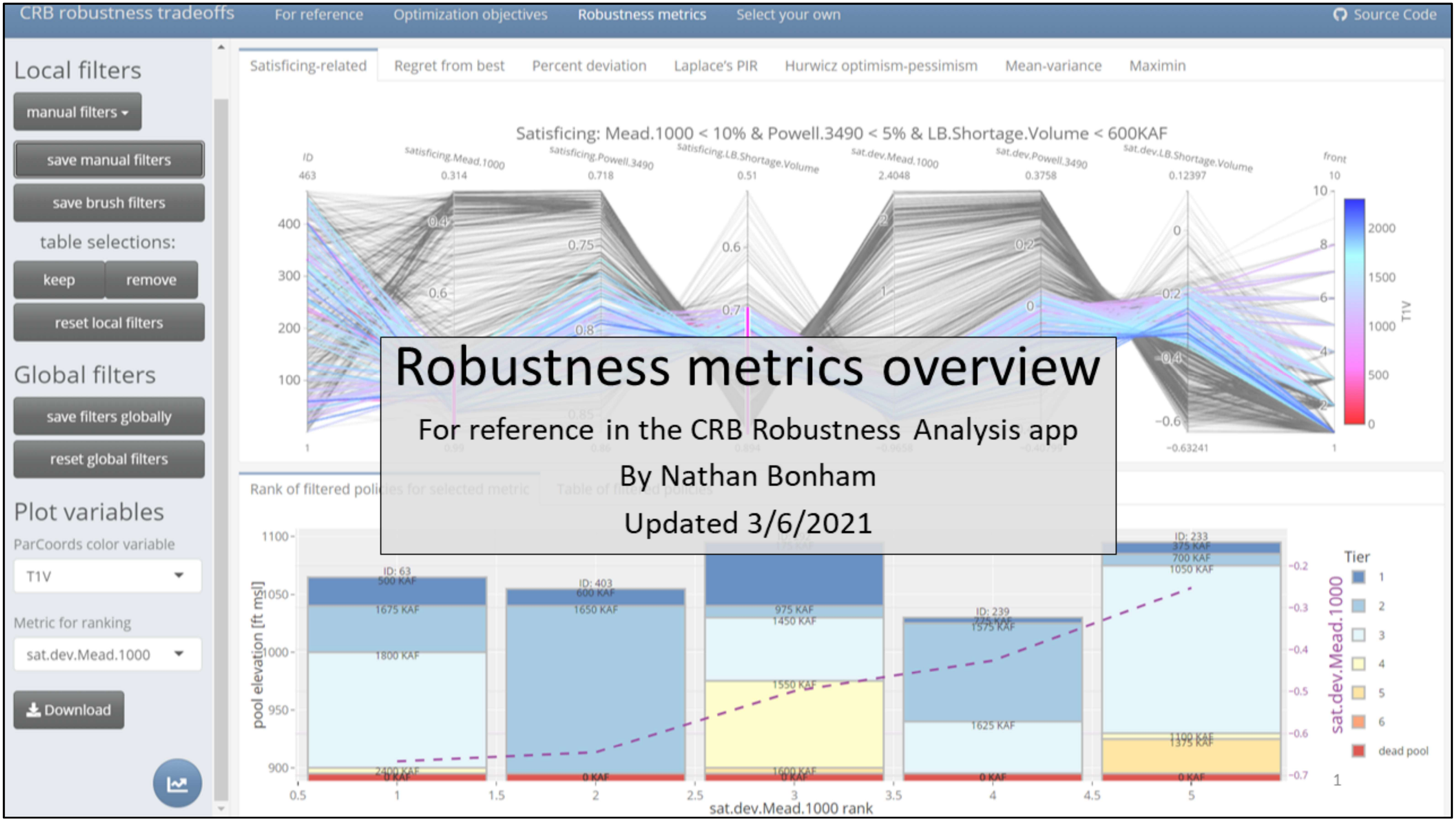
# Robustness metrics overview

For reference in the CRB Robustness Analysis app

By Nathan Bonham

Updated 3/6/2021




We recommend downloading this document and having it open alongside the app for reference.

# Contents

- Define robustness
- How robustness is assessed in the Colorado River Basin with Many Objective Robust Decision Making
- How robustness metrics are calculated
- How to choose robustness metrics

# What is robustness?

- A robust solution “perform(s) well under a range of plausible conditions” (McPhail et al. 2018)
- In the Colorado River Basin (CRB) context, we desire operation policies that perform ‘well’ in water supply and storage objectives when tested in many plausible realizations of hydrology and demand.

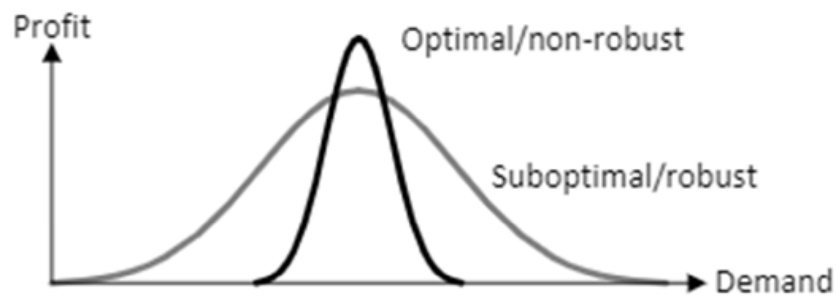



Fig. 1. Illustration of robustness.

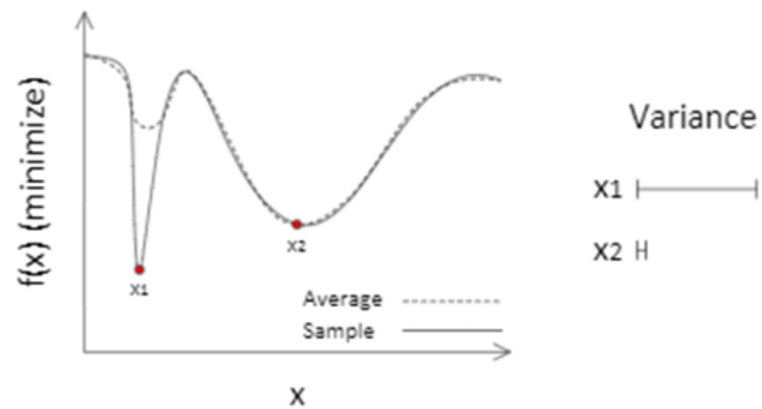
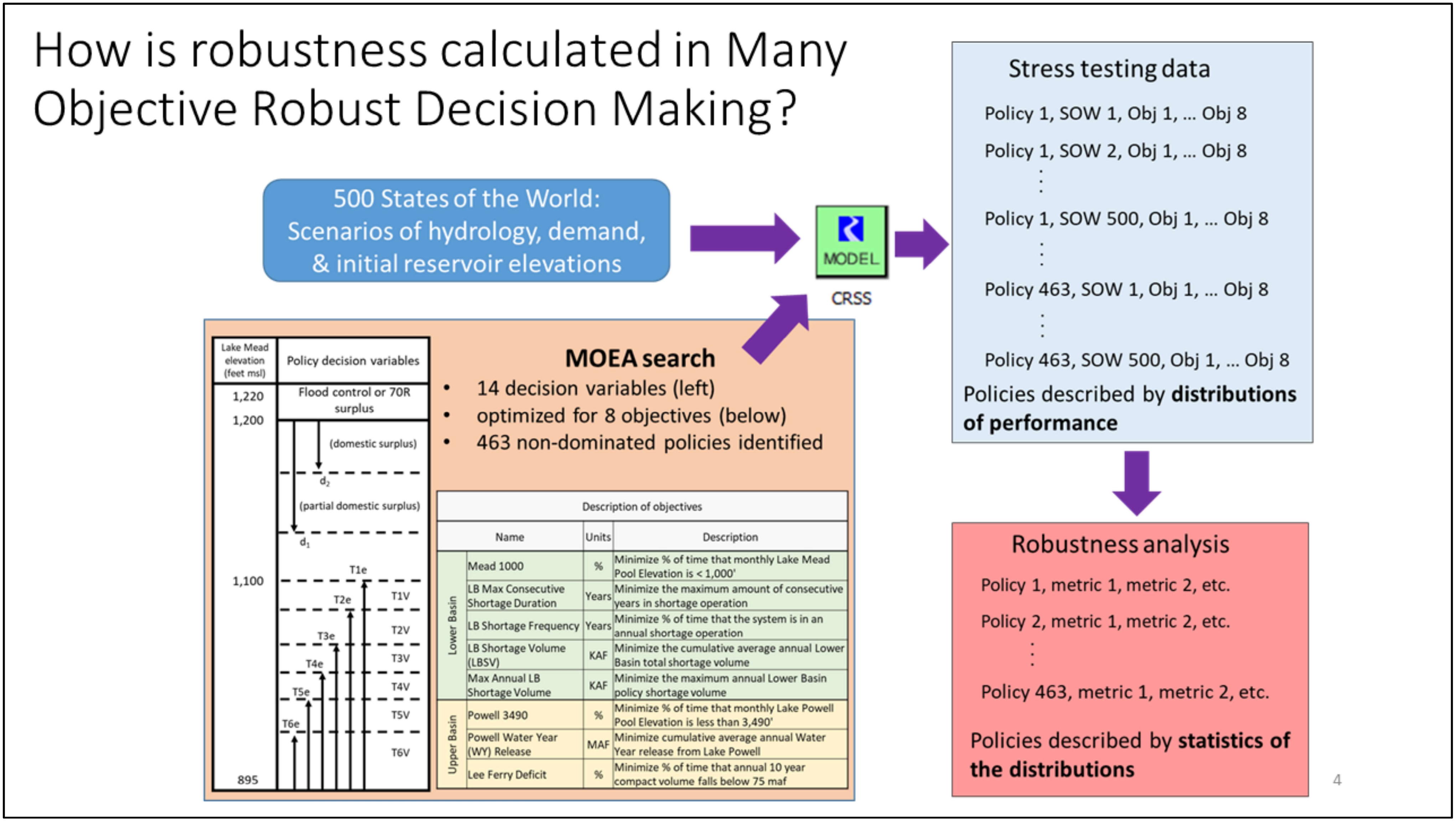



Fig. 2. Example of a sample versus a robust optimal solution for a function $f(x)$.

Figures from Syberfeldt and Gustavsson 2014

3

Both figures are examples where the globally optimal solution is highly sensitive to changes in another variable. In contrast, the performance of the robust solution suffers less when the uncertain variable changes.




Description of objectives

| | Name | Units | Description |
|---|---|---|---|
| Lower Basin | Mead 1000 | % | Minimize % of time that monthly Lake Mead Pool Elevation is < 1,000' |
| Lower Basin | LB Max Consecutive Shortage Duration | Years | Minimize the maximum amount of consecutive years in shortage operation |
| Lower Basin | LB Shortage Frequency | Years | Minimize % of time that the system is in an annual shortage operation |
| Lower Basin | LB Shortage Volume (LBSV) | KAF | Minimize the cumulative average annual Lower Basin total shortage volume |
| Lower Basin | Max Annual LB Shortage Volume | KAF | Minimize the maximum annual Lower Basin policy shortage volume |
| Upper Basin | Powell 3490 | % | Minimize % of time that monthly Lake Powell Pool Elevation is less than 3,490' |
| Upper Basin | Powell Water Year (WY) Release | MAF | Minimize cumulative average annual Water Year release from Lake Powell |
| Upper Basin | Lee Ferry Deficit | % | Minimize % of time that annual 10 year compact volume falls below 75 maf |

We provide an example of how robustness is assessed in the Many Objective Robust Decision Making framework using our Colorado River Basin (CRB) case study. 463 operation policies are each simulated in 500 plausible states of the world (SOW) using the Colorado River Simulation System (CRSS). The operation policies were identified in previous research using the Borg Many Objective Evolutionary Algorithm coupled with CRSS. The result of simulating all policies in all SOW is that each policy is described by distributions of performance in the objectives. Robustness analysis is the process of summarizing the distributions of performance as scalar values by applying one or more statistical functions. The statistical functions are called robustness metrics. The figure describes the process of evaluating policy robustness, but keep in mind that copious robustness metrics exist and the analyst should consider how they are calculated and what the implications might be on robustness magnitude and ranking of the policies.

***Note: LB stands for Lower Basin

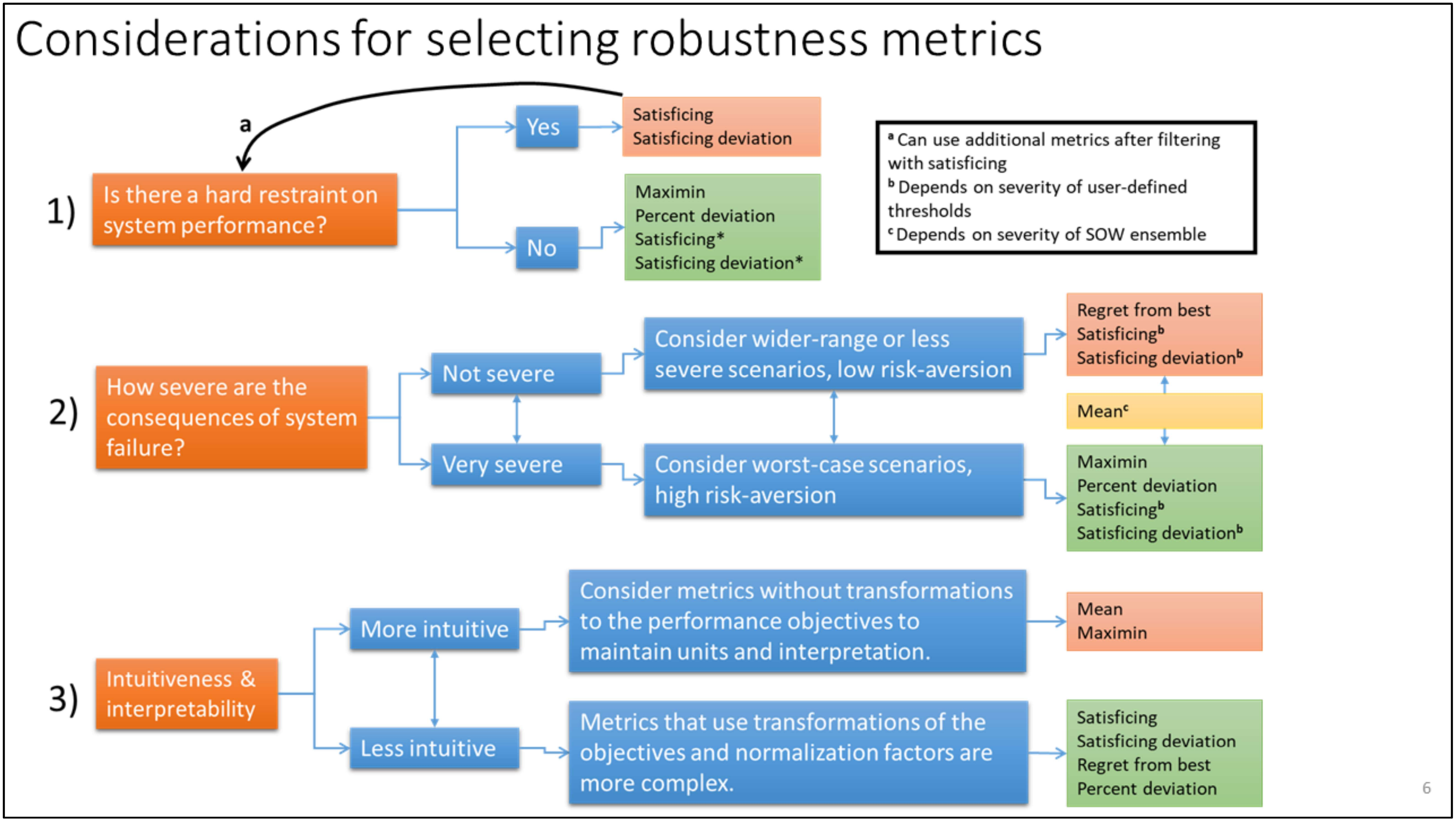
# Robustness metric summary table

| Description | | | | Calculation | | | |
|---|---|---|---|---|---|---|---|
| **Category** | **Name** | **Definition** | **Interpretation** | **Transformation of objective** | **SOW used** | **Normalization factor** | **Summary statistic** |
| threshold | Satisficing | The fraction of SOW where a given policy satisfies user-defined performance thresholds. | 1 indicates the performance thresholds have been satisfied in every SOW, whereas 0 indicates the thresholds are violated in every SOW. | Satisfactory (1) or Not Satisfactory (0) | All | None | mean |
| threshold & regret | Satisficing deviation | The average percent by which a policy satisfies a user-defined performance threshold. | Negative percentages indicate performance better than the threshold, and positive indicates performance worse than the threshold. | Deviation from satisficing threshold | All | Threshold | mean |
| regret | Regret from best | The average percent by which a solution's performance in each SOW deviates from the best performance obtained by any policy in each SOW. | 0 percent indicates the policy is the best performing policy in every SOW, and greater positive values indicate larger average regret compared to the best performing policies. | Deviation from best performance by any policy in each SOW | All | Global range of performance observed by all policies, unique to each SOW | mean |
| | Percent deviation from optimization | The percent by which the 90th percentile performance of a solution deviates from its performance during optimization. | Positive percentages indicate the policy performs worse during robustness simulations compared to optimization simulations, whereas negative values means the policy performs better (mathematically possible, but does't occur in our analysis). | Deviation from optimization performance | 90th percentile | Optimization performance | None |
| No objective transform | Mean (Laplace's Principle of Insufficient Reason) | The performance averaged over the SOW ensemble. | The units and interpretation of each performance objective are maintained (e.g. smaller values are desired for minimization objectives) | None | All | None | mean |
| | Maximin | The worst performance obtained by a solution in the SOW ensemble. Alternatively, other percentiles can be used (95th or 90th percentile performance, for example). | The units and interpretation of each performance objective are maintained (e.g. smaller values are desired for minimization objectives) | None | Worst-case (or user-defined percentile) | None | None |

**See the app's "Example calculations" tab for example calculations for each metric in Google Sheets**



This table serves as a "cheat sheet" for interpreting the robustness metrics supported in the app. The table is divided into the blue section (left), which provides practical information for understanding what each robustness metric is telling you, and the green section (right), which provides details about how the metrics are calculated. The "Transformation of objective" column indicates what, if any, mathematical transformation is performed to the performance objectives. Then, "SOW used" explains if all or a subset of the States of the World (SOW) are used in the calculation. Next, "Normalization factor" defines the denominator used to convert the metric to a percentage or fraction. Lastly, "Summary statistic" lists the statistical moment used to describe the distribution of performance as a scalar value, in the case more than one SOW is used. The app's "Example calculations" tab provides example calculations for each metric in Google Sheets, and it is helpful for understanding what each of the columns in the "Calculation" section means.




This diagram provides three considerations for which robustness metrics should be used in an analysis, and provides a road map for decision makers to translate their preferences into robustness metrics. First, if hard restraints on system performance have been identified, then satisficing-based robustness metrics are appropriate. However, additional metrics should be considered after filtering policies by satisficing and/or satisficing deviation. Second, if the consequences of system failure are severe, use metrics that are calculated with worst-case SOW or using severe performance thresholds (McPhail et al. 2018). Lastly,  metrics that use transformations to the performance objectives and normalization factors are inherently more complex than mean and maximin, which maintain the original units and meaning as the performance objectives.

# References


- McPhail, C., Maier, H.R., Kwakkel, J.H., Giuliani, M., Castelletti, A., Westra, S., 2018. Robustness Metrics: How Are They Calculated, When Should They Be Used and Why Do They Give Different Results? Earths Future 6, 169–191. https://doi.org/10.1002/2017EF000649
- Syberfeldt, A., Gustavsson, P., 2014. Increased Robustness of Product Sequencing Using Multi-objective Optimization. Procedia CIRP 17, 434–439. https://doi.org/10.1016/j.procir.2014.01.141

# Please email me with questions and comments

Nathan.Bonham@Colorado.edu