

Impact of Quality Control on Uncertainty in Solar PV Performance Evaluation

Hulik Jansova M., Blstak Catlosova K., Camara A., Cebecauer T., Jakubik M., Osvald O. | Solargis, Bratislava, Slovakia

Ground measured data is never perfect, but does it matter?

Instrumental faults, environmental and installation factors introduce biases that silently distort every downstream result. We quantify the full chain from raw measurements to final performance metrics across three key questions addressing the scale of the problem, the effectiveness of QC methods, and the propagation of residual uncertainty into EPI.

1 What Bias and MAE do undetected QC issues introduce into GHI and GTI measurements, and how often do they occur?

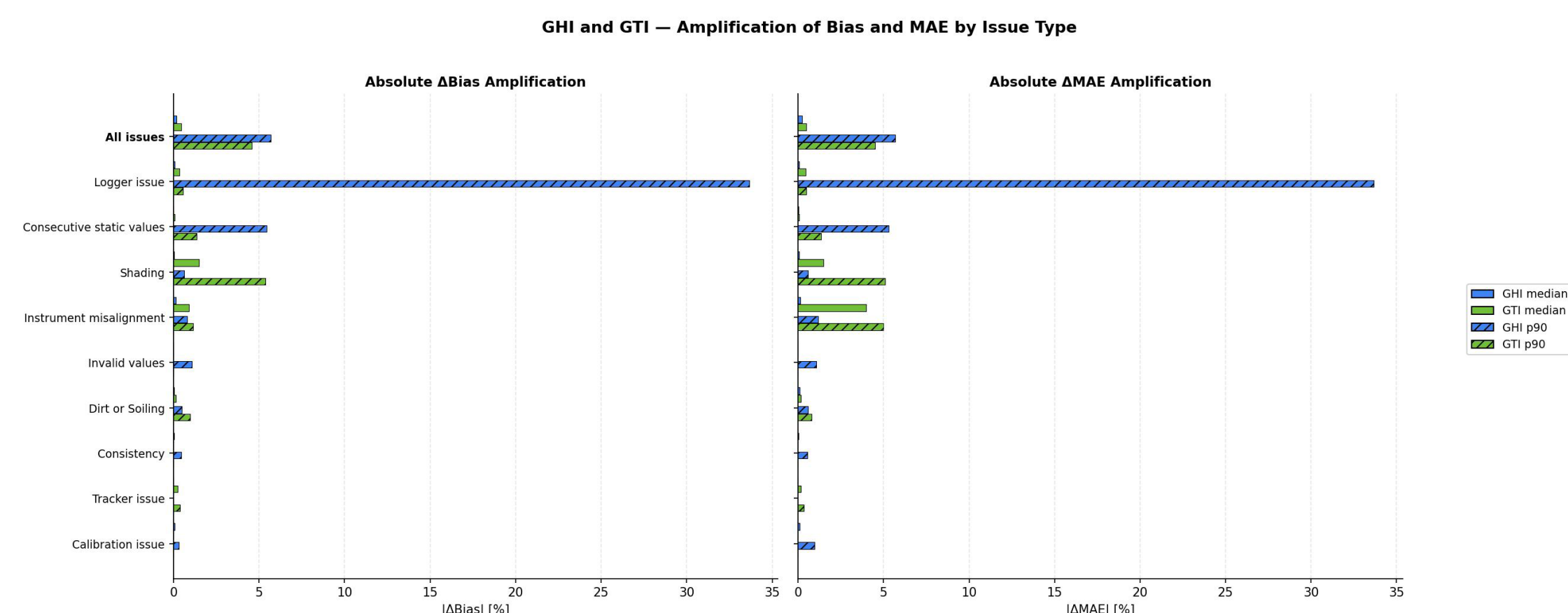


Figure 1: Top-impact quality issues in GHI and GTI measurements and their effect on Bias and MAE [%] at median and p90, computed only on datasets where each issue is present. Δ Bias and Δ MAE represent the difference in Bias and MAE between raw and QC'd measurements relative to the Solargis solar resource model used as reference.

Issue Type	GHI Proportion of Affected Datasets [%]	GHI Proportion of Affected Data Points - Median [%]	GHI Proportion of Affected Data Points - p90 [%]	GTI Proportion of Affected Datasets [%]	GTI Proportion of Affected Data Points - Median [%]	GTI Proportion of Affected Data Points - p90 [%]
Any issue	100.0	3.2	18.2	100.0	6.1	14.6
Logger issue	11.0	0.2	11.3	3.9	0.2	1.7
Consecutive static values	38.9	0.0	1.7	48.7	0.1	0.5
Shading	58.5	0.6	3.5	52.6	3.7	8.6
Instrument misalignment	9.2	4.4	46.5	5.3	17.8	32.1
Invalid values	56.8	0.2	6.1	64.5	0.1	2.7
Dirt or Soiling	61.5	0.7	4.4	32.9	1.3	5.4
Consistency	18.9	0.2	3.3	0.0	0.0	0.0
Tracker issue	0.0	0.0	0.0	5.3	1.5	3.3
Calibration issue	1.6	3.4	46.0	0.0	0.0	0.0

Table 1: Share of affected datasets and data points per selected quality issue type, for GHI and GTI. Note: Affected data point statistics are computed only from datasets where the respective issue is present.

Analysis of 509 GHI and 77 GTI measurements evaluated using Solargis expert QC procedures, covering two aspects: (1) occurrence of quality issues across datasets, and (2) their impact on Bias and MAE when left uncorrected.

Key takeaways

- Every dataset contains quality issues.** 100% of analyzed GHI and GTI time series were flagged for at least one issue.
 - Typical impact is small but tails are severe.** Median affected data points are low (GHI: 3.2%, GTI: 6.1%), but reach 18% (GHI) and 15% (GTI) at p90. Persistent calibration or misalignment issues can corrupt 47% (p90) of all data points.
 - Uncorrected issues significantly affect model evaluation.** Leaving all issues unflagged amplifies Bias by 5.7% (GHI) and 4.6% (GTI) at p90, with similar MAE increases.
- Conclusion:** Even small residual issues meaningfully increase uncertainty - consistent, well-defined QC procedures are essential.

Assume your data has quality issues - because it always does. 100% of analyzed measurements contained at least one flagged issue. While the median proportion of affected data points is modest (3.2% for GHI), in extreme cases effects are severe. Left uncorrected, the Bias is amplified by 5.7% (in 90th percentile).

2 How does automated QC performance vary with data resolution, record length, and sensor redundancy?

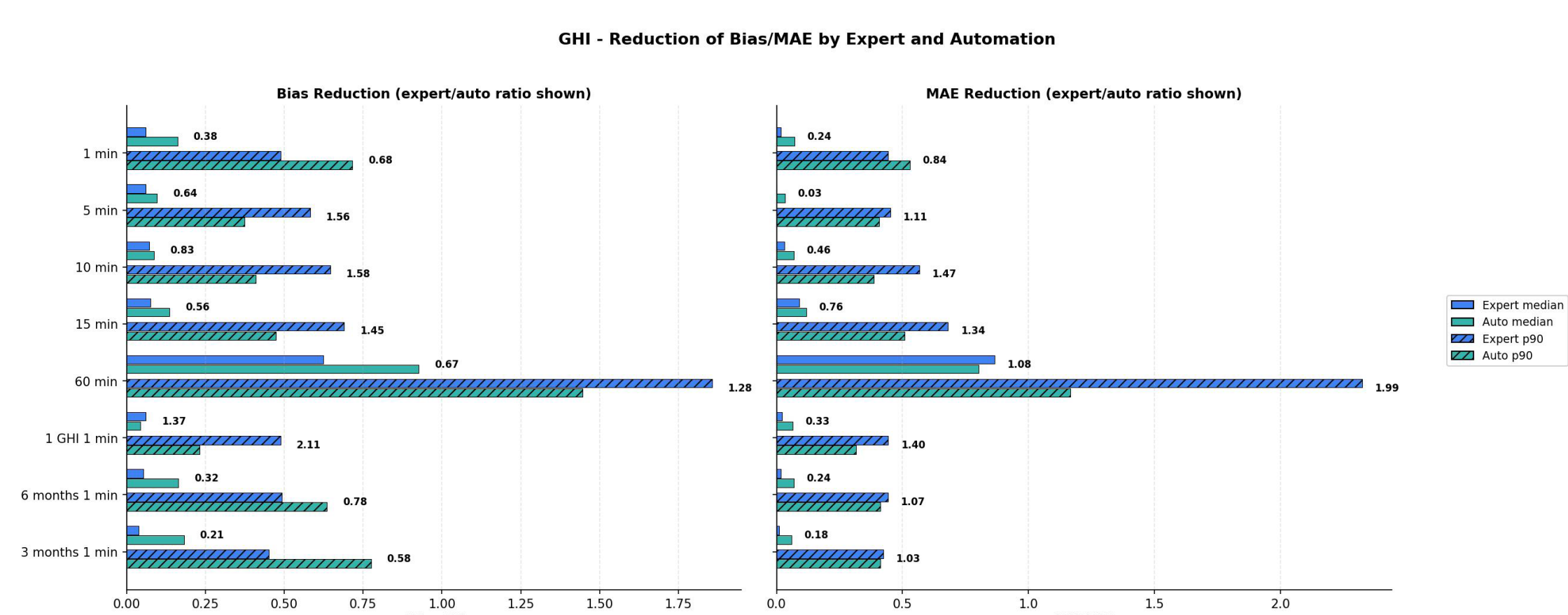


Figure 2: Δ Bias and Δ MAE [%] achieved by expert and automated QC relative to the raw measurements, using the Solargis solar resource model as reference, at median and p90. Numbers on bars show the expert/auto ratio - values below 1 indicate automation outperforms expert QC, above 1 the opposite.

Analysis of 114 one-year expert QC'd datasets with 2 GHI measurements. Each dataset was independently modified in 7 ways: aggregated to 4 coarser timesteps, reduced to a single GHI sensor, or shortened to 6 and 3 months. Quantifies the relative decrease in Bias and MAE when quality issues are removed by expert or automated methods.

	Precision (flagged class)							
	1 min	5 min	10 min	15 min	60 min	1 GHI 1 min	6 months 1 min	3 months 1 min
Median	0.39	0.43	0.40	0.35	0.50	0.31	0.30	0.28
Std	0.27	0.28	0.27	0.25	0.30	0.33	0.27	0.28

	Recall (flagged class)							
	1 min	5 min	10 min	15 min	60 min	1 GHI 1 min	6 months 1 min	3 months 1 min
Median	0.51	0.39	0.34	0.37	0.32	0.21	0.52	0.52
Std	0.26	0.22	0.22	0.22	0.26	0.28	0.25	0.27

Table 2-3: Precision and recall of automated QC flagging (median \pm std) across all dataset modifications. Flagged class = all points with QC issues.

Key takeaways

- Automation matches expert performance under optimal conditions.** Multi-sensor, 1-min resolution sites achieve expert/auto ratio below 1. The ratio grows substantially with coarser aggregation.
- Sensor redundancy protects against worst-case data quality issues.** Single-sensor setups maintain reasonable median performance but deliver only half the expert Bias reduction on difficult datasets (p90).

- Data length requirements are more forgiving than expected.** 6 months performs nearly identically to 1 year. Even 3 months is viable on typical datasets, though consistency suffers on challenging sites.
- Flagging performance degrades without redundancy and at coarser timesteps.** Single-sensor and short datasets show substantial overflagging (low precision). Real issues are also missed more often (low recall): half slip through without a redundant sensor, and more with growing timestep. Expert/auto ratio reaches up to 2 at p90.

Conclusion: Automated QC scales effectively for well-equipped sites with 6+ months of multi-sensor data, but reliability on difficult datasets requires redundant sensors and manual review.

Expert QC is the reliability baseline; automation is a complement, not always a replacement. Solargis automated QC can approach expert-level performance but only under favorable conditions: invest in multiple sensors, high temporal resolution, and at least 6 months of data. For high-stakes assessments or challenging sites, hybrid manual review remains essential.

3 How does residual measurement uncertainty propagate into energy system performance assessment (EPI)?

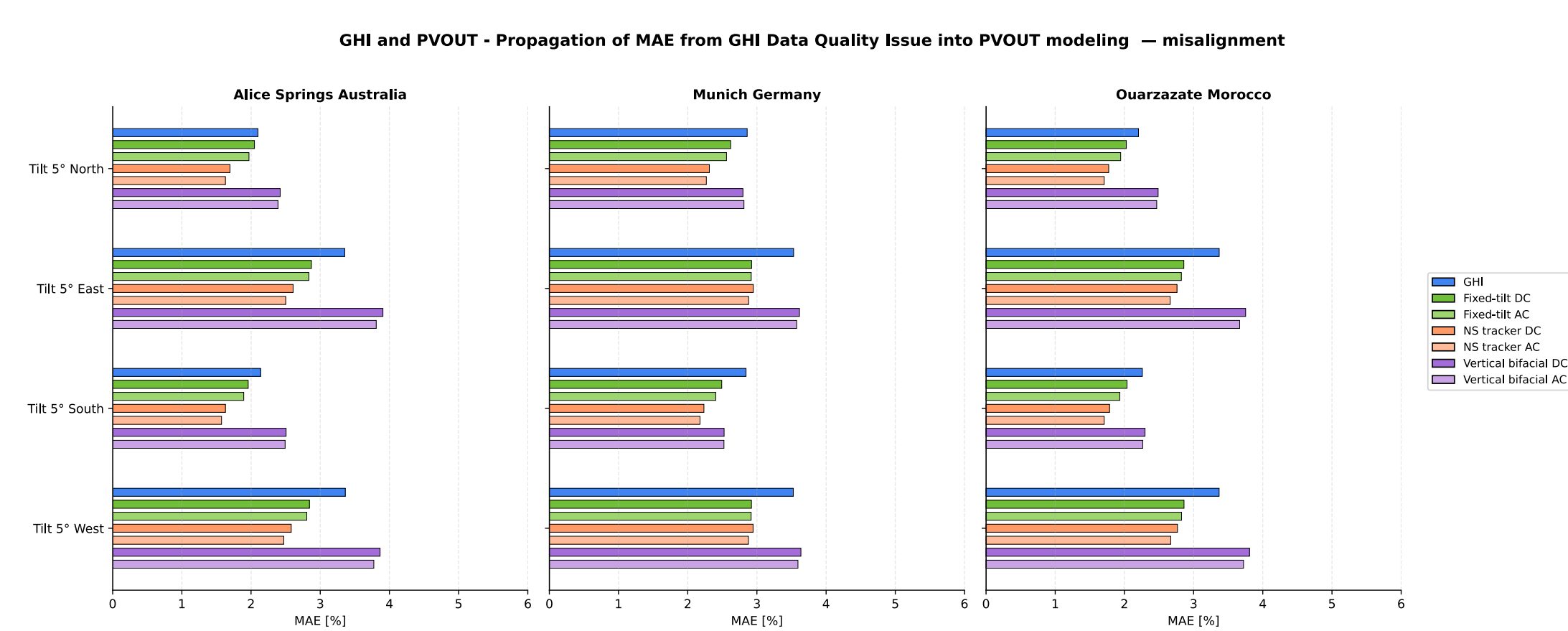
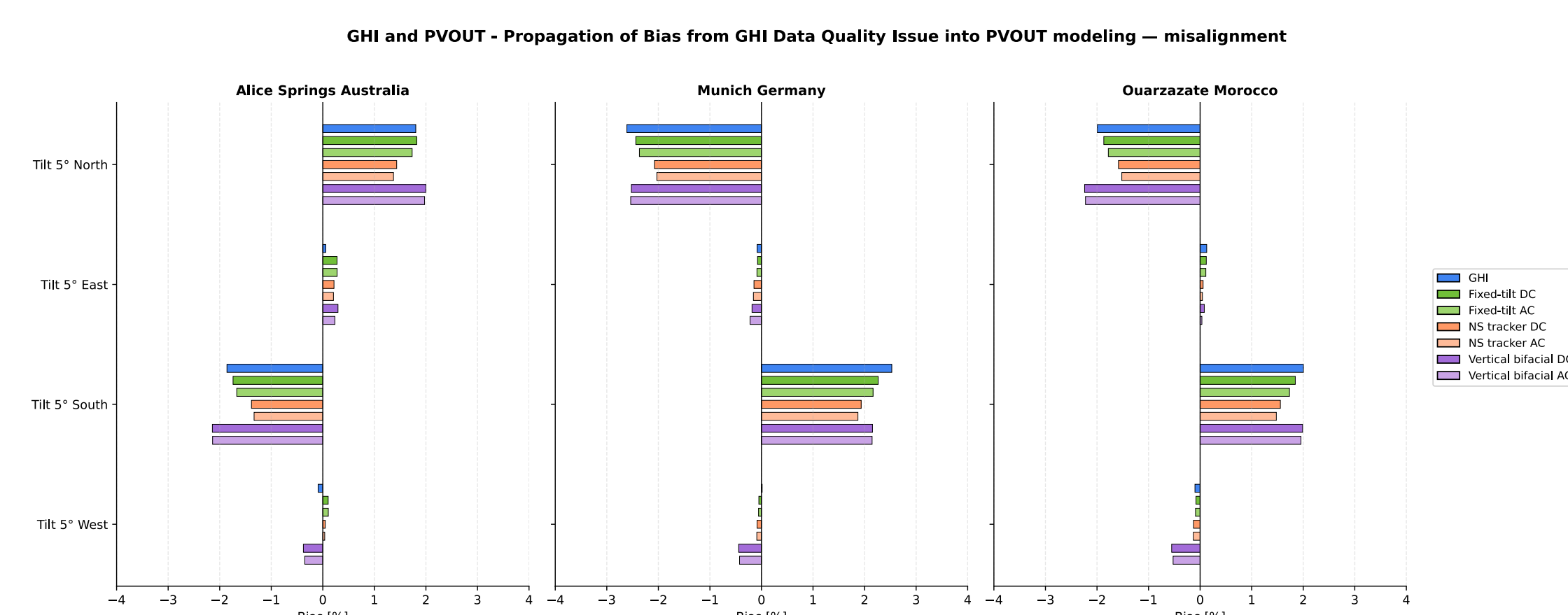


Figure 3-5: Bias and MAE [%] propagated into simulated PVOU for two quality issue types, three PV configurations, and three locations. Note: for soiling, |Bias| = MAE, as the effect is strictly unidirectional.

Label	Scope	Description		
Bird droppings	Apr–Jun, solar noon \pm 1.5h	GHI modified by 10% cosine-tapered U-shape dip around solar noon		
Dust accumulation	Apr–Jun, all daytime	GHI modified by linear soiling accumulation from 0% (April 1) \rightarrow 15% (June 30)		

Label	Scope	Tilt	Azimuth	Description
Tilt 5° North	Jan–Jun	5°	0° (North)	GHI replaced with tilted GTI - away from equator (NH) / toward equator (SH)
Tilt 5° East	Jan–Jun	5°	90° (East)	GHI replaced with tilted GTI - morning-enhanced, afternoon-deficient
Tilt 5° South	Jan–Jun	5°	180° (South)	GHI replaced with tilted GTI - toward equator (NH) / away from equator (SH)
Tilt 5° West	Jan–Jun	5°	270° (West)	GHI replaced with tilted GTI - morning-deficient, afternoon-enhanced

Table 5-6: Catalogue of quality issues injected into GHI time series: soiling (Table 5) and misalignment (Table 6).

One year of 5-min GHI time series injected with controlled quality issues, used to simulate PVOU via Solargis Evaluate with three representative PV configurations: fixed tilt (optimal angle), horizontal NS tracker, and vertical bifacial single-row. Simulations were performed for three climatically distinct sites.

Key takeaways

- Uncertainty propagation depends on system design.** Vertical bifacial setups are most sensitive to GHI Bias on annual aggregates, with transfer ratios (PVOU MAE / GHI MAE, similarly for Bias) that can exceed 1.
- PV simulation Biases mirror GHI Biases in magnitude.** PVOU Bias trends closely follow GHI data trends across all configurations, with AC and DC outputs differing only marginally.
- Misalignment impact is largest at the highest-latitude site.** NS tracker systems are least affected by sensor misalignment.
- A \pm 2% GHI Bias translates directly to EPI deviation.** A \pm 2% GHI Bias yields \sim 2% EPI deviation. Negative GHI Bias inflates EPI and hides underperformance; positive Bias deflates it.

Conclusion: The MAE transfer ratio from GHI to PVOU is typically 0.7-1 for monofacial fixed and tracker systems. Vertical bifacial setups can exceed 1. Crucially, skipping QC can inflate EPI, potentially masking real underperformance of the PV plant.

Don't omit QC before calculating EPI. Residual QC errors introduce Biases comparable in magnitude to the full simulation uncertainty chain. For bifacial/vertical systems the effect is amplified. An unchecked QC issue can make a struggling asset appear perfectly healthy.

PV yield uncertainty spans the full chain (from input data and simulation models to interannual variability), typically reaching a few percent at p90. Our results show that skipping measurement QC alone introduces errors of comparable magnitude. Under normally distributed assumptions ($Up90 \approx 1.6 \times MAE$), residual MAEs translate to p90-level uncertainties that rival the full simulation chain.

What does this mean for you?

- Audit your data:** Every dataset contains QC issues and in extreme cases, uncorrected Bias alone can reach the magnitude of the full simulation uncertainty chain.
- Invest in your setup:** Redundant sensors, high resolution, and 6+ months of data are prerequisites for reliable automated QC. For high-stakes assessments, complement with expert review.
- Never skip QC before performance evaluation:** An uncorrected issue can make a struggling plant look perfectly healthy on paper.