Comparing Metrics for Scenario-based Robustness Assessment of Building Performance

Linus Walker, Alexandra Kuhn, Illias Hischier, Arno Schlueter
Chair of Architecture and Building Systems, Institute for Technology in Architecture, ETH Zurich, Stefano-Franscini-Platz 1, 8093 Zurich, Switzerland
E-mail: walker@arch.ethz.ch

Abstract. To decrease greenhouse gas emissions of the Swiss building stock, effective retrofit strategies are necessary. Due to the long-term operation of buildings, future developments and uncertainties need to be considered, which calls for assessing the robustness of retrofit decisions. Existing studies propose robustness metrics for decisions under deep uncertainty to be coupled with a scenario-based simulation approach. We review these metrics and present a simulation approach that includes current and future operational energy, emissions, and costs. We apply the seven identified metrics to retrofit decisions of a multifamily house located in Zürich, where future scenarios in terms of climate, occupancy, decarbonization, and cost development are included. The metrics are based on different assumptions and positions towards risk. We further find that the discriminatory power is different, confirming the Minimax Regret metric to be most suitable for the building context when looking at individual buildings. For the case study, we find that deep retrofit seems to be a robust decision from an environmental perspective. From a cost perspective, the electrification of the heating system with heat pumps and the installation of PV without a complete envelope retrofit proves to be most robust.

1. Introduction
Reaching Swiss greenhouse gas emission reduction goals for the building sector requires effective renovation strategies [1]. Due to the energy performance gap (EPG), the building stock might not be decarbonised as expected [2]. Uncertain boundary conditions such as occupant behaviour, climatic conditions, and characteristics of energy system components lead to inaccurate predictions of operational emissions and decarbonization pathways. To account for such uncertainties, robustness assessment in the design stage is required. Robust performance means "the ability of a building to perform effectively and remain within acceptable margins under the majority of possible changes in internal and/or external environments." [3] Further, it is clear that the embodied impact also needs to be included in such an assessment [4]. In research, robustness assessments are carried out primarily using probabilistic or scenario-based approaches. Especially in the context of deep uncertainty, where probability distributions of future developments are unknown, a non-probabilistic, scenario-based approach is preferable [5]. This article aims to summarize and discuss the current state of scenario-based robustness assessment of buildings, identify suitable metrics, and present a case study where those metrics are applied.
2. Review of Scenario-based Methods and Metrics

In this work, we review studies that use scenario-based robustness assessment for building performance assessment in terms of decision-making goals and the case study context. Kotireddy et al. reviewed studies on building performance robustness assessment [6]. The main uncertainties analyzed in their reviewed articles were based on occupant behavior, climate, uncertainty in design parameters, and techno-economic parameters. They find that the robustness metrics mostly applied in scenario analysis are the Spread and the Minimax Regret. At the same time, Deviation and the Hurwicz Criterion are also used specifically in a building performance context. In another study, Kotireddy et al. apply the Minimax Regret metric because they say this is the ideal metric when the decision-maker is willingly accepting risk as a trade-off for better performance [5]. Gang et al. use Minimax Regret for the optimal design of building cooling systems. They claim that this metric is suitable because it is easy to implement and to understand by decision-makers [7]. Homaei and Hamdy review existing robustness metrics and compare them. They introduce their own metric, which considers performance targets and robustness margins. Their metric is built on multiple performance indicators, which are combined into a single decision metric [3]. Nik et al. investigate the robustness of envelope retrofitting measures using a scenario-based approach, using mean and standard deviation as robustness metrics [8, 9]. Rysanek and Choudhary investigate optimal building retrofits based on technical and economic scenarios also using the Minimax Regret metric [10]. Piccard et al. investigate the robust performance of net-zero energy buildings based on scenarios, but they do not use specific metrics. Instead, they use graphics for a visual assessment [11]. Further, a systematic review of scenario-based robustness assessment was carried out by Mc Phail et al. for infrastructure decisions [12]. In their analysis, they rank several metrics according to the respective inherent risk aversion. The metrics applied in the reviewed studies, and three additional metrics that we identify to potentially be suitable because they require little input from the decision-maker are described in Table 1. In addition, the basic properties of the metrics are compared in Table 2.

Table 1: Identified scenario-based robustness metrics used in literature

| Robustness Metric Definition                                                                 |
|---------------------------------------------------------------------------------------------|
| **Spread**                                                                                 |
| The Spread is calculated for each configuration by taking the difference between best and the worst performance over all scenarios. The configuration with the minimum Spread is then considered to be the most robust one. |
| **Deviation**                                                                              |
| The Deviation metric is very similar to the Spread. However, the difference is calculated between the worst-case value for each configuration and the best-case value over all configurations and scenarios. |
| **Minimax Regret**                                                                         |
| The Minimax Regret metric aims at minimizing possible regret. Therefore, at first, the performance matrix is transformed into a regret matrix. In a second step, minimax is applied to the regret matrix. |
| **Taguchi and other combinations of mean and standard deviation**                           |
| Here, the Taguchi metric is calculated by addition of the mean and standard deviation. The smallest value is considered the most robust one. |
| **Maximin**                                                                                |
| The Maximin metric is derived by identifying the worst-performing scenario for each configuration and then choosing the configuration with the best worst-case performance. |
| **Laplace’s Principle of Insufficient Reasoning**                                           |
| Equal probabilities for all scenarios are assumed, and the average performance is calculated. |
| **Starr’s Domain Criterion**                                                               |
| Here, the decision-maker has to define a performance acceptance threshold. The performance matrix is then transformed into a boolean matrix. For each configuration, the mean acceptance ratio is calculated. |
Table 2: Properties of robustness metrics

| Method          | Evaluates the absolute performance | Inherently pessimistic | Minimizes Variability | Compares all configurations against threshold | Needs decision maker input |
|-----------------|------------------------------------|------------------------|-----------------------|-----------------------------------------------|----------------------------|
| Spread          | no                                 | no                     | yes                   | no                                            | no                         |
| Deviation       | no                                 | no                     | yes                   | no                                            | no                         |
| Minimax Regret  | no                                 | yes                    | -                     | no                                            | no                         |
| Maximin         | yes                                | yes                    | -                     | no                                            | no                         |
| Taguchi         | yes                                | no                     | yes                   | no                                            | no                         |
| Laplace         | yes                                | -                      | no                    | no                                            | no                         |
| Starr           | yes                                | -                      | yes                   | yes                                           | yes                        |

3. Methods and Materials

3.1. Overview

For our own case study, the robustness of different renovation strategies is assessed by applying the identified robustness metrics on a multi-family house located in the city of Zurich. We assess the robustness in terms of greenhouse gas emissions (GHG), eco-points (UBP), and cost. Figure 1 shows the workflow of the analysis and Table 3 the basic building description.

![Robustness assessment workflow](image)

3.1.1. Configurations

A set of choices for each variable is called a configuration. All variables that a system designer can willingly choose, such as heating systems or thermal insulation, are considered part of a configuration. The configuration choices are shown in Table 3.

3.1.2. Scenarios

The scenarios for which the robustness of the configurations should be assessed are defined. This includes parameters of user behavior (heating set-points, shading factors, and ventilation), climatic conditions (weather files), parameters of technical matter (efficiency and U-value deviations), and parameters that influence the life cycle performance (lifetime variability, electricity decarbonization, energy cost, rate of interest). The future development of said parameters is based on expert predictions for 2050 that were collected in [13]. The scenario parameters are listed in Table 3.

3.1.3. System Models

**Heating and Cooling Demand:** An hourly, single-zone thermal simulation is carried out using an RC model to assess the building’s heating and cooling demand [14, 15]. The model is initially based on ISO 13790 and has a 5R1C configuration.

**Electricity and DHW Demand:** SIA 2024 provides quantities and schedules for electricity and domestic hot water demand.
Table 3: Case study info

| Building Description | Configuration Choices | Scenario Parameters |
|----------------------|-----------------------|---------------------|
| parameter | value | heating system | climate | current, 2050 RCP2.6 |
| footprint | 412 m² | air source heat pump | heating setpoint | 19°C, 23°C |
| wall area | 3300 m² | ground source heat pump | shading factor | 0.5, 1.0 |
| window area | 402 m² | wood | infiltration add on | -20%, +20% |
| roof area | 144 m² | pellets | thermal bridges | -25%, +25% |
|                  |         | natural gas | conversion efficiency | -20%, +20% |
|                  |         | district heating | PV technology | m-Si 2020, m-Si 2050 |
|                  |         | PV installation | envelope lifetime | -30%, +70% |
|                  |         | envelope renovation | system lifetime | -30%, +30% |
|                  |         | ground floor insulation | energy cost | current, NEP 2050 |
|                  |         | wall insulation | interest rate | 0.0%, 3.0% |
|                  |         | window renovation | all above | |

Energy Supply Models: Combustion systems are modeled using a predefined combustion efficiency. Heat pumps are modeled using a dynamic COP calculation according to the exergetic efficiency and the required temperature lift. PV yield is calculated based on photovoltaic efficiency and a performance ratio. The solar incident is simulated using pvlib, which uses a three-component irradiation model based on the location and weather file data [16].

Impact Modeling: To estimate the environmental impact, the embodied and operational impact are assessed. We consider life cycle stages A1-3 (production), B4 (replacement), B6 (operation), and C1-4 (disposal) according to EN 15978:2011. Materials not related to the building systems are omitted. For PV, the full embodied emissions are allocated to the building, and no negative emissions are accounted for when feeding electricity into the grid.

Cost Modeling: The total annual costs are calculated based on the annuity method [17]. Current and future cost data were collected from multiple sources and compiled in [13]. To assess PV from a conservative perspective, it is assumed that overproduction is supplied to the grid for free.

3.1.4. Robustness metric calculation
Using the simulated performance matrix, the robustness of a configuration is calculated for the metrics identified in Section 2. For Starr’s Domain Criterion, the chosen thresholds for GHG emissions, UBP, and cost are $7 \text{kg}^{-1} \text{CO}_2 \text{eq} \text{m}^{-2} \text{a}$, 15000 UBP m² a⁻¹, and 21 CHF m² a⁻¹.

4. Results and Discussion
4.1. Case Study Results
For the multifamily house case study described in Section 3.1 and Table 3, Figure 2 shows results regarding cost and environmental impact, including the variations for the different scenarios. It is interesting to see that heating systems powered by fossil fuels seem viable in terms of cost under current conditions but not under future conditions. Pure electric heating, which means direct conversion from electricity to heat, shows a bad environmental performance under current conditions and a poor cost performance under future conditions. However, heat pump systems are generally located in the low-impact/low-cost area with relatively little variation, which means that they behave robustly for both criteria. In Table 4, the most robust configurations according to the metrics are shown. For GHG emissions, most metrics identify a GSHP, installed PV, and full renovation as the most robust retrofit decision. When considering UBP, all metrics identify district heating, PV, and a full renovation to be most robust. In terms of cost, heat pump
systems with PV installation perform robustly without major envelope renovation measures for all metrics.

4.2. Comparison of Robustness Metrics

If we only look at the single most robust choice according to each metric, we can see that they mostly agree with some minor deviations. However, the discriminatory power varies among the different metrics. Figure 3 visualizes this by showing the number of configurations of the respective heating systems, where the robustness metric value is less than 10% worse than the most robust one. Minimax Regret, Starr, and the Spread tend to have a high selectivity, while Taguchi and Laplace seem to have less discriminatory power. High discriminatory power is beneficial for decision-making because it gives a clearer picture of which configurations perform most robustly.

![Figure 2: Scatter plot of total annualized cost and emissions with variation bars showing the 75% percentile grouped by heating systems with current and 2050 scenarios.](image)

| Metric       | GHG Emissions | UBP   | LCC   |
|--------------|---------------|-------|-------|
| Spread       | pellets yes all | district yes all | ASHP yes none |
| Deviation    | GSHP yes all | district yes all | GSHP yes none |
| Minimax Regret | GSHP yes wall | district yes all | GSHP yes floor |
| Taguchi      | GSHP yes all | district yes all | GSHP yes none |
| Maximin      | GSHP yes all | district yes all | GSHP yes none |
| Laplace      | GSHP yes wall | district yes all | GSHP yes none |
| Starr        | pellets yes all | district yes all | GSHP yes none |

5. Conclusion & Outlook

Scenario-based robustness assessment is considered a suitable method for assessing future performance under deep uncertainty. Based on literature, we identify a range of robustness metrics that have been successfully applied in research, some of them in the building energy domain. They differ, especially in their sensitivity to positive and negative outliers and the inherent risk aversion and their selectivity. Nonetheless, we observe that there are only minor disagreements between the metrics and higher disagreements between the different performance indicators, GHG emissions, UBP, and cost, for the presented case study. Our findings confirm that the Minimax Regret metric is most suitable for decision-making in the building energy domain.
Figure 3: Selectivity of the different robustness metrics exemplified for the heating system choice according to GHG emissions and cost. All options within 10% robustness value deviation of the most robust choice are included. There are a total of 60 configurations assessed.

context. Using the Spread itself is not recommended without a further metric that includes absolute performance, neither is the Deviation. Because of their low selectivity, Laplace and Taguchi seem more suitable to assess the general trends caused by a scenario but less valuable for analyzing a single building. Starr is dependent on user inputs and can therefore not easily be classified. To conclude on a more general level, representative buildings of the Swiss building stock need to be assessed in a next step. Furthermore, multi-criteria analysis can be applied to combine the environmental and cost dimension.

References
[1] Cozza S, Chambers J and Patel M K 2020 Energy Policy 137 111085 ISSN 0301-4215
[2] Frei B, Sagerschnig C and Gyalistras D 2017 Energy Procedia 122 421–426 ISSN 18766102
[3] Homaei S and Hamdy M 2020 Applied Energy 267 114868 ISSN 03062619
[4] Röck M, Saade M R M, Balouktis M, Rasmussen F N, Birgisdottir H, Frischknecht R, Habert G, Lützkendorf T and Passer A 2020 Applied Energy 258 114107 ISSN 03062619
[5] Kotireddy R, Hoes P J and Hensen J L 2018 Applied Energy 212 428–442 ISSN 03062619
[6] Kotireddy R, Loonen R, Hoes P J and Hensen J L 2019 Energy and Buildings 202 109362 ISSN 03787788
[7] Gang W, Wang S, Yan C and Xiao F 2015 Science and Technology for the Built Environment 21 789–799 ISSN 2374-4731, 2374-474X
[8] Nik V M, Mata E and Kalagasidis A S 2015 Energy Procedia 78 955–960 ISSN 18766102
[9] Nik V M, Mata nd Sasic Kalagasidis A 2015 Energy and Buildings 88 262–275 ISSN 03787788
[10] Rysanek A and Choudhary R 2013 Energy and Buildings 57 324–337 ISSN 03787788
[11] Picard T, Hong T, Luo N, Lee S H and Sun K 2020 Energy and Buildings 224 110251 ISSN 03787788
[12] McPhail C, Maier H R, Kwakkel J H, Giuliani M, Castelletti A and Westra S 2018 Earth’s Future 6 169–191 ISSN 23284277
[13] Kuhn A 2021 Scenario Based Robustness Analysis of Building Systems’ Life Cycle Impact in the Swiss Context Master Thesis Chair of Architecture and Building Systems, ETHZ
[14] Jayathissa P 2020 RC Building Simulator URL https://github.com/architecture-building-systems/RC_BuildingSimulator
[15] Jayathissa P, Luzzatto M, Schmidli J, Hofer J, Nagy Z and Schlüter A 2017 Applied Energy 202 726–735 ISSN 03062619
[16] F Holmgren W, W Hansen C and A Mikolski M 2018 Journal of Open Source Software 3 884 ISSN 2475-9066
[17] enbau-online 2020 13.4 Berechnung der Kosten