SEP 22-26
AIM for Scale
AI Weather Training Program

AIM FOR SCALE | WEATHER PACKAGE

Data Quality Assurance and Quality Control(QA/QC)

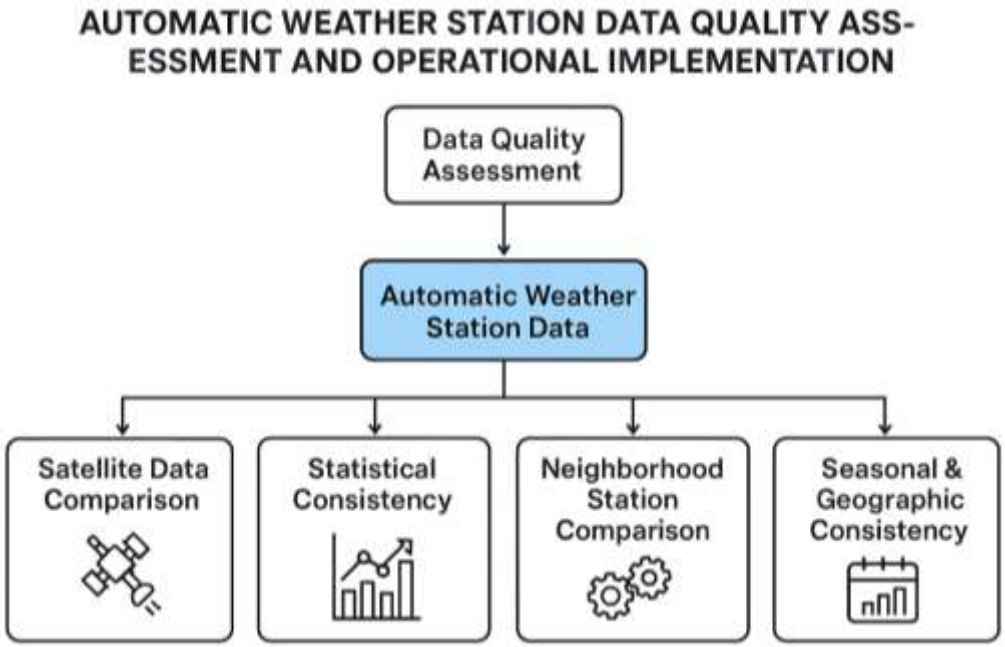
Introduction

Precipitation Data Quality Assessment and Operational Implementation

- AI models (e.g., GraphCast) depend on high-quality input data
- Precipitation is simple to measure, but hardest to ensure accuracy for both automatic and manual instruments.
- Ground stations validate & calibrate satellite products (CHIRPS, ERA5)
- Satellites then provide reliable large-scale coverage

QA/QC Tailored to the approach used by TAHMO

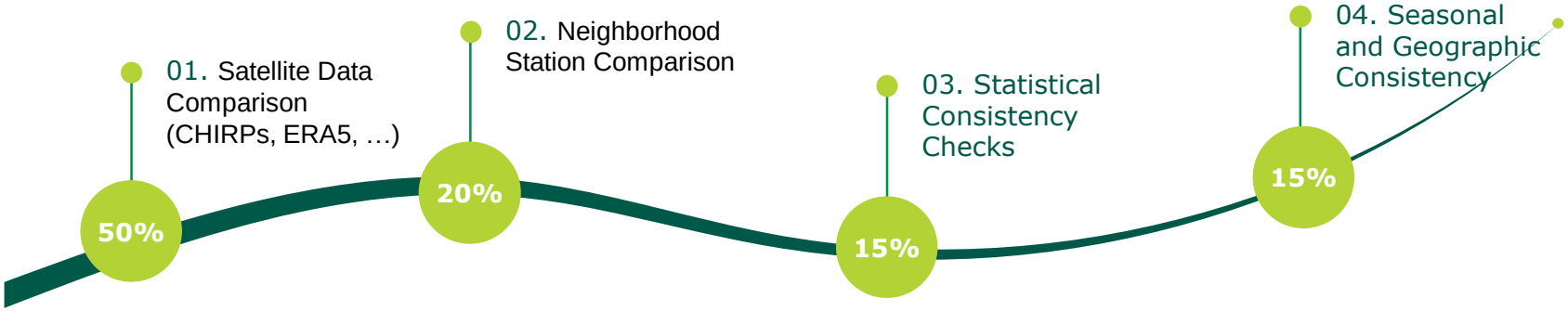
- **Observational datasets are prone to errors:** instrument drift, environmental contamination, transmission loss, and observer bias
- **Robust QA/QC procedures** are required to ensure accuracy, reliability, and usability of data in operational forecasting
- **Integrated frameworks (e.g., TAHMO) apply multi-tiered approaches:** Internal consistency checks (range, step, and persistence tests), Spatial consistency checks, Climatological/homogeneity checks



AIM FOR SCALE

AWS Data Quality Assessment

(Ecosystem of Tools and Approaches)



Satellite Data Comparison

- Detects major biases in sensor observations
- Identifies inconsistencies in precipitation reporting (under/overestimation in satellite products and under/over-reporting from ground stations)

Neighborhood Station Comparison

- Ground truthing for Satellite
- Reduce false flags from satellite QAQC
- Limited by geography and local sensor reliability
- Most effective in dense networks with high spatial resolution

Statistical Consistency Checks

- Detect spikes, unrealistic values, and sensor drifts.
- Multi-Parameter Correlation Checks
- Enables real-time anomaly detection
- Applies automated threshold rules for parameter ranges

Seasonal and Geographic Consistency

Ensure data reflects expected climatic/seasonal patterns.

- Compare seasonal averages with historical climatology
- Detect deviations (climatological thresholds, history, neighbors)

AIM FOR SCALE | WEATHER PACKAGE

Example: AWS Data Quality Assessment

(Satellite products + neighbor comparison + Statistical Consistency)

Positive Bias - List of stations with suspected fault

Station	Date	Station Precipitation total mm (monthly)	min precipitation mm (CHIRPs & GSMap)	max precipitation mm (CHIRPs & GSMap)
TA00784	**2025-08**	**2416**	**252.1**	**333.6**
TA00416	2025-08	1270.9	13.4	86.5
TA00813	2025-08	399	37.2	102.6
TA00436	2025-08	332	41.5	73.2
TA00119	2025-08	115	0.2	40.7
TA00679	2025-08	103	2.9	41.9
TA00196	2025-08	99.5	1.1	36.1
TA00588	2025-08	97.4	21.5	30.1
TA00057	**2025-08**	**55.6**	**0.9**	**22.2**

Expected Sensor records count (persistence)	8064	28 days month
	8352	29 days month
	8640	30 day month
	8928	31 days month

Satellite Comparison

Date	TA00784 Recorded Precipitation (Monthly)	Sensor records count (Monthly)	GSMaP	CHIRPS
2025-01	0	8928	0.679	0.78
2025-02	0	8064	0.508	0.025
2025-03	0	8913	0.284	0
2025-04	0	8640	0	0
2025-05	0	8928	3.344	2.563
2025-06	21.569	8640	22.349	39.553
2025-07	88.352	8928	101.566	159.242
2025-08	2415.999	8347	333.589	252.053

Date	TA00057 Recorded Precipitation (Monthly)	Sensor records count (Monthly)	GSMaP	CHIRPS
2025-01		0	27.462	96.138
2025-02		0	5.154	4.924
2025-03	141.168	4172	42.532	160.046
2025-04	280.296	8639	59.582	172.11
2025-05	205.887	8928	13.673	96.093
2025-06	31.253	8635	0.364	23.069
2025-07	12.176	8906	0.309	7.618
2025-08	55.635	8353	0.892	22.237

Persistency and Statistical Consistency (internal and external consistency tests)

Neigbors to TA00784	Precipitation total mm (08-2025)	count	GSMaP	CHIRPS
TA00908	200.192	8353	289.86	287.49
TA00785	171.128	6955	323.93	224.51
TA00782	21.176	8353	319.03	311.1

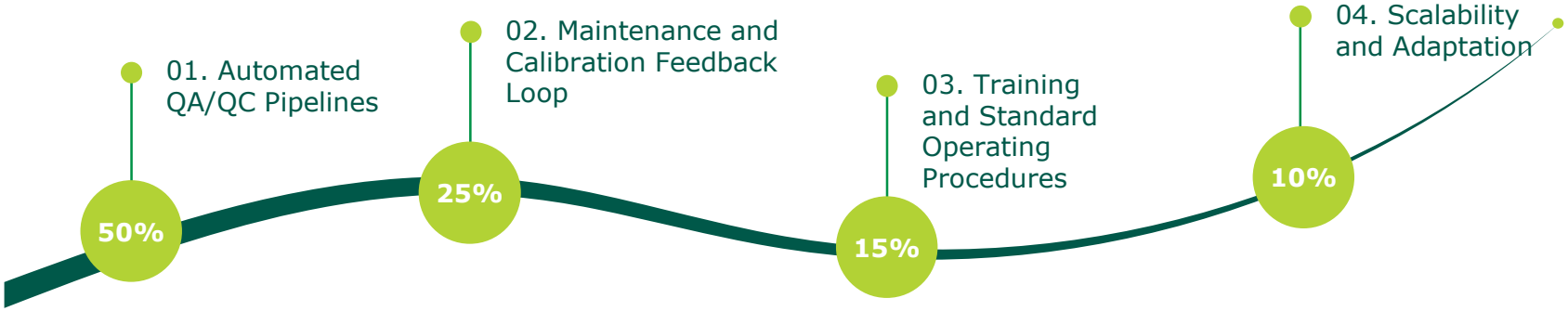
Positive flag - Sensor fault

Neigbors to TA00057	Precipitation total mm (08-2025)	count	GSMaP	CHIRPS
TA00025	42	8353	2.882	40.054
TA00066	41.367	8353	2.402	48.261
TA00080	39.744	8329	3.798	42.12
TA00024	50.267	6235	0.421	25.149

False Precipitation flag - under estimation of CHIRPs

neighbor comparison

Operational Implementation Considerations



Automated QA/QC Pipelines

- Implement software frameworks for real-time data validation
- Data Flagging and Metadata
- Integration with Satellite and External Datasets

Maintenance and Calibration Feedback Loop

Maintain historical records for predictive maintenance

Training and Standard Operating Procedures

- Train personnel on QA/QC protocols and flag interpretation
- Define clear SOPs for data handling

Scalability and Adaptation

- Incorporate ML for advanced anomaly detection

AIM FOR SCALE | WEATHER PACKAGE

QAQC based on Parallel measurements

Parallel measurements

Manual rain gauge at ground level,
Automatic gauge at 0.3m height,
TAHMO at 2m

May be used for bench marking

Sensor Redundancy

AWS with duo sensors for precipitation

Most effective approach for anomaly detection in real time from AWS precipitation data through automated QAQC

Parallel measurements With multiple technologies

Primary sensor 's QAQC may benefit from other sensors that utilize different measurement technology for the same weather parameter.

- Tipping bucket vs Drip count
- Ultrasonic vs Mechanical wind sensors

AIM FOR SCALE | WEATHER PACKAGE

Contact Us

TAHMO

Email : info@aimforscale.org
Web : www.aimforscale.org

AIM FOR SCALE | WEATHER PACKAGE