Imbalance Reserves (IR) Mosaic Parameter Results and Summary: Stakeholder Session 2

Tobiah Steckel
Short Term Forecasting
February 24th, 2025

Outline

- Session 1
  - Draft decisions
  - Stakeholder Comments
  - Data Correction
  - Alternate Sample Scheme
- Background
- Supplementary Background (10 mins)
  - Equal-weighted Hscore
  - Quality Assurance Method
  - Thresholds
- Mosaic parameter testing approach
- Mosaic parameter results
  - Quality Assurance (Ensemble 3)
  - Dynamic Thresholds (Ensemble 4)
  - Static Thresholds (Ensemble 5)
- Final STF Recommendation

Session 1 Review and Recap

BACKGROUND

STF Draft Recommendation November Call

Mosaic Parameter	Order of Evaluation	Current Value	Session 1 Recommended Value
Historical Days / Split Window (Sample Scheme) November Call	1	Sample scheme 4	Sample scheme 4
Historical Period Sample Days November Call	2	180 sample days	150 sample days

Data Correction needed following Session 1 results

Why:

Trade Hour shift from aggregated "external" DA forecasts and CAISO RTPD forecasts for ~1/2 data set

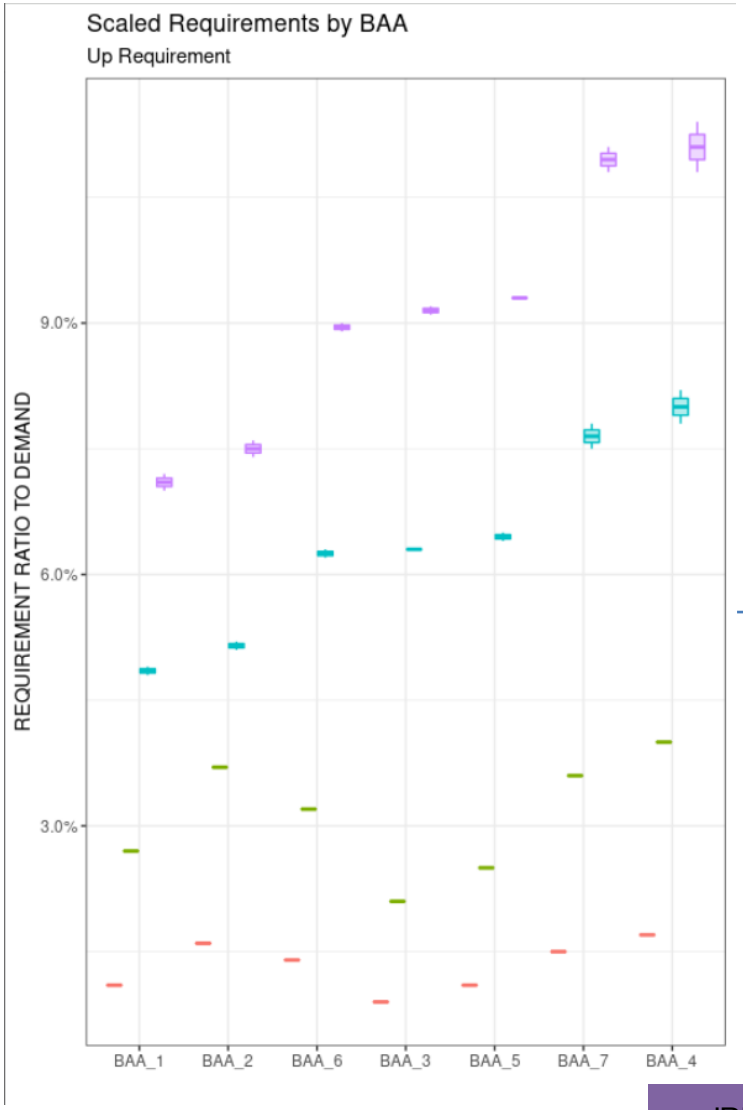
Impact:

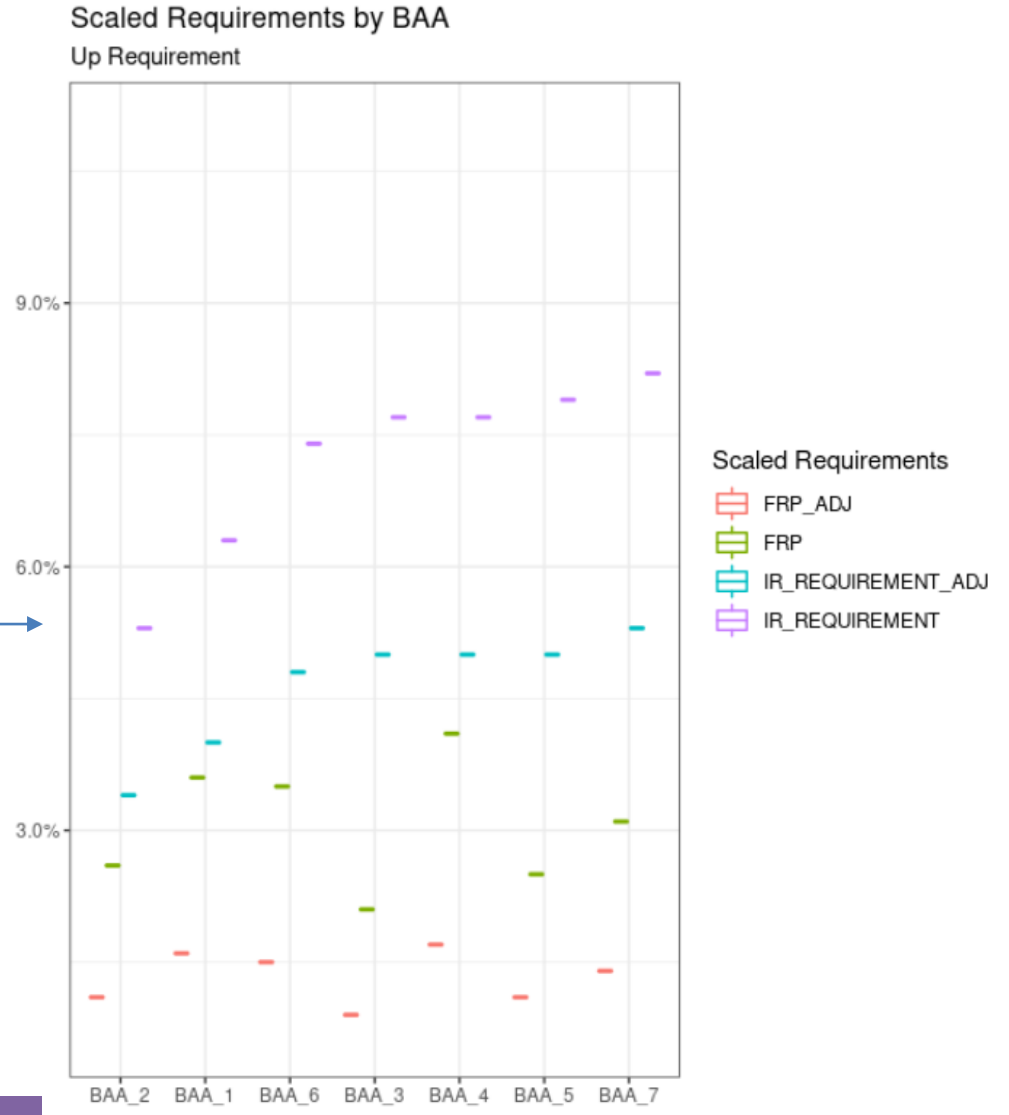
Corrected input data led to the following:

- Requirement reduction
- IR highest requirement being in the Summer

No changes to recommendations from session 1 due to data correction. Please see updated slides from session 1.

Data Correction Reduces Requirement Across BAAs

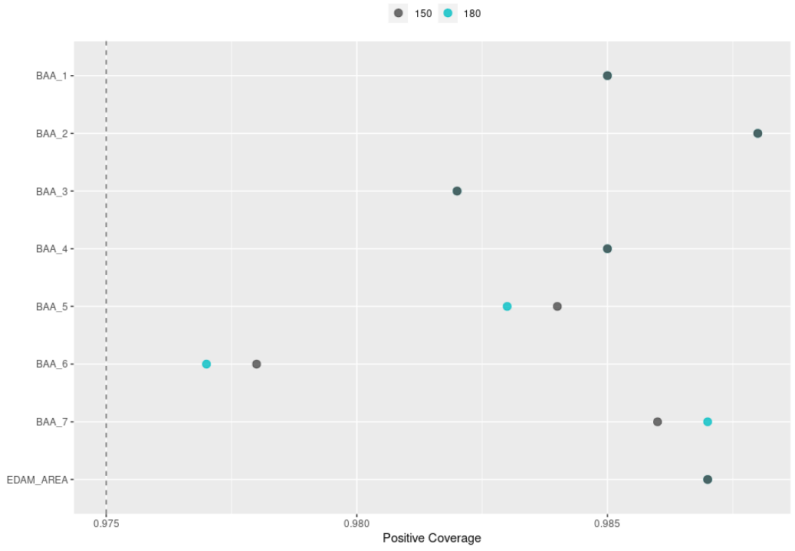




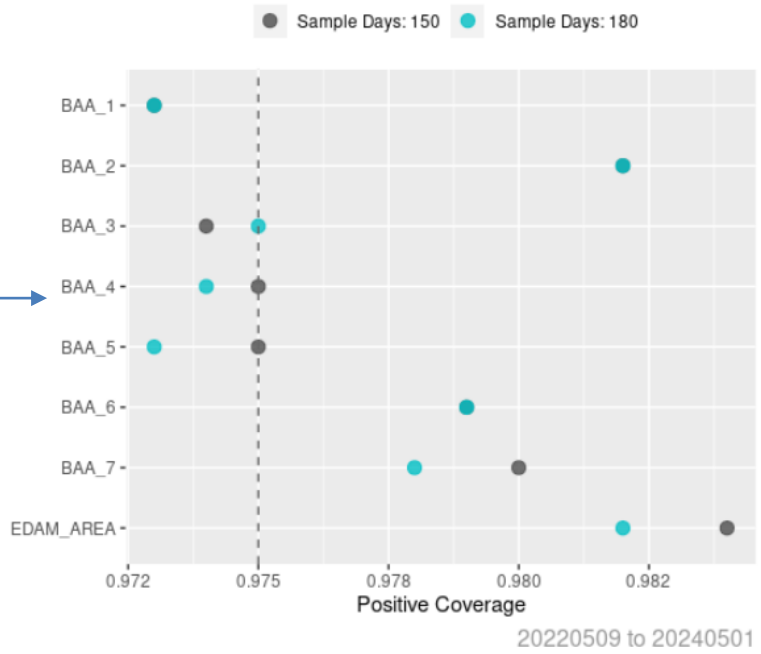
IRU

Data Correction provides results closer to target coverage with some instances below across BAAs

Before Data Correction



After Data Correction



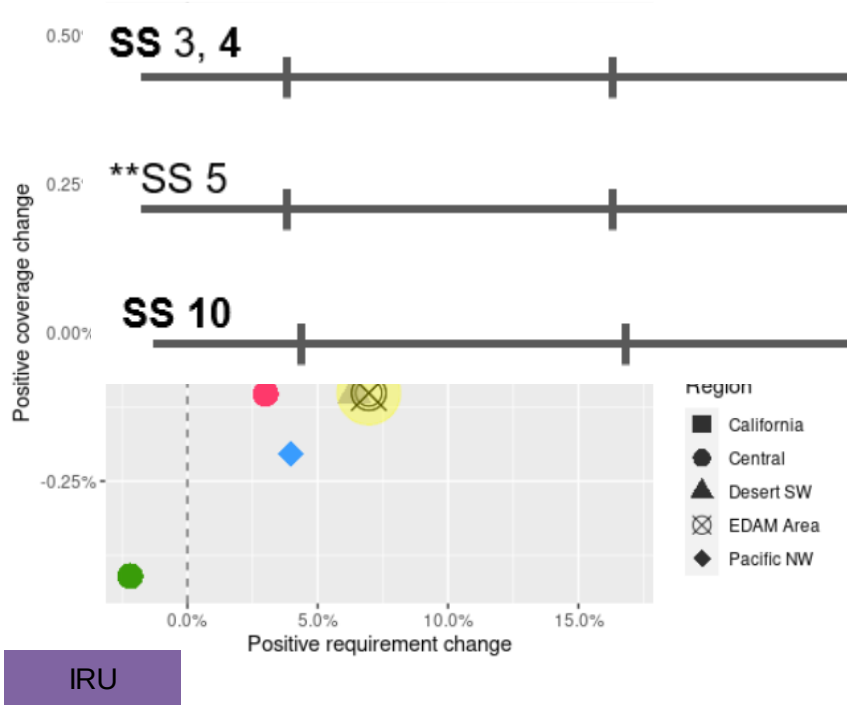
High level Summary of Stakeholder Comments

- Elaboration or consolidation of Hscore
  - Supplementary Background
- Provide exceedance metrics
  - Highlight magnitude of non-covered observations
  - Mosaic Parameter Results
  - Additionally we propose putting a supplementary repository of results
- Explore alternative sampling schemes
  - Affirm results from session 1

Ensemble 1

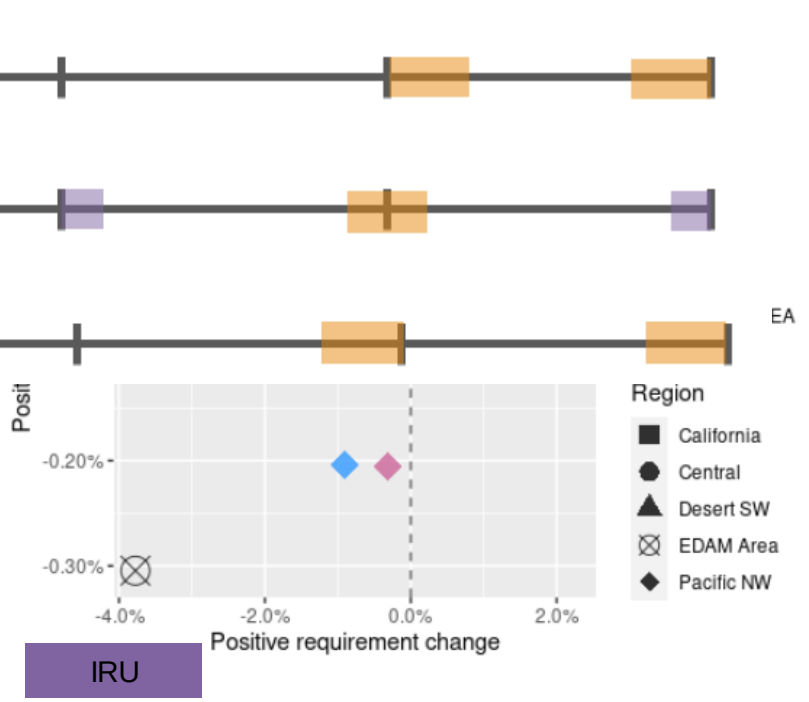
Alternative Samples: Mixed Results for future enhancement MSC-suggested (SS5) and for WPTF-suggested (SS10)

SS 4 → SS 10



IRU

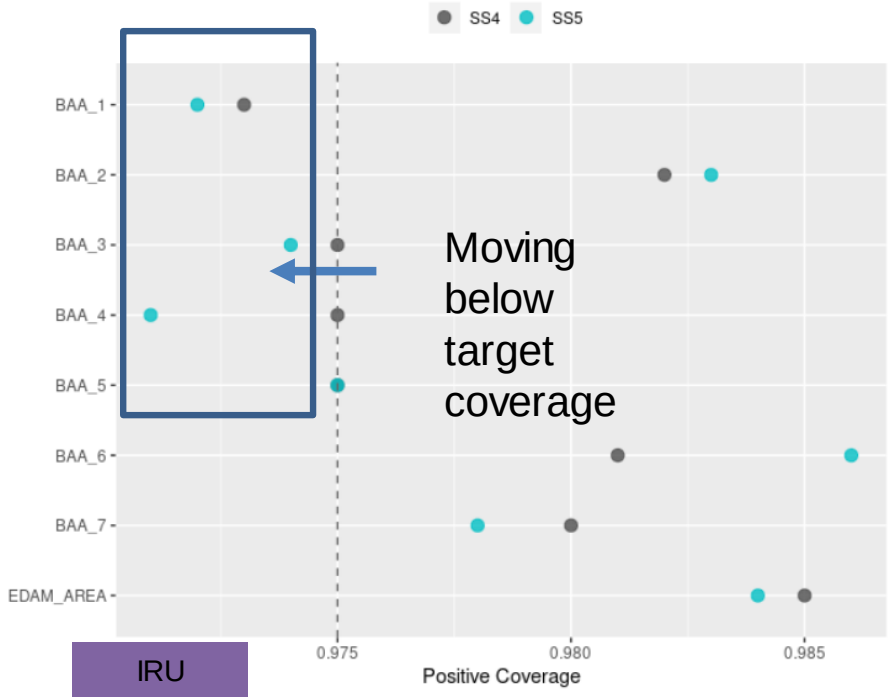
SS 4 → SS 5

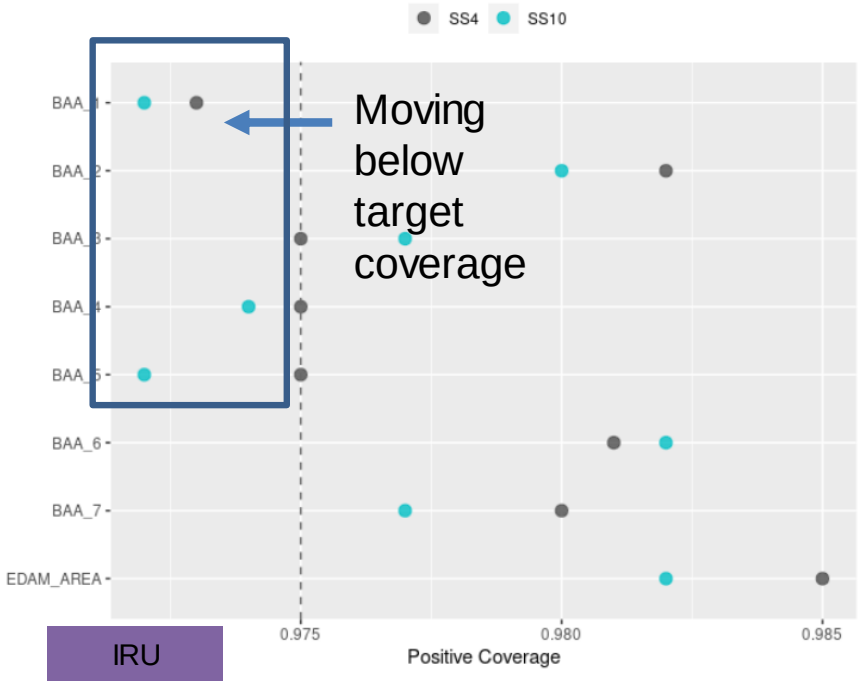


IRU

Ensemble 1

Sample scheme 5 and 10 show mixed results where multiple BAAs observe coverage divergence from target 97.5 percentile.





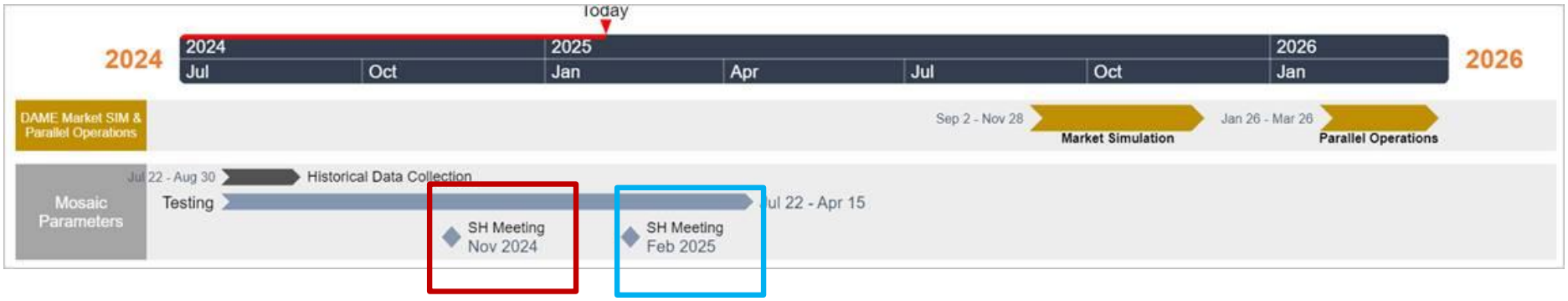
Background

- [Purpose] Ongoing evaluation of IR mosaic quantile regression parameters. Expectations of mosaic established from FRP may diverge regarding IR

- [Plan] STF team will trial mosaic parameter configurations published in external BRS, as well as alternative configurations, for overall performance. STF will then present on findings and offer recommendation.

- [Goal] Market participants will get a chance to evaluate and provide comments and optimized values will be deployed in DAME Market Simulation.

Timeline



November SH Call – Discuss initial findings regarding mosaic parameters and diversity benefit results for entities that have signed implementation agreement as well as for expanded EDAM

February SH Call – Discuss any follow-up or overflow from November stakeholder call 1. Present draft parameter recommendations for data quality, dynamic and quarterly thresholds steps. Solicit final recommendation with respect to November and February presentations, and open up for external comments.

What are the Mosaic Parameters?

Mosaic Parameter	Order of Evaluation	Current Value (→)	Items Evaluated
Historical Days / Split Window (Sample Scheme) November Call	1	Sample scheme 4 → 4	Performance of IR with sampling schemes: 4, 7, 9 (5,10)
Historical Period Sample Days November Call	2	180 sample days → 150	Performance of IR with samples days in the historical period: 150, 180, 210
Quality Assurance (QA) Binary February Call	*3*	QA not applied	Performance of IR with QA applied and QA not applied
Dynamic Threshold Percentile February Call	*4*	99% (1%)	Performance of IR with 99% (1%), 98.5% (1.5%), and 98% (2%) dynamic threshold percentiles applied
Static Threshold Sample February Call	*5*	90 days, sample scheme 1 [ss1,90]	Performance of IR with [ss1,90], [ss1,150], [ss4,90], and [ss4,150] static threshold applied

Supplementary Background

EQUAL-WEIGHTED HIERARCHICAL SCORE (HSCORE)

Hscore Background

Hscore is an aggregate metric, designed to efficiently capture multifaceted performance considerations.

While component metrics per BAA are presented and reviewed, Hscore helps holistically simplify and evaluate performance across mosaic parameter configurations.

In Sample Metrics

Metrics (e.g., coverage and requirement) [Out of Sample]

Hscore

Production cost model (counterfactual)

Less difficult to construct
Less compute
Less information
Less efficient to communicate

More difficult to construct
More compute
More information
More efficient to communicate

Hscore Construction

Step	Description	Specifics used for this analysis	Additional considerations
Identify metrics	Define the component metrics.	• Coverage • Requirement • Exceedance	There are many more options for metrics to consider (e.g. breakpoint, AIC, pinball loss etc.)
Identify scenarios	Define the relevant scenarios and their respective values.	• Direction [up, down] • Period of Day [all hours, peak] • BAA [BAA 1 - 7]	There are many more potential scenario values like summer or winter months, ramp hours, day type, etc.
Calculate	Calculate the component metric per scenario (e.g. upward coverage at peak for BAA 1).		
Normalize	Prepare metrics for aggregation by centering, scaling, and standardizing to ensure lower is better.		
Add weights	If desired, add weights for critical areas of performance.	No weights used in an equal-weighted Hscore ($w_i = 1$).	One could choose to more heavily weight requirement performance over coverage and exceedance and assign weights like ($w_{\text{requirement}} = 0.6, w_{\text{coverage}} = 0.2, w_{\text{exceedance}} = 0.2$) for example.
Aggregate	Combine all weighted (in this case equal-weighted) component metrics across scenarios to achieve a single Hscore per mosaic parameter configuration.	Considering 3 metrics, 2 directions, 2 periods, and 7 BAAs, there are 84 metrics per mosaic parameter configuration. These 84 values are averaged and presented as Hscore.	There are other options for aggregating these values like nesting under reliability, cost etc.

In summary, $\text{Hscore} = \frac{1}{mn} \sum_j^m \sum_i^n (w_i x_i)_j$ where i indicates metric and j indicates scenario. For this analysis, $m = 28$ (2 directions * 2 scenarios * 7 BAAs) and $n = 3$ for 3 metrics considered. This formulation is a slight simplification as the full algorithm includes some further mapping and constraints on w_i, but this is a sufficient representation of the steps outlined above. We plan to provide a reference document with greater detail on Hscore in Q2 of 2025.

Hscore Evaluation

- Lower is better (golf score)
  - Normalization step includes orienting all metrics such that lower is better (e.g. minimize requirement magnitudes, minimize exceedance magnitudes, minimize deviation from target coverage).
- Relative scoring metric
  - Hscore is dependent on its particular construct and the metrics and scenarios considered.
  - Hscore is not universally comparable and get its meaning from comparing Hscores of the same construct across different models or across different parameter configurations.

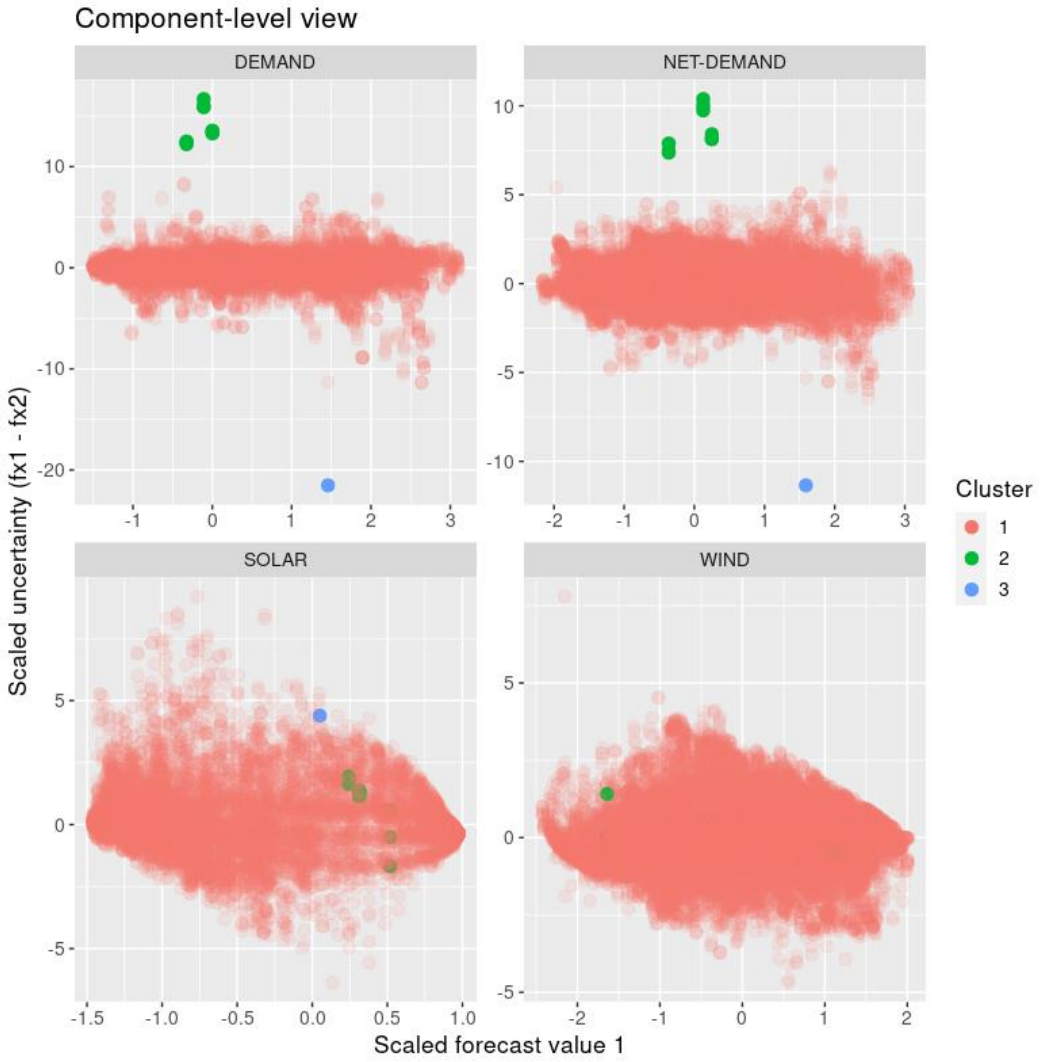
Supplementary Background

QUALITY ASSURANCE (QA) METHOD

Quality Assurance (QA) is needed to protect historical data quality

- Methodology: DBSCAN clustering
- Input: scaled net demand uncertainty and scaled DA forecast and trade hour
- Output: Normal (cluster = 1) or abnormal (cluster ≠ 1)
- Objective: Errors caused by IT issues are captured within the algorithm –
  - **Forecasting Team Member will still assesses methodology outputs and potentially overrides designation**

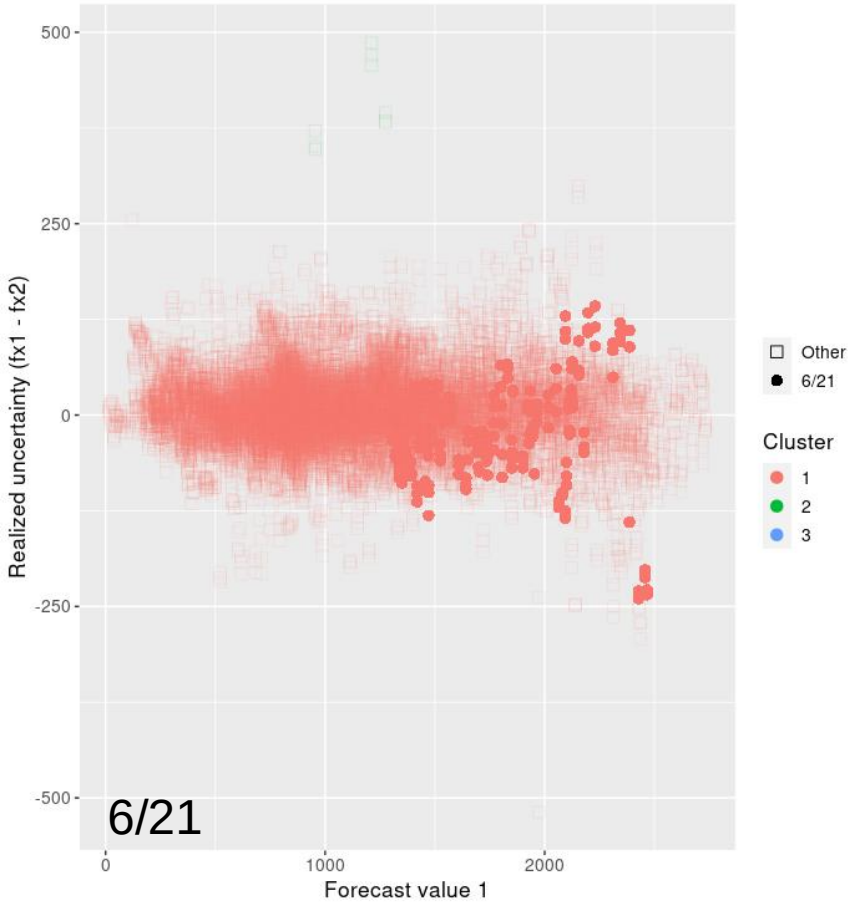
- Qualities that indicate higher likelihood of need to discard samples:
  - Anomaly can be traced back to an individual component (solar, wind, or demand)
  - Time period clustering, consecutive timestamps tend to indicate an event rather than noise
  - Identifiable cause like stale forecasts or telemetry quality issues

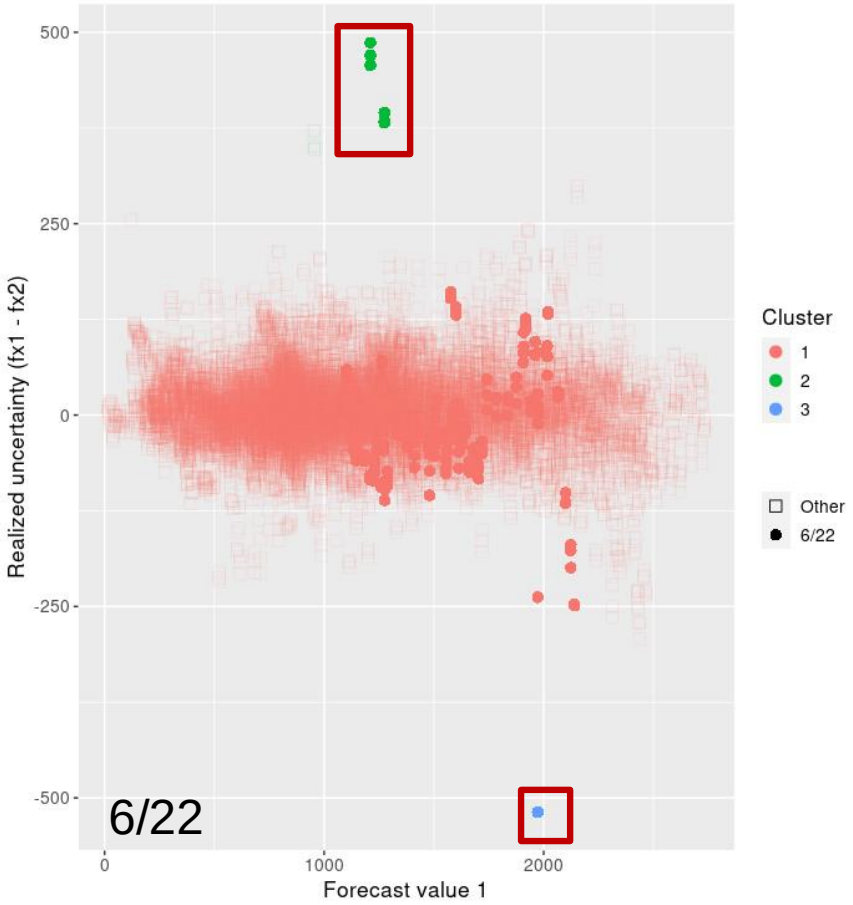


Example anomaly with identifiable cause

6/21 and 6/22 had monsoon conditions in this region

- 6/21 (left) had monsoon conditions without anomaly
- 6/22 (right) had a known IT issue captured by anomaly detection in net demand samples





Bulk QA Application vs. “Online” QA Application

Simulation / Onboarding

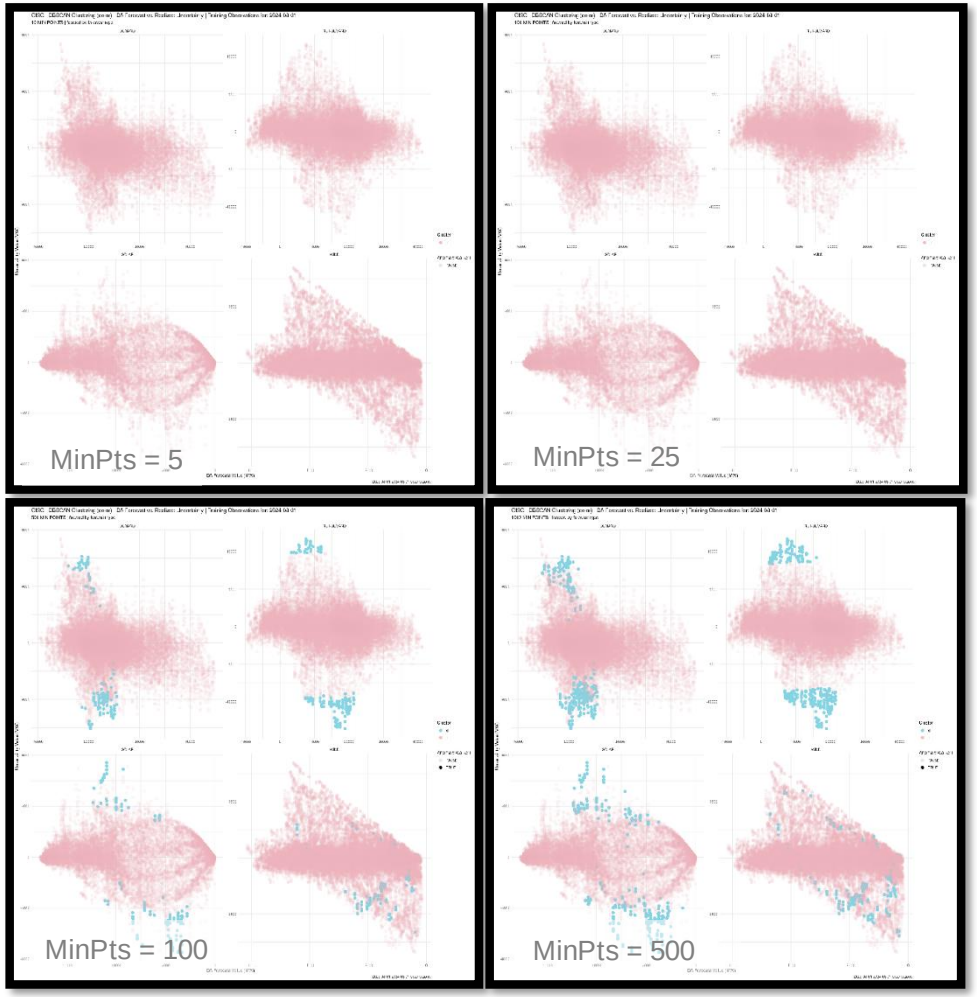
- Bulk process
- Decide one shared set of DBSCAN parameters and apply for whole training set
- Provide further information in BPMs as IR product goes live.

Production IR

- “Online” process
- Manual Review, discard ranges and effective dates reported quarterly – aggregating with existing review
- Provide further information in BPMs as IR product goes live.

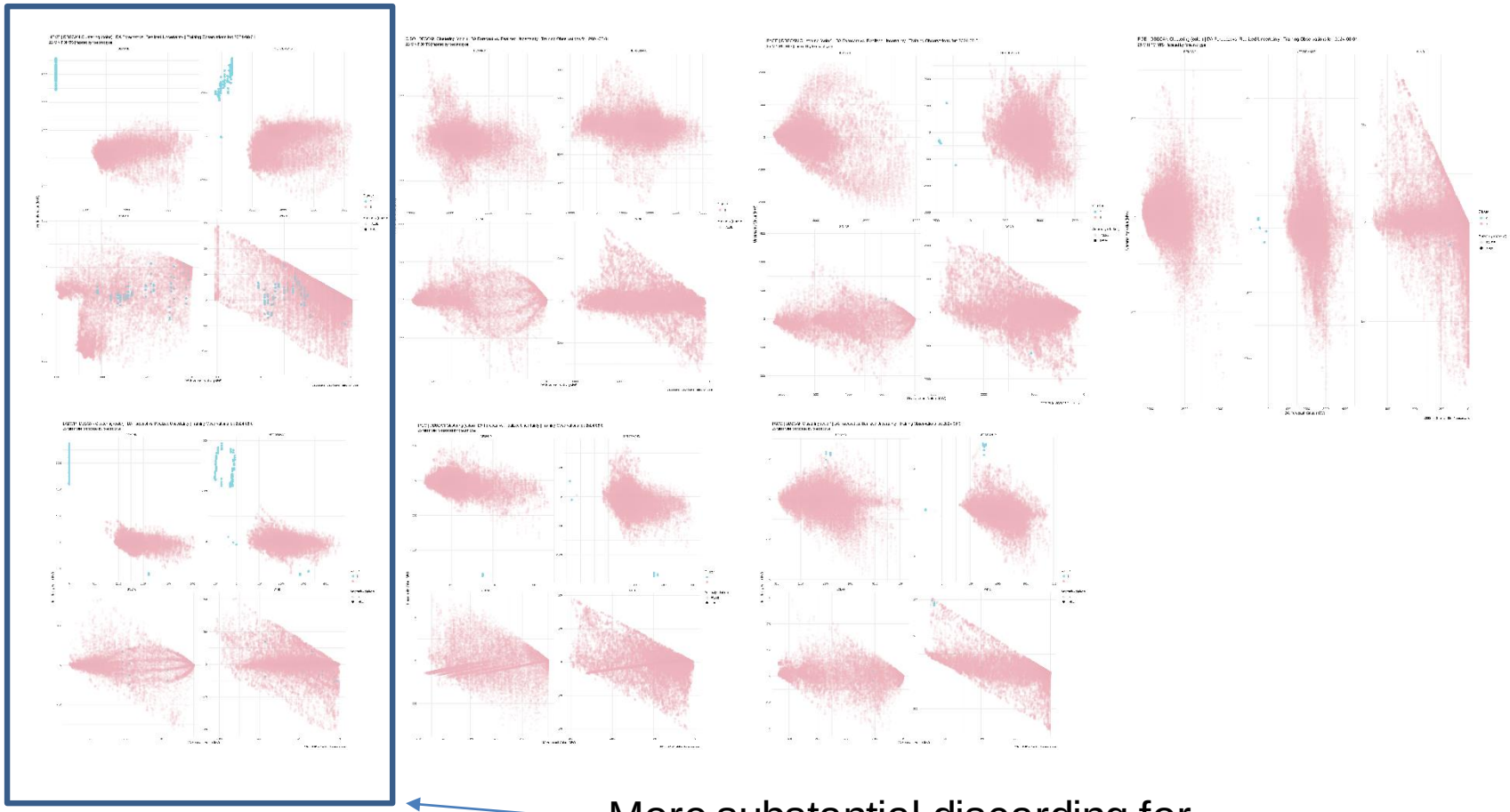
Fixing DBSCAN Parameters for Simulation

- DBSCAN parameters are decided prior to simulation run
- Changing *MinPts* parameter until universal representation of guidelines for "online" review
- As we increase *MinPts* parameter (all else equal) the observations that will be "discarded" increase.
- Bottom two figures are discarding too many observations (e.g., *constructive* uncertainty between forecast type)



Discarded observations (blue) for the participating BAAs

MinPts = 25 provided best universal representation for "online" review across 7 BAAs, thus this parameter was utilized to populate the QA applied set for Ensemble 3



More substantial discarding for BAA_3 and BAA_7

Supplementary Background

THRESHOLDS

Thresholds overview (IR)

Threshold	Granularity	Update frequency	Value *
Floor	Constant	None	0.1 MW
Dynamic	Hourly	Daily	1st and 99th percentile from a sample that mirrors the mosaic calculation sample
Static	Constant	Quarterly	1st and 99th percentile from 90-day rolling sample

The raw mosaic results in IRU and IRD are compared to and constrained by static and dynamic thresholds. At an absolute minimum, flex ramp requirements are 0.1 MW. Assuming down requirements and thresholds are expressed as negative values, thresholds are applied as follows:

Flex ramp up (IRU) requirement = $\max(\min(IRU_{raw\_mosaic}, threshold_{dynamic}, threshold_{static}), 0.1)$

Flex ramp down (IRD) requirements = $\min(\max(IRD_{raw\_mosaic}, threshold_{dynamic}, threshold_{static}), 0.1)$

* Under the current configuration, the dynamic threshold is a 180-day day symmetric sample and the static threshold is a 90-day rolling sample. A rolling sample includes N days preceding the trade date. A symmetric sample is composed of an N/2 days preceding the trade date and N/2 days succeeding the trade date from one year prior.

Summary of steps: Dynamic
Differences from static threshold calculation in blue

- Create historical sample of realized uncertainty
  - Calculate net demand forecasts
  - Calculate realized uncertainty from difference in advisory to binding net demand forecasts
  - 150 day symmetric sample (SS4)
  - For IR, keep only minimum and maximum sample per DA interval (this will eliminate ~1/2 of data)
- Group uncertainty samples by hour and calculate 1st and 99th percentile
  - Hourly percentiles used as histogram thresholds

Summary of steps: Static Thresholds

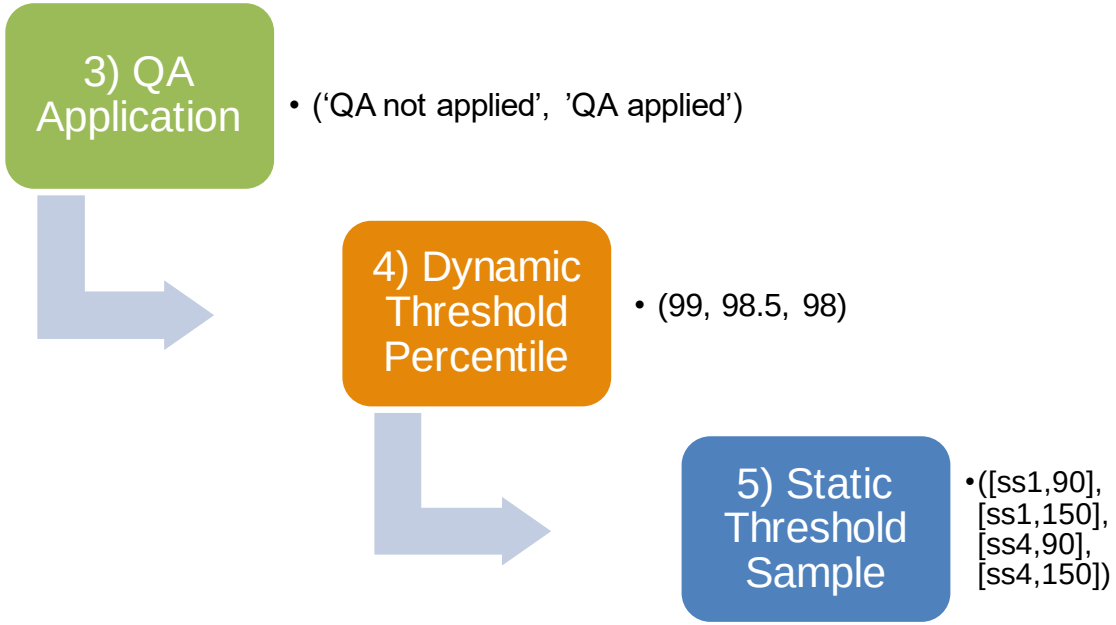
- Create historical sample of realized uncertainty
  - Calculate net demand forecasts
  - Calculate realized uncertainty from difference in advisory to binding net demand forecasts
  - 90 day rolling sample (SS1)
  - For IR, keep only minimum and maximum sample per DA interval (this will eliminate ~1/2 of data)
- Group uncertainty samples by hour and calculate 1st and 99th percentile
  - Take the min and max of all hours to get static thresholds

February Meeting

MOSAIC PARAMETER TESTING APPROACH

Testing Approach

- Simulate 2 years worth of Imbalance Reserve Requirements (~70,000 intervals per BAA)
  - (May 2022 to May 2024)
- Assess **performance** within ensembles and *feed the optimal result forward*
- If change is suggested at summary level, assess rolling/seasonal summaries and weighted summaries



Three approaches to assessing performance

- **(1) *Period summary**
  - **BAA level**
  - **EDAM AREA**
- (2) Rolling or Seasonal summaries
  - BAA level
  - EDAM AREA
- (3) Hscore
  - All BAAs are equally weighted

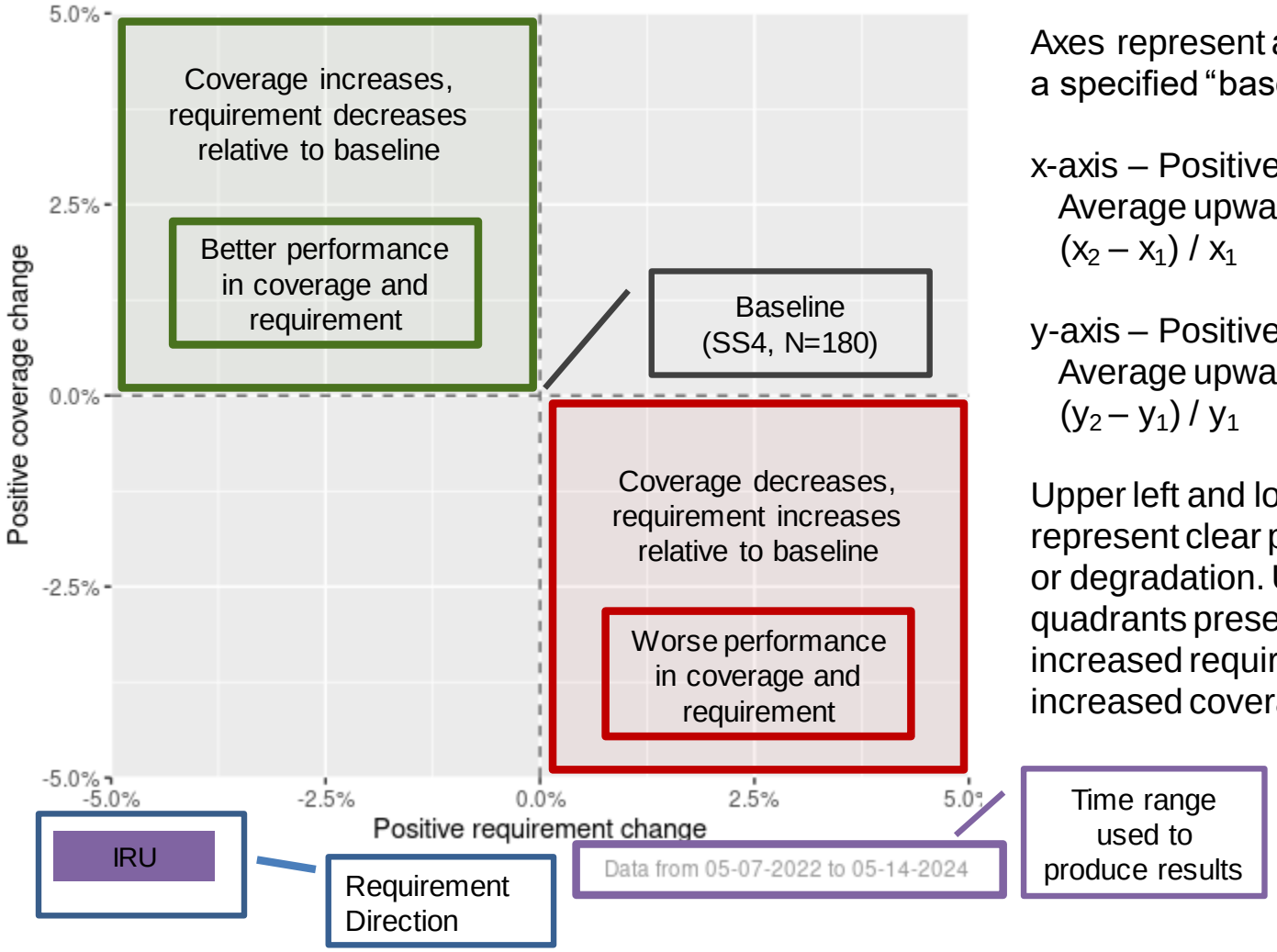
*2 year average

Hscore

- (H)ierarchical score
- Holistic performance metric
- Combination of coverage and requirement, plus consideration of time of day, requirement direction, and sample period length

Meetings and events > Market Surveillance Committee > Uncertainty Performance – Presentation – Apr 11, 2024

Explainer for pareto-type period summary plot



Axes represent a scaled change relative to a specified "baseline" performance

x-axis – Positive requirement change
Average upward requirement (x)
$(x_2 - x_1) / x_1$

y-axis – Positive coverage change
Average upward coverage (y)
$(y_2 - y_1) / y_1$

Upper left and lower right quadrants represent clear performance improvement or degradation. Upper right and lower left quadrants present mixed results. Typically, increased requirements coincide with increased coverage.

Results are presented in their order of evaluation

Ensemble 1	Sample Scheme
Ensemble 3	QA Application
Ensemble 4	Dynamic Threshold Percentile
Ensemble 5	Static Threshold Sample

Results are presented by requirement direction

IRU	Upward Imbalance Reserve
IRD	Downward Imbalance Reserve
BOTH	Both directions

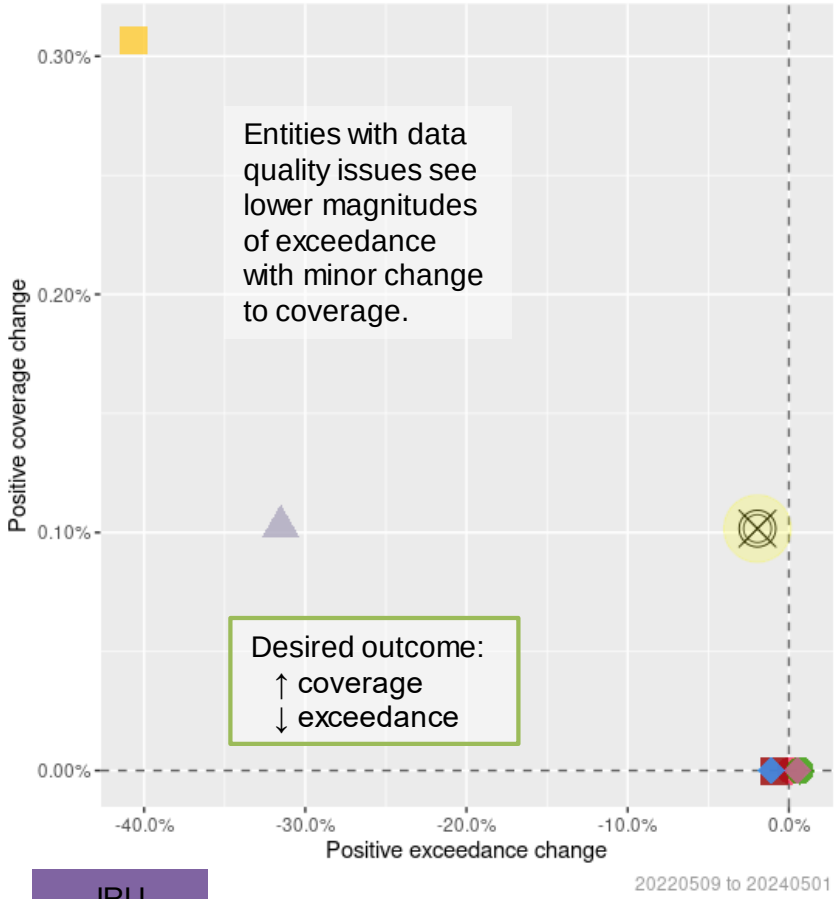
February Meeting

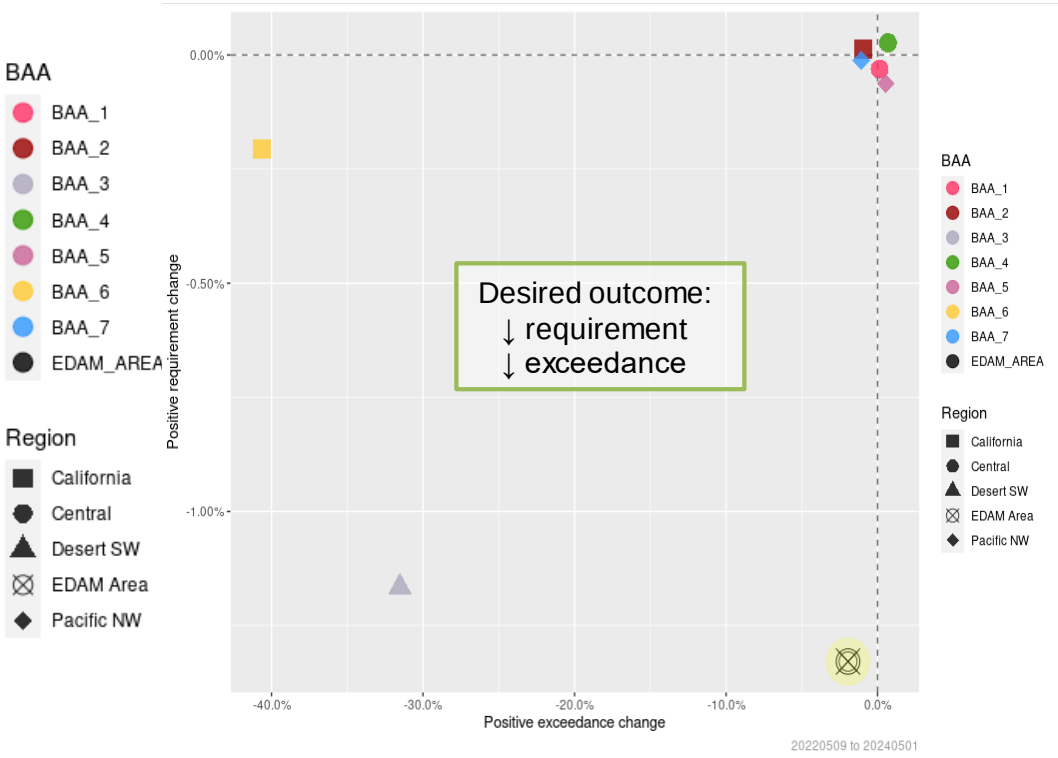
MOSAIC PARAMETER RESULTS

Ensemble 3

Application of QA method improves performance, specifically benefiting BAAs with data quality issues

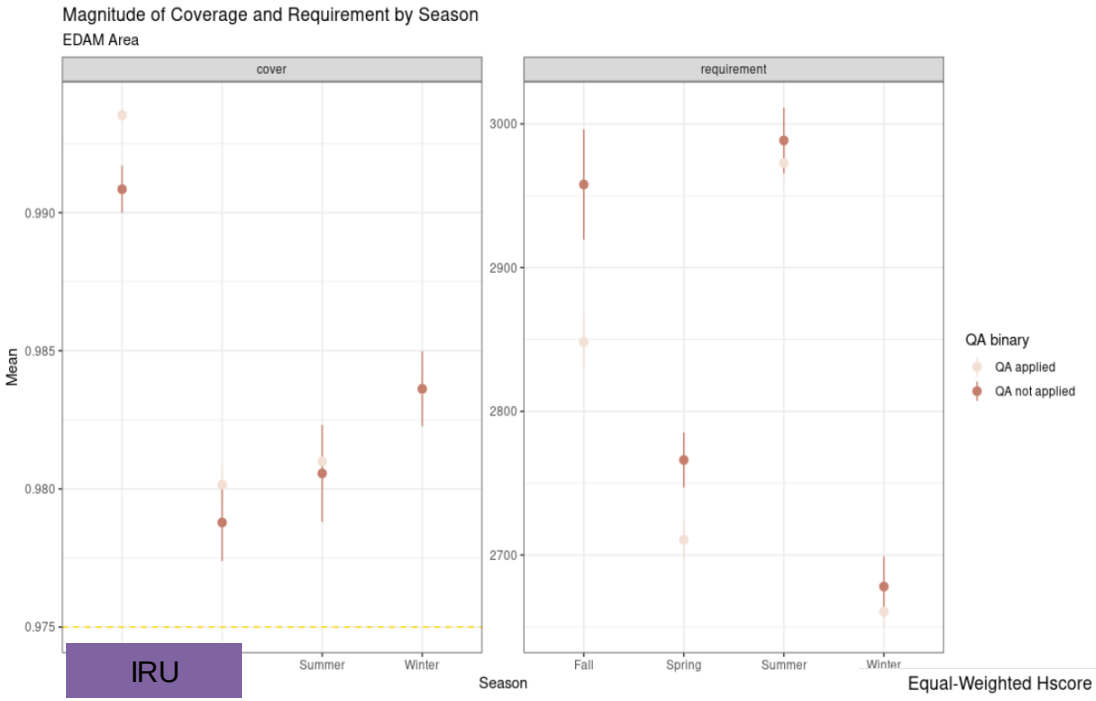
Upward coverage change vs upward exceedance change for simulation results without QA applied → with QA applied





IRU

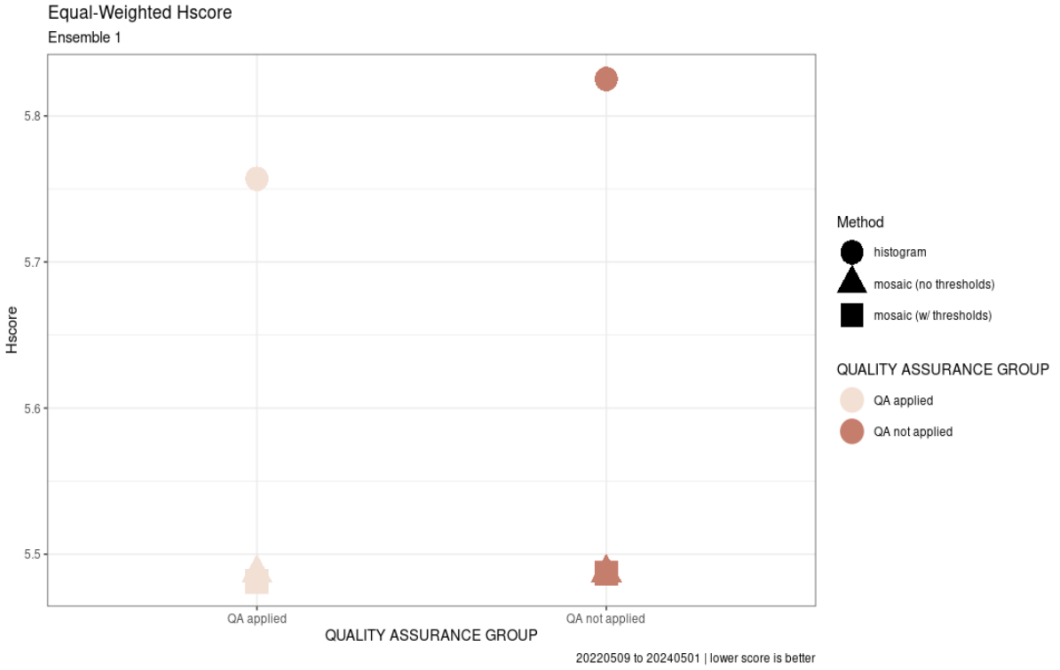
Ensemble 3



IRU

Coverage is increased across seasons while requirement is generally decreased

Hscore decreases when QA applied



BOTH

California ISO

Ensemble 3 / QA
Mosaic Parameter Recommendation

- Recommending change (Applying QA method)
  - As anticipated, most BAAs show small changes (commensurate with amount of values discarded). The benefit, however, is in a couple of BAAs as dramatically decreasing exceedance values and increasing coverage. Alternatively, this provides assurance that adverse data won't affect other BAAs in future scenarios.

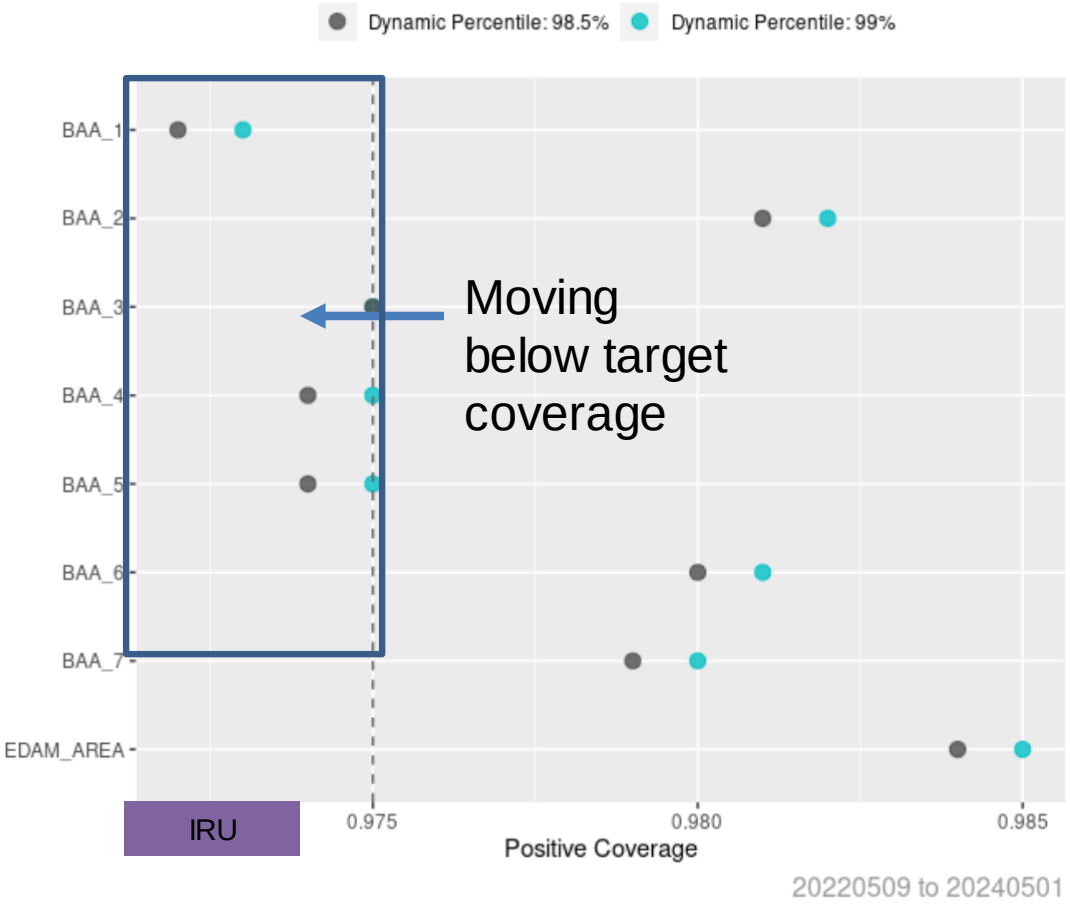
Ensemble 4

Lowering the dynamic threshold percentile reduces coverage, below target, for multiple BAAs

Upward coverage per BAA for dynamic threshold percentile values of 98.5% vs 99% at right.

Multiple BAAs see coverage reduced further below target with reduction in threshold percentile.

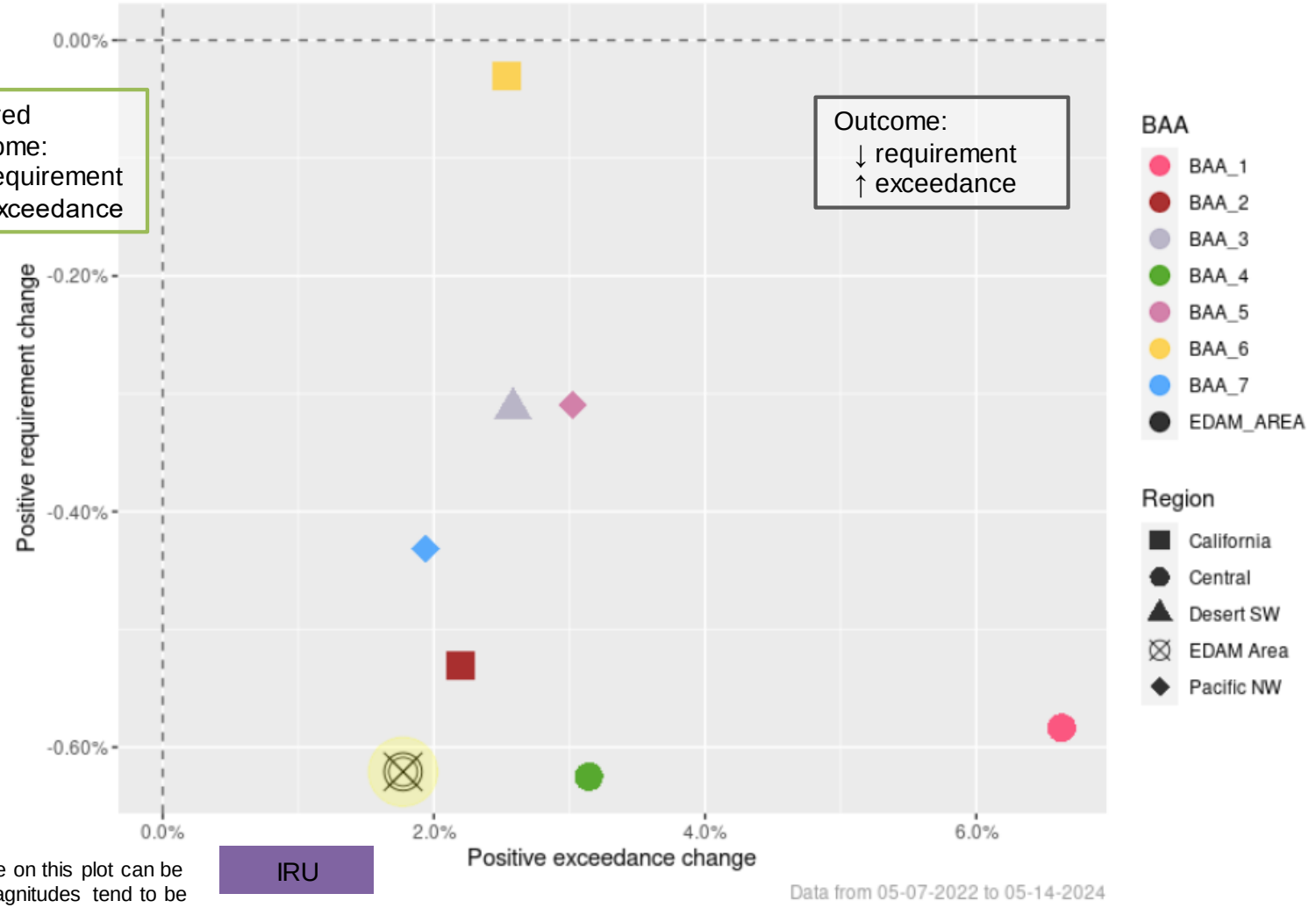
The following slide shows requirement and exceedance.



Ensemble 4

Mixed results show trade off in requirement and exceedance metrics when reducing threshold percentile

Change in requirement and exceedance for dynamic threshold percentile change from 99% → 98.5%.



Note: The scale of relative change on this plot can be misleading. Since exceedance magnitudes tend to be much smaller than requirement magnitudes, exceedance changes will appear proportionally larger.

Ensemble 4
Recommendation

Ensemble 4 / Dynamic Thresholds
Mosaic Parameter Recommendation

- Recommending no change (99% dynamic threshold)
  - Some BAAs starts to drop coverage below 97.5% target when relaxing
  - Most BAAs experience disproportionate increase in exceedance vs. requirement decrease

Focusing on rolling summaries for static threshold

- (1) *Period summary
  - BAA level
  - EDAM AREA
- **(2) Rolling or Seasonal summaries**
  - **BAA level**
  - **EDAM AREA**
- (3) Hscore
  - All BAAs are equally weighted

*2 year average

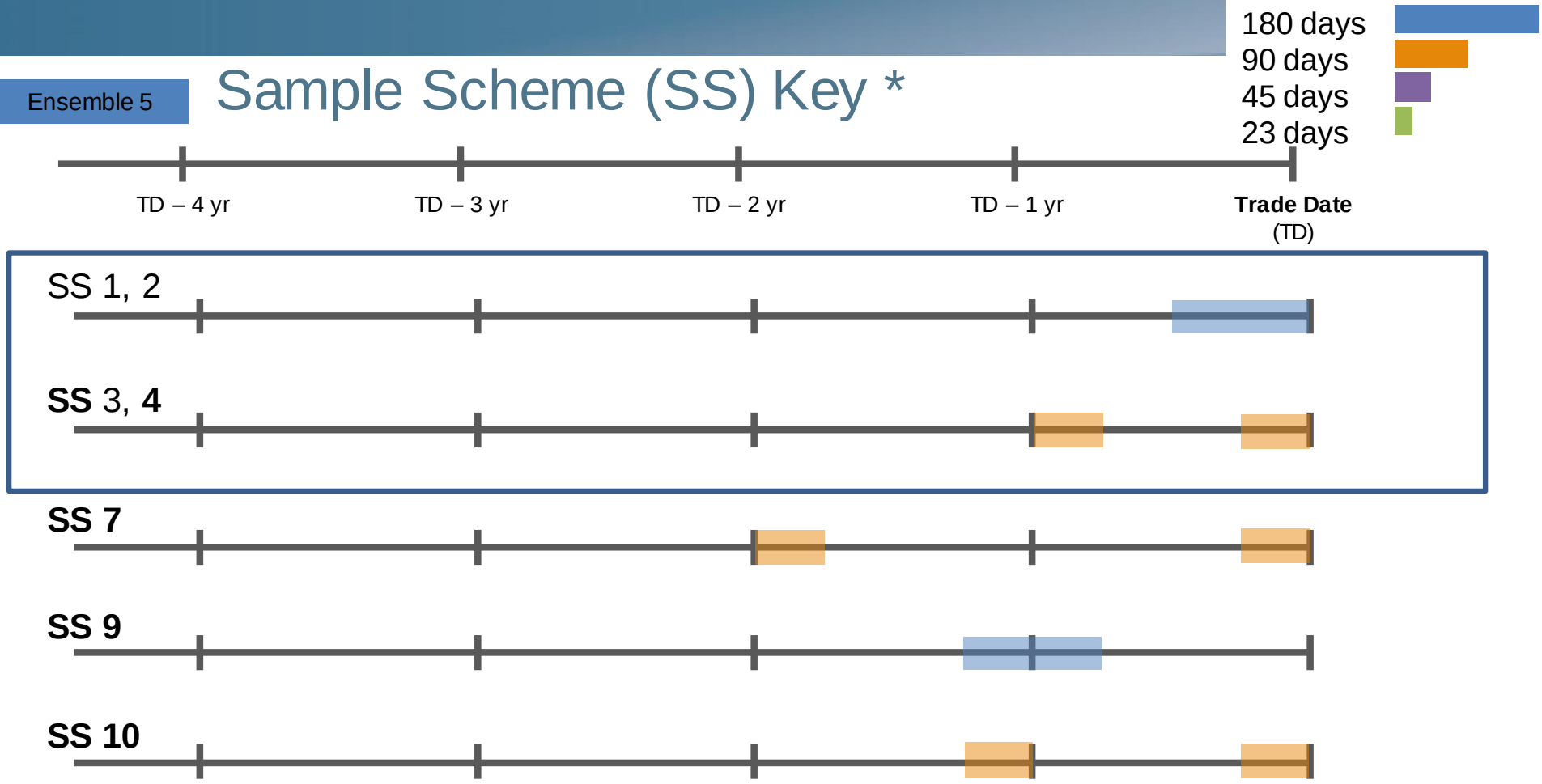
Hscore

- (H)ierarchical score
- Holistic performance metric
- Combination of coverage and requirement, plus consideration of time of day, requirement direction, and sample period length

Meetings and events > Market Surveillance Committee > Uncertainty Performance – Presentation – Apr 11, 2024

Ensemble 5

Sample Scheme (SS) Key *

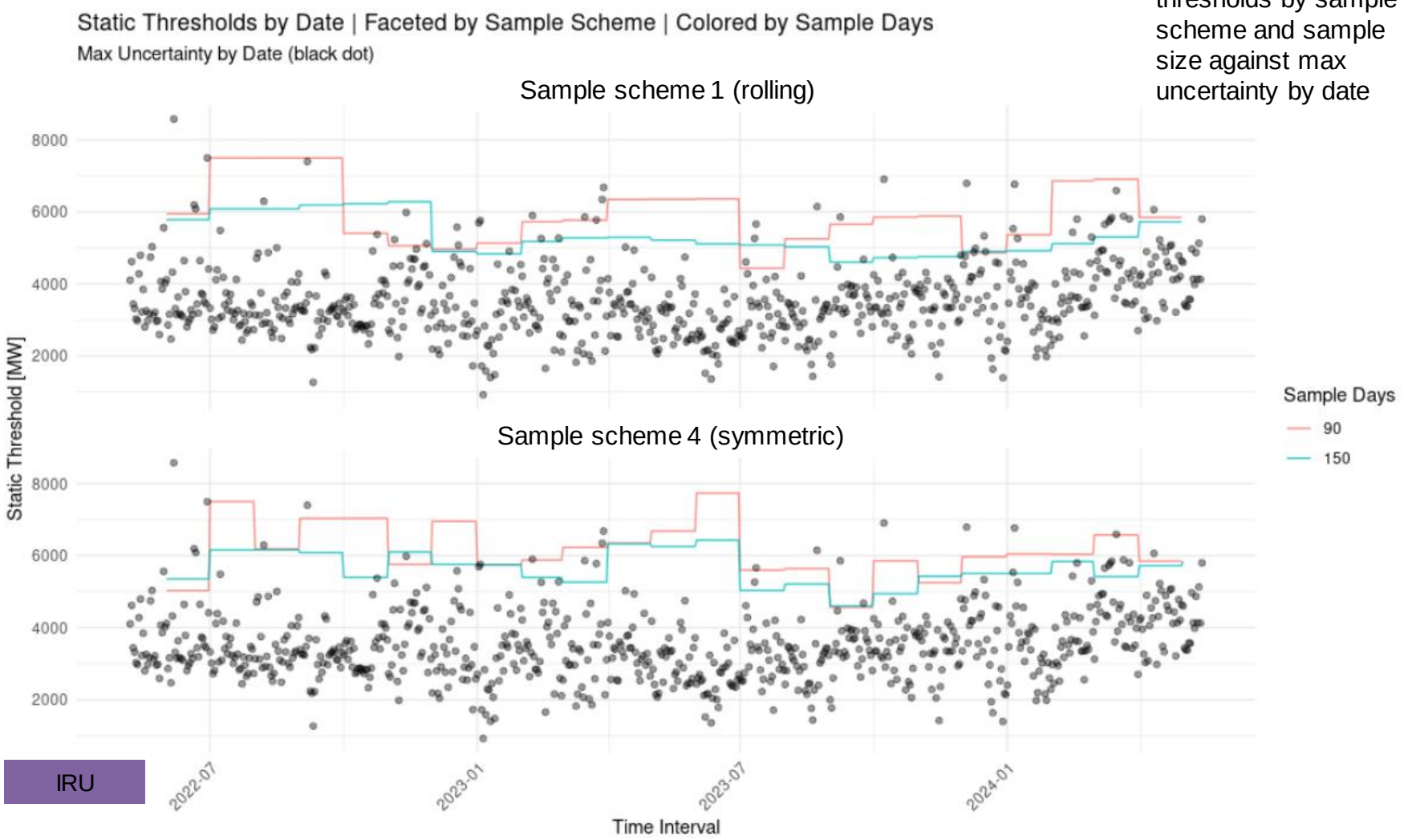


* Illustrative examples. Periods and timelines not to scale.

Ensemble 5

Larger sample size provides better stability and efficiency for static thresholds

Compare static thresholds by sample scheme and sample size against max uncertainty by date



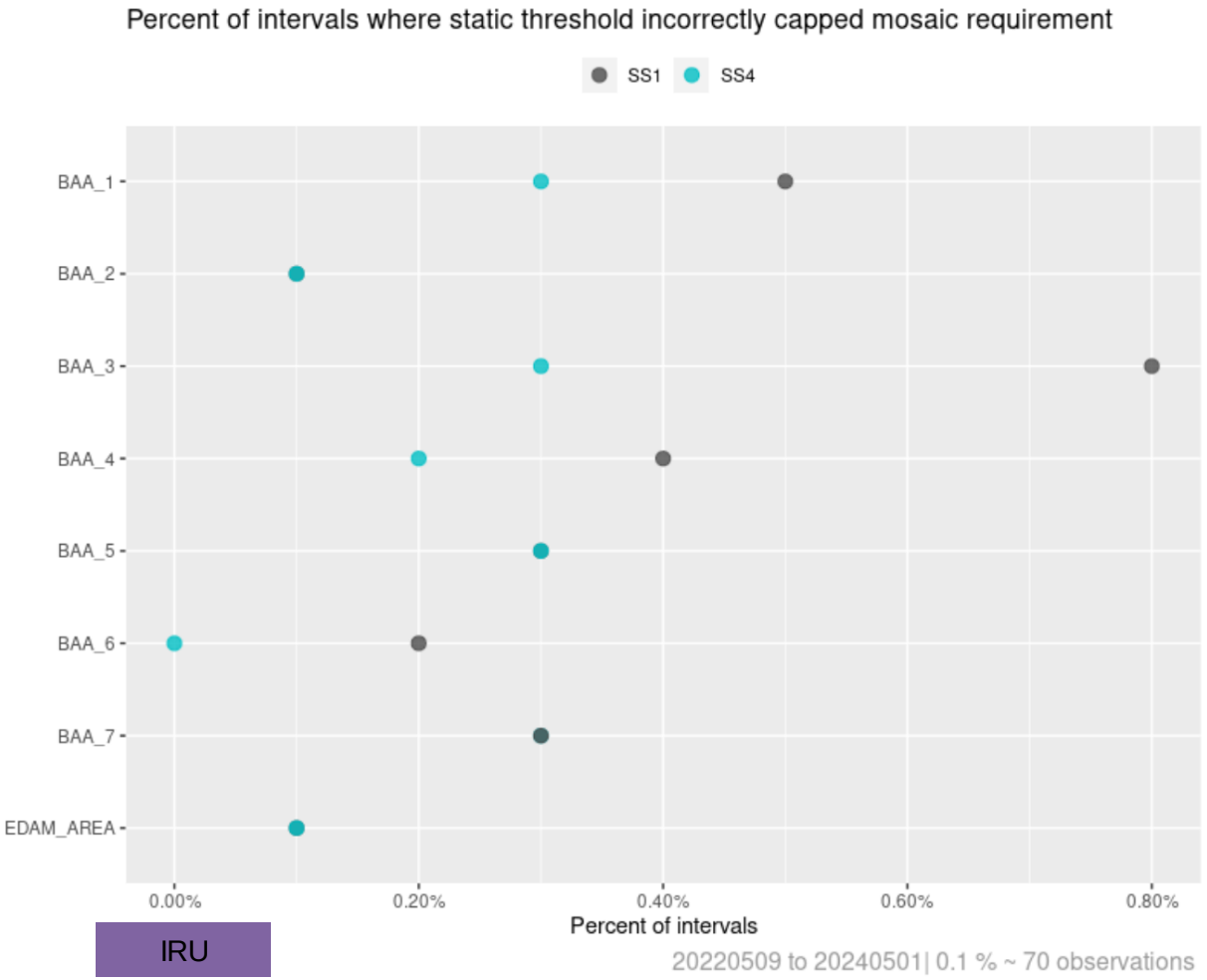
IRU

Ensemble 5

Symmetric sampling reduces static threshold use

Rolling sampling scheme (SS1) has more instances than symmetric sampling scheme (SS4) of the static threshold constraining requirements when realized uncertainty exceeds thresholds.

Static thresholds protect against data anomalies, but are purposefully conservative (99th percentile) to allow requirements in the tails of the distribution to cover realized uncertainty.



Ensemble 5 Recommendation

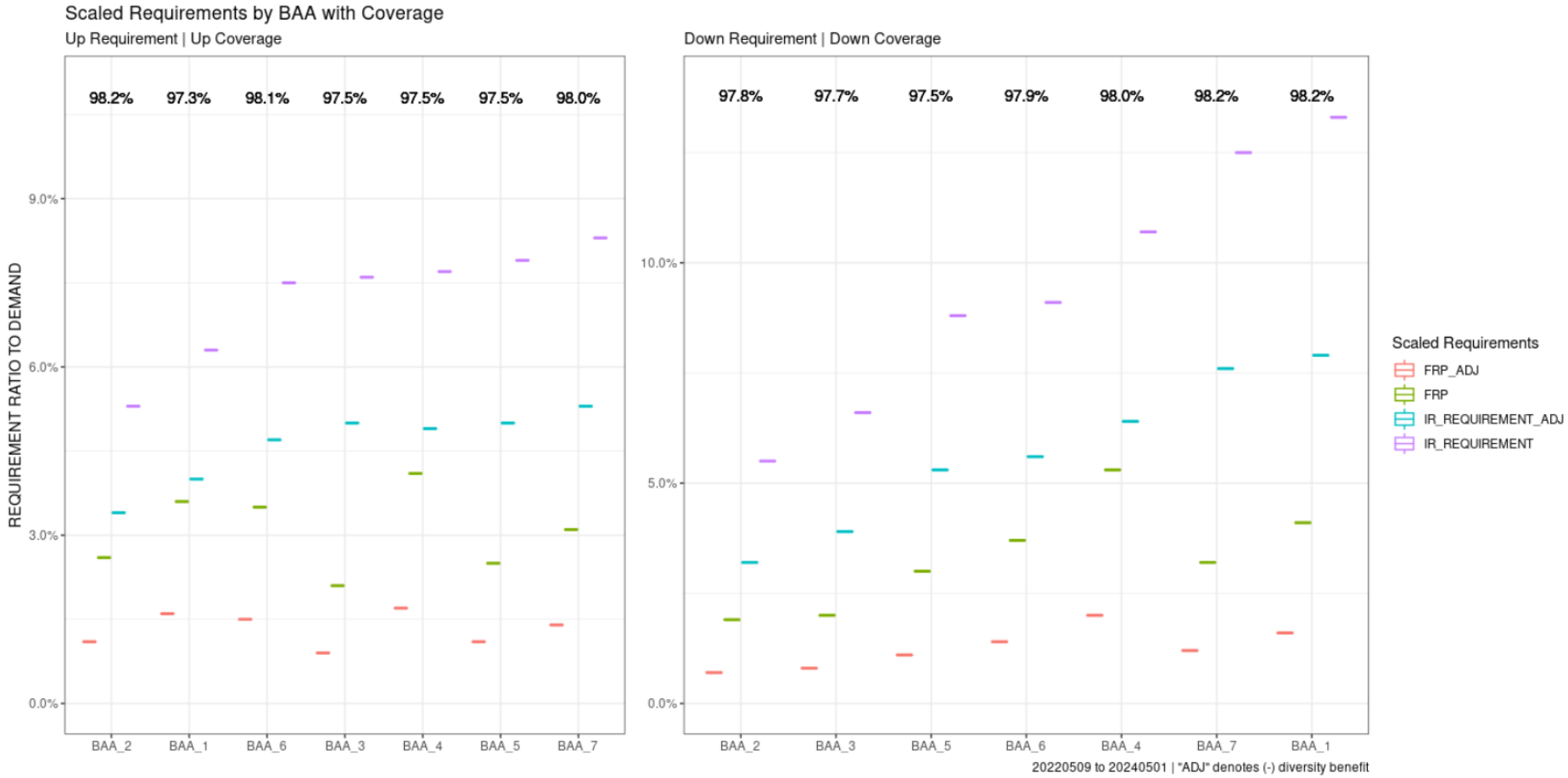
Ensemble 5 / Static Thresholds
Mosaic Parameter Recommendation

- Recommending change (SS4, 150)
  - Results show less volatility (month to month, quarter to quarter), especially within season changes
  - SS4 results in less overall threshold hits and about ½ as many hits where uncertainty exceeded static threshold

Final Recommendation

Mosaic Parameter	Order of Evaluation	Current Value	Final Recommendation	Items Evaluated
Historical Days / Split Window (Sample Scheme) November Call	1	Sample scheme 4 (symmetric)	**Sample scheme 4 (symmetric)**	Performance of IR with sampling schemes: 4, 7, 9 (5,10)
Historical Period Sample Days November Call	2	180	**150**	Performance of IR with samples days in the historical period: 150, 180, 210
Quality Assurance (QA) Binary February Call	*3*	QA not applied	**QA applied**	Performance of IR with QA applied and QA not applied
Dynamic Threshold Percentile February Call	*4*	99% (1%)	**99% (1%)**	Performance of IR with 99% (1%), 98.5% (1.5%), and 98% (2%) dynamic threshold percentiles applied
Static Threshold Sample February Call	*5*	90 days, sample scheme 1 [ss1,90]	**150 days, sample scheme 4 [ss4, 150]**	Performance of IR with [ss1,90], [ss1,150], [ss4,90], and [ss4,150] static threshold applied

Scaled Requirements and Coverage with Final Recommendation



QUESTIONS / SUGGESTIONS

SUPPLEMENTARY

DBSCAN

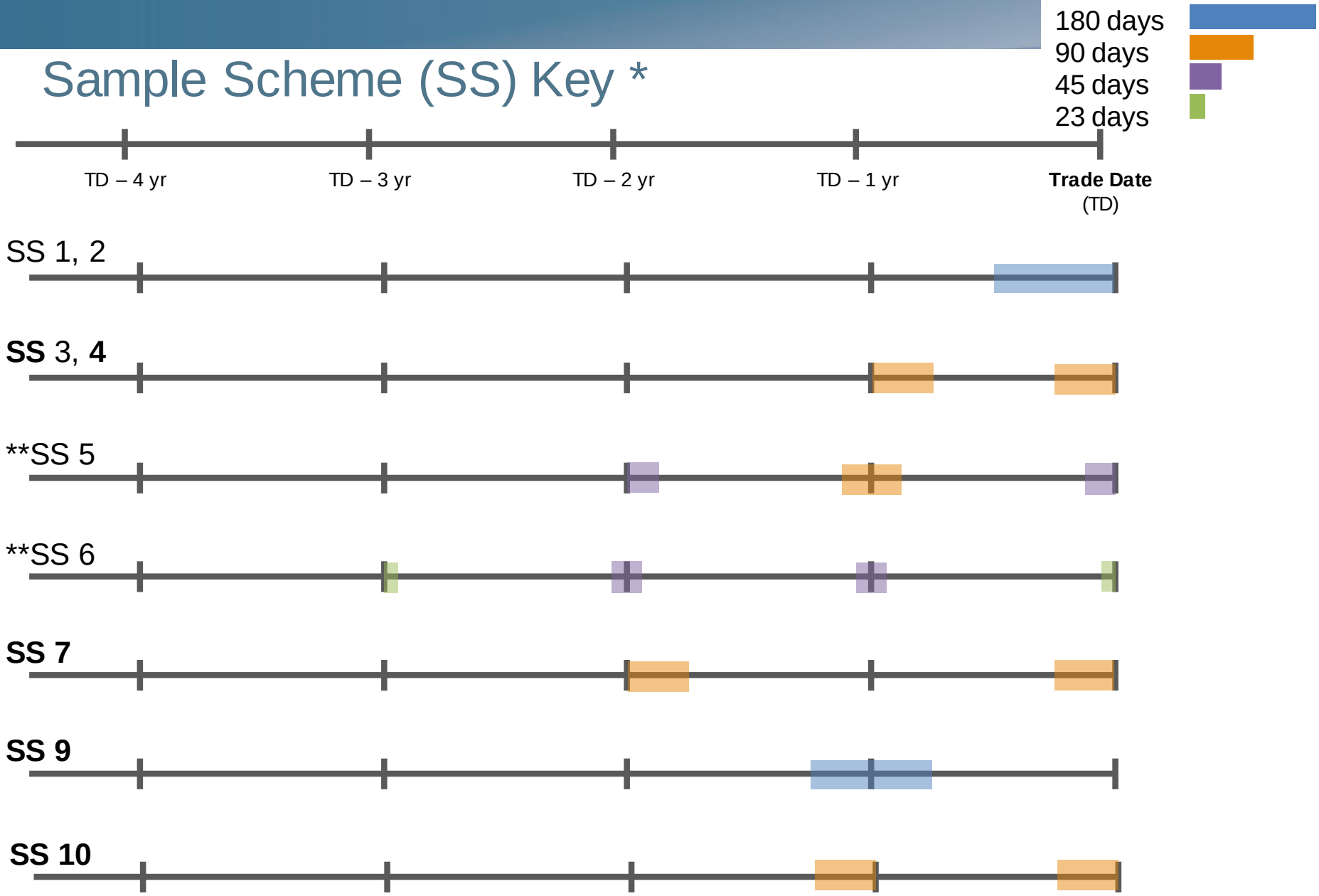
Why DBSCAN ?

- Was efficient for the amount of data provided
- Good at detecting anomalies
- Readily accessible algorithm

Main Parameters

- **Epsilon** – largest radius of neighborhood around a given point
- **MinPts** – Minimum number of points that constitute neighborhood

Sample Scheme (SS) Key *



* Illustrative examples. Periods and timelines not to scale. ** Future enhancement

Example Communication

Hello,

CAISO is preparing for a workshop regarding configurable parameters within the DAME/EDAM initiative. Part of that initiative will be to trial different sampling (longer sample periods) and performance to prepare for the IR deployment for all participating/interested entities. To assist in this we are requesting to receive data from your entity starting from 1/1/2020 - present day (to facilitate the study of longer sample periods).

Through this we will also be able to provide estimates of IRU and IRD for your entity.

You can find some details on similar analysis we have done regarding configurable parameters in FRP in the following stakeholder links as well.

Nov. 29, 2023 Market Surveillance Committee (MSC) presentation

Feb. 7 2024 Board of Governors presentation

The only requirement we have is to receive your hourly DA forecasts for demand, solar and wind in the attached format (solar and wind aggregations done prior to sending to CAISO and stacked "longways" – attached format should help calibrate). We will provide the FMM values for demand, solar and wind.

Thank you,

Format constraint must follow template

	A	B	C	D	E	F	G
1	Timestam	TRADE_DT	TRADE_HE	FORECAST	FORECAST	LEAD_IND	BAA_NAME
2	2021-01-0	1/1/2021	1	0	SOLAR	DA	BAA
3	2021-01-0	1/1/2021	2	0	SOLAR	DA	BAA
4	2021-01-0	1/1/2021	3	0	SOLAR	DA	BAA
5	2021-01-0	1/1/2021	4	0	SOLAR	DA	BAA
6	2021-01-0	1/1/2021	5	0	SOLAR	DA	BAA
7	2021-01-0	1/1/2021	6	0	SOLAR	DA	BAA
8	2021-01-0	1/1/2021	7	0	SOLAR	DA	BAA
9	2021-01-0	1/1/2021	8	92.98754	SOLAR	DA	BAA
10	2021-01-0	1/1/2021	9	404.9092	SOLAR	DA	BAA
11	2021-01-0	1/1/2021	10	452.752	SOLAR	DA	BAA
12	2021-01-0	1/1/2021	11	413.9338	SOLAR	DA	BAA
13	2021-01-0	1/1/2021	12	448.4426	SOLAR	DA	BAA

Energy Matters blog provides timely insights into ISO grid and market operations as well as other industry-related news.

https://www.caiso.com/about/news/energy-matters-blog

Story | Operations

ISO now has expanded training and emergency operations facilities

By Tricia Johnstone

02/20/2025

Story | Inside the California ISO

You can now follow the ISO on Bluesky

01/27/2025

Story | Operations

The growing importance of winter readiness and the outlook for 2025

By Dede Subakti

12/20/2024

Subscribe to Energy Matters blog monthly summary