Modeling and Regulatory Support Workgroup Meeting July 7, 2020

Remote Access Only (see next slides)

Remote Access Options

Equipment Type	Access Information	Notes
Computers with microphones and speakers	Join Microsoft Teams Meeting **Please mute your microphone** unless you want to provide input.	Press control and click on this link to bring up Microsoft Teams through the internet. You can view the screen share and communicate through your computer's speakers and microphone
Computers without audio capabilities, or audio that is not working	Join Microsoft Teams Meeting (888) 404-2493 Passcode: 371 817 961# **Please mute your phone** unless you want to provide input.	Follow instructions above **Turn down your computer speakers, mute your computer microphone,** and dial the toll-free number through your phone and enter the passcode
Phone only	(888) 404-2493 Passcode: 371 817 961# **Please mute your phone** unless you want to provide input.	Dial the toll-free number and enter the passcode

Remote Access Guidelines

- This meeting will open 30 minutes prior to the official meeting start time to allow users to test equipment and ensure communication methods are working
- If you dial in through your phone, mute your microphone and turn down your speakers to avoid feedback
- Unless you are speaking, please mute your computer or device microphone and phone microphone to minimize background noise

Agenda

- Discussion of model output formats
  - Definitions of common model terms using volcano analogy
  - Application of model terms to UNRBA watershed model
  - Model performance examples
  - Discussion relative to UNRBA modeling
- Discuss potential training topics for MRSW
- Modeling and Regulatory Support status

Discussion of Model Output Formats

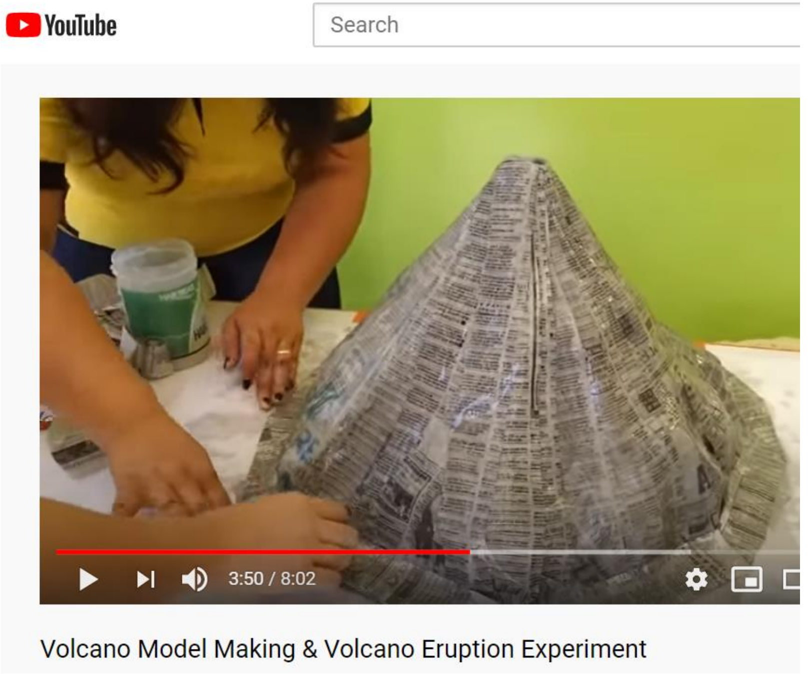
Definitions of Common Model Terms Using Volcano Analogy

Model

- A simplified representation of a system or process(es)
- Example – a volcano that erupts as high as the mountain is tall
- Reality – a complex set of conditions and interactions that cause the eruption and result in a specific eruption height
- Different types of models
  - Paper mache volcano with vinegar and baking soda
  - 2-L of soda with Mentos candy
  - USGS monitoring and computer modeling to predict eruptions and extents of ash plumes

Model Development

- Setting up the model so it will complete a simulation
  - Configure the model – build the paper mache volcano and let it dry
  - Obtain the inputs needed for the model – vinegar and baking soda
  - Initialize the model inputs – measure and record the amounts of vinegar and baking soda used for your first model run
  - Set up the model performance criteria – set up a video camera and a yard stick to measure the height of the eruption



Model Calibration

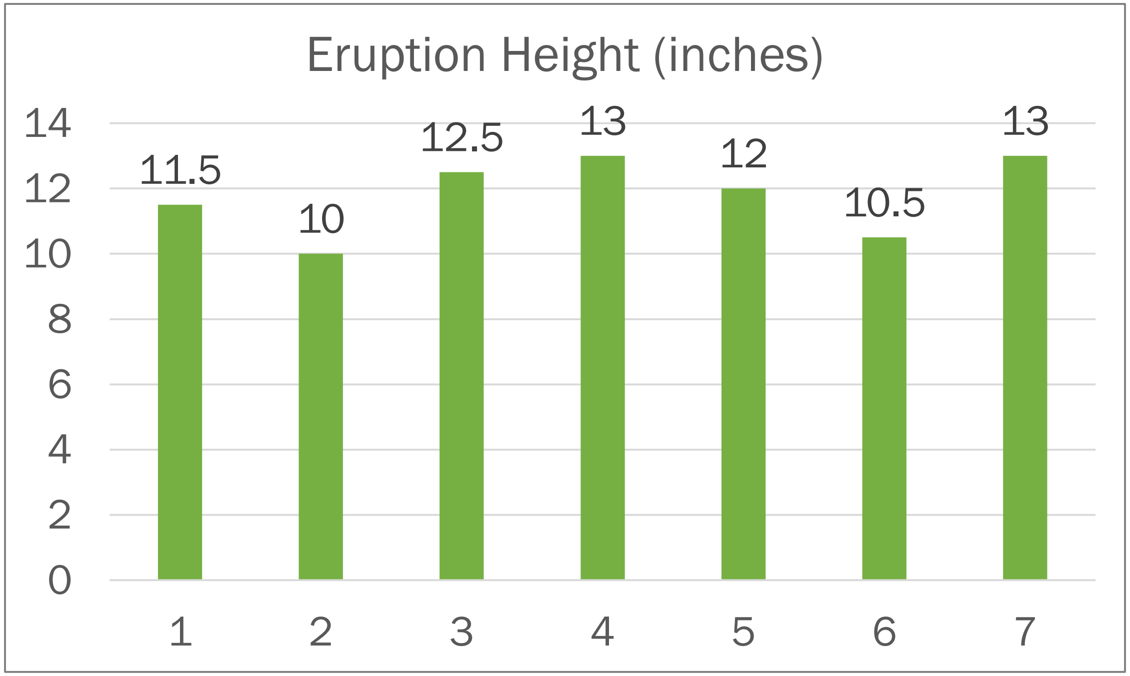
- Adjustment of model “knobs” (constants and rates) so simulated values match observed
  - Adjust the amount of vinegar
  - Adjust the amount of baking soda
- Continue adjustments until simulations match observations
  - Until the height of the eruption is equivalent to the height of the mountain

Model Validation

- Testing the model for an independent period to confirm it matches observations without adjustment of model "knobs"
  - Use the same temperature and amount of vinegar and baking soda
  - Repeat the eruption
  - Measure the height
  - Confirm it is still equivalent to the height of the mountain

Model Performance

- An assessment of how well the model simulates conditions relative to observations.

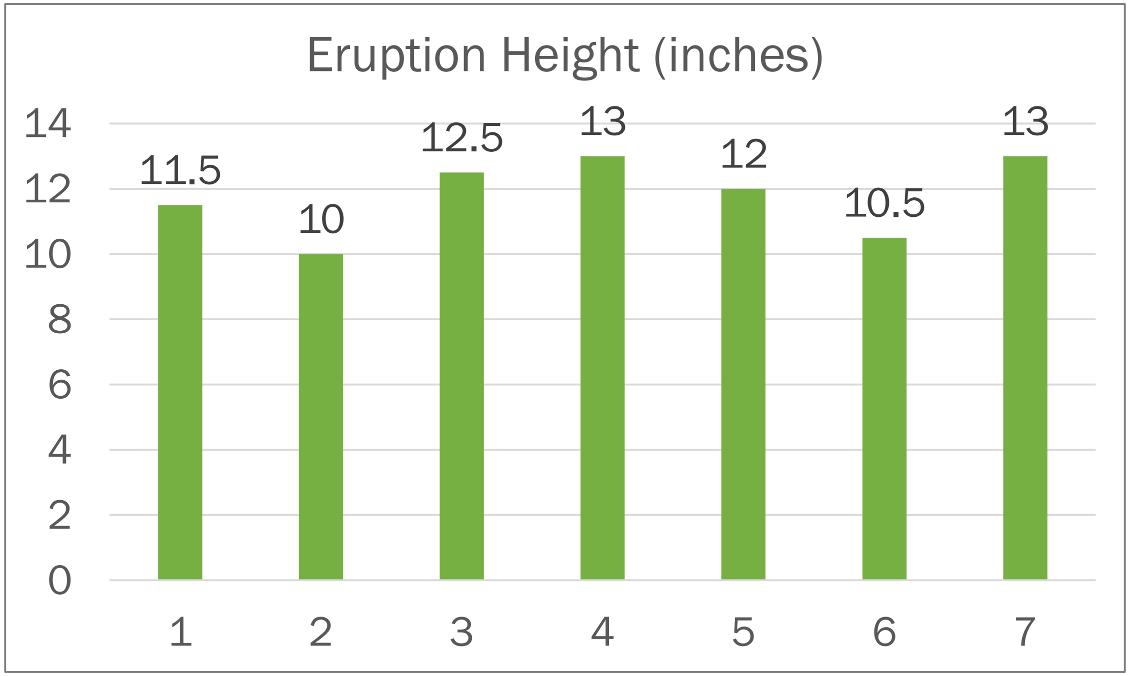


If we had built a 12 inch mountain

If we had built a 40 inch mountain

Model Output

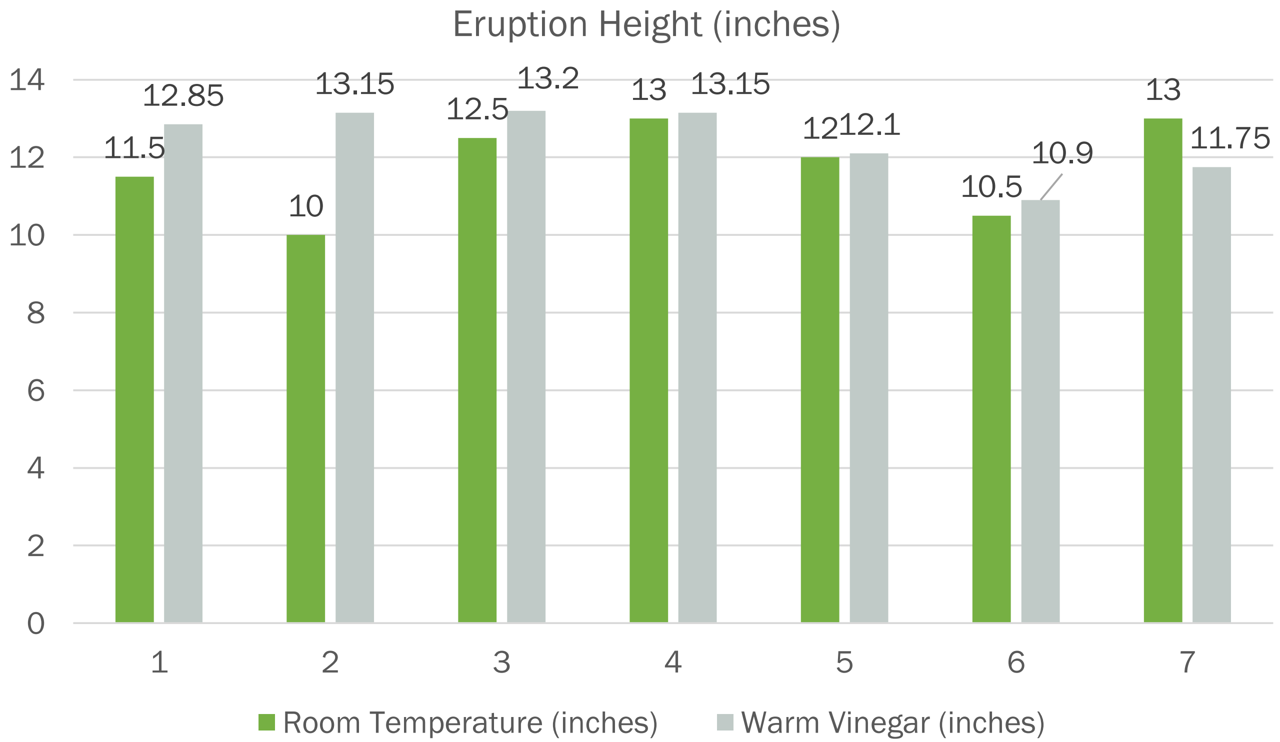
- The results of the model simulations that can be evaluated to answer questions
  - Using the calibrated amounts of vinegar and baking soda, what was the average height of the eruption after seven tests?



The average height of the eruptions from these seven tests is 11.8 inches.

Scenario

- A model run where a key input is changed and model output is evaluated for changes
  - **What happens if** the vinegar is warmed from room temperature?

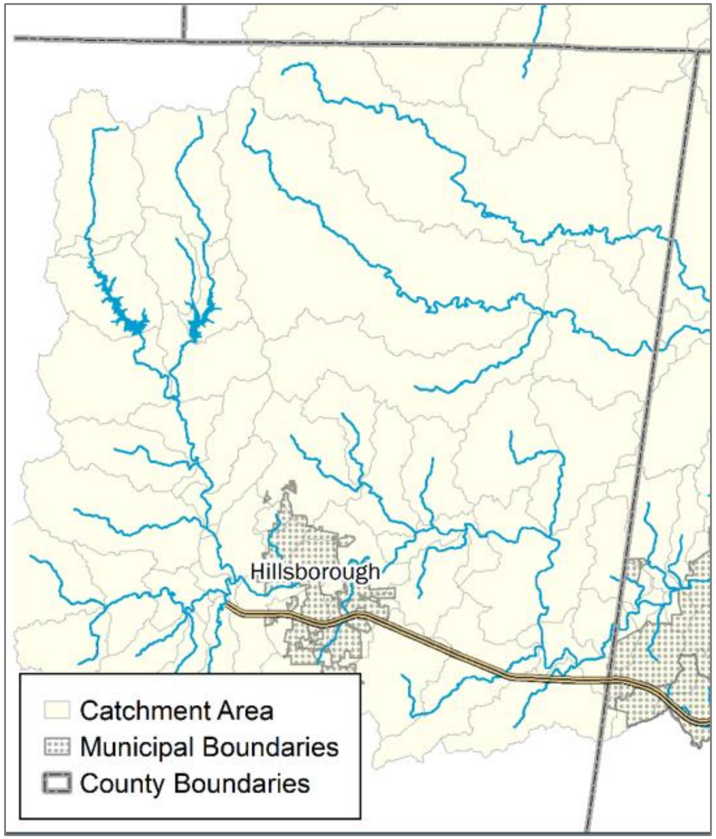


When the vinegar is warmed, the average height of the eruptions from seven additional tests is 12.4 inches.

Application of Model Terms to UNRBA Watershed Model

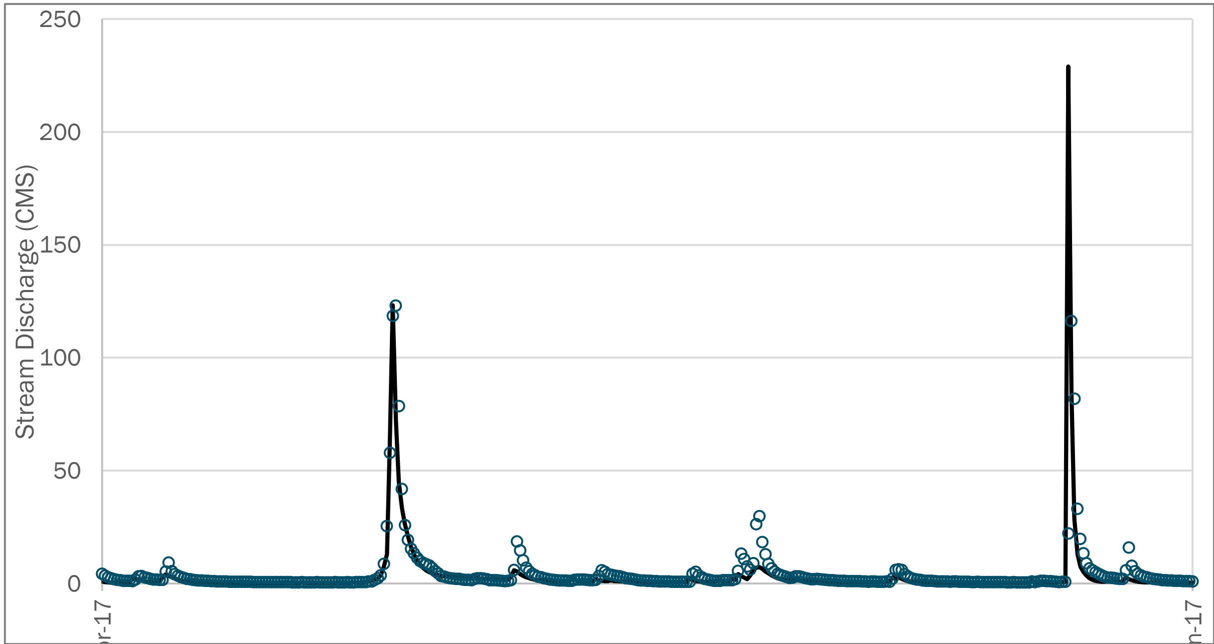
Current UNRBA Watershed Modeling

- Model: Falls Lake Watershed Analysis Risk Management Framework (WARMF)
- Development
  - Set up watershed modeling catchments
  - Acquired and formatted inputs
    - Rainfall
    - Land use
    - Soil characteristics
    - WWTP discharges
    - Etc.



Current UNRBA Watershed Modeling

- Calibration (2015 to 2016)
  - Adjusted hydrologic model parameters like evaporation magnitude and skewness so simulated stream flows match USGS observations
- Validation (2017 to 2018)
  - Ran the model for an independent period with out adjustment



Upcoming Evaluation of Model Output

- So far has been used to evaluate model performance using calibration and validation runs
  - Visual comparisons
  - Performance criteria (statistics)
- Additional output summaries can be developed for current status (hydrologic calibration)
  - Annual stream flow volumes by tributary
  - Monthly stream flow volumes by geologic basin
  - How stream flow volumes vary by storm size
- Primary focus is nutrient loading to Falls Lake, which would be evaluated after water quality calibration
- Work with the MRSW to determine what output summaries should be included in the model report

Upcoming Evaluation of Scenarios

- So far, model scenarios have not been run (models need to be calibrated first)
- UNRBA is tracking potential scenarios based on internal and external feedback
- Scenarios will be used to compare model output and answer questions, for example

How does nutrient loading change if street sweeping is implemented within all municipal boundaries every quarter?

How do seasonal nutrient loads change if lawn fertilizer application rates are reduced by 20 percent?

How might technology improvements at minor WWTPs affect nutrient loading?

Primary Types of Model Output

Calibration and Validation Output (Focus Today)

- Summarizes how well the models perform when compared to observations
- Good calibration and validation provides more confidence that the model simulations are relatively accurate and can be used to inform decisions

Model Output Summaries; e.g., Nutrient Loading

- By period (e.g., monthly, annual)
- By location (e.g., county, subwatershed)
- By source (e.g., atmospheric deposition, point sources)

Scenario Comparisons

- Compares the results of “what if” questions or different conditions
- Informs management decisions

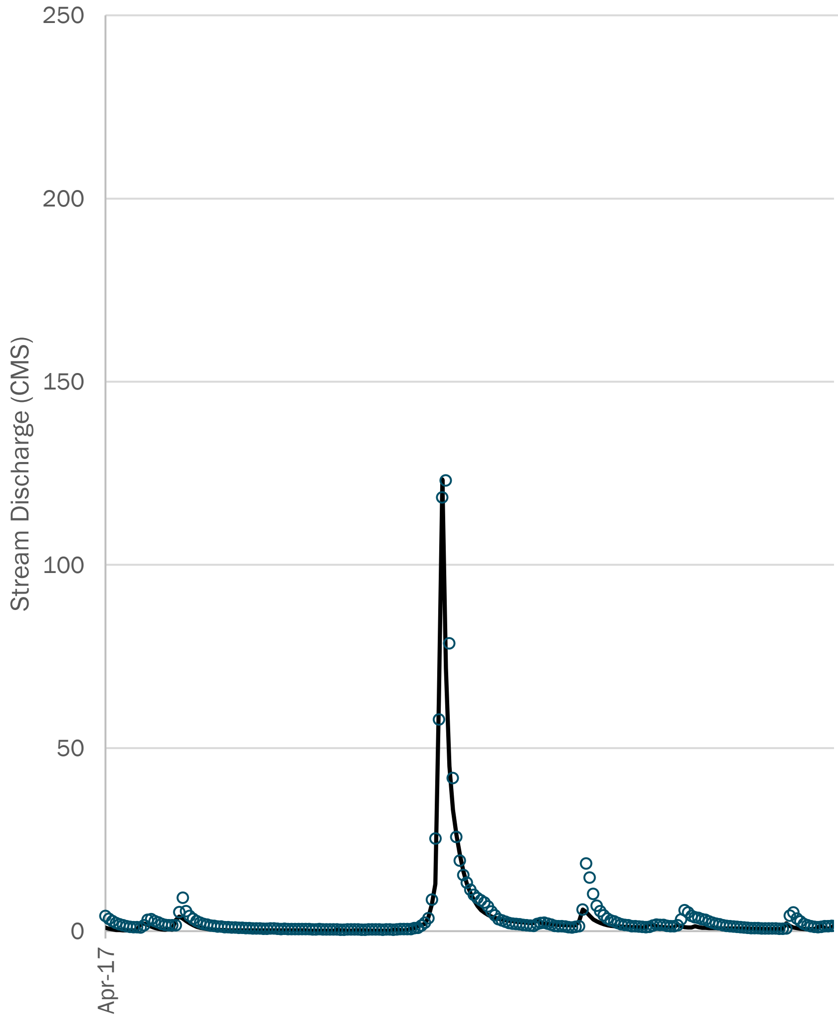
Model Performance Examples

UNRBA Modeling Quality Assurance Project Plan (QAPP)

- Specifies how model performance will be evaluated
  - Requires reporting of performance criteria
    - Quantitative and objective
    - 10 flow gages
    - 7 water quality monitoring stations
  - Visual comparisons can be made at other locations
    - Qualitative and subjective
    - Lists several examples of graphical outputs
    - Does not specify the type to be used
  - Graphical comparisons are particularly helpful when observed data are
    - Non-continuous
    - Consistently near zero (or the detection limit)
    - Low to zero variability

Example Time Series Chart

Comparison from WARMF Model, 6-hr time step



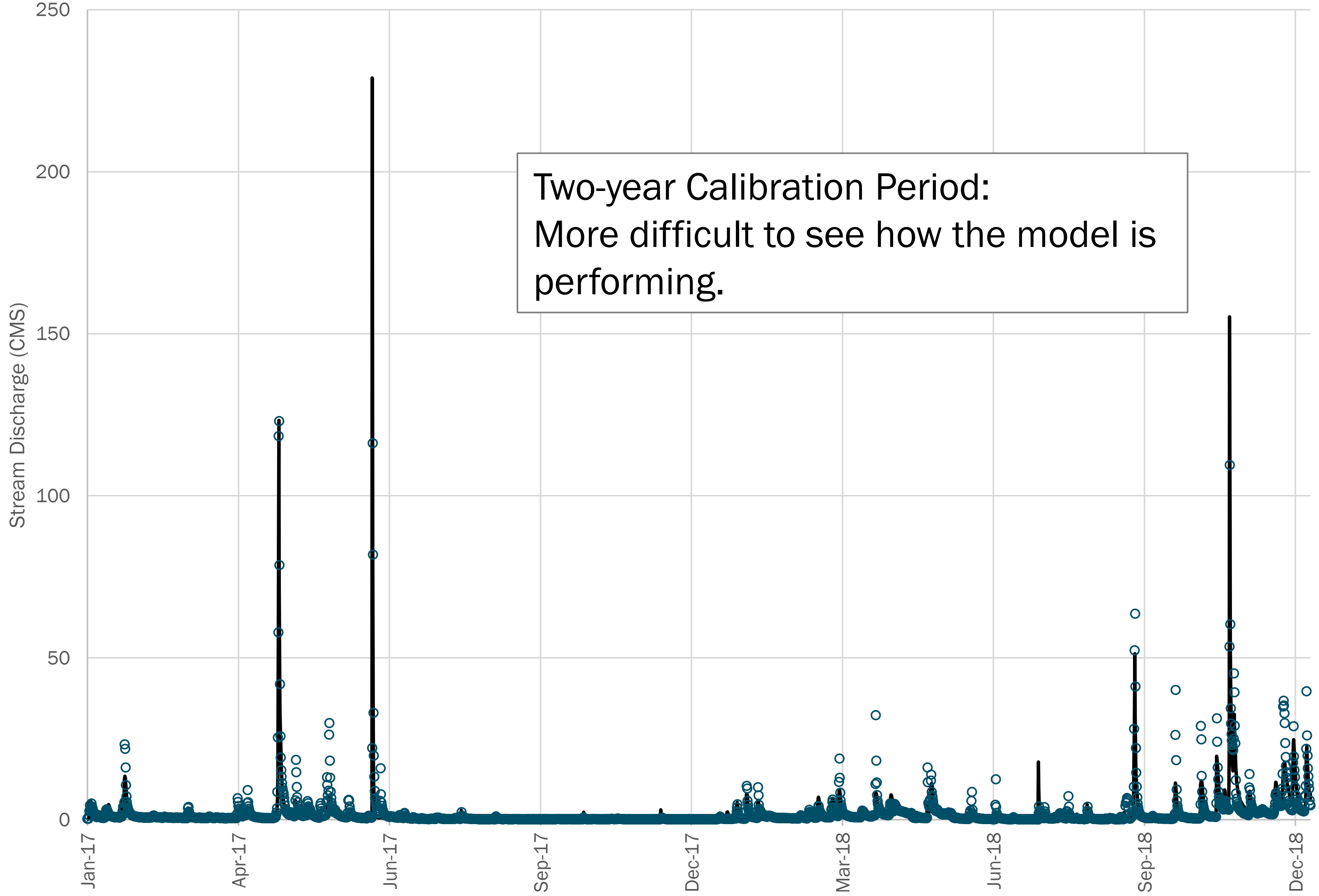
Zoomed into a 3 month period.

- Good for frequent observations
- Provides visual comparison at the model time step or aggregated
- Relatively easy, intuitive to interpret
- Difficult to read for longer periods
- Does not convey a quantitative assessment of overall performance

This figure shows that overall the model does a good job of simulating the timing and magnitude of stream discharge peaks. It shows some storm peaks are underestimated, some are over estimated, and others track fairly well, but it does not convey a quantitative summary of model performance.

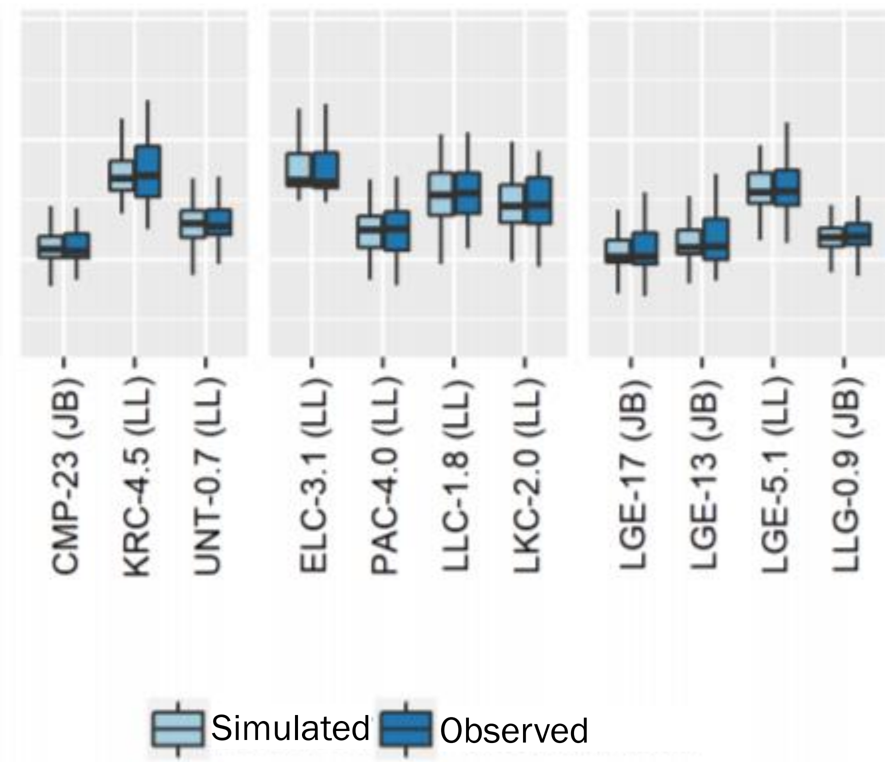
Example Time Series Chart

Comparison from WARMF Model, 6-hr time step



Example Box Plot

Comparison of LOADEST Model from Annual Report

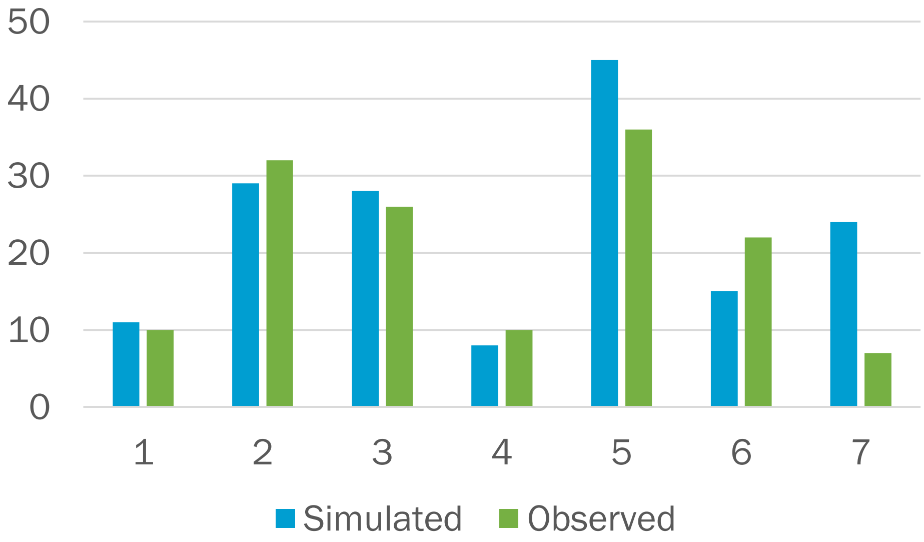


- Appropriate for infrequent observations (e.g,, grab samples)
- Provides visual comparison of the distribution of values aggregated over some time step; time step can be altered (monthly, annual, etc.)
- More complex to understand
- Potentially hides differences or mismatches in predicted timing as long as the totals match up

This figure shows that overall, the model simulates values that are similar to those observed, but it does not show details of timing.

Example Bar Chart

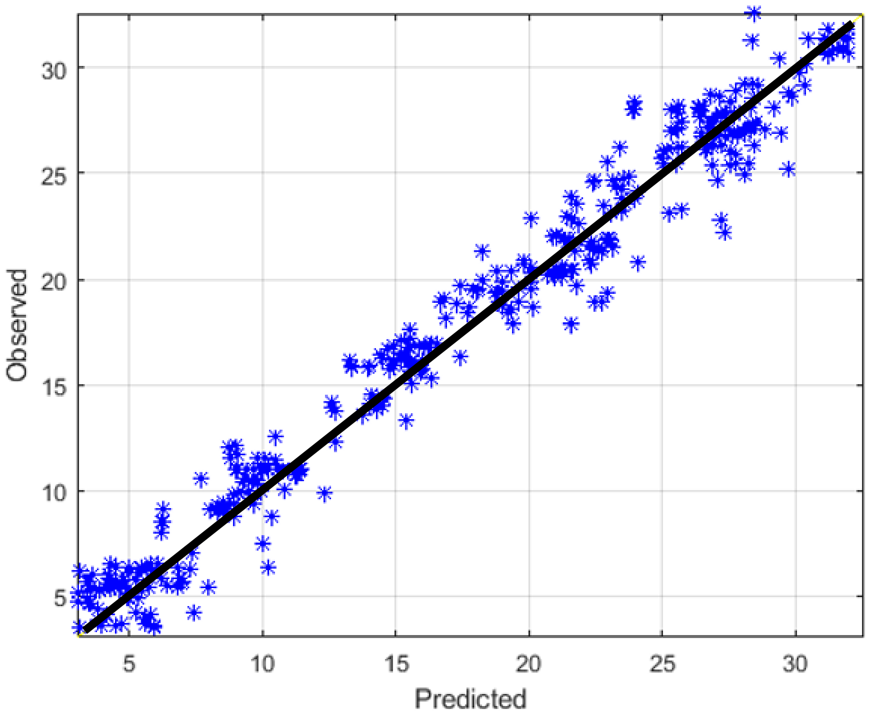
Generic Comparison for Seven Simulation Years



- Provides visual comparison at large time scale like monthly or annual
- May be appropriate for some parameters; lacks resolution
- Easy to understand
- Difficult to quantify performance

This generic example shows that the simulated values track well with observations for the first four periods, but do not perform as well for the last three periods.

Example Scatter Plot
from Jim Bowen's Presentation to the PFC

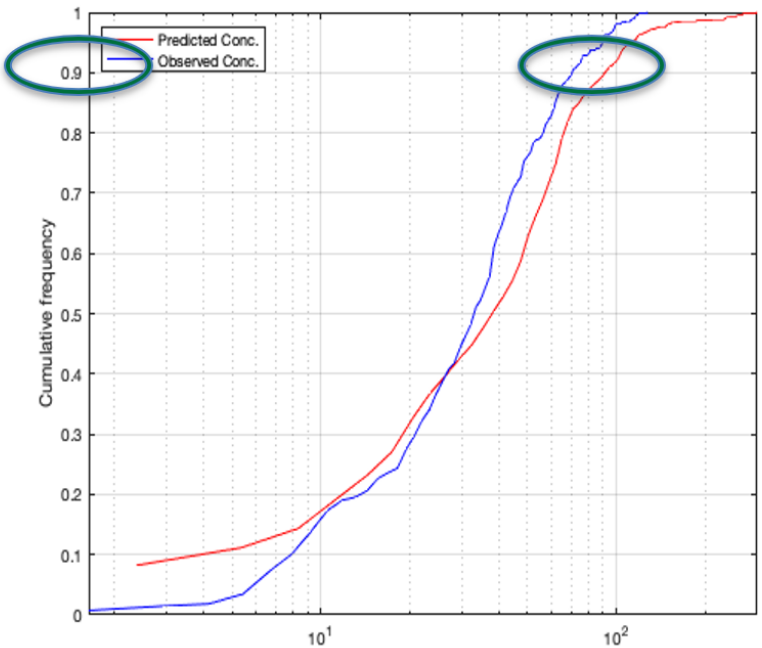


Credit: Dr. Jim Bowen, UNC Charlotte

- Provides visual comparison
- More complex to understand than time series
- Provides better visual assessment of performance
- Easier to see patterns of over or underestimation (i.e., bias)
- Can add statistics (e.g., R^2) to include a quantitative assessment

This figure shows that the model generally predicts observed temperatures within 2-3 degrees Celsius.

Cumulative Frequency Distribution
from Jim Bowen's Presentation to the PFC



Credit: Dr. Jim Bowen, UNC Charlotte

- Provides visual comparison
- Complex to understand
- Compares the percentage of values above or below a certain threshold like 10 percent exceedance
- Sampling regime can impact interpretation
- May be important consideration for chlorophyll-a simulations

This figure shows that the model predicts that chlorophyll-a concentrations would exceed 90 µg/L ten percent of the time. The observations indicate that chlorophyll-a concentrations are above 90 ug/L only five percent of the time. At this location and time period, the model overestimates chlorophyll-a.

Example Percent Difference Statistics

from WARMF Model, 6-hr time step

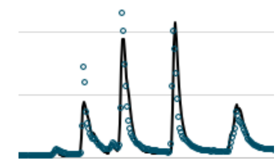
	Time Period	
	Calibration (2015-2016)	
	Statistic	Interpretation (QAPP)
Total Volume	7.7%	Good
Peak Flow	8.3%	Very Good
High Flow	9.1%	Very Good
Low Flow	-10.2%	Good
Winter	23.1%	Good
Spring	-29.5%	Good
Summer	-6.7%	Very Good
Fall	29.4%	Good

- Provides quantitative evaluation
- Complex to understand
- Color coding, etc., can be used to improve understanding
- Format to include more statistics
  - Average error
  - Coefficient of determination (R^2)
  - Ratio of Root Mean Square Error to Standard Deviation of the Observed Data RMSE/STDEVobs
  - Nash-Sutcliffe Efficiency

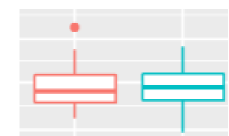
Discussion Relative to UNRBA Model Performance

Types of Graphical Output for Comparing Simulations to Observations

- The best type of output graphics will vary based on frequency of observations and key questions
- Time series comparisons
  - Good for frequent observations

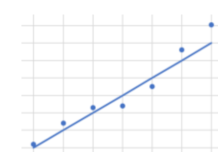


- Box-and-whisker plots
  - Appropriate for infrequent observations

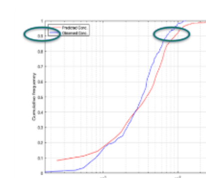


- Bar charts
  - Provide comparisons for large time steps
  - Do not show variability
  - Commonly used for comparing scenarios

- Scatter plots
  - All points would plot on the 1:1 line if in perfect agreement



- Cumulative frequency distributions
  - Compare the percentage of values above or below a certain threshold



Statistics Used to Compare Simulated to Observed Values

- Percent difference – how different are the simulations compared to the mean of the observations (0)

$$\%Diff = \frac{\sum P - O}{\sum O} \times 100$$

- Average error – how different are the simulations relative to the number of paired values (0)

$$AE = \frac{\sum (P - O)}{N}$$

- Coefficient of determination (R^2,1); less useful for infrequent sampling
- Ratio of Root Mean Square Error (RMSE) to Standard Deviation of the Observed Data ($STDEV_{obs}$) (0)

$$RSR = \frac{RMSE}{STDEV_{obs}} = \frac{\sqrt{\sum (P - O)^2}}{\sqrt{\sum \left(O - \bar{O}\right)^2}}$$

- Nash-Sutcliffe Efficiency (1)

$$NSE = 1 - RSR$$

Numbers in parentheses are best result

Considerations for Determining Formats

Observations are collected at different frequencies depending on the parameter	Stream flows - 15 minutes
	Water quality concentrations - monthly
Measurements have varying degrees of accuracy	Stream flows are estimates based on rating curves
	Many water quality parameters are often near reporting limits or highly variable
Graphical outputs (QAPP does not specify type)	Provide a good visual comparison, but not a quantitative one
	Can be supplemented with statistical comparisons

The QAPP specifies the number of stations that should include statistical comparisons and recommends that other stations use visual ones

Calibration of the Watershed Model

The QAPP focused on percent difference statistics

- Included a table of performance criteria and interpretative language
- Other statistics were specified for inclusion in the model reports

To date, the MRSW has reviewed percent difference statistics

- The watershed modelers evaluated other statistics during calibration, but focused on minimizing percent differences

Final reporting should include all of the statistics

Potential Calibration Reporting Formats: Stream Flow, Tabular

- Tables of statistics as specified in the QAPP
- 10 locations co-located with USGS flow gages
- Report for calibration (2015-16), validation (2017-18), and combined periods

	Percent Difference	Statistic 2	Statistic 3	Statistic 4	Statistic 5
Total Volume	#	#	#	#	#
Peak Flow	#	#	#	#	#
High Flow	#	#	#	#	#
Low Flow	#	#	#	#	#
Winter	#	#	#	#	#
Spring	#	#	#	#	#
Summer	#	#	#	#	#
Fall	#	#	#	#	#

Very good
Good
Fair

Potential Calibration Reporting Formats: Stream Flow, Graphical

- QAPP requires that performance criteria be provided for 10 locations
- 38 locations for visual comparison
  - 10 are co-located with a USGS stream flow gage
  - 28 would be compared to flow estimates
- Two to three periods could be compared
  - Calibration (2015-2016)
  - Validation (2017-2018)
  - Combined (2015-2018)
- Modeling Team is mindful of past feedback
  - Shorter, condensed reports are preferable
  - Information needs to communicate well
- Following slides will discuss pros and cons of different graphical formats (including ballpark page counts)

Potential Stream Flow Calibration Reporting Formats, Graphical

- Time series (38+ pages for all locations)
  - Provides direct comparison of simulated to observed values
  - May be difficult to read 2015 to 2018 on one figure
  - Will require many (~38) pages to display all locations
  - Could limit to the 10 locations with performance criteria
- Scatter plots (9+ pages for all locations)
  - Provides comparison of specific measurements across the flow regime, but does indicate specific storms
  - Can lay out many stations per period on one page
  - Could add R^2 or other statistic for 10 locations
  - Could limit scatter plots to 10 locations (3+ pages)
- Box plots (3 to 6+ pages for all locations)
  - Provide comparisons across the 38 monitoring stations in a more condensed format
  - Would show distributions of low and high flows, but not comparisons to specific measurements or storms

MRSW Discussion of Potential Reporting Formats for Stream Flow Calibration

- Tables of statistics (required) for 10 stations for calibration, validation, and combined period
- Graphics
  - Certain graphics may be more useful for the 10 stations for which we are generating statistics
  - Does the MRSW have a preference for one or more of the following at these 10 locations?
    - Time series (more detail, more pages (10+))
    - Scatter plots (moderate detail, moderate pages (3-6))
    - Box plots (less detail, less pages (3+))
  - When generating comparisons to flow estimates for other 28 stations with estimated flows, modelers recommend summarizing with box plots

MRSW Discussion of Potential Reporting Formats for Lake Level Calibration

- Observations at Falls Lake Dam and Beaverdam Lake Dam
- Tables of statistics as specified in the QAPP
  - Periods: calibration, validation, combined
  - WARMF lake is segment based; comparisons are not direct
  - EFDC is a gridded model so statistics are more relevant for comparison to a specific location
- Modelers suggest time series to evaluate performance visually

Discuss Potential Training Topics for MRSW

Potential Training Topics for MRSW

- Objectives
  - Increase understanding of model development and application
  - Answer questions and hear feedback
  - Increase comfort level for using the models to inform management decisions
  - Improve information sharing
  - Provide training on running the models for those interested
- Trainings can occur during model development and when models are being used to answer questions
- Suggested resource, particularly minutes 37:30 to 1:00:04:

https://www.epa.gov/waterdata/surface-water-quality-modeling-training

MRSW discuss potential training topics and thoughts on timing.

Modeling and Regulatory Support Status

MRS Status

- Executed FY2021 contracts
  - Prime contract between BC and UNRBA
  - Subcontracts between BC and
    - Systech Water Resources
    - Dynamic Solutions
    - KDV Analytics
- FY 2021 contracts to be developed
  - Ken Reckhow
  - Ashley Abernethy (economist)
- The Executive Director is reviewing a preliminary, interim draft for the WARMF watershed hydrologic modeling
- DEQ is in the process of finalizing the 319 grant contract with the UNRBA

Closing Comments
Additional
Discussion