\text{-value} = 0.28$).

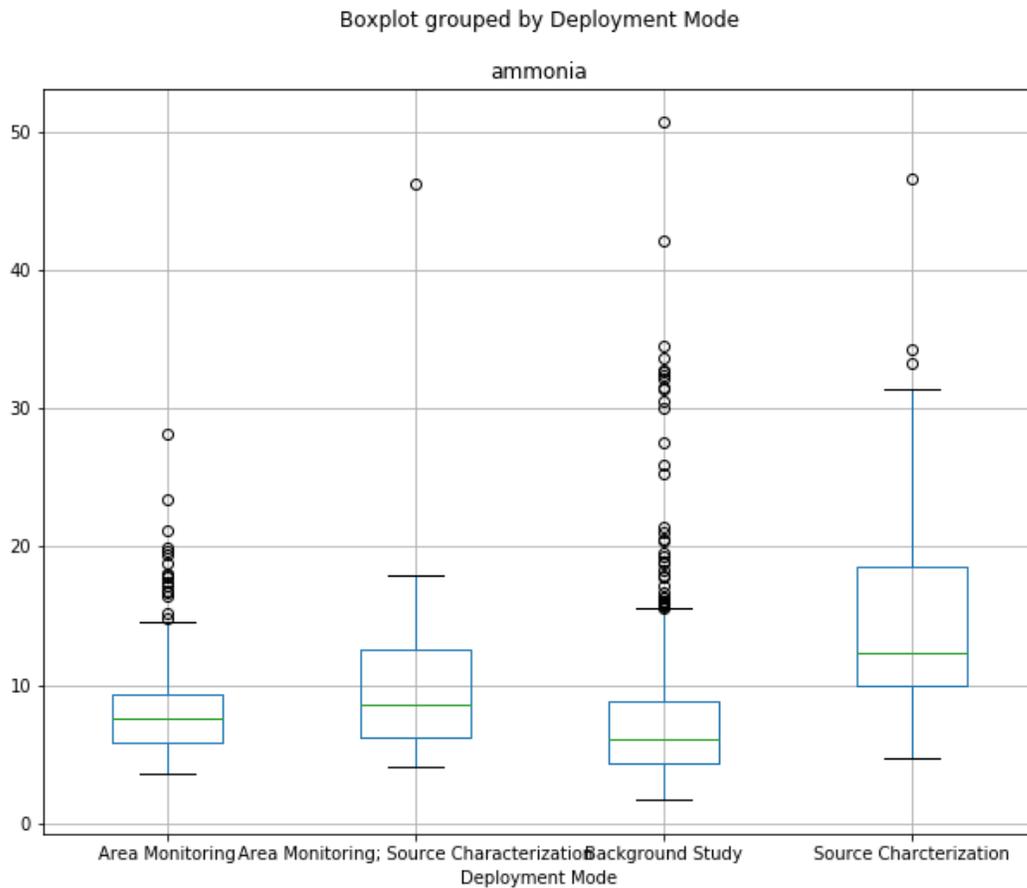


Figure 14. Boxplot and ANOVA analysis of ammonia data compared across sample activity ($F = 22.46$, $P\text{-value} < 0.01$)

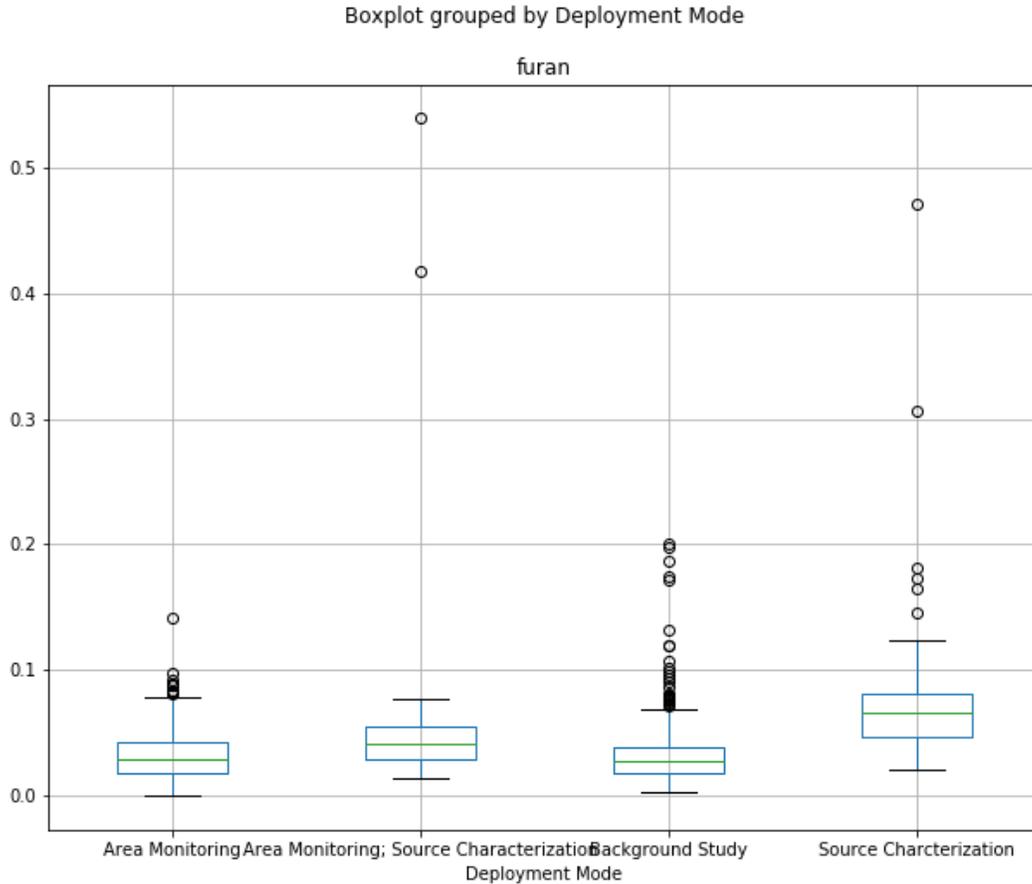


Figure 15. Boxplot and ANOVA analysis of furan data compared across sample activity ($F = 1.24$, p -value=0.28)

Interestingly, ammonia values were significantly different ($p < 0.01$) by sample activities, but furan was not ($p=0.28$). Most of the variance between concentration in ammonia can be attributed to the difference between source characterization and other types activities.

The entire corpus of data was treated as a continuous single sample event for the purpose of characterizing relationship between variance in concentration values and spatial autocorrelation with respect to tank farms and potential sources. It was designed this way with the understanding that potential implementations of the code could apply date filters to review the same analyses for specific dates. Additional analysis – forthcoming – will take into account that these were separate sample events and will be treated independently.

Prior to analysis, concentration data distribution was tested for normality using D’Agostino’s K-squared test, which calculates summary statistics from the data, namely kurtosis and skewness, to determine if the data distribution departs from the normal distribution. Data was transformed for analysis based on the outcome of this using log10 and/or z-scale transformation and tested again for normality. Importantly, if methods are used that assume a Gaussian distribution, and the data are drawn from a different distribution, the findings may be misleading or plain wrong. There are several techniques that can check if

the data sample is Gaussian or sufficiently Gaussian-like to use the standard techniques, or sufficiently non-Gaussian to instead use non-parametric statistical methods.

Eighty-four different chemicals from sampling campaign were analyzed, in addition to meteorological data, which including wind direction, wind speed, barometric pressure, and ambient temperature. Analyses for all chemicals was completed; however, most of the results shown are limited to an identified subset of chemicals that are proposed as leading indicators or contaminants of concern. Log transformed correlation coefficients matrix was plotted for a subset of chemical of interest (ammonia, furan, formaldehyde, butanol, 2-methylfuran, and NDMA; Correlation coefficients for log transformed data were estimated for all chemicals, show for the same subset (Figure 16 and Figure 17), also available as an Excel document for the entire dataset. Examples of chemicals exhibiting moderate positive correlation were plotted together using z-scale transformation. In addition, abiotic data exhibiting moderately negative correlation with ammonia where plotted. The distribution of correlation coefficients with wind speed, wind direction and barometric pressure for call chemicals were also charted

Spatial analysis involved 1) estimated measurement distance from tank and modeling best fit for concentration as a function of distance, 2) Hot Spot Analysis (Getis-Ord G_i^*) global indicator of spatial autocorrelation and hotspot analysis, and 3) Ripley-K function to characterize data pattern as a function of distance.

Concentration as a function of distance from tank was evaluated for testing the assumption that measurements at distances closer to source would have higher, or more varied concentration values, as opposed to measurements further away from tanks. A distance field was generated for every measurement location by calculating the distance to the centroid of nearest tank farm. In this approach there is unique pairing for every chemical measurement and tank farm, such that a single measurement may be close to several tank farms, but the distance returned assumes it is closest to one specific tank farm. Concentration values were plotted as a function of distance from tank farm centroid, and best fit model was developed for distances of 100 m, 500 m, and 5000 m (Figure 15).

Getis-Ord G_i^* hotspot statistic (pronounced G-i-star) analysis provide z-scores and p-value that indicate where values at geographic locations with either high or low values cluster spatially. This analysis works by looking at each measurement location within the context of neighboring locations. A measurement with a high value is interesting but may not be a statistically significant hot spot. To be a statistically significant hot spot, a measurement will have a high value and be surrounded by other features with high values as well. The local sum for a feature and its neighbors is compared proportionally to the sum of all features; when the local sum is very different from the expected local sum, and when that difference is too large to be the result of random chance, a statistically significant z-score results. When the FDR correction is applied, statistical significance is adjusted to account for multiple testing and spatial dependency. P-values were mapped for Getis-Ord G_i^* hotspot statistic, the symbol gradation assumes a significant level a 0.05, suggesting departure from statistical norm is more likely in areas closer to tank farms (Figure 18). Statistically significant clusters are shown with respective cluster type (high-high, low-low, low-high) for 200 area and close up.

10 minute moving average concentration behavior of selected leading indicators or contaminants of potential concern in 200 East, and at an offsite onion processing plant were plotted in Tableau to illuminate on trends relevant to known sources and potential fugitive emission sites (Figure 19. , Figure 20.)

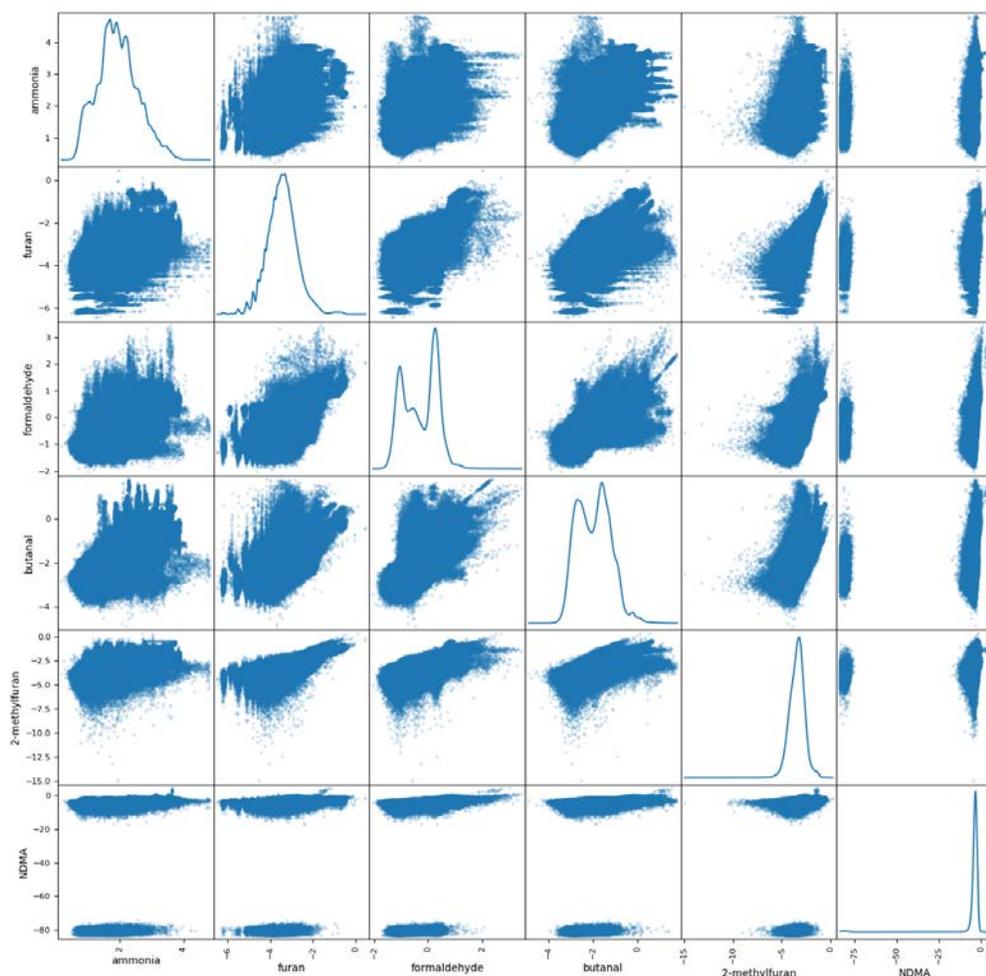


Figure 16. Scatter correlation matrix with select chemical values (log transformed).

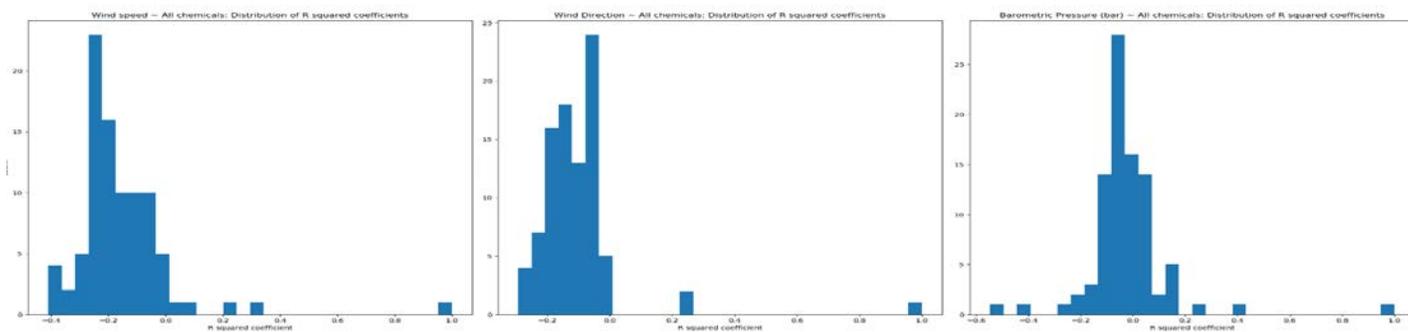


Figure 17. Distribution of correlation coefficients for Wind speed, wind direction, and barometric pressure against all chemicals

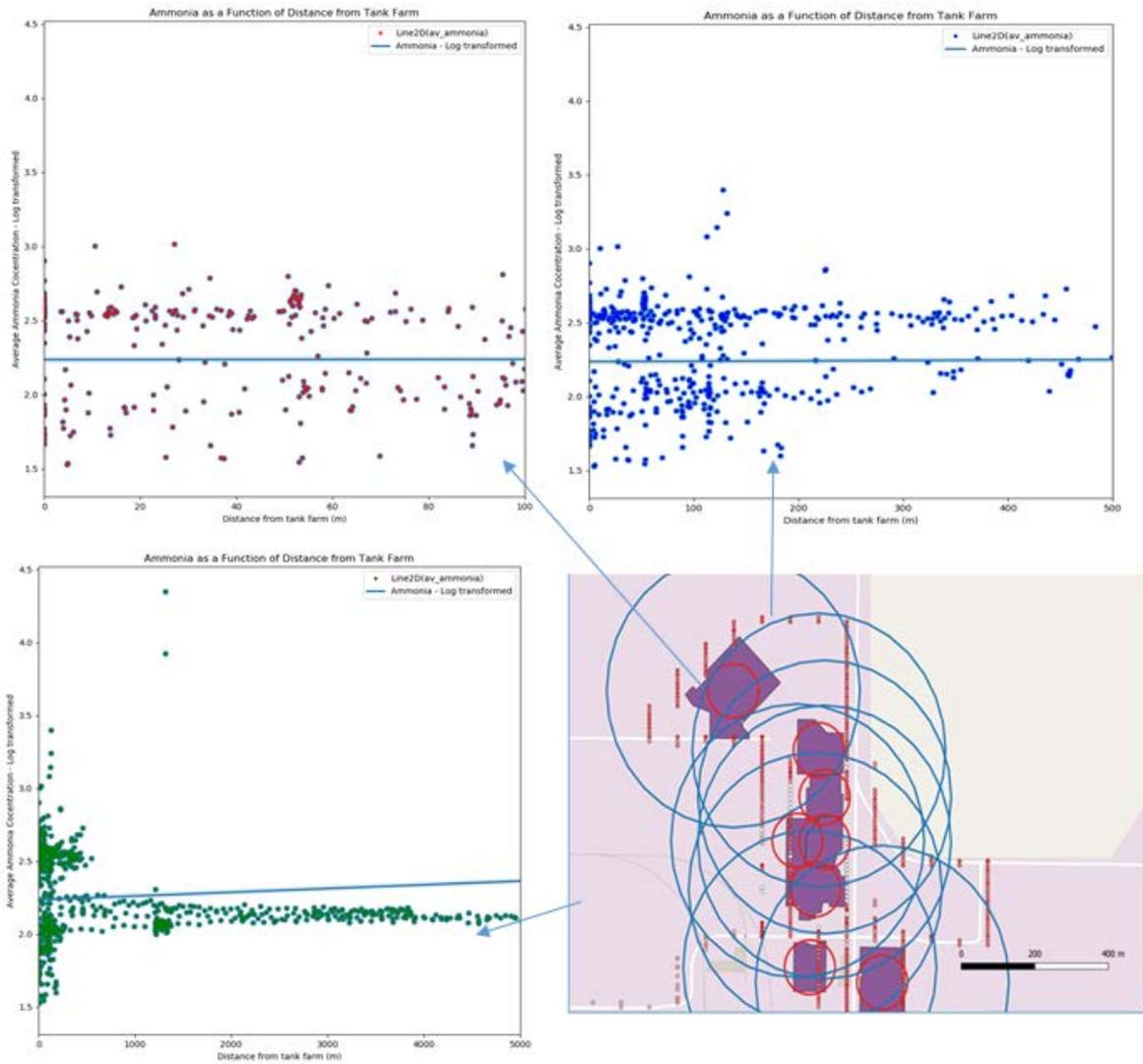


Figure 18. Log-transformed scatter plots and best fit model, depicted as a distance from nearest tank farm for 100 m (red), 500 m (blue), and 5000 m (green) distances. Circular buffers around tank farms indicate buffer distance from tank farm centroid.

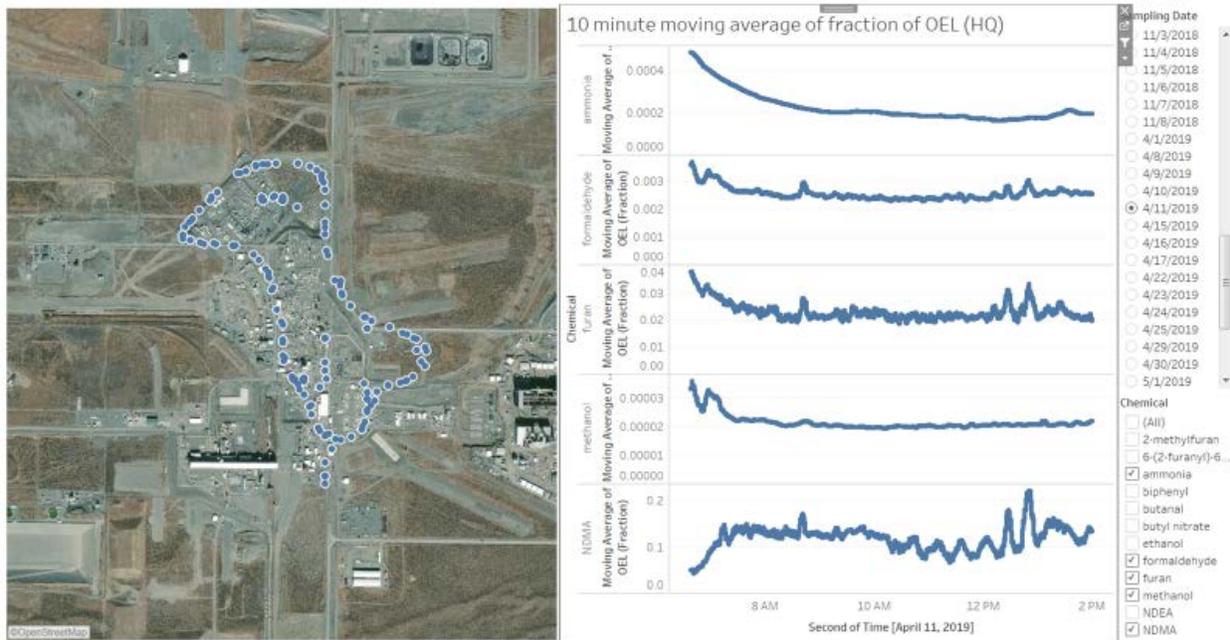


Figure 19. 10 minute moving average concentration behavior of selected leading indicators or contaminants of potential concern in 200 East.

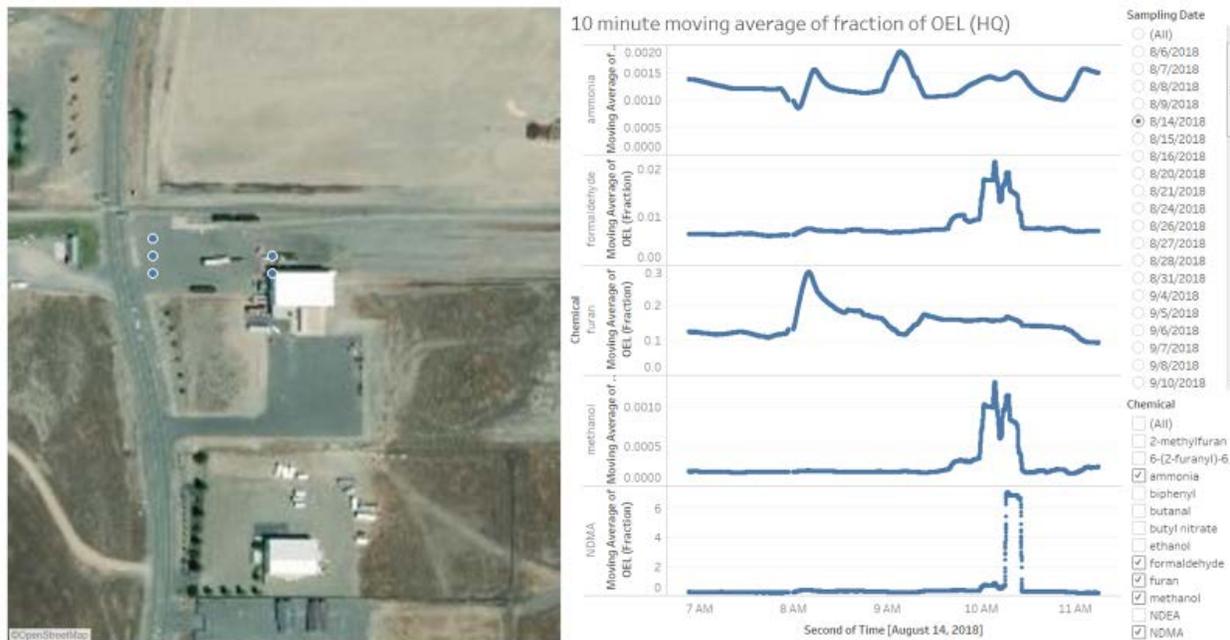


Figure 20. 10 minute moving average concentration behavior of selected leading indicators or contaminants of potential concern at an off-site onion processing plant upwind from Hanford. Note the behavior of NDMA in the 10 am to 11 am window.

9.0 Conclusion

This report outlined an approach for developing a modular and scalable web-based application suite, Internal Data Access and Visualization (I-DAV) Tools, that would be developed to meet the specific analytic requirements for Industrial Hygienist (IH) analysts and tank vapors Subject Matter Expert's (SME) at the Hanford site. The process of requirements elicitation, workshops, and one-on-one interviews with IH staff shed tremendous light on current organization pain points and provided valuable insight as to which automated processes would offer the most time and effort saved. Different from the DAV, the I-DAV would include mission critical, objective based logic. Importantly, the application would not infer outcome or information generated from autonomous predictive analytics, but would be designed to map closely with day to day tasks of the analyst.

Appendix A

Excel-based Workflow for Exposure Risk Rating Calculator

APPENDIX A

Excel-based Workflow for Exposure Risk Rating Calculator

Step 1. Data are ingested from SWIHD, selected by TWA sample type, and filtered by tank, species, and timeframe (<5 yrs). There are 177 tanks, 60+ COC/COPCs, and more sensors/survey data anticipated to come. If there are fewer than 8 samples, no further analysis is done. Data are manually copy-pasted into the worksheet (Col A to Col N) to begin the workflow. Formulas, macros and conditional formatting are then used to perform calculations and identify errors.

Step 2. Once the data is entered, the workflow proceeds by doing data conditioning duplicates are identified and removed. Data are sorted to identify the order statistic (the top values are removed to conduct the non-parametric tests). In parametric testing, data below the reporting limit are replaced by the value of the reporting limit/sqrt(2)—This conditioning step is in Col P. Defective samples (volume, time, units, etc.) are identified and removed (Col R and Col S).

Step 3. Data are often left-censored, resulting in few values above the detection limit. If > 10% of the data are censored, a non-parametric method (QNP) is used to evaluate the data. If < 10% of the data are censored, parametric methods using normal and lognormal distributions are evaluated for suitability using a Shapiro-Wilk test and a separate software program (IHSTAT+) is used. If the data cannot be evaluated using the selected parametric distributions, the QNP is used. Where sample size is between 8 and 59, the QNP test is based on Wambaugh 2015 and is based on sample count. Where sample size is between 59 and 809, the QNP test is based on Beal 2012, and uses a ratio basis (TWA/OEL).

Appendix B

Attached as Separate Document

Appendix C

Chemical Mixtures Model Method

Chemical Mixtures Methodology (CMM) Application to Personal Sample Results

- Enables assessment of hazard posed to workers from exposure to chemical mixtures.
- Developed for DOE Subcommittee on Consequence Assessment and Protective Actions to provide exposure guidelines for emergency response situations (acute exposures).
- CMM adapted by PNNL to assess hazards from chronic exposures – assessment of routine exposures for Hanford tank farm workers.
- Assesses additive effects for chemical mixtures, not synergistic or antagonistic effects.
- Includes 3,100+ chemicals in the data set (currently using tank vapor COPCs and 73 non-COPCs with available HCNs)
- Characterizes health effects on target organs and systems
 - Identifies adverse health effects using a set of Health Code Numbers (HCNs)
 - Chronic and Acute HCNs are identified through extensive literature surveys
 - Includes 25 effects: Specific Target Organ Effects (STOE) and Target Organ System Effects (TOSE)
- Calculation:
 - Hazard Quotient (HQ_i) is Concentration/OEL for each chemical_(i) in a mixture.
 - Hazard Index ($HI_{(STOE)}$) for a target organ effect = $\sum HQ_i * BinaryEffect_{(i, STOE)}$
 - ✓ $BinaryEffect_{(i, STOE)}$ is 1 if a specific chemical_(i) in a mixture that has an effect on the selected organ, 0 otherwise)

Specific Target Organ Effect (STOE)

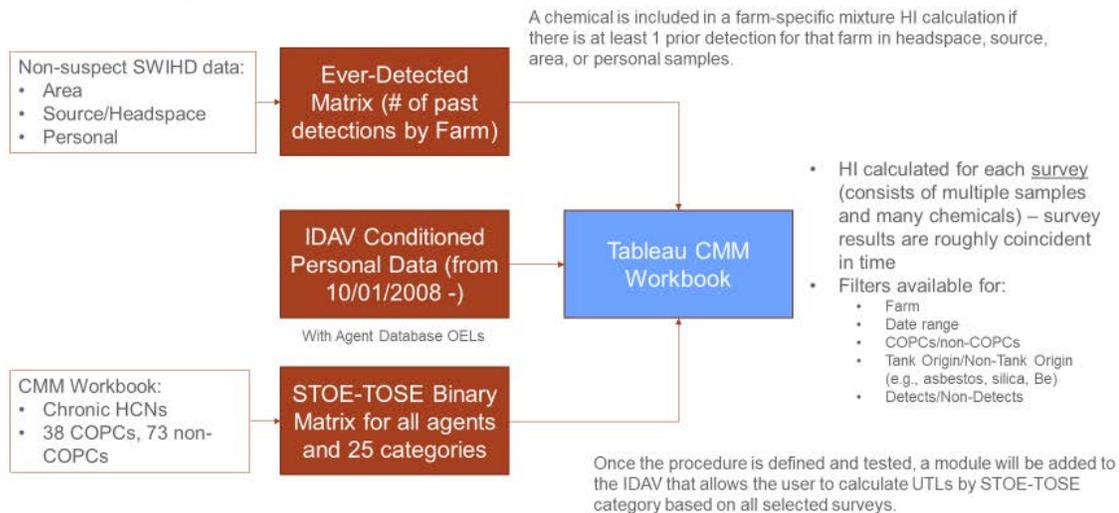
Bladder
Blood
Bone
Carcinogen STOE
Central Nervous System
Eyes
Gastrointestinal Tract
Heart
Kidney
Liver
Reproductive Organ
Respiratory Tract
Skin
Unspecified Nervous System

Target Organ System Effect (TOSE)

Carcinogen TOSE
Cardiovascular System
Digestive System
Hematological System
Integumentary System
Nervous System
Reproductive System
Respiratory System
Skeletal System
Urinary System
Vision System



Tableau Implementation of CMM for analysis of personal samples



11

CMM Implementation Approach for IH Exposure Assessments:

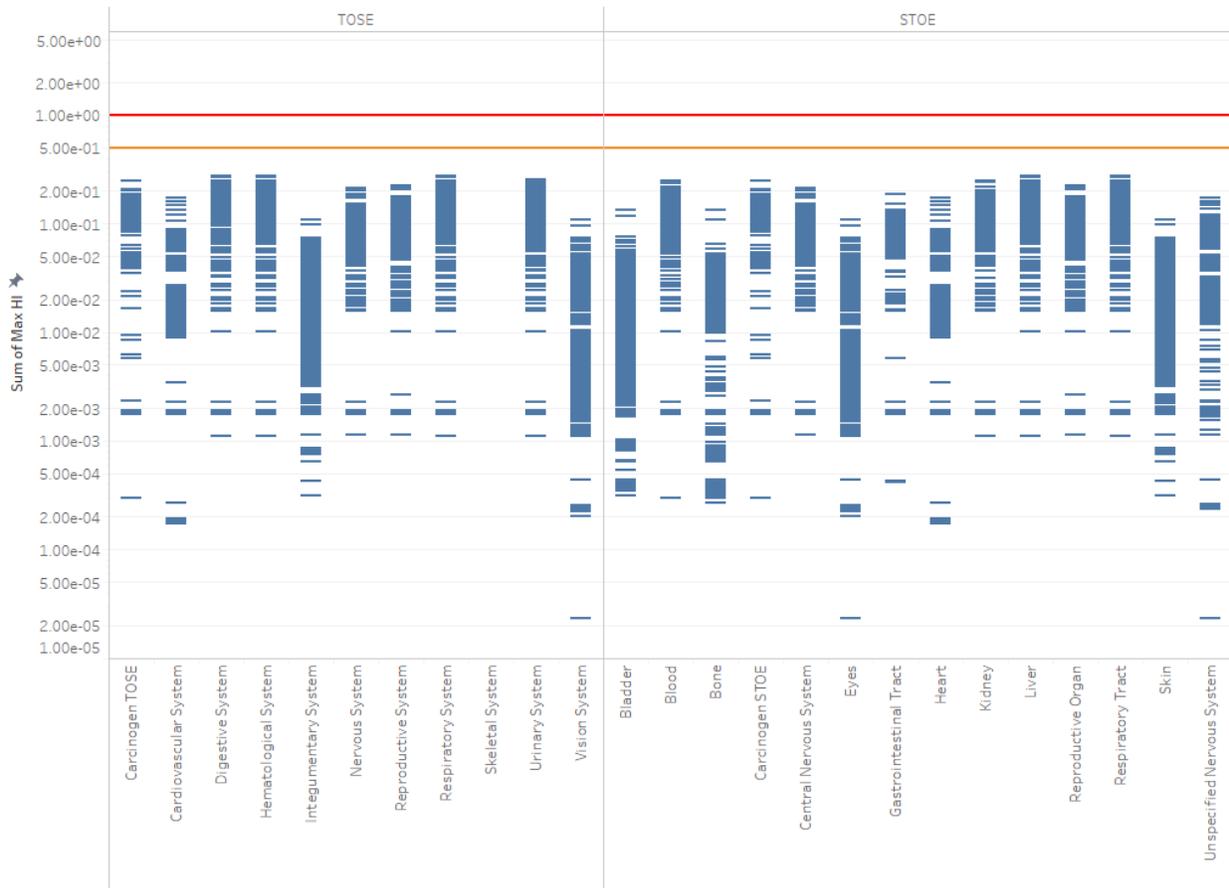
Personal samples will be used to calculate an aggregate Hazard Index for each STOE and TOSE for each individual survey. Sample results for agents within the survey will be represented by an individual HI (TWA Exposure/OEL). Agents will be included in the aggregate HI calculation only if they have been detected in source or headspace samples for the selected farm or set of farms, except all detections in personal samples will be included (even if never detected in headspace/source sampling).

Past detections in source and headspace will be identified from the full data set from 10/1/2008 to the present. A table will be created that can then be accessed by the CMM application. It is probably only necessary update this review approximately annually. Because a single survey contains multiple samples, there are cases where one survey will have multiple samples for the same agent. Rather than sum all results for a single agent to contribute to the HI for the sample, only the maximum value for each agent within the survey will be used to compute the HI for the sample.

To calculate the HI for each survey, both detect and non-detect results will be used (but only for those chemicals that have been actually detected in source or headspace results). For AX Farm only 2 furans have been detected in headspace and source samples: 2,3-Dihydrofuran and Tetrahydrofuran. Only detect and non-detect results for those two furans will be included in the aggregate HI calculations. Results can be plotted similar to the following chart which includes only the furans that have been detected at AX farm.

Personal Samples:

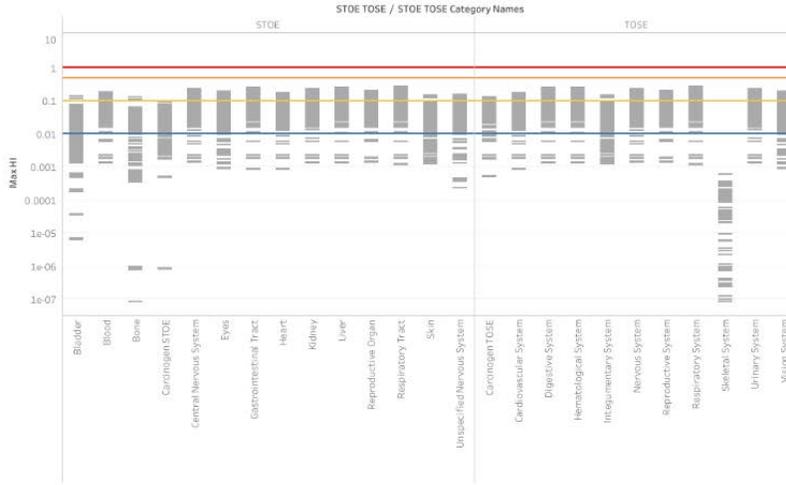
Target Organ System Effect (TOSE) and Specific Target Organ Effect (STOE) Hazard Index (HI) Sums for all Surveys
 Farm = AX FARM, AX FARM, AY FARM, AX FARM, AZ FARM



An UTL for each STOE-TOSE will be computed using the standard methodology but based on the aggregate HI for each survey. PNNL will review the data to determine if there are situations where two surveys cover the same location/time/job and may have been reported separately. These may need to be combined. If this situation is very rare, it may not be worth the effort to try to combine them. We will want the ability to also compute STOE-TOSE aggregate HIs using Area data. Area samples will be examined in the same manner as personal samples with HIs computed for all results within a survey. HI's will be based on an 8-hour TWA calculation taking into account the length of time that the sample was taken.

Mixtures with Hazard Index Plotted for all Surveys

Select survey data point to see detail; Farm: AX FARM



Select Date Range
All values

Select Tank Farm
Contains "ax farm"

Minimum Number of Detections
1

COPC Non-COCP
All

Detection
 Detect
 Non-Detect

Tank Origin
 Non-Tank Origin
 Tank Origin

Tank Origin
■ Tank Origin

COPC Non-COCP
■ COPC
■ Non-COCP

COPC Origin Contribution Bar

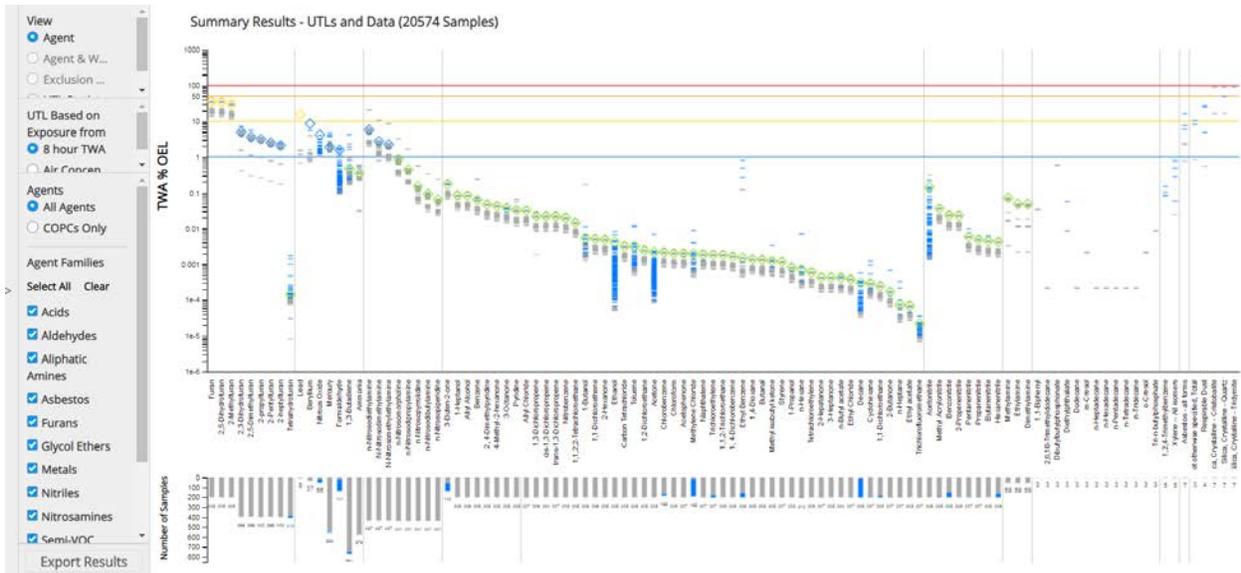
TOSE	Max HI
Bladder	~10
Blood	~50
Bone	~10
Carcinogen STOE	~30
Central Nervous System	~100

Tank Origin Contribution Bar

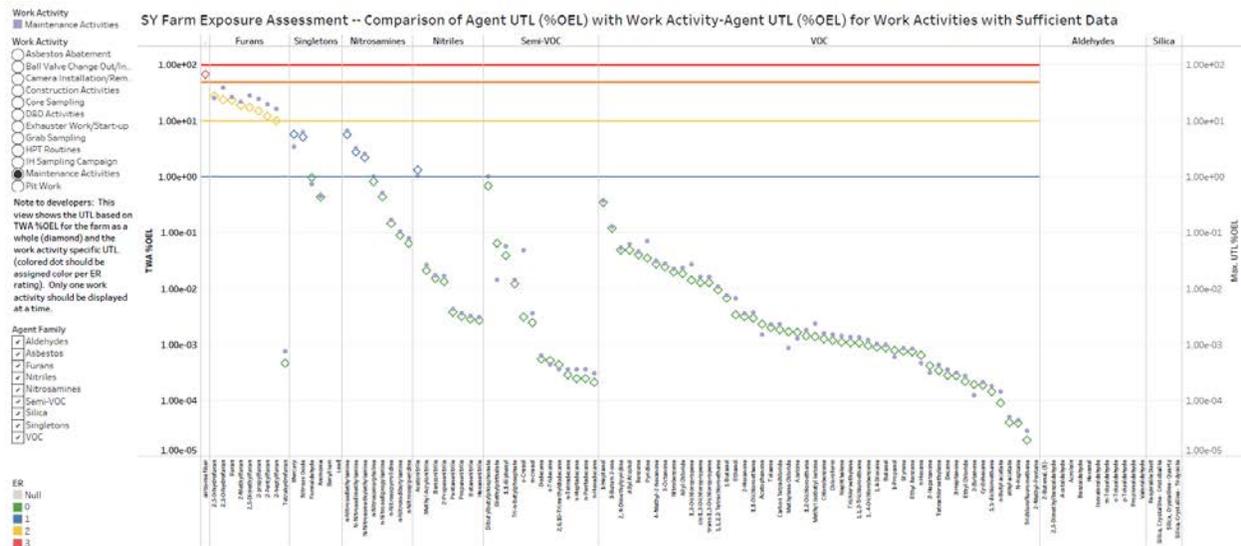
TOSE	Max HI
Bladder	~10
Blood	~50
Bone	~10
Carcinogen STOE	~30
Central Nervous System	~100

Appendix D

IDAV EA Results Views –and Multiple Detailed Results (Download only)



Graphical UTL summary results may be selected by Agent, 8-hour TWA or Air Concentration, All agents or COPCs, Agent Families



Graphical UTL results may be selected by Agent & Work Activity, 8-hour TWA or Air Concentration; All agents or COPCs, and Agent Families

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
CAS	AGENT	TWA_OEL	OBSERVATION	DETECT	PERCENT_DETECT	TWA_UTIL	TWA_UTIL_METHOD	TWA_UTIL_PERCENT_OF	TWA_HIGH	TWA_HIGH_PERCENT_OF	TWA_ER	TWA_ER_JUSTIFICATION				
71-36-3	1-Butanol	20	189	17	8.99	0.002490968	Binomial Look-up Table	0.00254534	0.033647364	0.16823682	0	<1% of OEL				
111-70-6	1-Heptanol	0.5	206	0	0	0.000423362	Binomial Look-up Table	0.0846724	0.000505431	0.0108662	0	<1% of OEL				
71-23-8	1-Propanol	100	207	4	1.93	0.000800437	Binomial Look-up Table	0.000800437	0.000977233	0.000977233	0	<1% of OEL				
106-46-7	1,4-Dichlorobenzene	10	205	0	0	0.000163622	Binomial Look-up Table	0.00163622	0.000199762	0.000199762	0	<1% of OEL				
92-52-4	1,1-Biphenyl	0.2	3	0	0	#N/A	Insufficient sample size to estimate L	#N/A	0.0000686	0.0343	#N/A	Data Needed				
75-34-3	1,1-dichloroethane	100	205	0	0	0.000243059	Binomial Look-up Table	0.000243059	0.000296745	0.000296745	0	<1% of OEL				
75-35-4	1,1-Dichloroethene	5	205	0	0	0.000250553	Binomial Look-up Table	0.0001106	0.000305894	0.00011788	0	<1% of OEL				
79-00-5	1,1,2-Trichloroethane	10	205	0	0	0.000182221	Binomial Look-up Table	0.00182221	0.000224669	0.000224669	0	<1% of OEL				
79-34-5	1,1,2,2-Tetrachloroethane	1	205	0	0	0.000144898	Binomial Look-up Table	0.0144898	0.000176903	0.0176903	0	<1% of OEL				
107-06-2	1,2-Dichloroethane	10	205	0	0	0.000243207	Binomial Look-up Table	0.00243207	0.000296925	0.00296925	0	<1% of OEL				
95-63-6	1,2,4-Trimethylbenzene	25	5	3	60	#N/A	Insufficient sample size to estimate L	#N/A	0.036958781	0.147835124	#N/A	Data Needed				
106-99-0	1,3-Butadiene	1	772	22	2.83	0.004695674	Binomial Look-up Table	0.4695674	0.041939706	4.1939706	0	<1% of OEL				
142-75-6	1,3-Dichloropropene	1	205	0	0	0.000219508	Binomial Look-up Table	0.0219508	0.000286974	0.0286974	0	<1% of OEL				
128-91-1	1,4-Dioxane	20	205	0	0	0.000273331	Binomial Look-up Table	0.001366655	0.000333703	0.001666515	0	<1% of OEL				
78-93-3	2-Butanone	200	207	4	1.93	0.000377894	Binomial Look-up Table	0.000168947	0.001322755	0.000661378	0	<1% of OEL				
110-43-0	2-Hepatanone	50	205	0	0	0.000210992	Binomial Look-up Table	0.000421984	0.000257955	0.00051591	0	<1% of OEL				
3777-71-7	2-heptylfuran	1	402	1	0.25	0.020730222	Binomial Look-up Table	2.0730222	0.02462193	2.462193	1	<10% of OEL				
591-78-6	2-Hexanone	5	207	0	0	0.000240147	Binomial Look-up Table	0.00480294	0.000293189	0.00586378	0	<1% of OEL				
534-22-5	2-Methylfuran	1	205	0	0	0.292966406	Binomial Look-up Table	29.2966406	0.357675125	35.7675125	2	10-50% of OEL				
3777-69-3	2-Pentylfuran	1	398	0	0	0.024932012	Binomial Look-up Table	2.4932012	0.027900113	2.7900113	1	<10% of OEL				
107-13-1	2-Propenenitrile	2	205	0	0	0.000453832	Binomial Look-up Table	0.0226916	0.000554073	0.02770165	0	<1% of OEL				
4229-91-8	2-propylfuran	1	402	0	0	0.01283287	Binomial Look-up Table	3.1283287	0.035807487	3.5807487	1	<10% of OEL				
108-47-4	2,4-Dimethylpyridine	0.5	205	0	0	0.000232299	Binomial Look-up Table	0.04646598	0.000234829	0.04646598	0	<1% of OEL				
1191-99-7	2,3-Dihydrofuran	1	398	2	0.5	0.04933553	Binomial Look-up Table	4.933553	0.071235472	7.1235472	1	<10% of OEL				
1708-29-8	2,5-Dihydrofuran	1	205	0	0	0.343169844	Binomial Look-up Table	34.3169844	0.418967375	41.8967375	2	10-50% of OEL				
625-86-5	2,5-Dimethylfuran	1	398	1	0.25	0.035853595	Binomial Look-up Table	3.5853595	0.05069134	5.069134	1	<10% of OEL				
3891-98-3	2,6,10-Trimethyldecane	200	3	0	0	#N/A	Insufficient sample size to estimate L	#N/A	0.0004329	0.0004329	#N/A	Data Needed				
78-94-4	3-Buten-2-one	0.2	207	2	0.97	0.000343271	Binomial Look-up Table	0.1716355	0.000418967	0.2094835	0	<1% of OEL				
106-35-4	3-Heptanone	50	207	0	0	0.000210992	Binomial Look-up Table	0.000421984	0.000257955	0.00051591	0	<1% of OEL				
106-68-3	3-Octanone	0.5	206	3	1.46	0.000187604	Binomial Look-up Table	0.0375208	0.000229041	0.0458082	0	<1% of OEL				
105-42-0	4-Methyl-2-hexanone	0.5	205	0	0	0.000216941	Binomial Look-up Table	0.0421282	0.000257166	0.0514332	0	<1% of OEL				
67-64-1	Acetone	250	205	189	92.1	0.000368223	Binomial Look-up Table	0.002147252	0.01673745	0.00694898	0	<1% of OEL				
75-05-8	Acetonitrile	20	142	83	58.45	0.027996531	Binomial Look-up Table	0.139982655	0.061952275	0.309761375	0	<1% of OEL				
98-96-2	Acetophenone	10	205	1	0.49	0.000200192	Binomial Look-up Table	0.00200192	0.00024441	0.0024441	0	<1% of OEL				
107-18-6	Allyl Alcohol	0.5	204	0	0	0.000414138	Binomial Look-up Table	0.0828276	0.00050561	0.101132	0	<1% of OEL				
107-05-1	Allyl Chloride	1	205	0	0	0.00031442	Binomial Look-up Table	0.031442	0.000383867	0.0383867	0	<1% of OEL				
7664-41-7	Ammonia	25	576	2	0.35	0.085924894	Binomial Look-up Table	0.343699576	0.14836535	1.4836535	0	<1% of OEL				

Tabular summary results can be selected by 8-hour TWA or Air Concentration, and All agents or COPCs, Exposure Assessment tables analogous to those in Appendices B, C, and D are available as downloadable MS Excel workbooks.

Appendix B - AY AZ Tank Farms Number of Workers Sampled by Activity

Work Activity	Type of Worker																	
	Contractor	Electrician	First Line Supervisor	Health Physics Technician	Helper-Laborer (General)	Industrial Health/Safety Tech	Instrument & Control Tech	Millwright	Nuclear Waste Process Operator	Other-Type Craft Worker	Other-Type Laborer/ Service Wkrs	Other-Type Professional	Other-Type Technician	Painter	Plant Engineer	Plumber/ Pipefitter	Safety Engineer	Utility System Operator
Area/RAE Checks						10												
Asbestos Abatement	1														1			
Beryllium Abatement						1												
Breather Filter Work/Maintenance									4									
Cam Replacement							5											
Camera Installation/Removal	34			29		47			201						23			
Confined Space Entry									1									
Construction Activities	25	14		55	16	29			12	488	34	30	7			12		
Core Sampling									6									

Appendix C -- Exposure Risk Ratings by Agent

Agent	Work Activity																							
	Area/RAE Checks	Asbestos Abatement	Beryllium Abatement	Breather Filter Work/Maintenance	Cam Replacement	Camera Installation/Removal	Confined Space Entry	Construction Activities	Core Sampling	Drilling	Exhaustor Work Start-Up	Hammable Gas Checks	Grab Sampling	Grinding	Hose in Hoses	Hot Surfaces	Hot Work	Hot Work	Maintenance Activities	Operator Routine	Painting	Plt Work	Refinery Operations	Transfer Activities
1-Butanol	2	0	0	0	0	22	0	17	1	0	0	0	0	0	17	20	3	27	18	0	2	18	18	16
1-Heptanol	2	0	0	0	0	22	0	17	1	0	0	0	0	0	17	20	3	27	18	0	2	18	18	26
1-Propanol	2	0	0	0	0	22	0	17	1	0	0	0	0	0	17	20	3	27	18	0	2	18	18	27
1,4-Dichlorobenzene	2	0	0	0	0	22	0	17	1	0	0	0	0	0	17	20	3	27	18	0	2	18	18	25
1,4-Biphenyl	2	0	0	0	0	22	0	17	1	0	0	0	0	0	17	20	3	27	18	0	2	18	18	25
1,1-dichloroethane	2	0	0	0	0	22	0	17	1	0	0	0	0	0	17	20	3	27	18	0	2	18	18	25

Appendix D - AY AZ Tank Farms Comprehensive Exposure Results

Hazard	Exposure Rating (0-4)	Basis for the ER	OEL	Exposure Judgment	Number of Observations	Health Effect Rating (0-5)	Uncertainty Rating (0-2)	Priority Rating (Pe = ER*HE*U) (0-8)	Priority	Assessment (ER+HE*U) (0-12)	Assessment Interval
1-Butanol	0	<1% of OEL	20 ppm	Compliant	189	3	0	0	Low Priority	0	As Needed
1-Heptanol	0	<1% of OEL	0.5 ppm	Compliant	206	3	0	0	Low Priority	0	As Needed
1-Propanol	0	<1% of OEL	100 ppm	Compliant	207	2	0	0	Low Priority	0	As Needed
1,4-Dichlorobenzene	0	<1% of OEL	10 ppm	Compliant	205	4	0	0	Low Priority	0	As Needed
1,1-Biphenyl	--	Data Needed	0.2 ppm	Compliant	3	4	2	--	--	--	--

Additionally, exclusion reports, extensive statistical results, and data records by agent can be extracted.

Excluded Data

Survey ID	Agent1	TWA 30	Units OK	Suspect not VOC	Furan	Volume Errors
Grand Total						10,624
12-04821	2-Methylfuran	OK	OK	Suspect	OK	1
	2,5-Dihydrofuran	OK	OK	Suspect	OK	1
	Furan	OK	OK	Suspect	OK	1
13-00029	1,3-Butadiene	Not OK	OK	OK	OK	1
	Ammonia	Not OK	OK	OK	OK	1
	Formaldehyde	Not OK	OK	OK	OK	1
	Mercury	Not OK	OK	OK	OK	1
13-00329	1,3-Butadiene	Not OK	OK	OK	OK	1
	Ammonia	Not OK	OK	OK	OK	1
	Formaldehyde	Not OK	OK	OK	OK	1
	Mercury	Not OK	OK	OK	OK	1
13-00542	1,3-Butadiene	Not OK	OK	OK	OK	1
	Ammonia	Not OK	OK	OK	OK	1
	Formaldehyde	Not OK	OK	OK	OK	1
	Mercury	Not OK	OK	OK	OK	1
13-00584	1,3-Butadiene	Not OK	OK	OK	OK	1
	Ammonia	Not OK	OK	OK	OK	1
	Formaldehyde	Not OK	OK	OK	OK	1
	Mercury	Not OK	OK	OK	OK	1
13-00635	1,3-Butadiene	Not OK	OK	OK	OK	1
	Ammonia	Not OK	OK	OK	OK	1
	Formaldehyde	Not OK	OK	OK	OK	1
	Mercury	Not OK	OK	OK	OK	1
13-00711	1,3-Butadiene	Not OK	OK	OK	OK	1
	Ammonia	Not OK	OK	OK	OK	1

Statistical Results

Agent: All

Agent Family: All

Agent	UTL TWA	# of Samples	MEDIAN TWA	MEAN TWA	GEOMETRIC MEAN TWA	GEOMETRIC SD TWA	NORMAL SW P-VALUE TWA	LOGNORMAL SW P-VALUE TWA
1-Butanol	0.001649068	189	0.000333538	0.000584752	0.000391126	1.580779464	Null	Null
1-Heptanol	0.00423362	206	0.000210949	0.000234558	0.000227812	1.2498802775	Null	Null
1-Propanol	0.00220537	207	0.000405837	0.000429514	0.000428888	1.178851130	Null	Null
1,4-Dichlorobenzene	0.002203622	205	8.31E-05	8.724592220E-05	8.59320494E-05	1.178921712	Null	Null
1,1-Biphenyl	Null	3	6.15E-05	6.64616667E-05	6.444967508E-05	1.002810120	Null	Null
1,4-Dichlorobenzene	0.00243038	205	0.00123411	0.00129567	0.00127651	1.171604495	Null	Null
1,1-Dichloroethane	0.00230553	205	0.00127217	0.00133593	0.00131587	1.171606500	Null	Null
1,1,2-Trichloroethane	0.00232221	205	9.25E-05	9.715787317E-05	9.569960452E-05	1.171601965	Null	Null
1,1,2,2-Tetrachloroethane	0.00214458	205	7.38E-05	7.76038512E-05	7.610000574E-05	1.171595992	Null	Null
1,2-Dichloroethane	0.00243207	205	0.00113408	0.00123676	0.00127728	1.171607180	Null	Null

Agent

1-Butanol

UTL Results for 1-Butanol

twa oel	twa high	twa high perc	twa utl	twa utl perc	exp rat	no. of samp
20	0.033647364	0.16823682	0.001049068	0.00524534	0	189

TWA Data Used in UTL Calculations for 1-Butanol

Sample Number	Total Time (min.)	Volume (L)	Detection	Air Concentration (Tableau)	Air Concentration (Tableau) UOM	TWA Exposure (Tableau)	TWA Exposure (Tableau) UOM
17-00632-5-021	42	8.5806	Non-Detect	0.3845413	ppm	0.033647364	ppm
14-01389-3-011	55	2.7544	Detect	0.01916698	ppm	0.002196216	ppm
16-07540-1-1	90	4.75938	Detect	0.006932826	ppm	0.001299905	ppm
16-03254-1-001	34	1.76766	Detect	0.01735981	ppm	0.001229653	ppm
16-04795-4-11	155	7.719	Detect	0.003248727	ppm	0.001049068	ppm
14-01431-3-011	76	3.90564	Detect	0.006167247	ppm	0.000976481	ppm
14-01458-3-011	60	2.4978	Non-Detect	0.006340803	ppm	0.0007926	ppm
16-03337-1-001	37	1.851295	Detect	0.009624514	ppm	0.00074189	ppm
14-02513-2-003	35	1.70345	Detect	0.01007244	ppm	0.000734449	ppm
14-03020-1-002	255	13.188575	Detect	0.001325985	ppm	0.00070443	ppm
16-01042-1-001	60	2.9196	Non-Detect	0.005424736	ppm	0.000678092	ppm
16-02188-1-001	141	6.93579	Non-Detect	0.002283526	ppm	0.000670786	ppm
16-02402-1-001	42	2.08509	Non-Detect	0.007595864	ppm	0.000664638	ppm
14-03540-2-003	58	2.82054	Detect	0.005498273	ppm	0.000664375	ppm
16-02100-1-001	141	7.0277	Non-Detect	0.003260004	ppm	0.000662003	ppm



**Pacific
Northwest**
NATIONAL LABORATORY

www.pnnl.gov

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99352
1-888-375-PNNL (7665)

U.S. DEPARTMENT OF
ENERGY