CONNECTING THE DOTS TO INNOVATION

CORPS SHOALING ANALYSIS TOOL (CSAT) VALIDATION

SYSTEMATIC ACCURACY ASSESSMENT FOR END-USERS

Rachel Bain and Michael Hartman

FY25 and FY26 CSAT contributors: Enrique Mendez, Brooke Walker, Doyle Jones, Kate Richards, Kate Smith, and Jose Hernandez

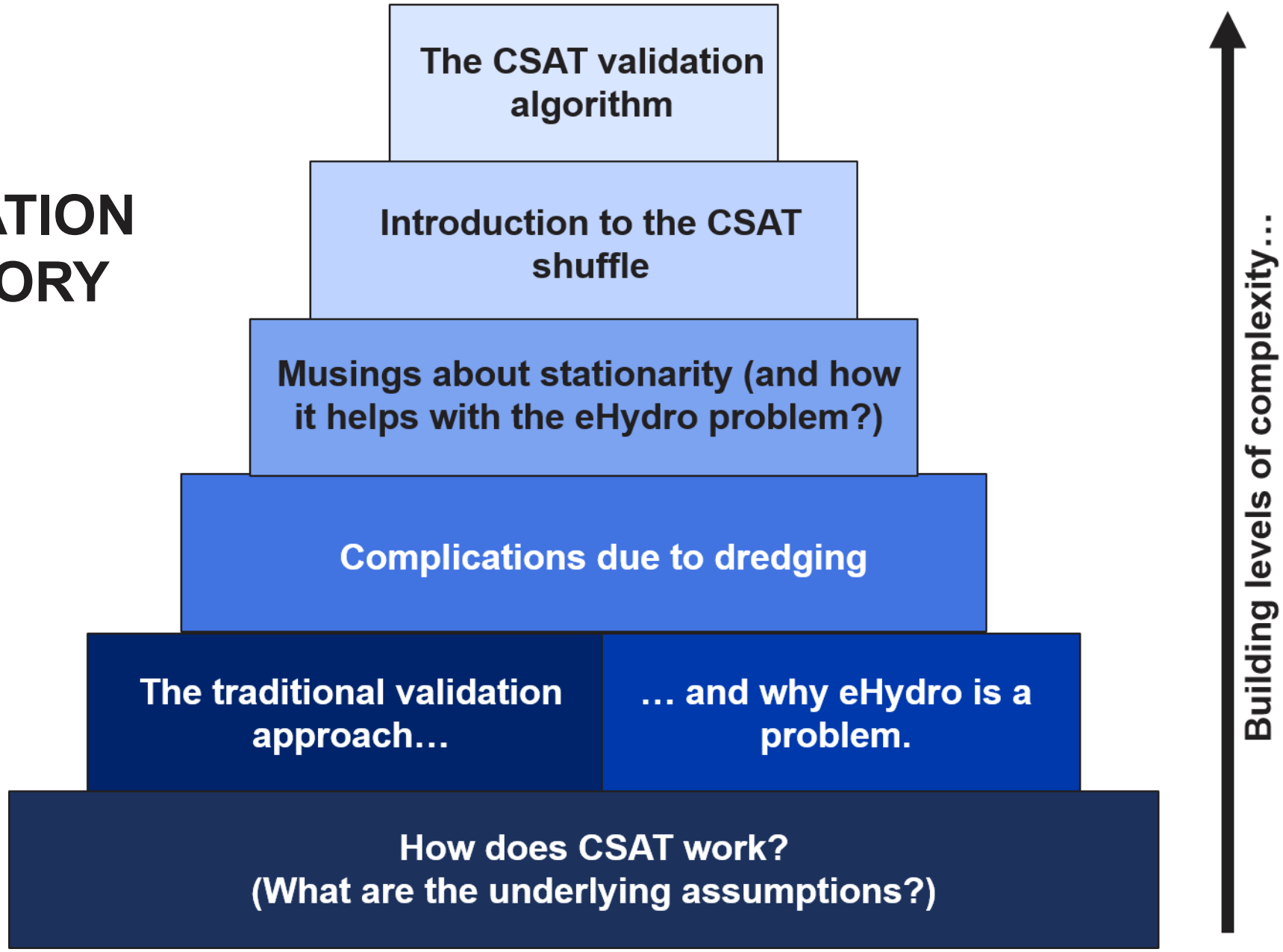
Coastal Inlets Research Program (CIRP) Technical Discussion • September 22, 2026

U.S. ARMY

US Army Corps of Engineers®

ERDC
ENGINEER RESEARCH & DEVELOPMENT CENTER

PRESENTATION TRAJECTORY



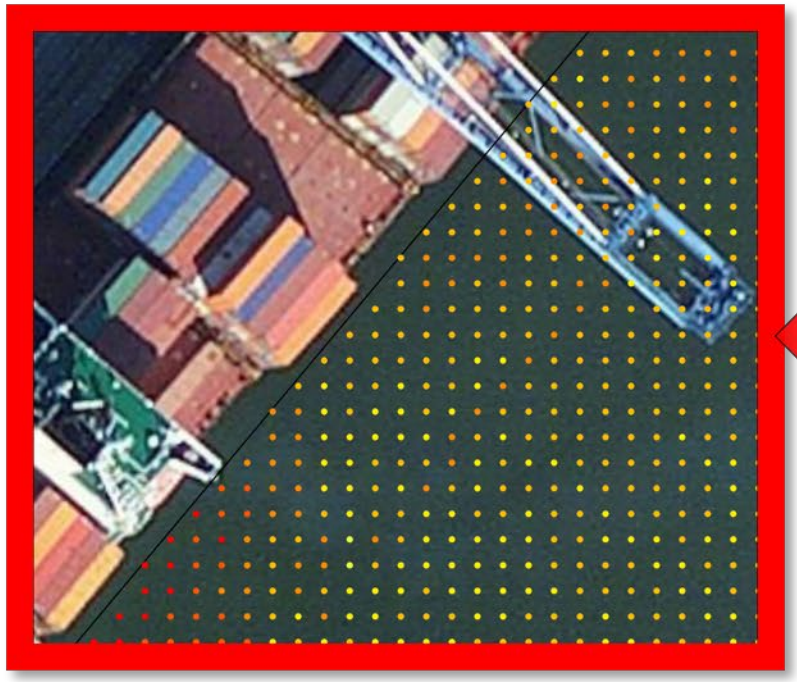
BACKGROUND

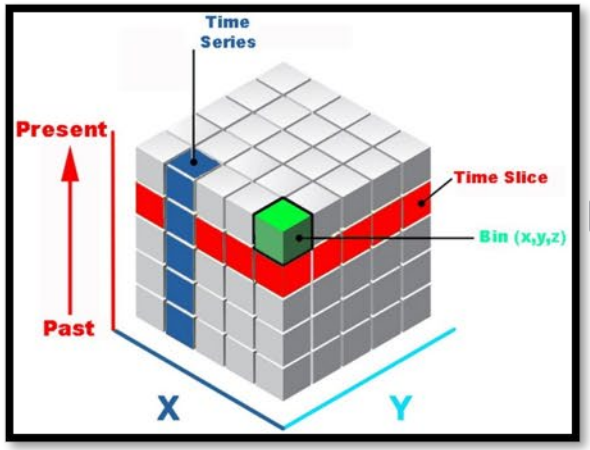
— the space-time cube

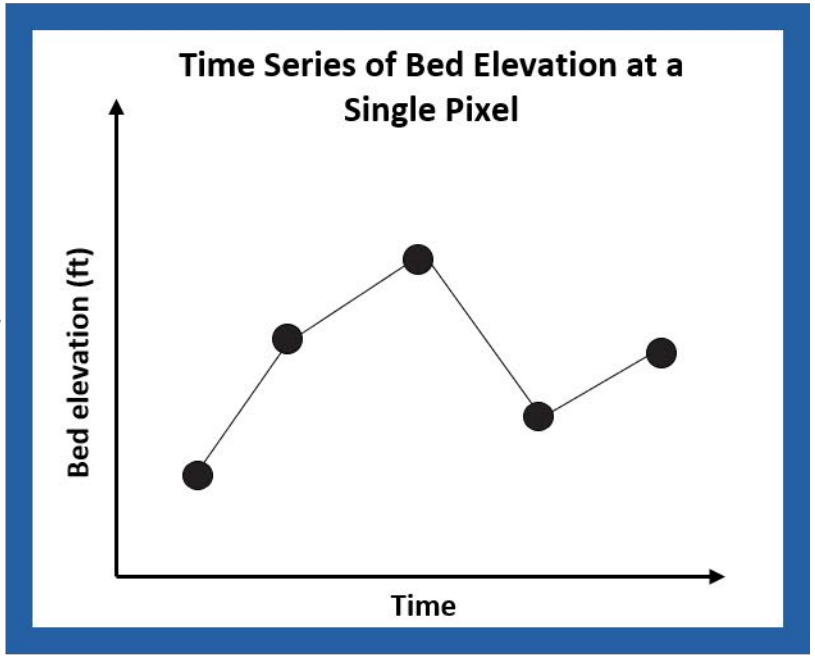
CSAT is a geospatial analysis application which quantifies shoaling rates using a space-time cube of measured bed elevations from the eHydro database.

Slice horizontally → map of bathymetry at a single timestamp.

Slice vertically → time series of bed elevation at a discrete point in space.



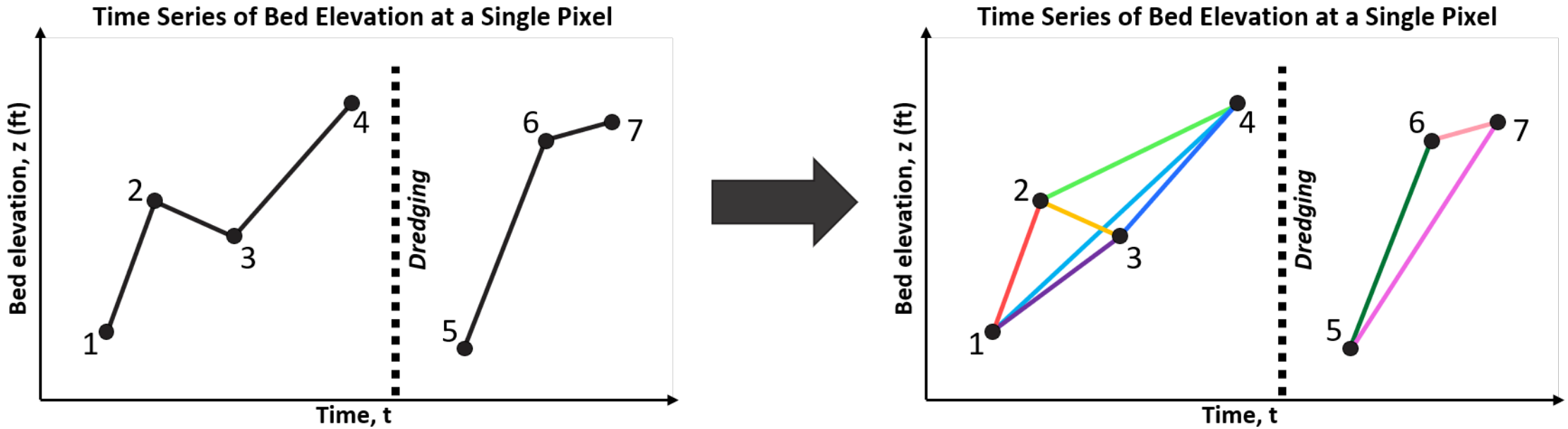




BACKGROUND

— pairing every survey (ideally)

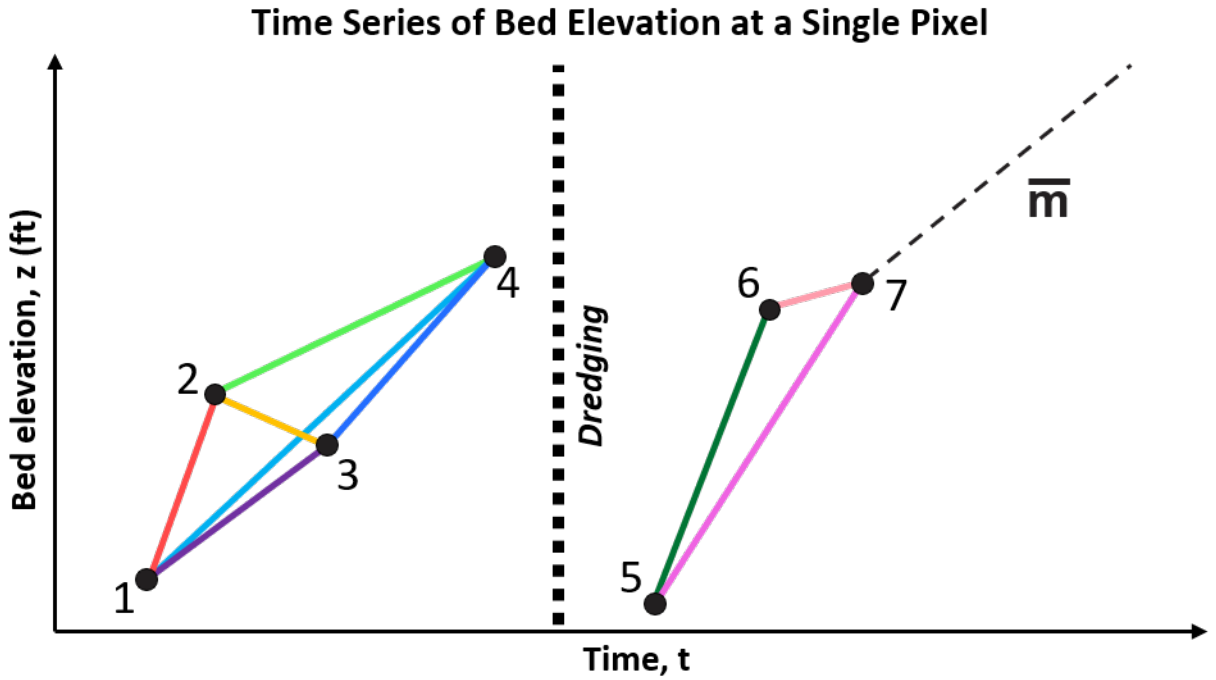
Within the CSAT algorithm, every pair of observations which are not separated by dredging is treated as a separate observation of the shoaling rate.



BACKGROUND

— averaging paired observations into a shoaling rate

To forecast future bed elevations, all paired observations are averaged to determine a single representative shoaling rate $\bar{m}$, which is then projected forward in time beginning from the last survey.



To survey...

From survey...	2	3	4	5	6	7
1	m_{12}	m_{13}	m_{14}			
2		m_{23}	m_{24}			
3			m_{34}			
4						
5					m_{56}	m_{57}
6						m_{67}

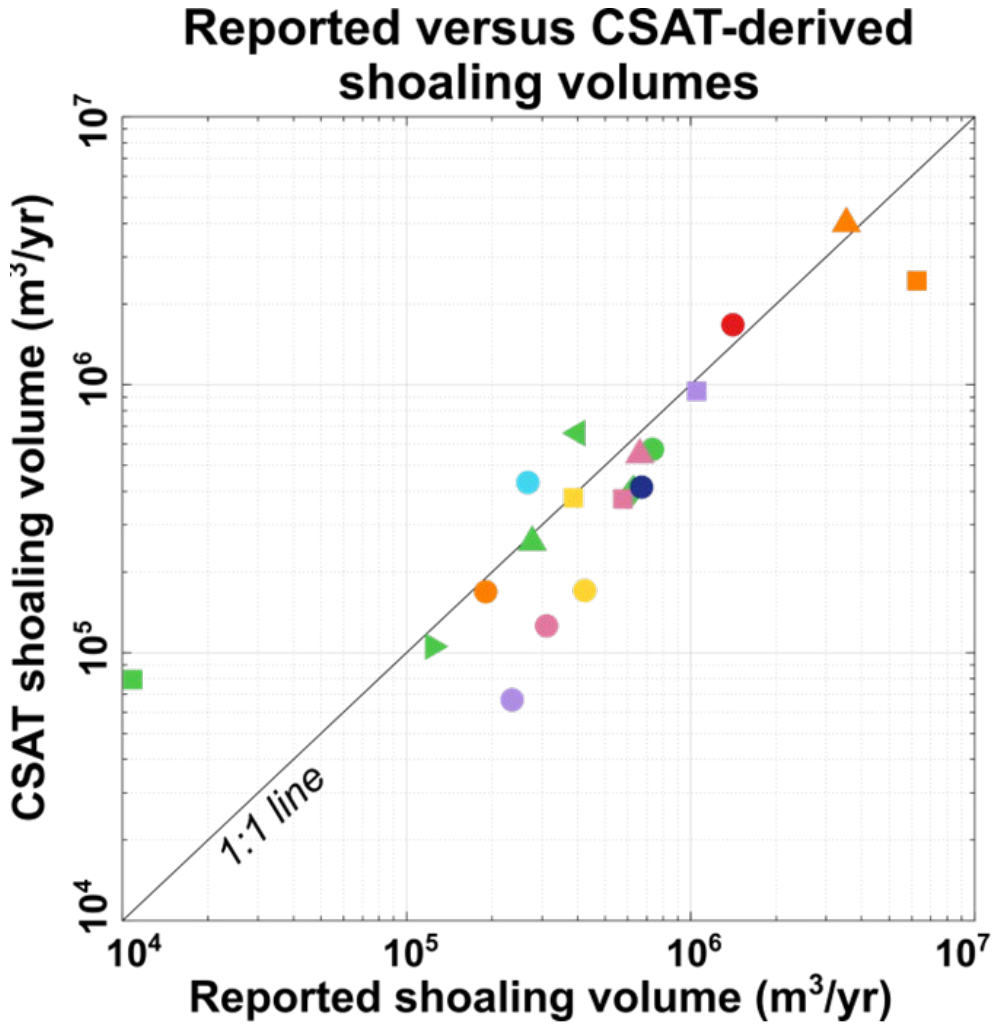
$$m_{ij} = \frac{\Delta z_{ij}}{\Delta t_{ij}} = \frac{z_j - z_i}{t_j - t_i}$$

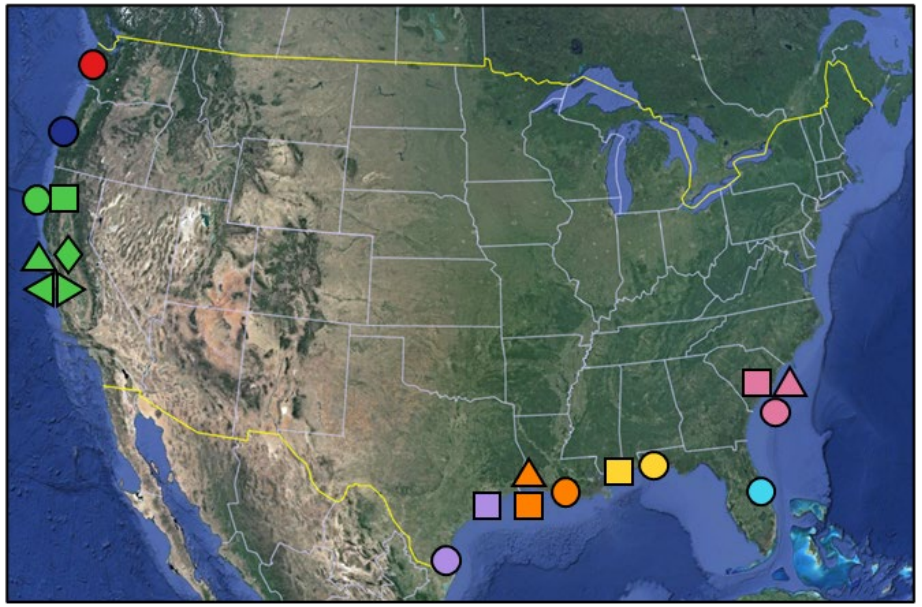
$\bar{m}$ = average*(m_{12}, m_{13}, m_{14}, m_{23}, m_{24}, m_{34}, m_{56}, m_{57}, m_{67})

**using one of several averaging methods*

HOW ACCURATE IS CSAT?

— validation results across 16 sites



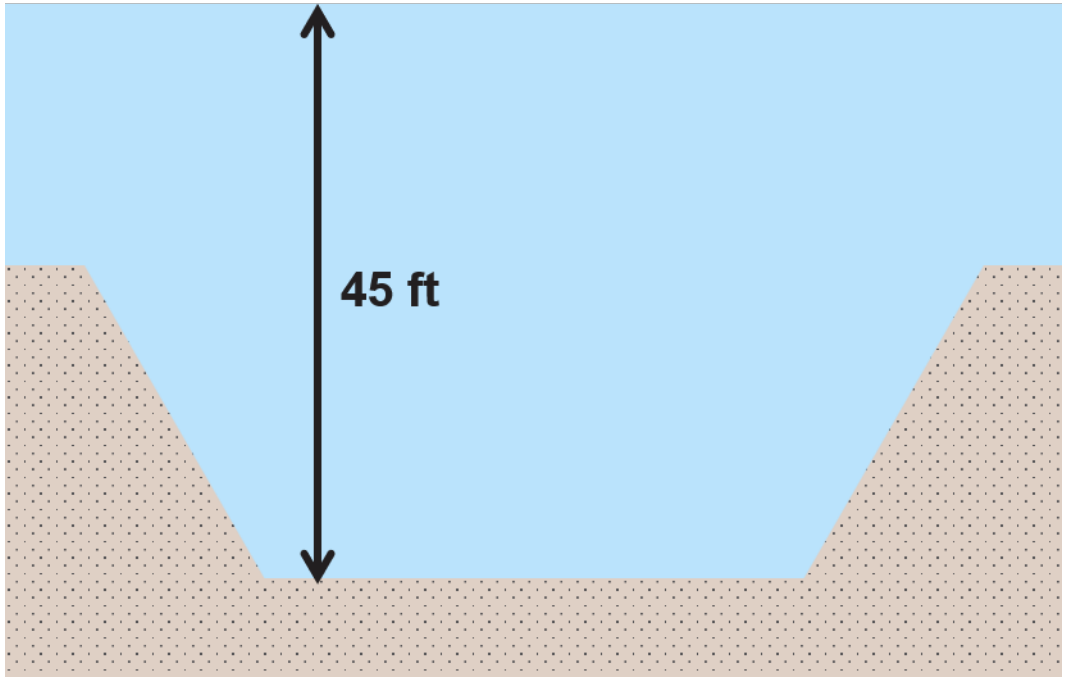


- Grays Harbor (entire project)
- Coos Bay (entire project)
- San Francisco entrance
- Oakland Harbor
- Pinole Shoal channel
- Richmond (CA) Harbor
- Humboldt bar and entrance
- Humboldt interior
- Brazos Island entrance
- Freeport entrance
- Houma bar channel
- Calcasieu bar channel
- Calcasieu interior
- Gulfport bar channel
- Pensacola entrance
- Port Canaveral (entire project)
- Charleston entrance
- Charleston Lower Harbor
- Charleston Upper Harbor

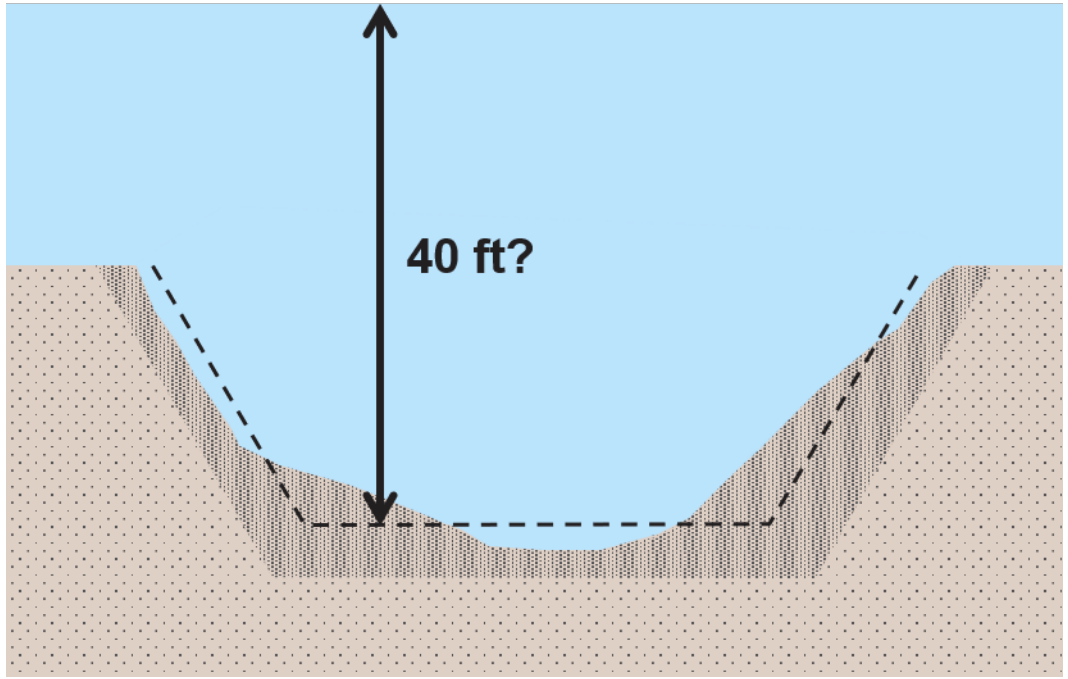
HOW ACCURATE IS CSAT?

— what a forecast promise means

If I measure the channel to be 45 ft deep in January…



… and CSAT predicts that it will be 40 ft deep next June…



… will the channel actually be 40 ft deep next June?

What level of trust should I place in this prediction?

HOW ACCURATE IS CSAT?

— open questions about real-world use

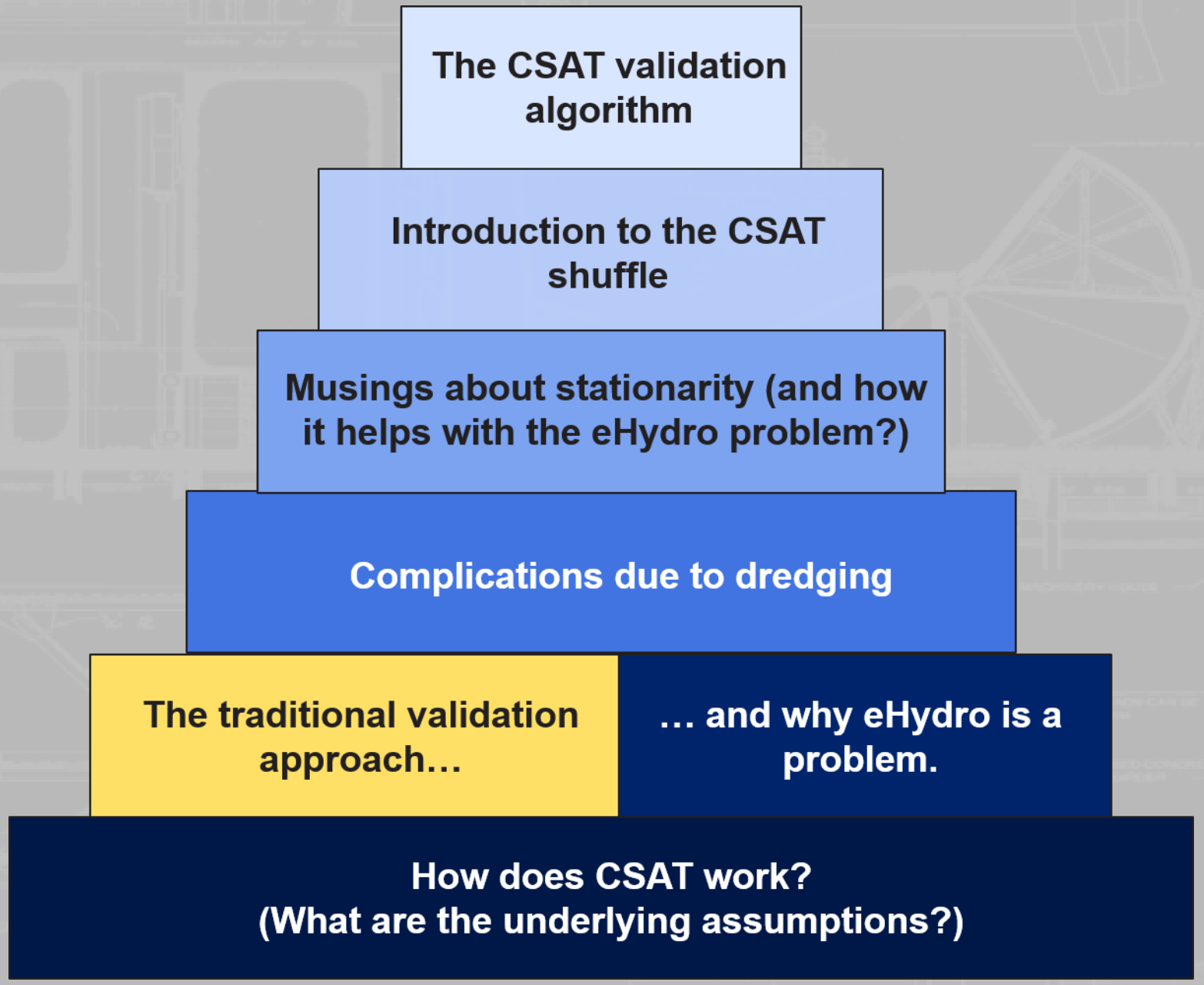
What if I upload extra surveys to eHydro?

What if I use a different averaging method?

What if I want to predict 2 years into the future instead of 1?

What if I cherry-pick which surveys I plug into CSAT?

THE STARTING POINT



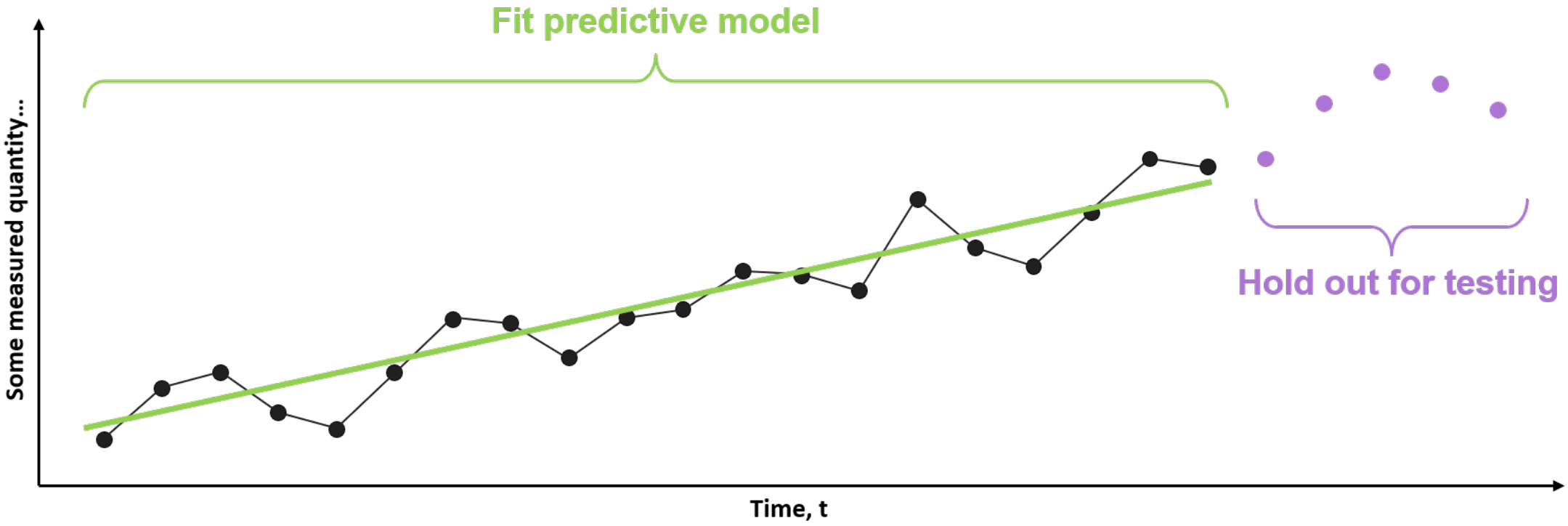
U.S. ARMY

US Army Corps of Engineers®

ERDC
ENGINEER RESEARCH & DEVELOPMENT CENTER

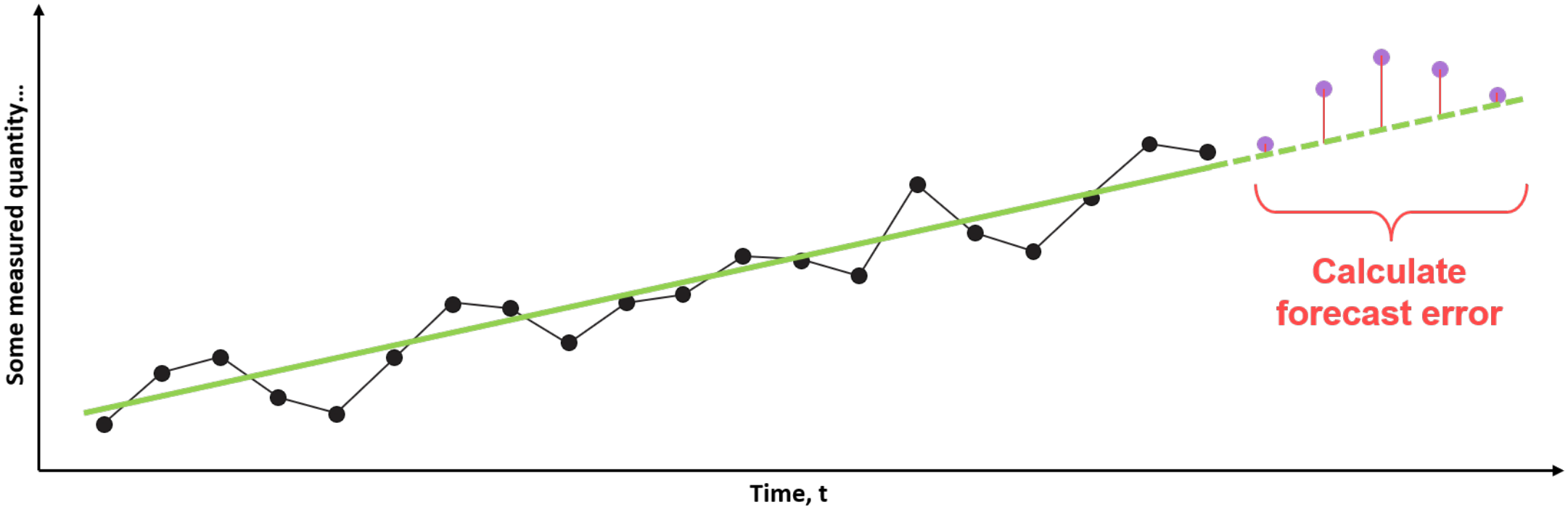
TIME SERIES CROSS-VALIDATION

(the traditional approach) — fit model, hold out for testing



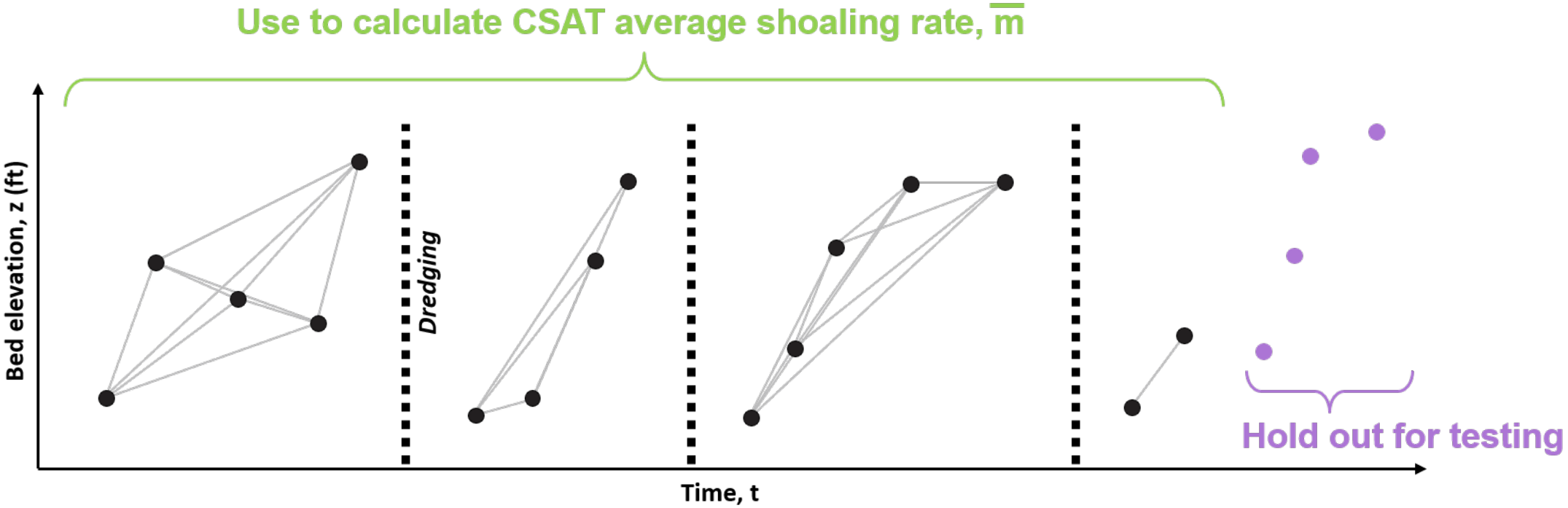
TIME SERIES CROSS-VALIDATION

(the traditional approach) — calculate the forecast error



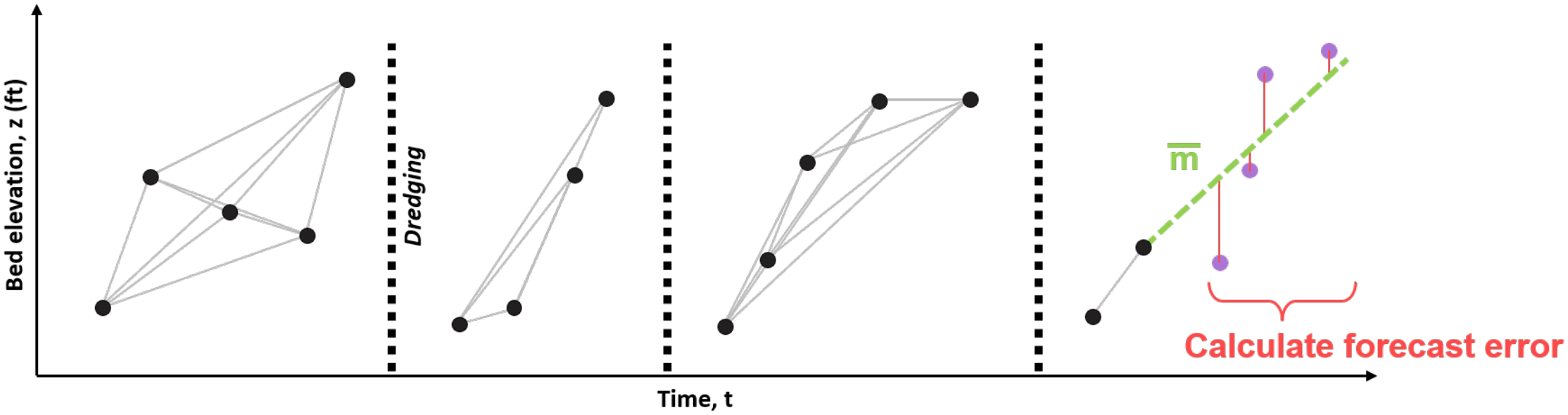
WHAT WOULD THAT LOOK LIKE FOR CSAT?

(ideally!) — hold out surveys for testing

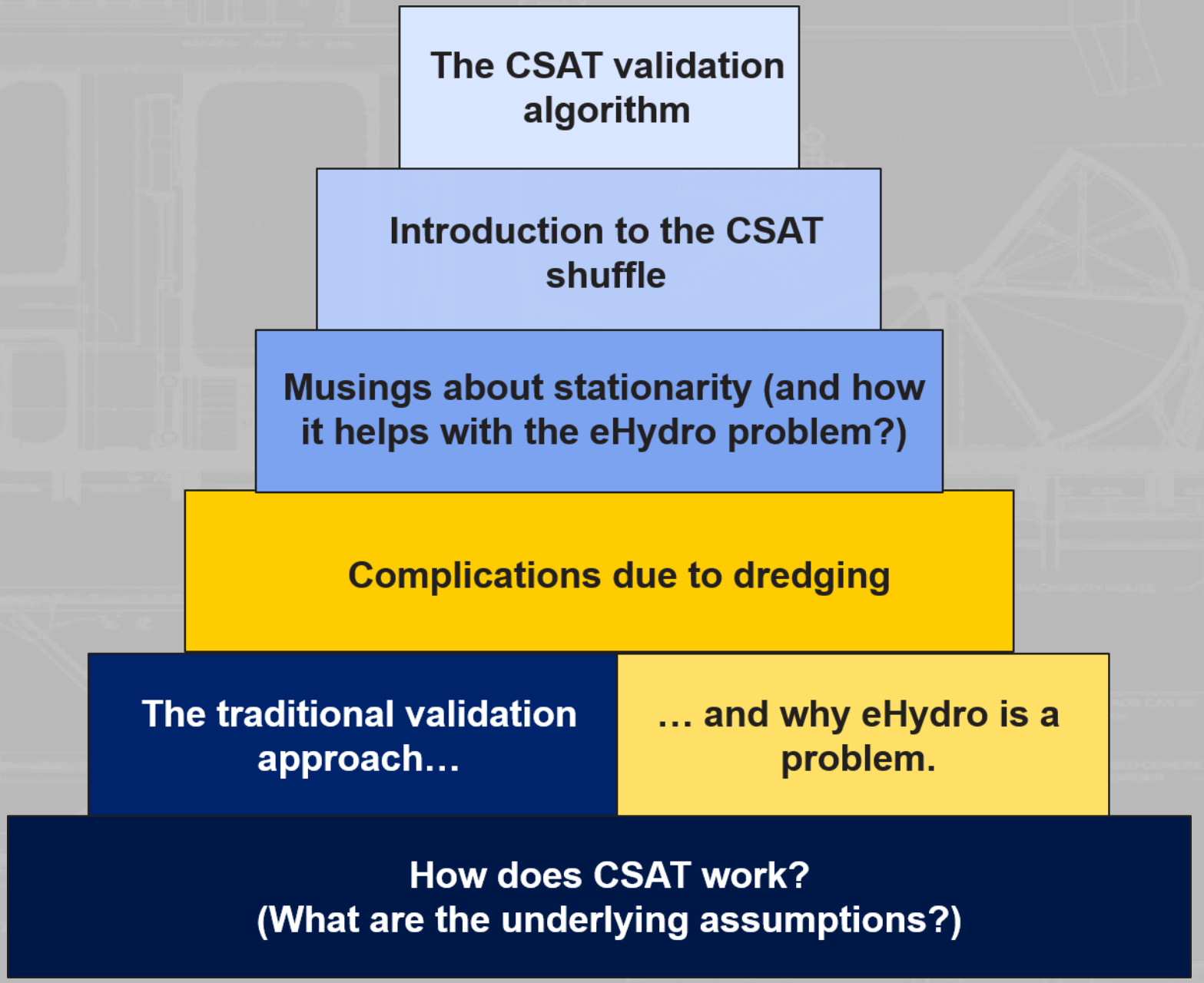


WHAT WOULD THAT LOOK LIKE FOR CSAT?

(ideally!) — calculate the forecast error



WHAT GOES WRONG?



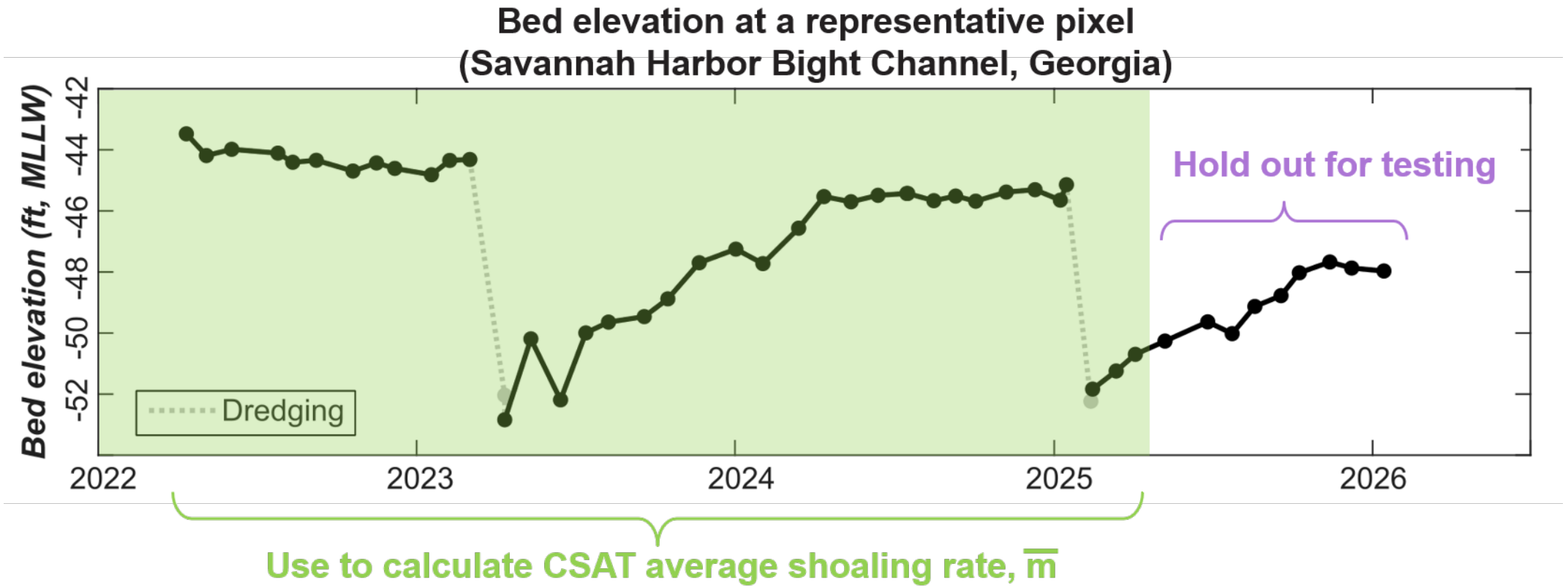
U.S. ARMY

US Army Corps of Engineers®

ERDC
ENGINEER RESEARCH & DEVELOPMENT CENTER

REAL-WORLD IMPLEMENTATION

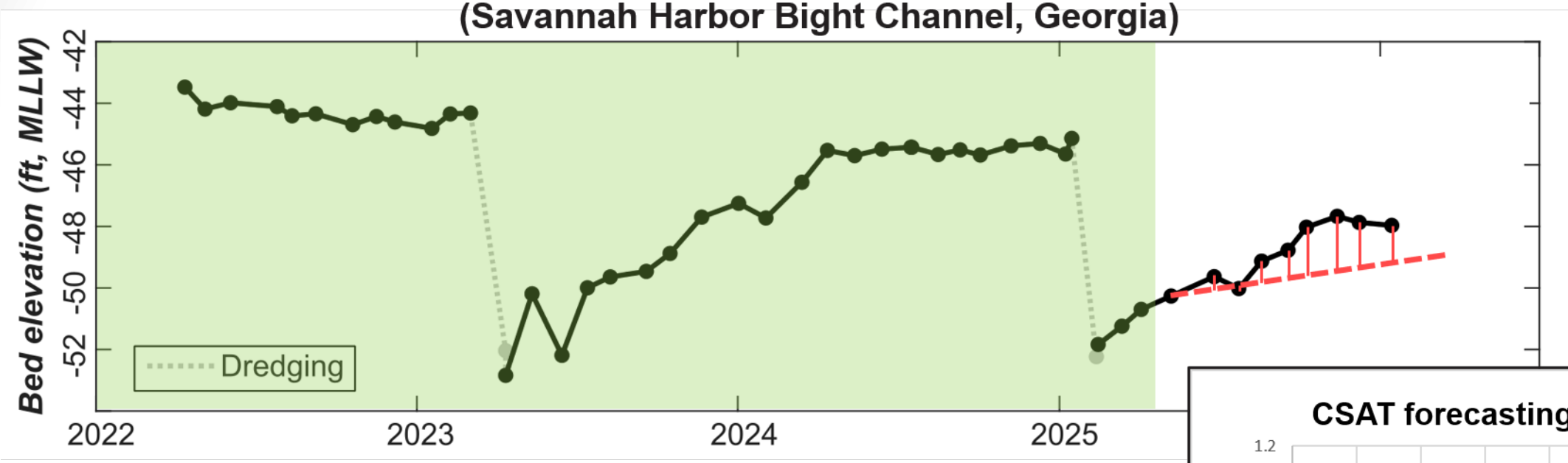
— holding out data for testing

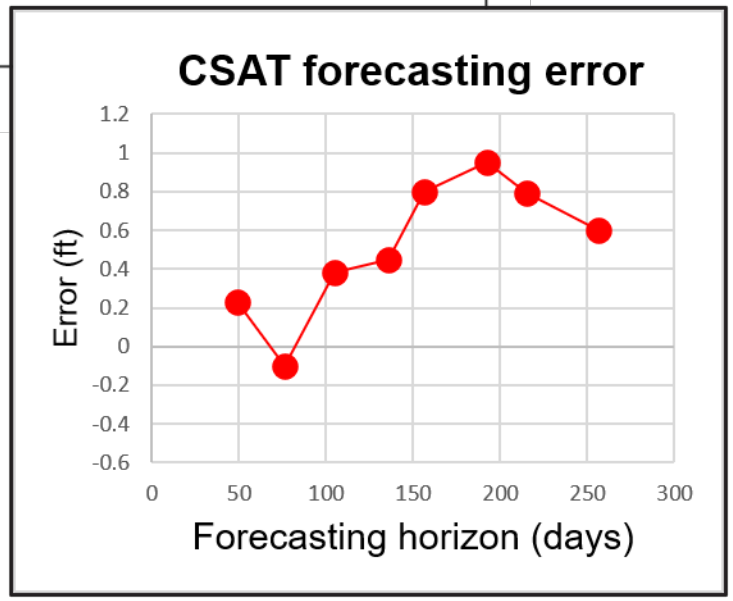


REAL-WORLD IMPLEMENTATION

— forecast from one held-out survey

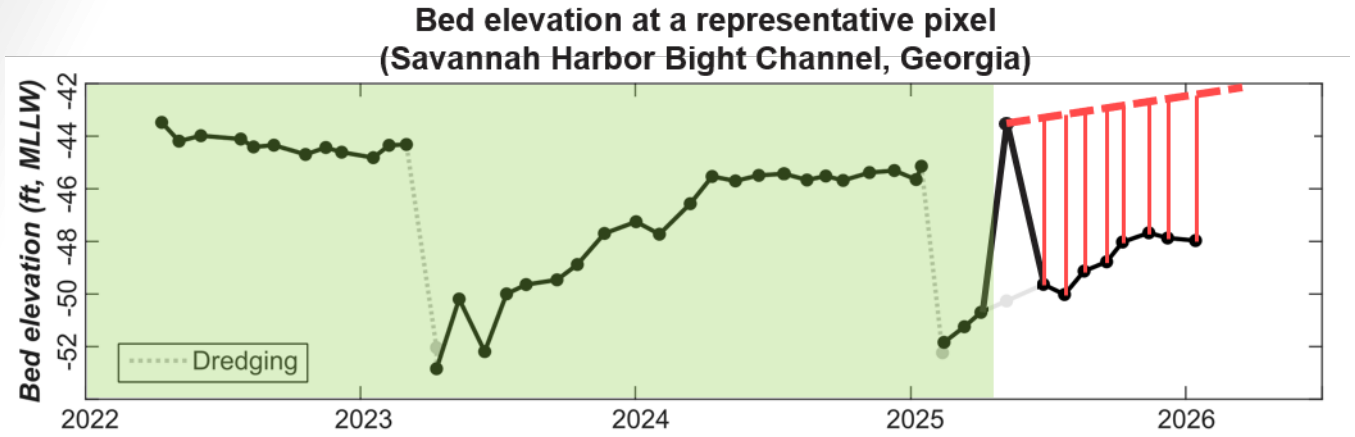
Bed elevation at a representative pixel
(Savannah Harbor Bight Channel, Georgia)

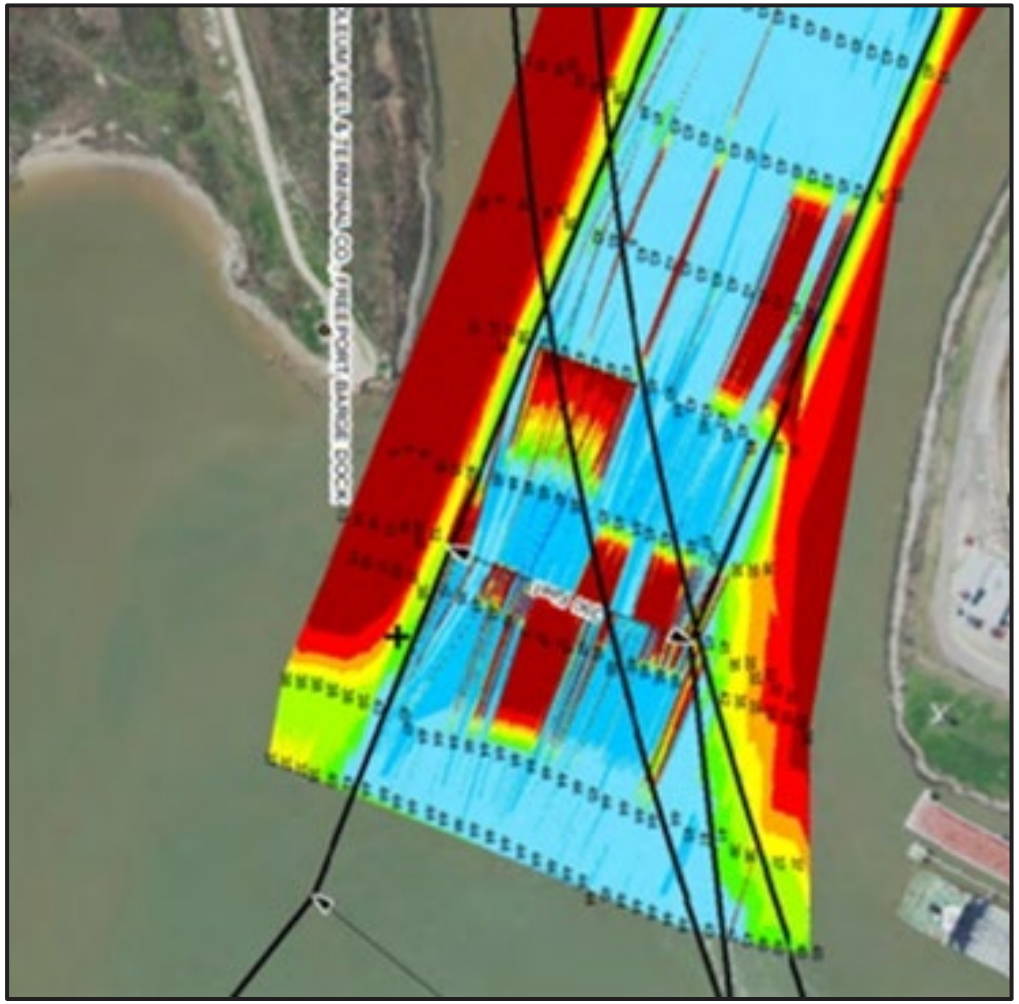




REAL-WORLD IMPLEMENTATION

— when a held-out survey is unusual



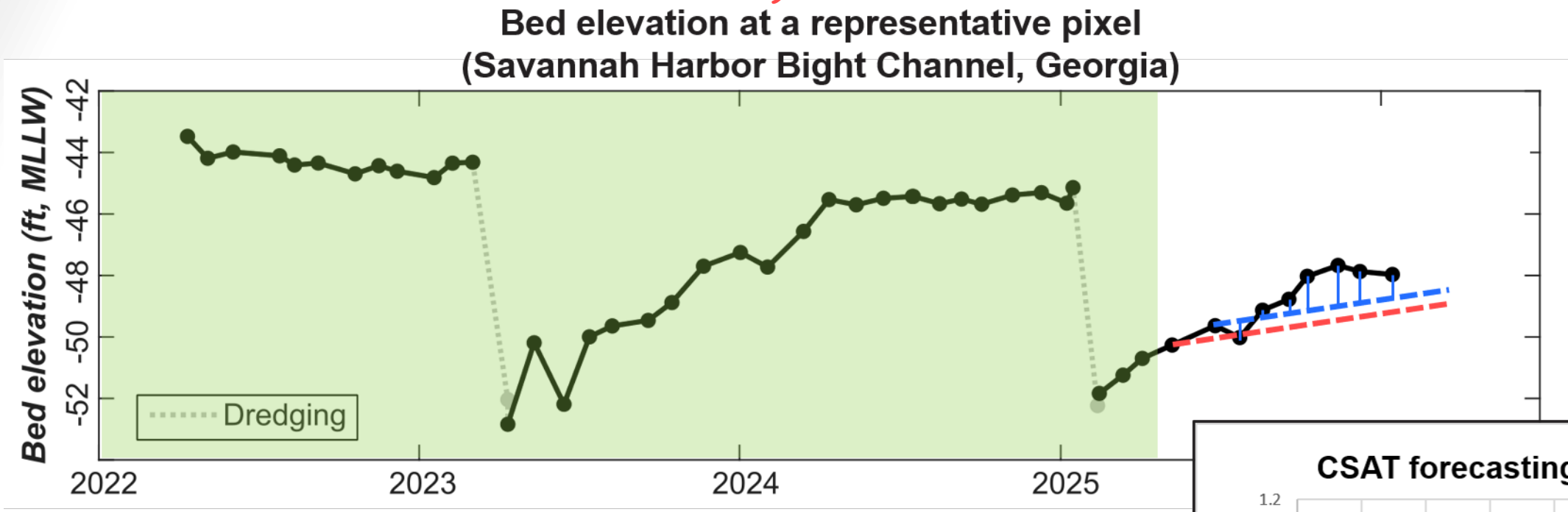


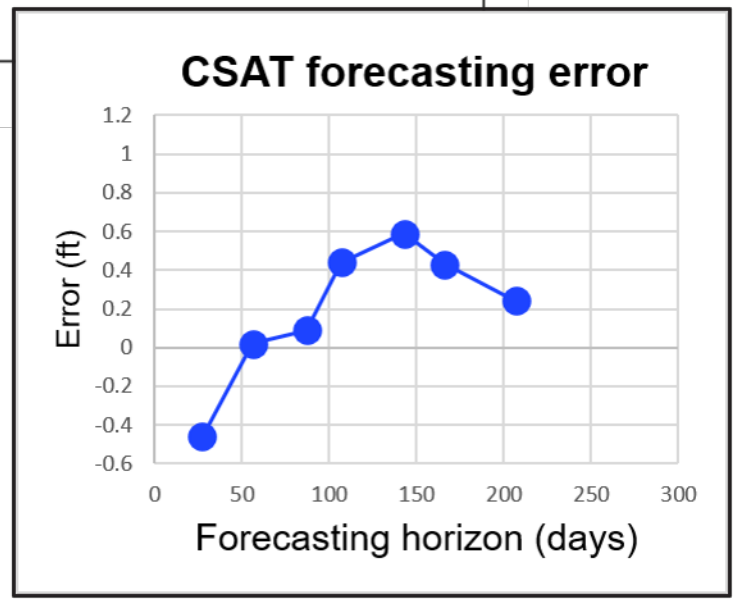
What if one of the "held out" surveys happens to be bad?

How can we prevent one bad survey from having an outsize influence on the validation results?

REAL-WORLD IMPLEMENTATION

— recalculating at every held-out survey

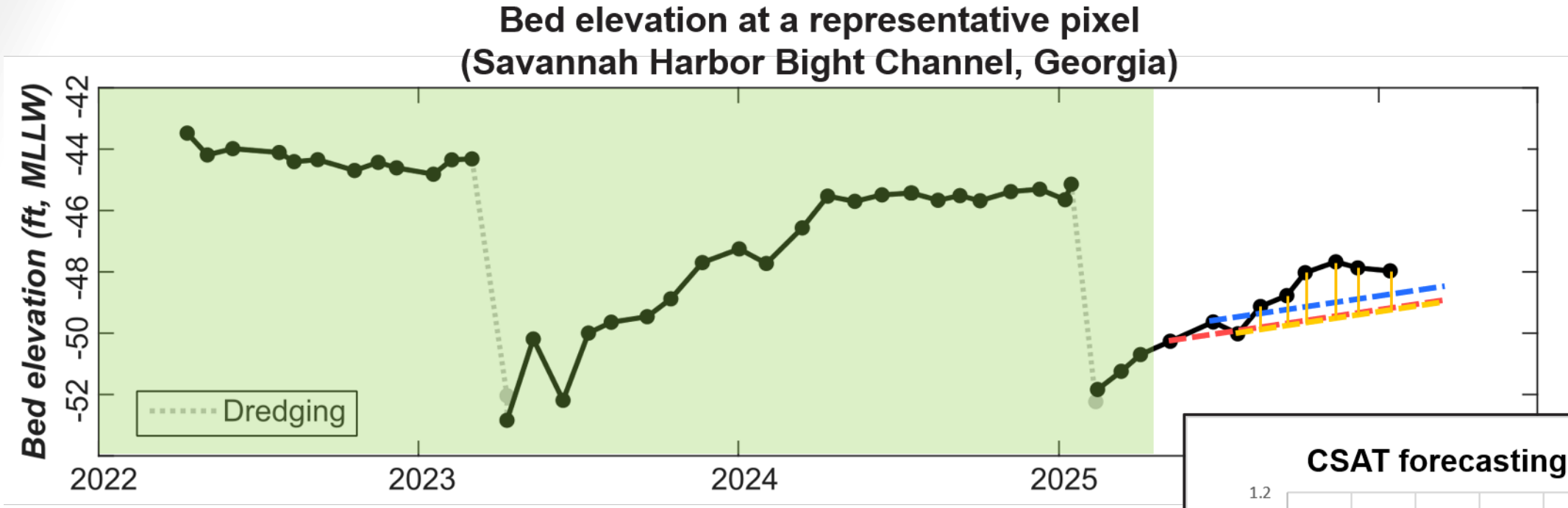


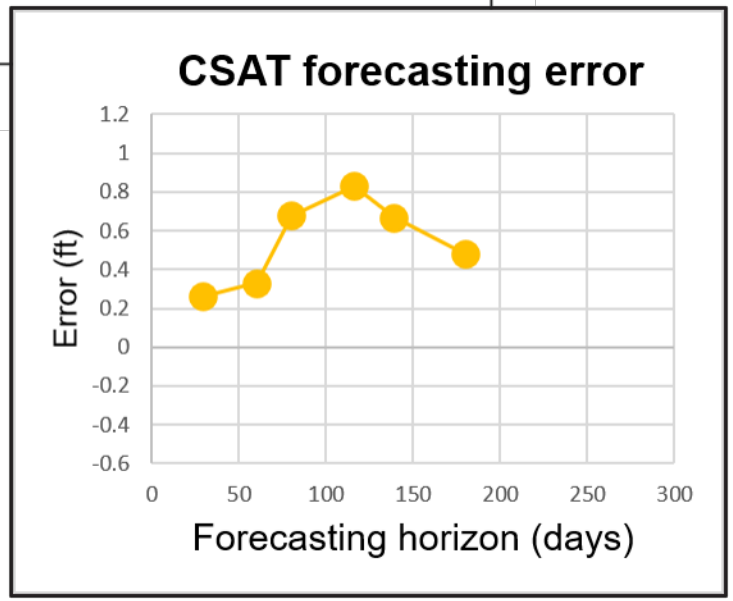


Solution: begin the forecast at EVERY "held out" survey and systematically recalculate the forecasting error.

REAL-WORLD IMPLEMENTATION

— a second recalculated forecast

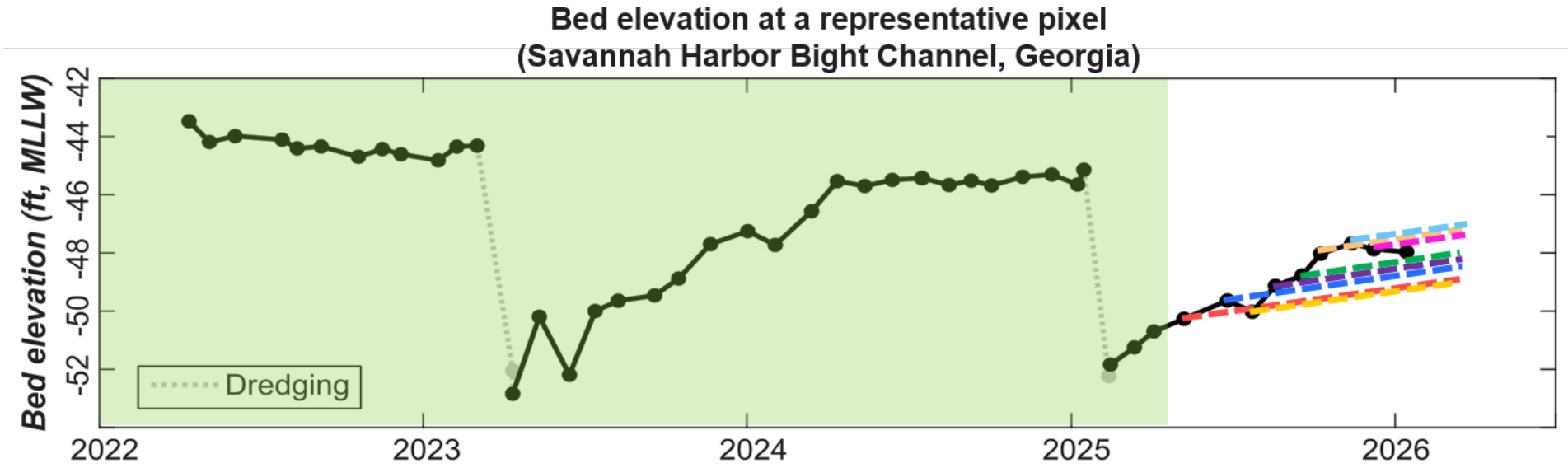


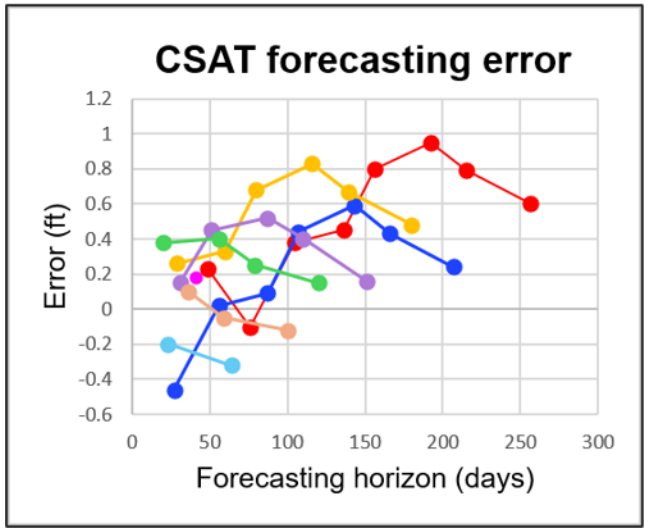


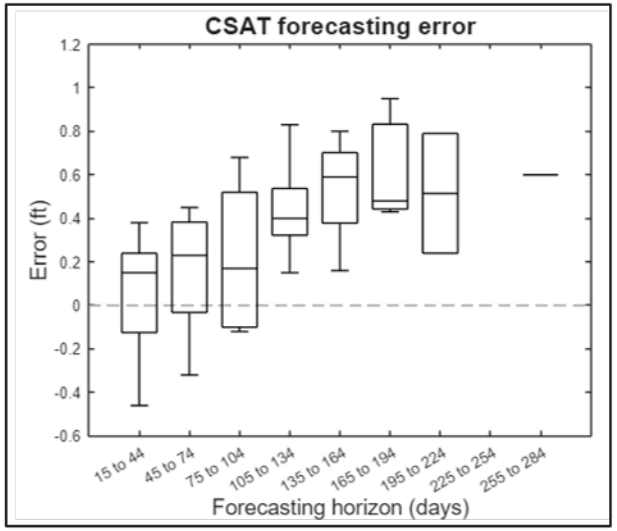
Solution: begin the forecast at EVERY "held out" survey and systematically recalculate the forecasting error.

REAL-WORLD IMPLEMENTATION

— building a distribution of forecast errors



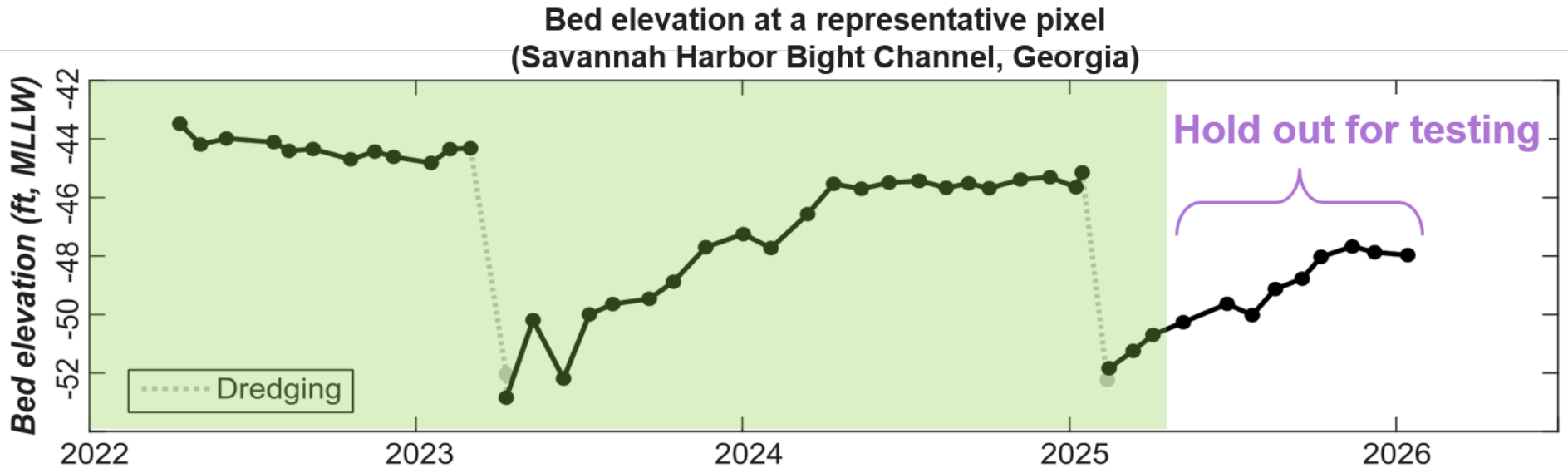




This gives us a <u>distribution</u> of possible error outcomes for a given set of forecasting parameters.

WHAT ABOUT LONGER FORECASTING HORIZONS?

— the 8.5-month test limit

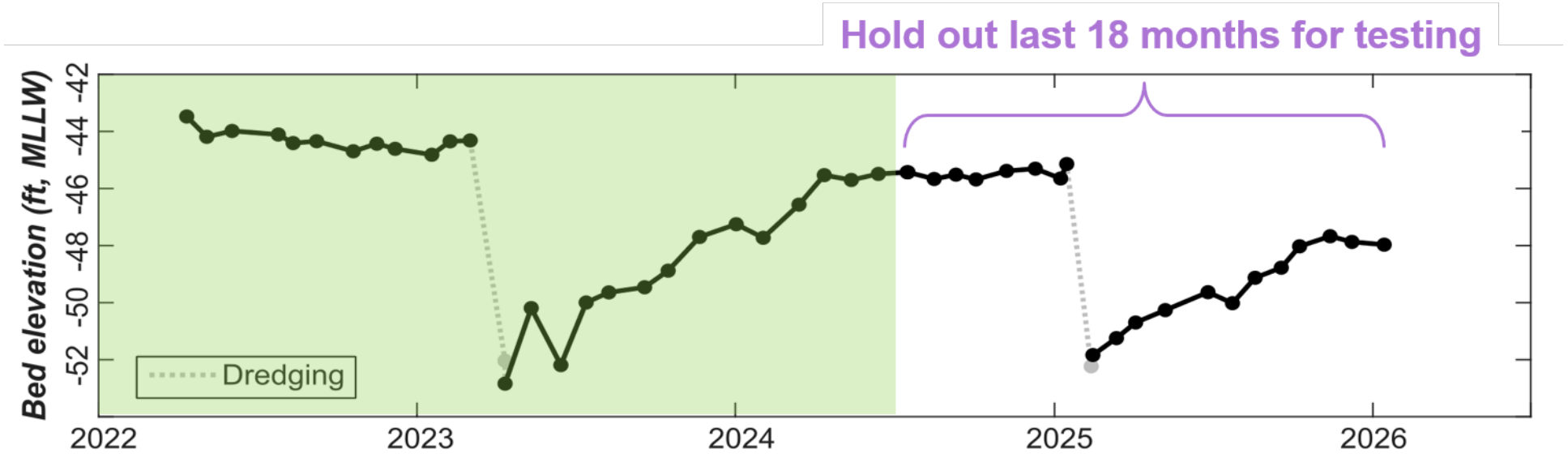


- **Holding out 20% of the data for testing allows us to evaluate CSAT's performance a maximum of 8.5 months into the future.**

  (For this example—actual horizon will vary based on site-specific data configuration)

- **What if we want to know how well CSAT performs (*e.g.*) 18 months into the future?**

WHAT ABOUT LONGER FORECASTING HORIZONS?

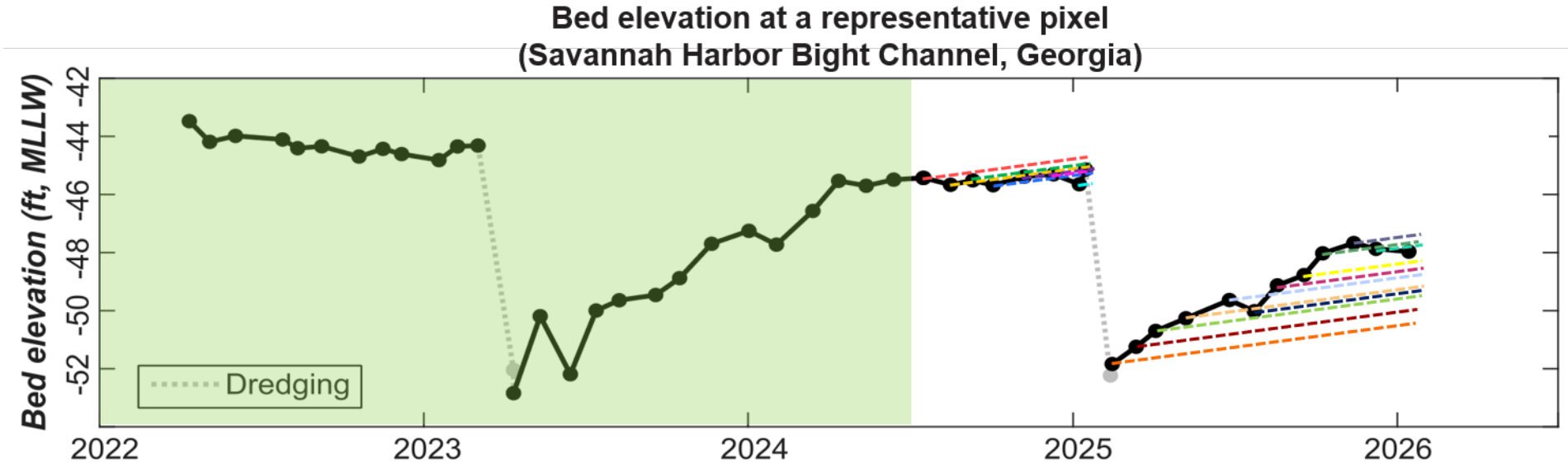


- **Let's try holding out 18 months of data at the end of the time series. Does this give us the information we need?**

(Remember, CSAT treats dredging events as hard discontinuities!)

WHAT ABOUT LONGER FORECASTING HORIZONS?

— an 11-month test result



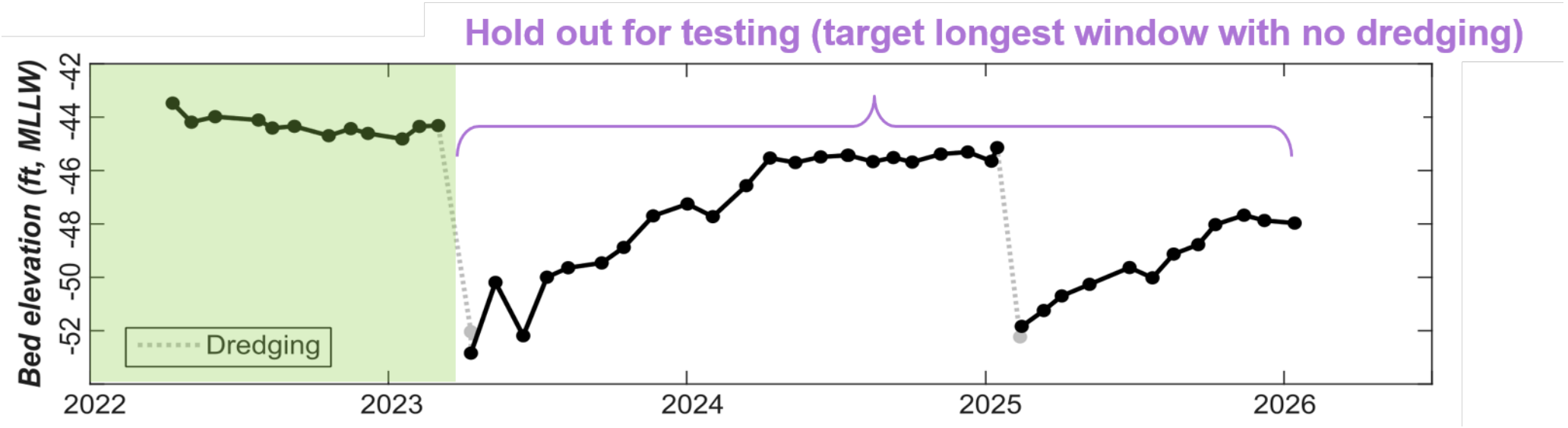
- Let's try holding out 18 months of data at the end of the time series. Does this give us the information we need?

(Remember, CSAT treats dredging events as hard discontinuities!)

- **Longest available test horizon for this example: 11 months.**

WHAT ABOUT LONGER FORECASTING HORIZONS?

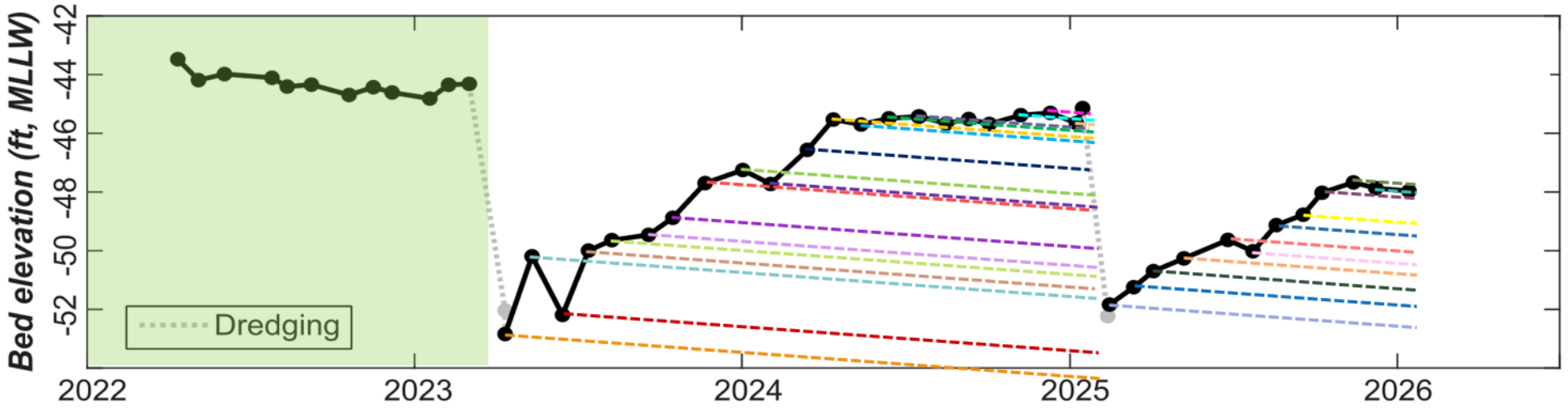
— targeting the longest dredge-free window



- What if we intentionally hold out the longest window of data with no dredging?

WHAT ABOUT LONGER FORECASTING HORIZONS?

— the tradeoff with less input data

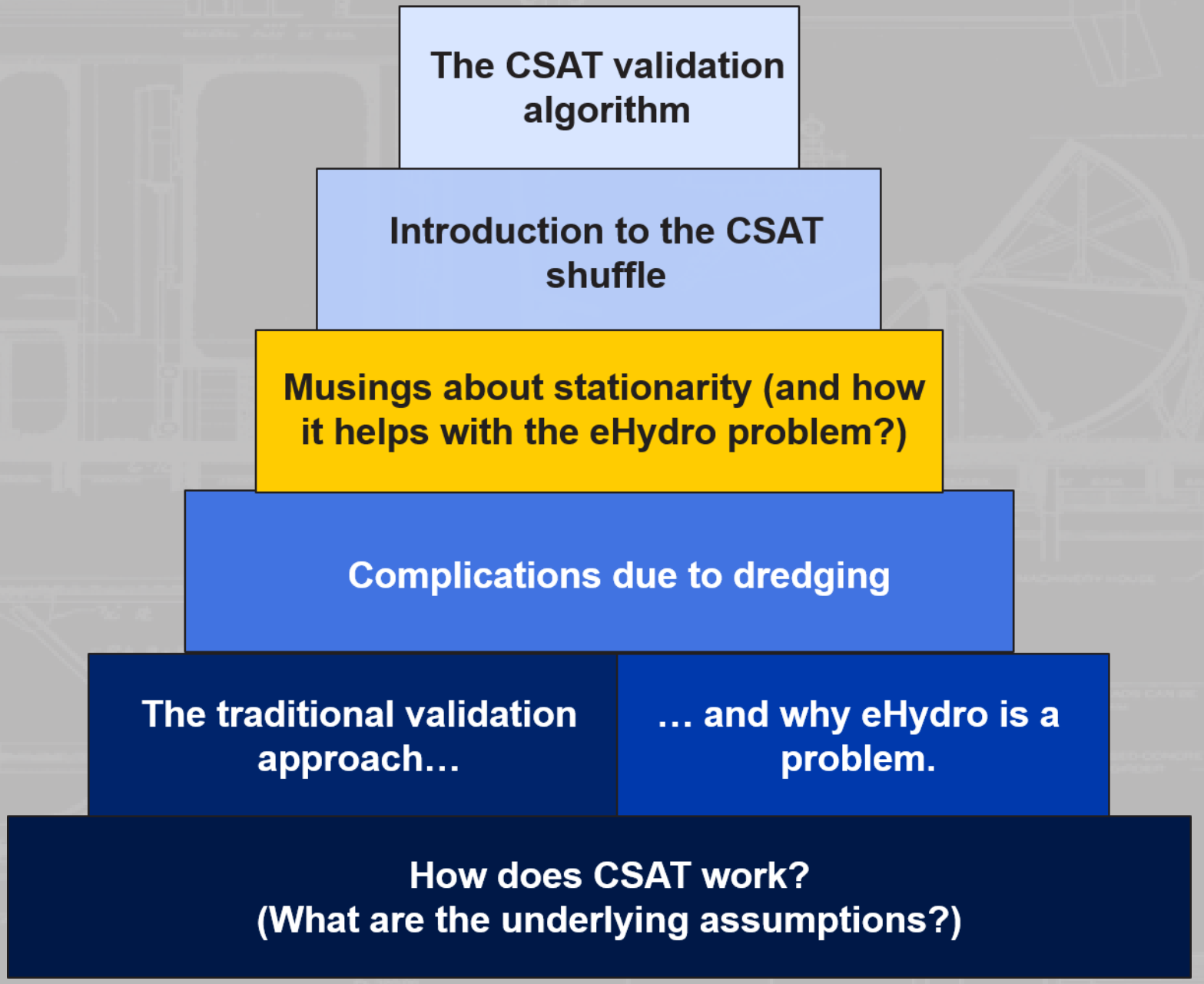


- What if we intentionally hold out the longest window of data with no dredging?

- **This severely reduces the amount of input data used to calculate the CSAT shoaling rate, $\bar{m}$ → gives the appearance of really bad forecasting performance!**

(Not representative of operational CSAT usage—what is the likelihood that a user restricts their input data to the green period, in practice?)

HOW DO THE INPUT DATA AFFECT PERFORMANCE?

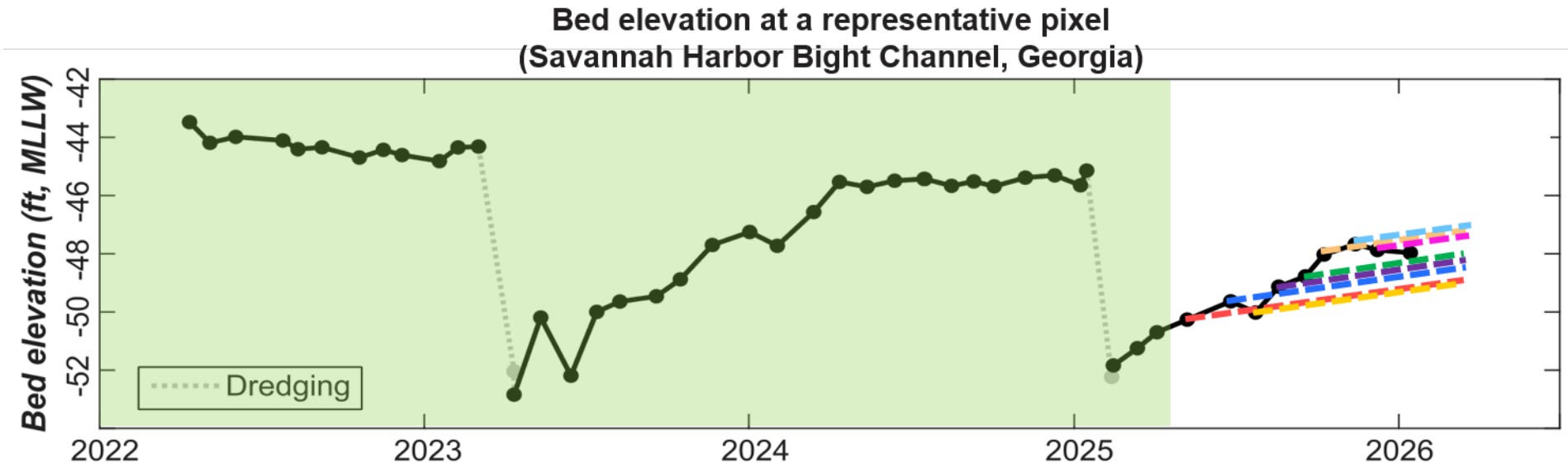


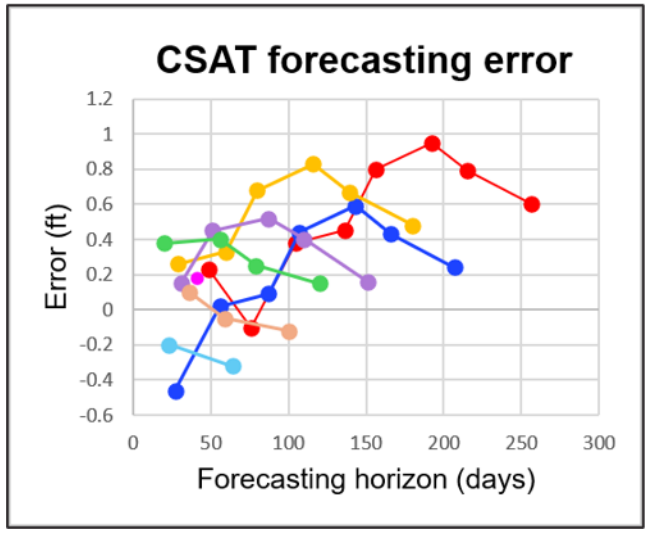
U.S. ARMY

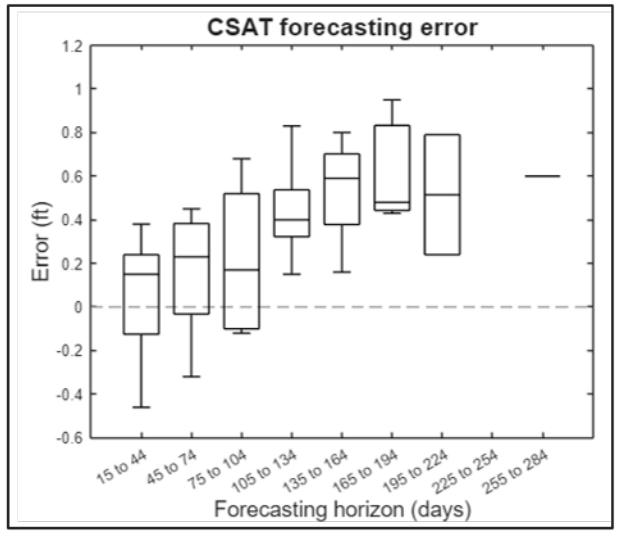
US Army Corps of Engineers®

ERDC
ENGINEER RESEARCH & DEVELOPMENT CENTER

RECALL…





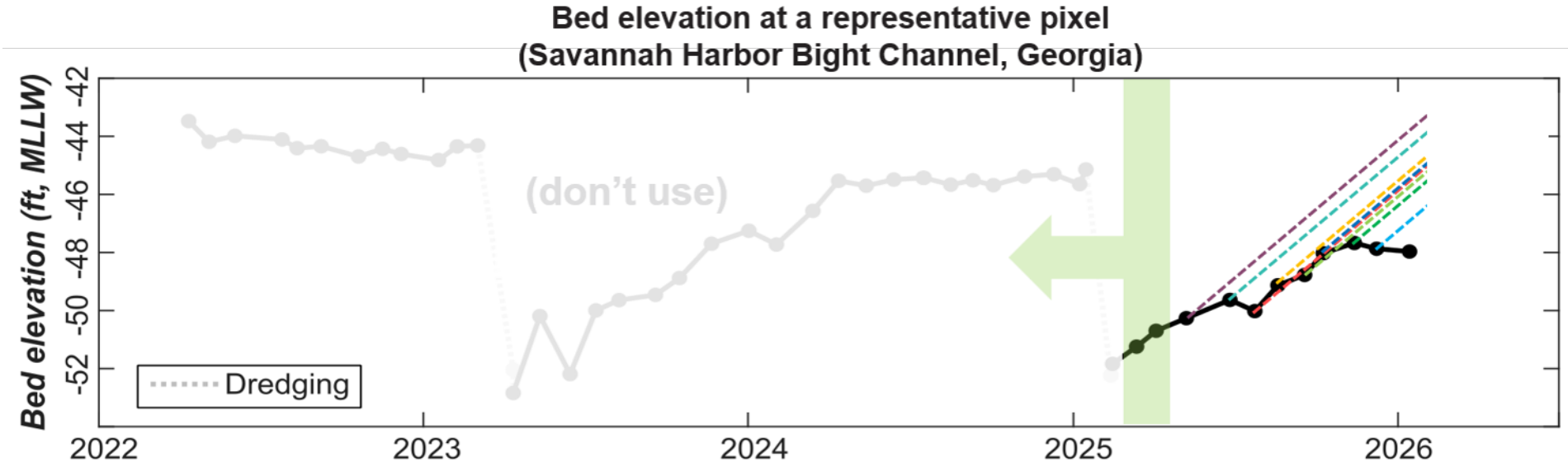


These results are for a fixed set of input data (the green box).

What if we vary the input data quantity?

THE "WALK BACKWARDS" APPROACH

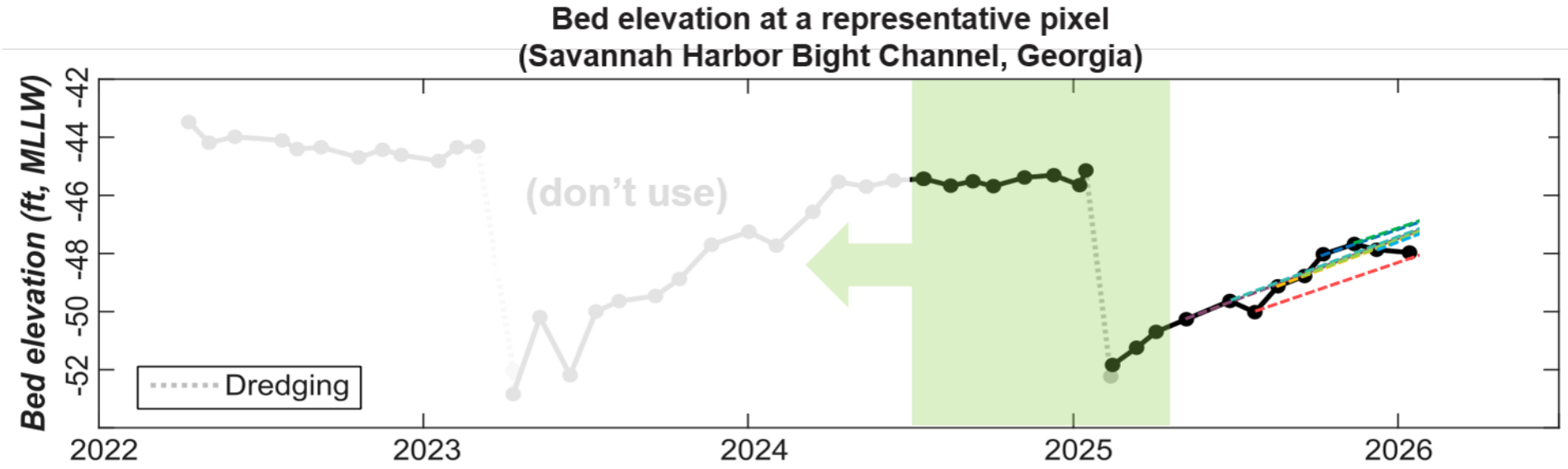
— starting the walk



Strategy: systematically increase the number of input surveys, moving backwards from the most recent.

THE "WALK BACKWARDS" APPROACH

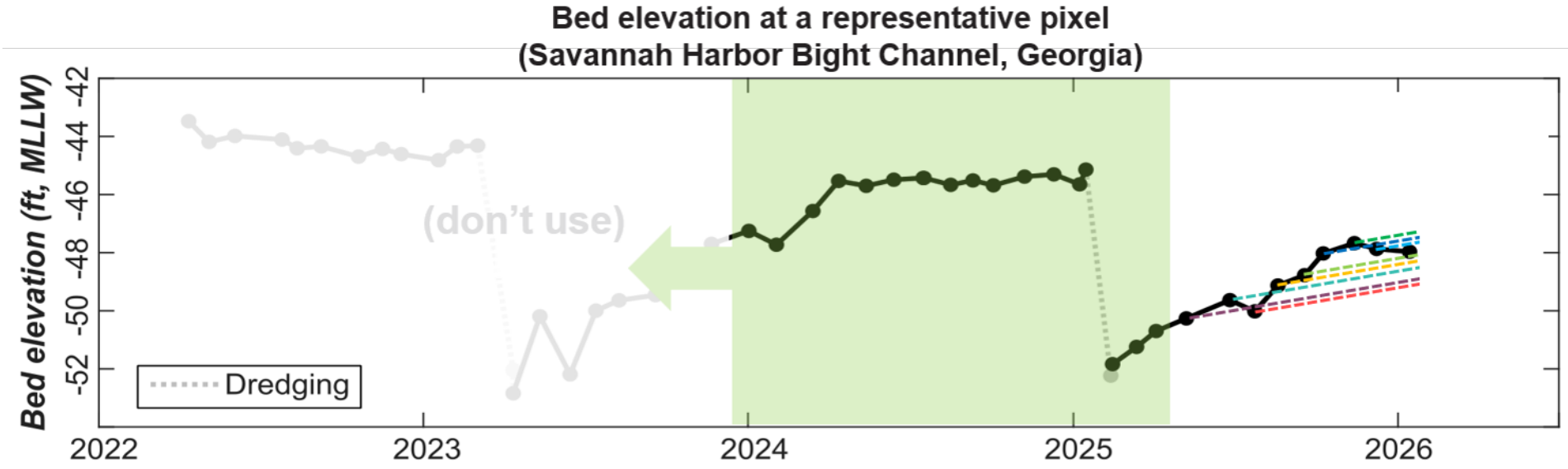
— widening the input window



Strategy: systematically increase the number of input surveys, moving backwards from the most recent.

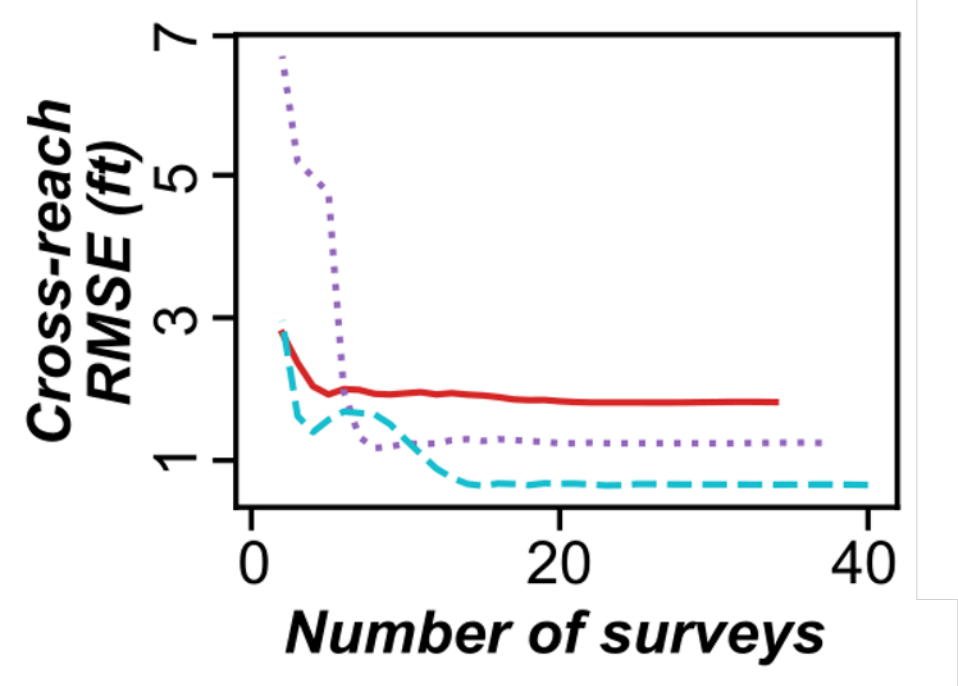
THE “WALK BACKWARDS” APPROACH

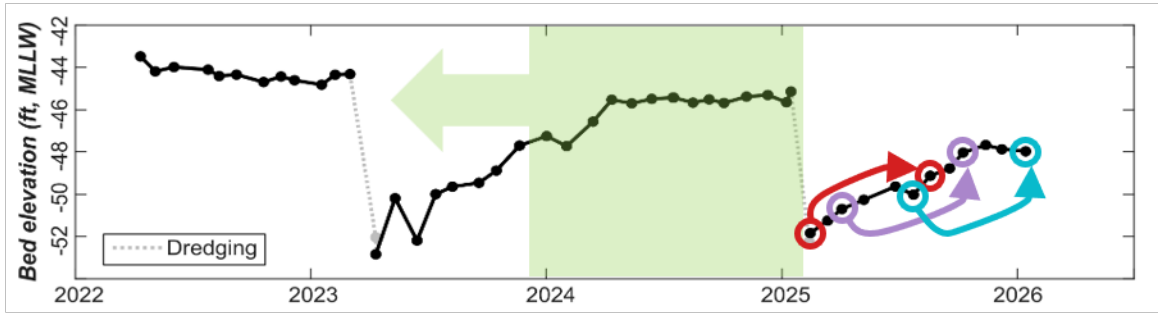
— widening it further



Strategy: systematically increase the number of input surveys, moving backwards from the most recent.

RESULTS – "WALK BACKWARDS" METHOD





IC: 1-Feb-2025, Val: 19-Aug-2025.

IC: 4-Apr-2025, Val: 9-Oct-2025.

IC: 24-Jul-2025, Val: 14-Jan-2026.

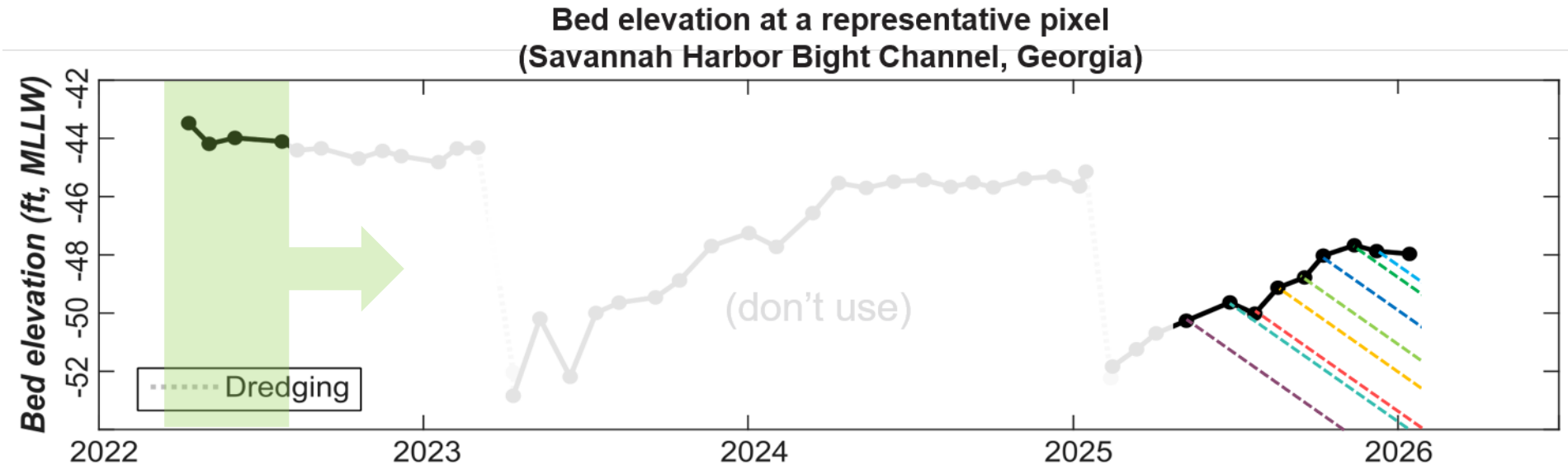
For a fixed forecast starting point…

…and a fixed validation point…

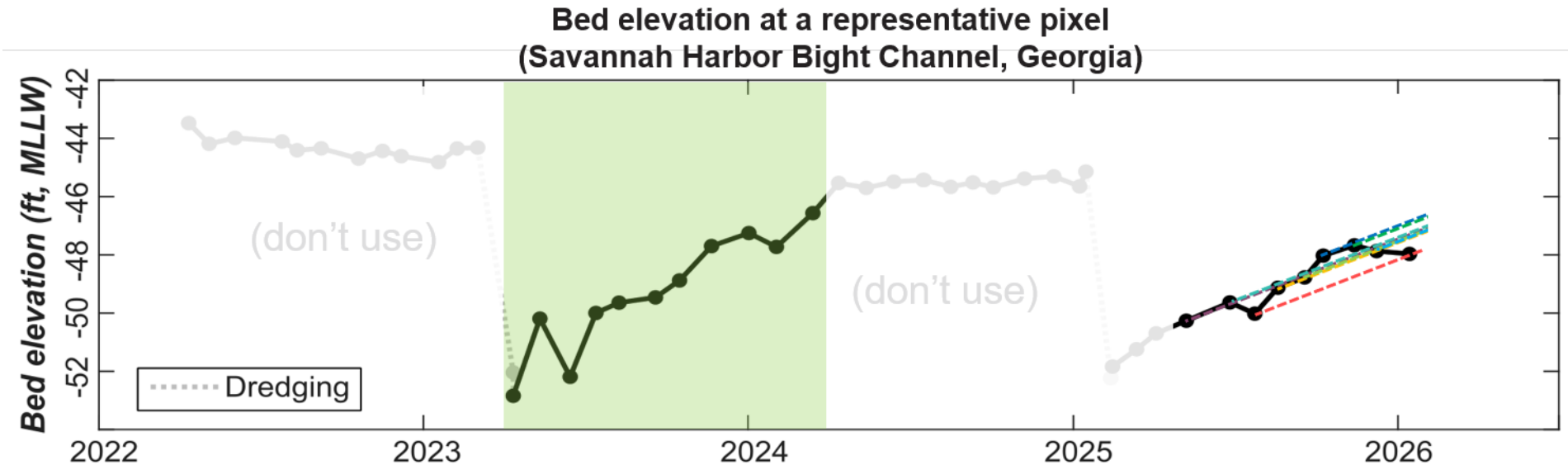
…increasing the number of CSAT input surveys eventually causes the error to "converge".

(but not necessarily to zero!)

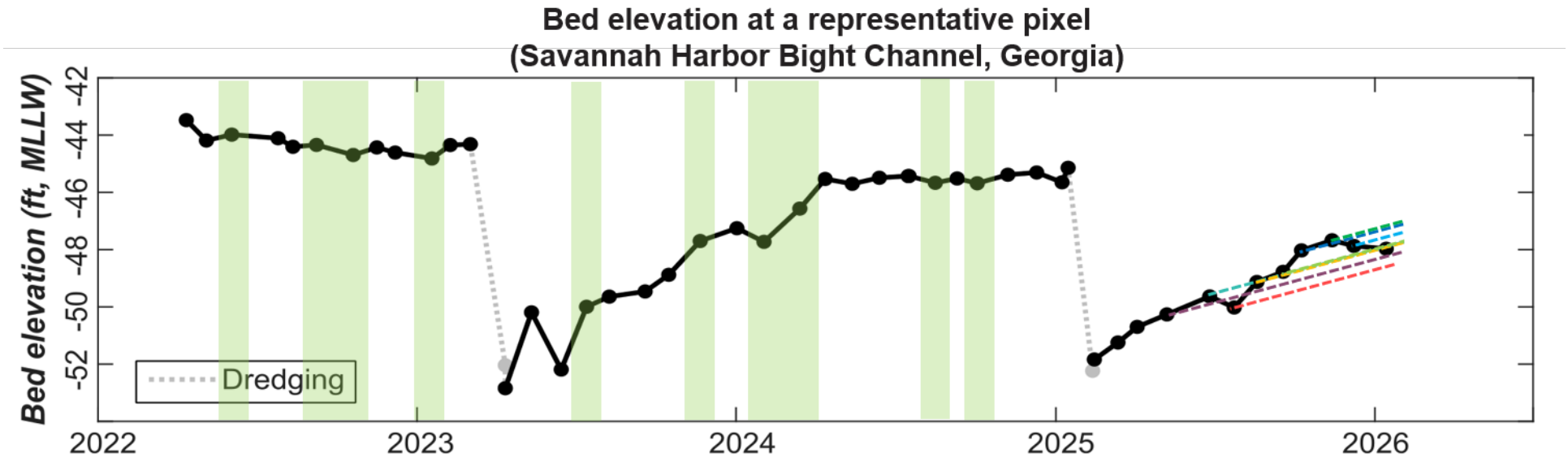
WHAT ABOUT A “WALK FORWARDS” METHOD?



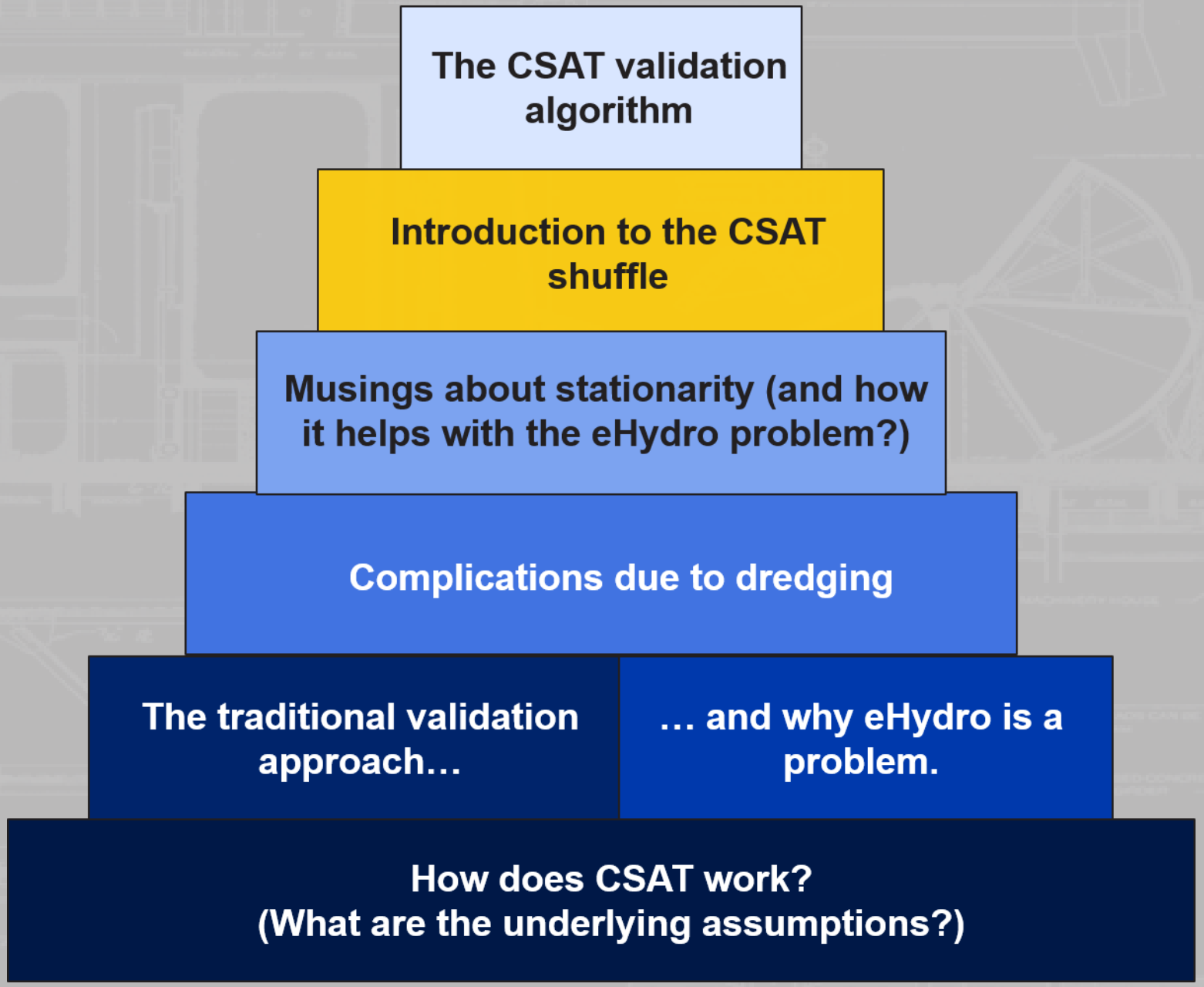
OR A SLIDING WINDOW METHOD?



WHY DON'T WE JUST PICK SURVEYS AT RANDOM?



THE CSAT SHUFFLE



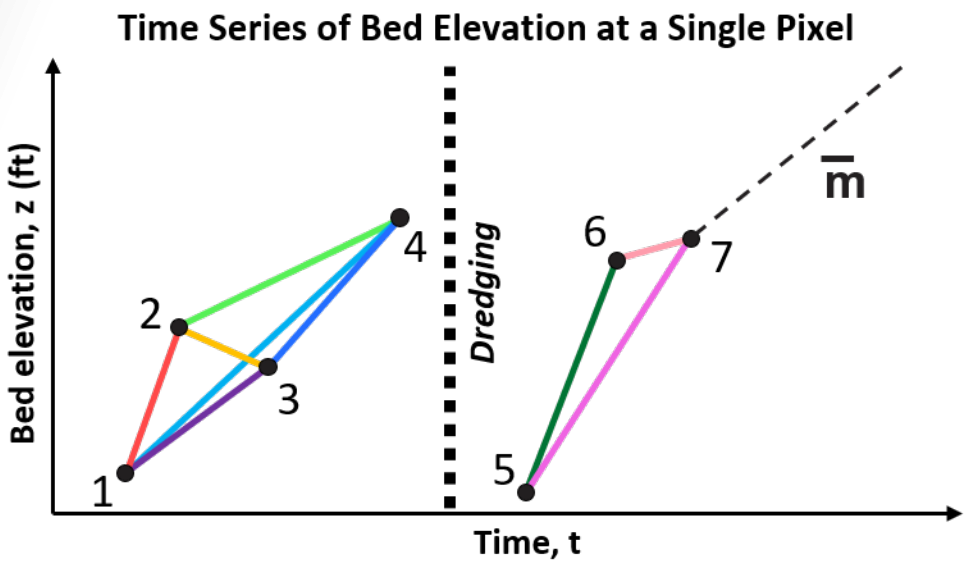
U.S. ARMY

US Army Corps of Engineers®

ERDC
ENGINEER RESEARCH & DEVELOPMENT CENTER

DOES CSAT CARE ABOUT DATA ORDERING?

— averaging treats order as irrelevant

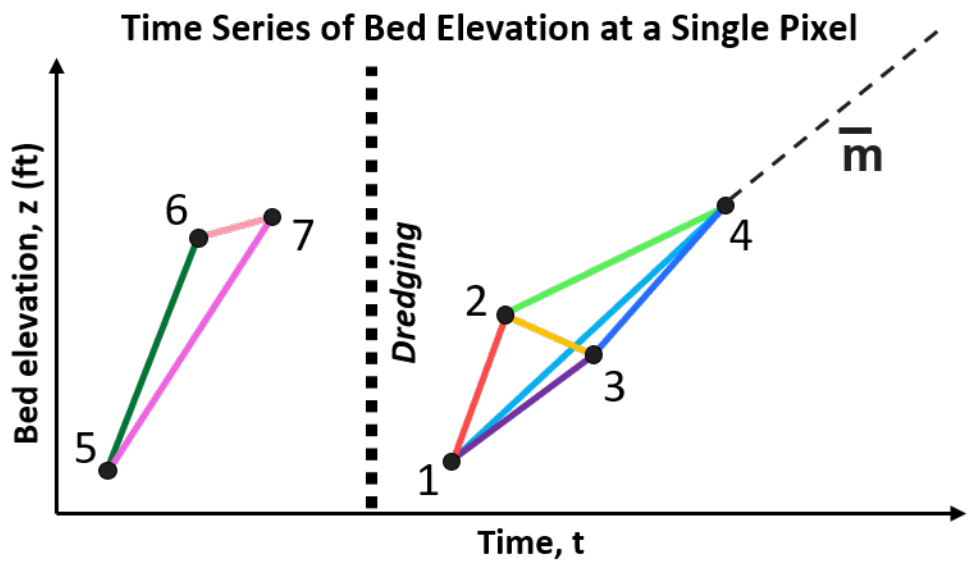


As far as CSAT is concerned,

THIS

is the same as

THIS



To survey...

From survey...	2	3	4	5	6	7
1	m_{12}	m_{13}	m_{14}			
2		m_{23}	m_{24}			
3			m_{34}			
4						
5					m_{56}	m_{57}
6						m_{67}

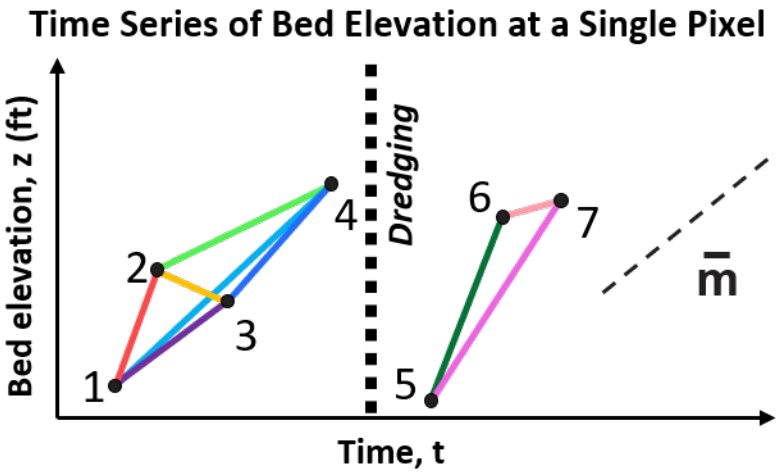
$$m_{ij} = \frac{\Delta z_{ij}}{\Delta t_{ij}} = \frac{z_j - z_i}{t_j - t_i}$$

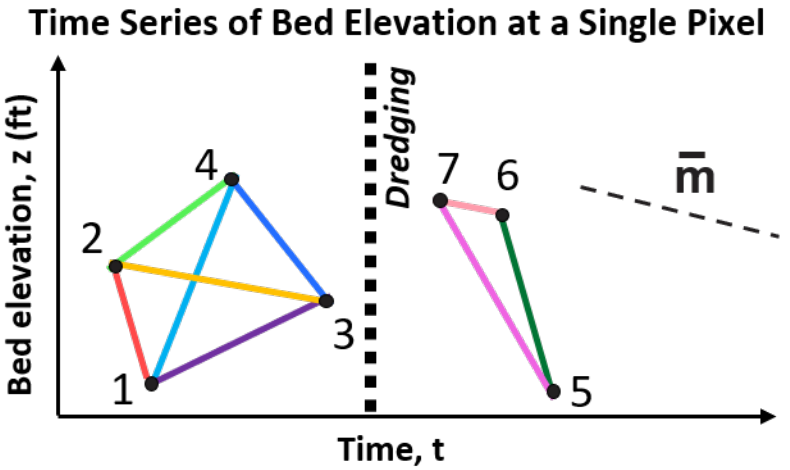
$\bar{m}$ = average*(m_{12}, m_{13}, m_{14}, m_{23}, m_{24}, m_{34}, m_{56}, m_{57}, m_{67})

**using one of several averaging methods*

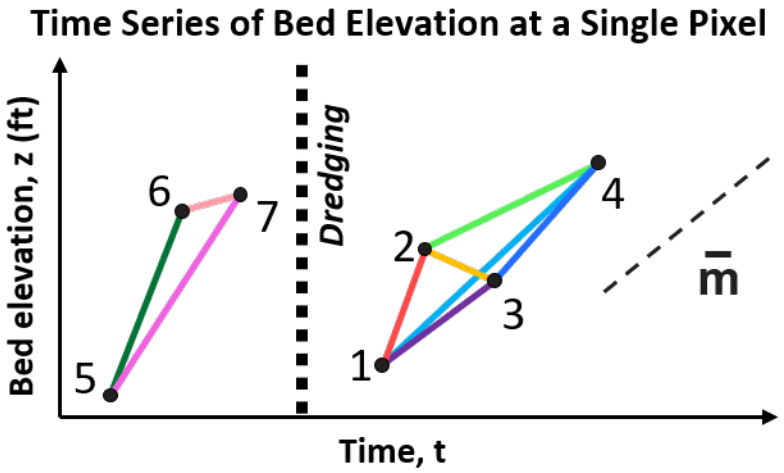
DOES CSAT CARE ABOUT DATA ORDERING?

— which reorderings are valid



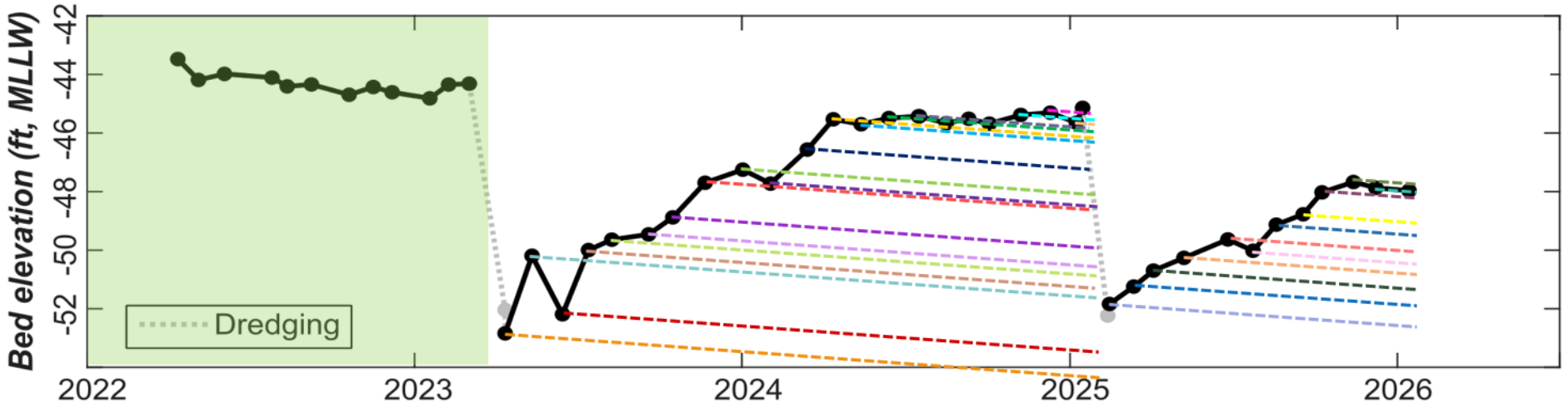


Change the order of surveys between dredging events → alter $\bar{m}$



Change the order the dredge-separated "blocks"→ no effect on $\bar{m}$

Back to our earlier example!



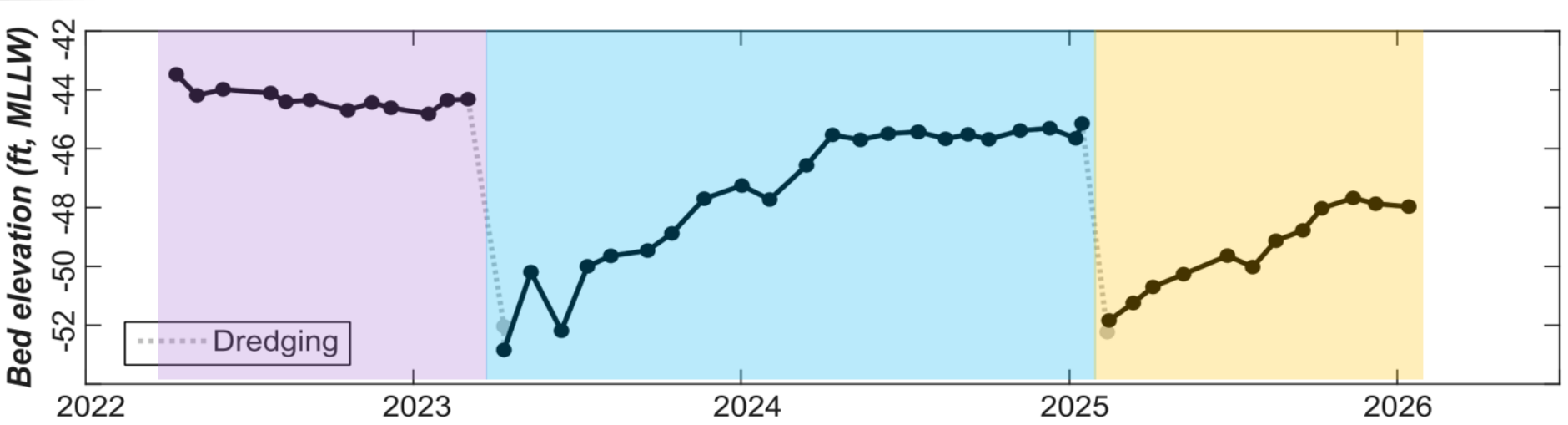
- What if we intentionally hold out the longest window of data with no dredging?

- **This severely reduces the amount of input data used to calculate the CSAT shoaling rate, $\bar{m}$ → gives the appearance of really bad forecasting performance!**

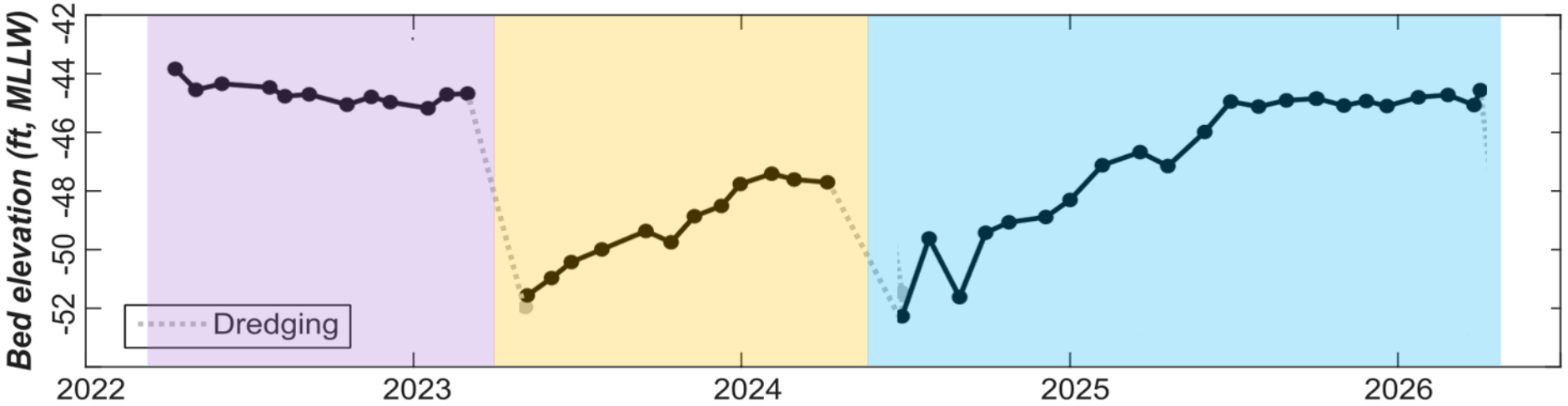
(Not representative of operational CSAT usage—what is the likelihood that a user restricts their input data to the green period, in practice?)

DOES CSAT CARE ABOUT DATA ORDERING?

— shuffling the longest data block



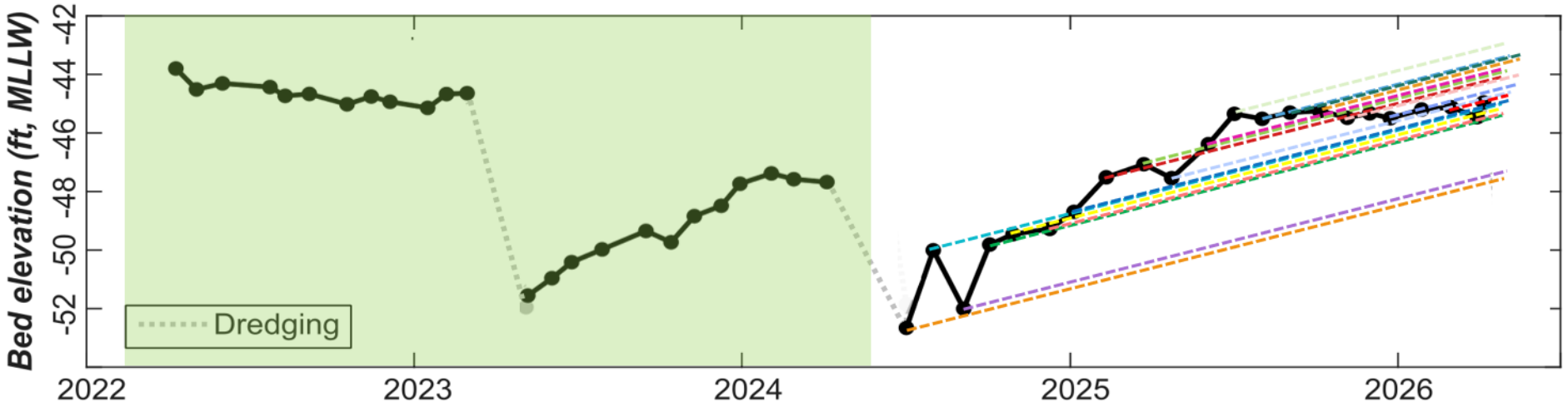
Original data configuration



“Shuffle” the longest continuous block of data to the end of the time series.

DOES CSAT CARE ABOUT DATA ORDERING?

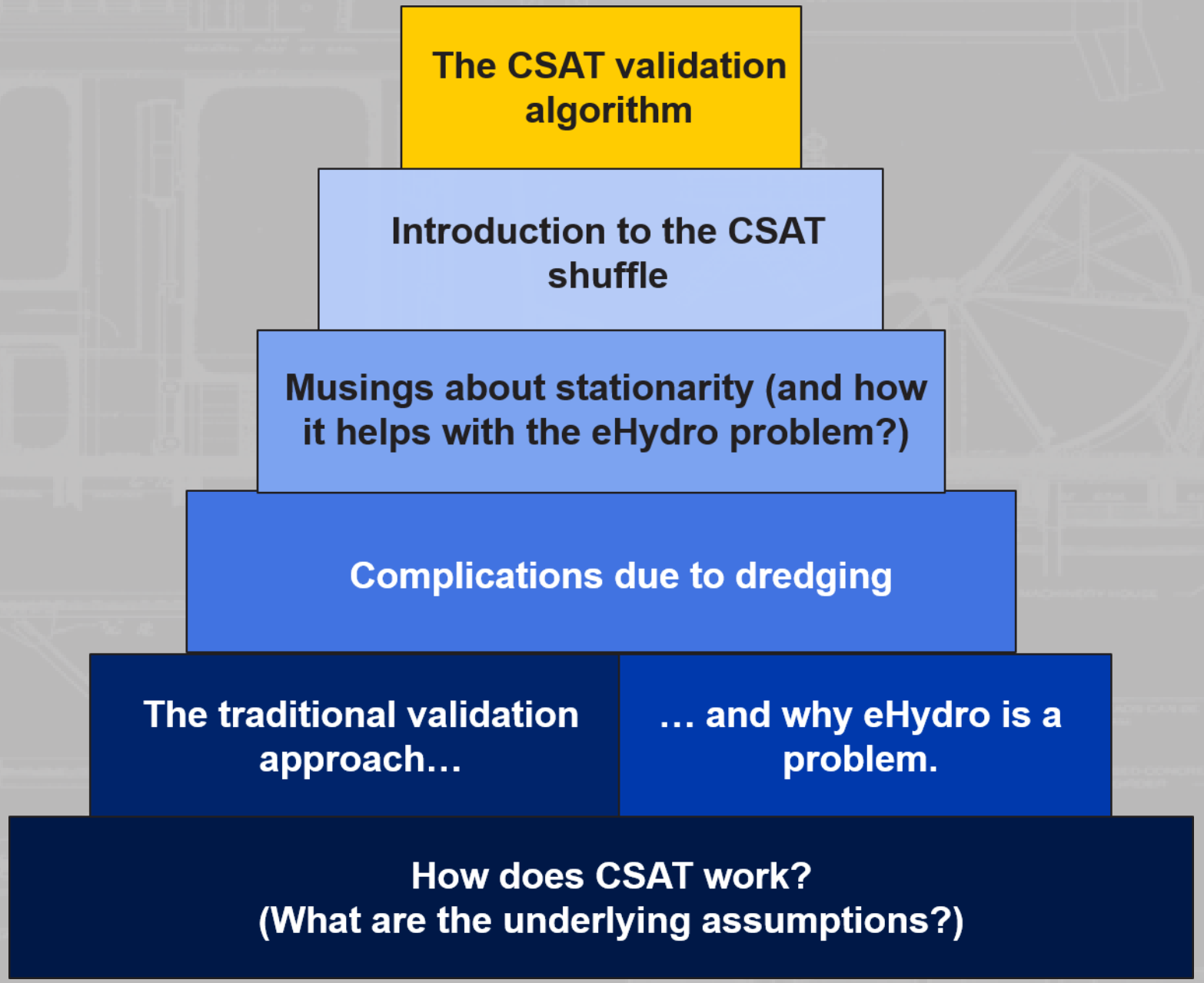
— testing a longer horizon with shuffled data



By moving our longest block of continuous data to the end of the time series…

… we can test CSAT's performance over a longer forecasting horizon while maintaining a large(r) quantity of input data.

PUTTING IT ALL TOGETHER



U.S. ARMY

US Army Corps of Engineers®

ERDC
ENGINEER RESEARCH & DEVELOPMENT CENTER

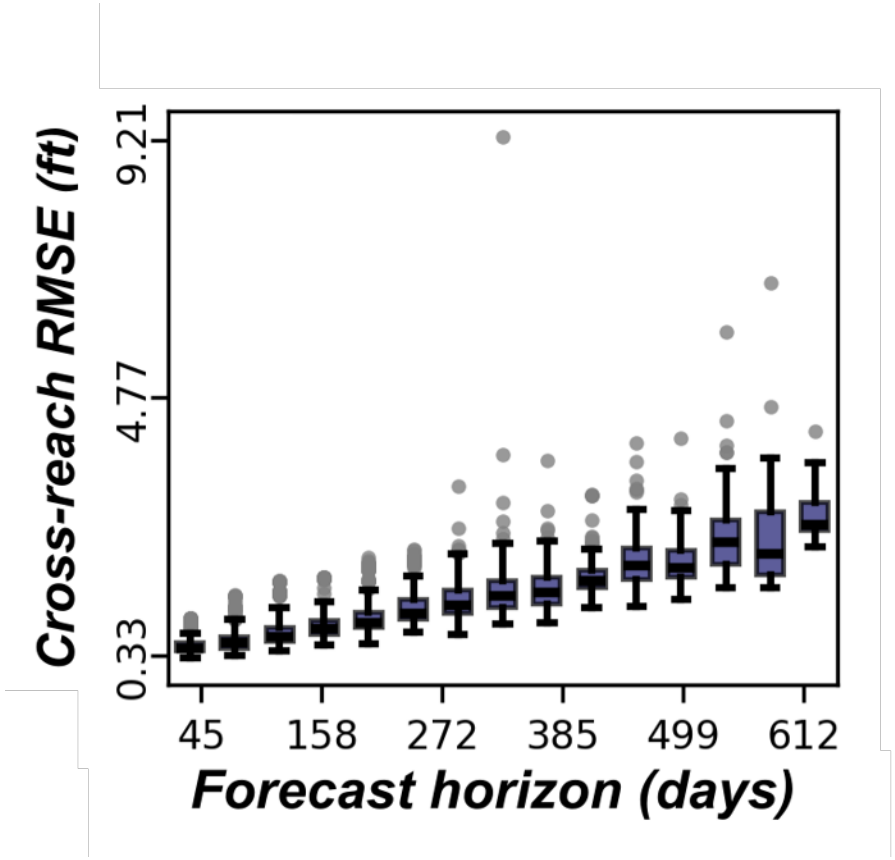
FINAL VALIDATION WORKFLOW

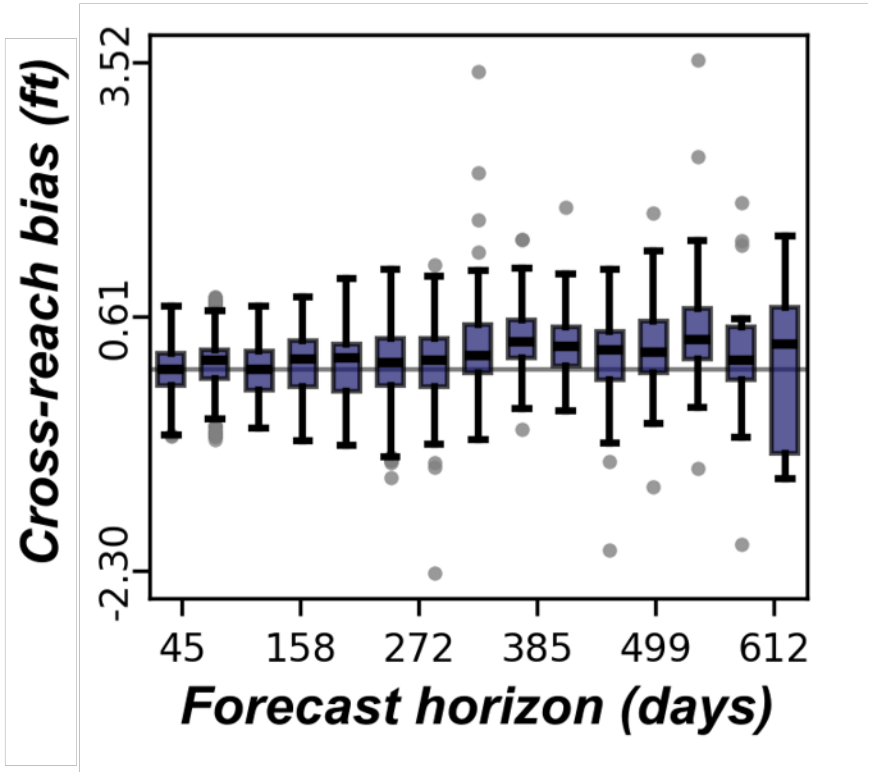
Observation	Handling within Workflow
The CSAT shoaling rate $\bar{m}$ is independent of dredge event ordering.	**Shuffle the blocks of continuous data to maximize…** **(1) the forecasting horizon** **and** **(2) the number of input surveys.**
CSAT assumes that shoaling is linear. • *Statistically stationary process—rate assumed to be independent of when the measurement occurred.* • *There is nothing "special" about using the most recent N surveys as input. Choosing any N surveys from the available input data gives a valid realization of shoaling observations.*	**Randomly select combinations of input surveys when calculating $\bar{m}$.** • *Initially attempted to create ALL possible combinations of input surveys—too computationally intensive for typical ACE-IT laptop.* • *Monte Carlo approach maintains practical runtime.*
Certain (not necessarily identified) eHydro surveys have terrible data quality. • *Depending on position within the time series, potential for outsize influence on the validation results.*	**Test $\bar{m}$ against all possible combinations of forecast start and end timestamps from the "held out" data.**

EXAMPLE RESULTS

— error vs. forecast horizon

How much does the forecasting error increase as I project further into the future?

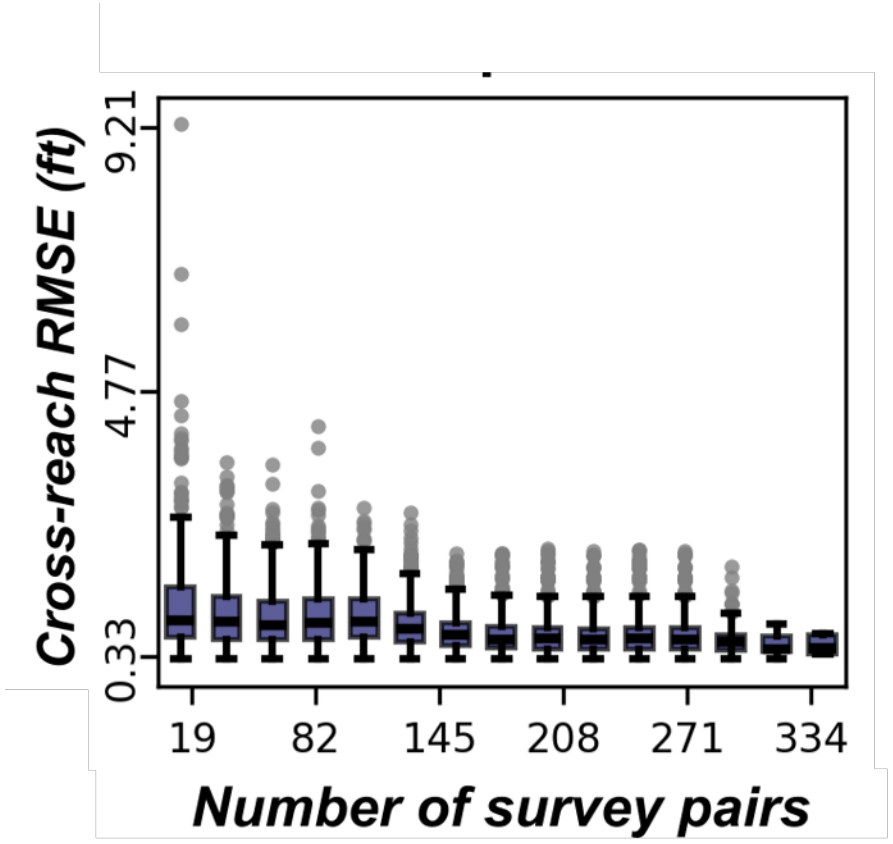


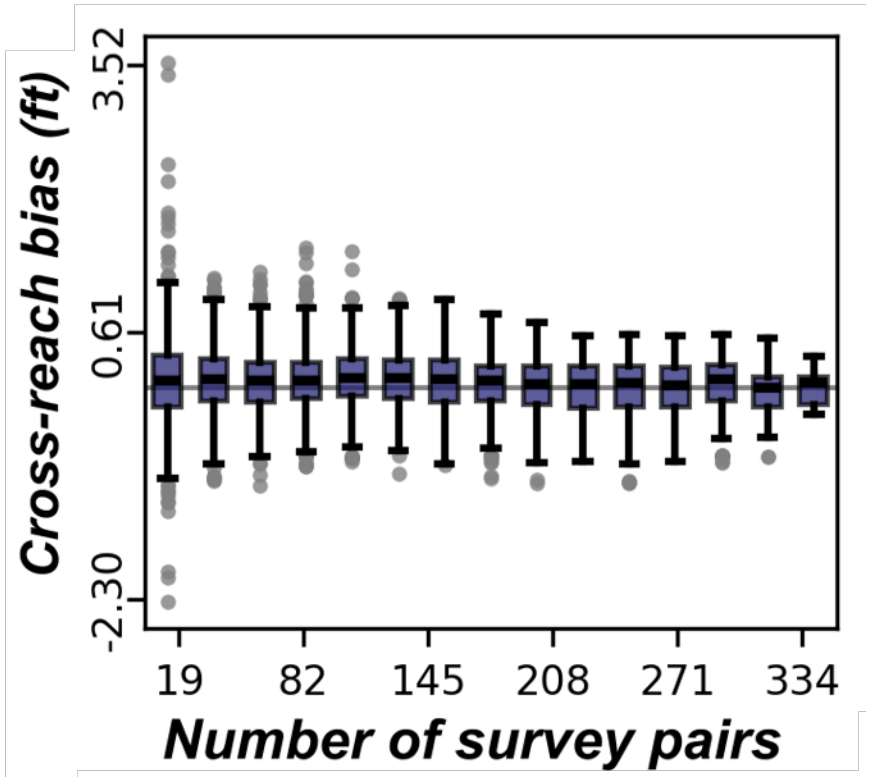


EXAMPLE RESULTS

— error vs. number of input surveys

How does the number of input surveys affect the forecasting accuracy?

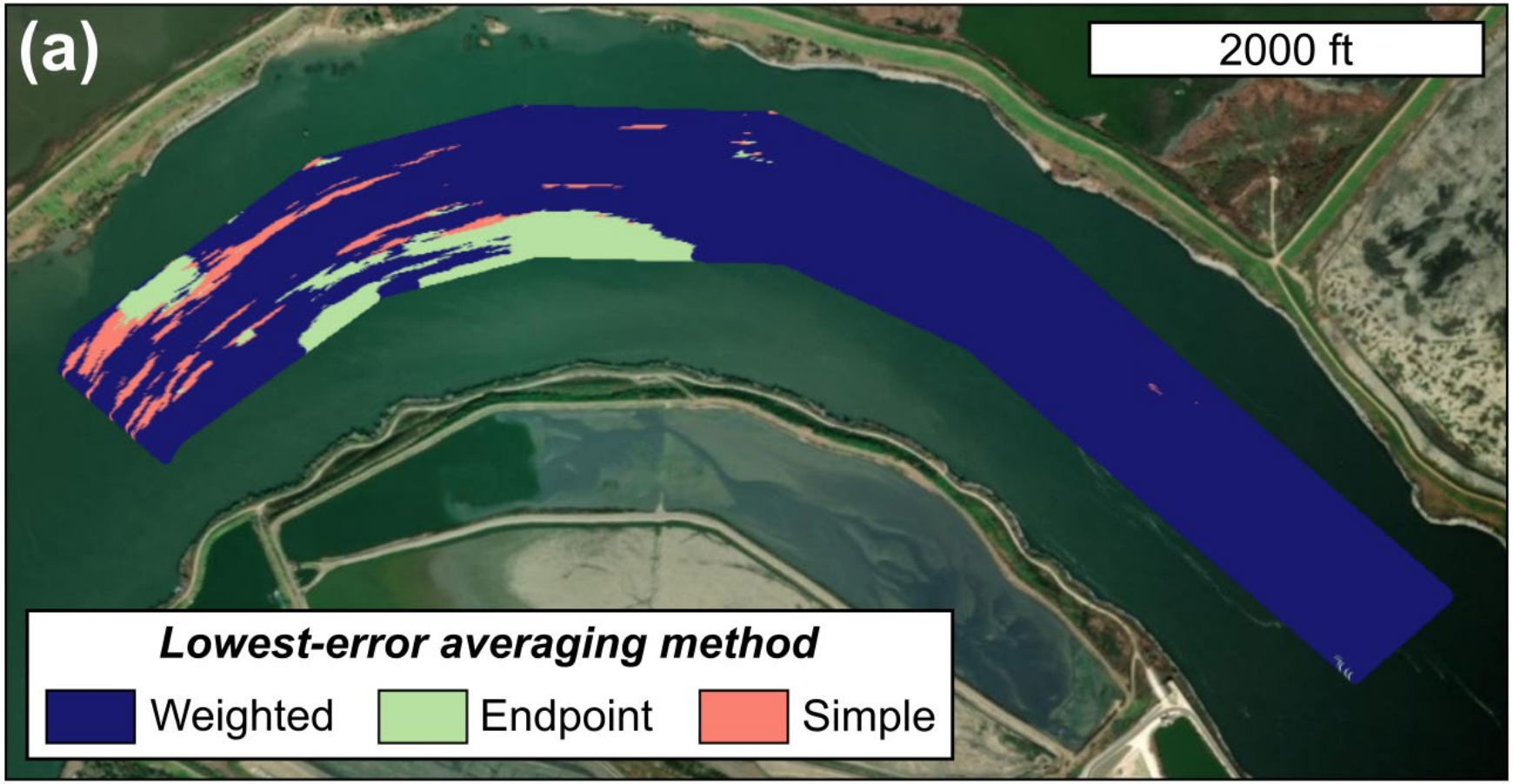




EXAMPLE RESULTS

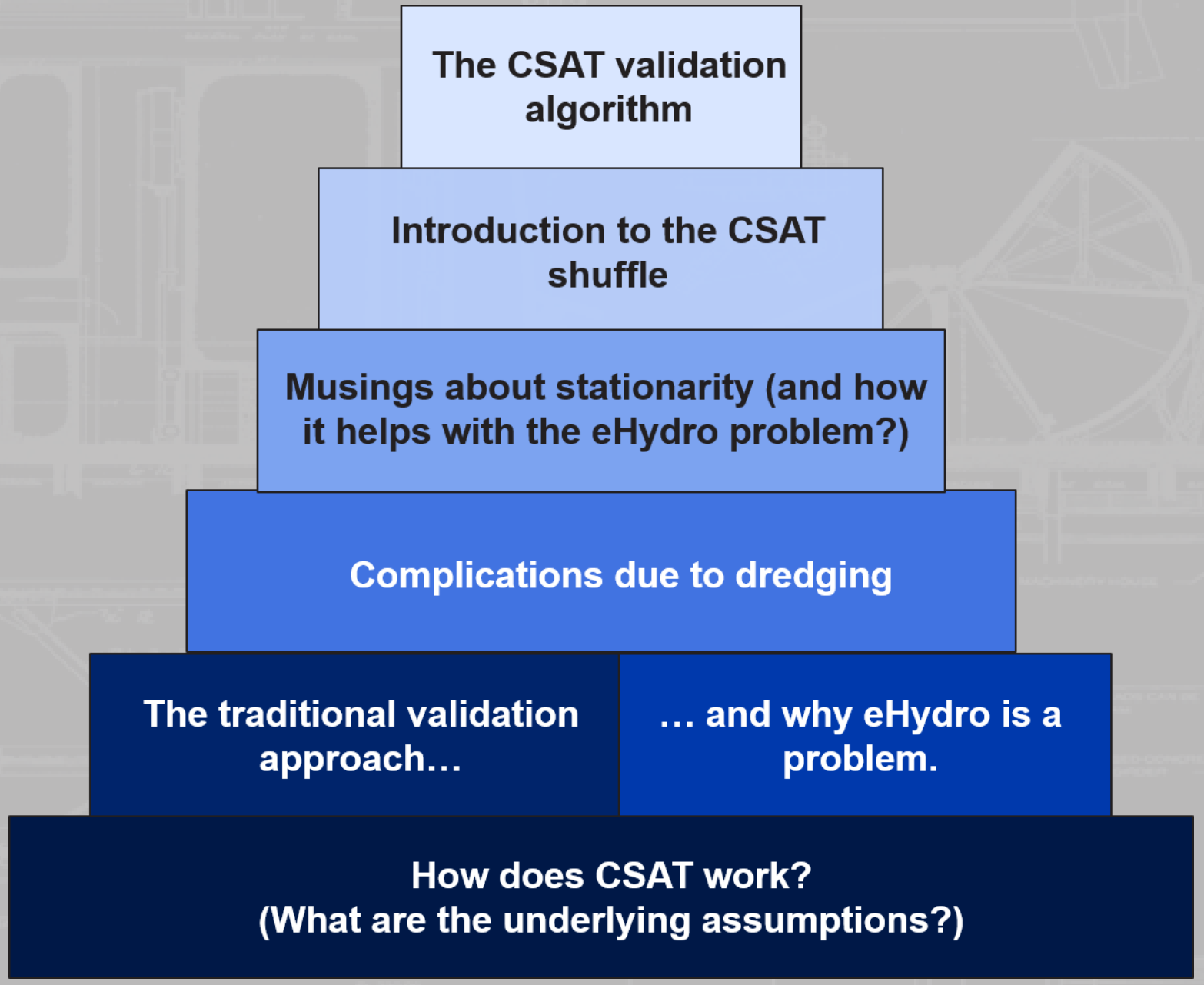
— comparing averaging methods spatially

Which CSAT averaging method is most accurate?



Basemaps: Esri, i-cubed, USDA, USGS, AEX, GeoEye, Getmapping, Aerogrid, IGN, IGP, UPR-EGP, and the GIS user community

RECAP



U.S. ARMY

US Army Corps of Engineers®

ERDC
ENGINEER RESEARCH & DEVELOPMENT CENTER

CONCLUSION

- **Time series validation is nothing new.**
  - *Validating CSAT should be straightforward… right?*
  - *What about when dredging gets in the way?*
  - *What about eHydro data quality issues?*
  - *What if I want to know how <THIS SPECIFIC THING> influences data accuracy?*
- **Our approach to handling the unique "character" of the eHydro time series is a random sampling approach which includes shuffling the blocks of data between dredging events to maximize the windows of usable data.**
  - *Unorthodox? Yes!*
  - *Necessary? We argue, also yes!*
- **Coming soon to your CSAT release!**