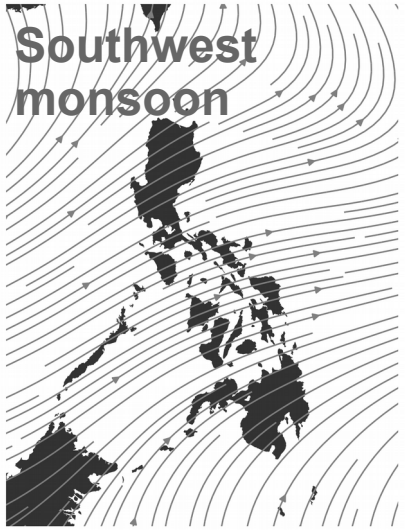
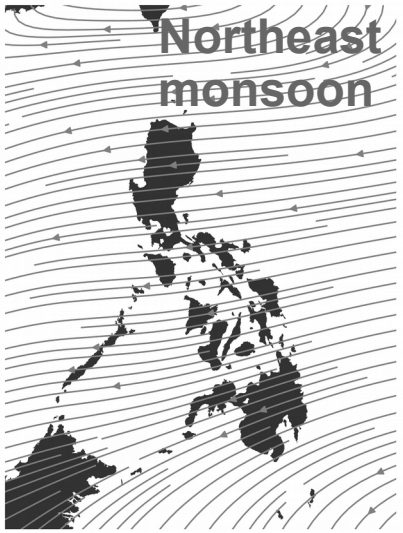
Data Quality Control on ASTI Automated Weather Station (AWS) Measurements

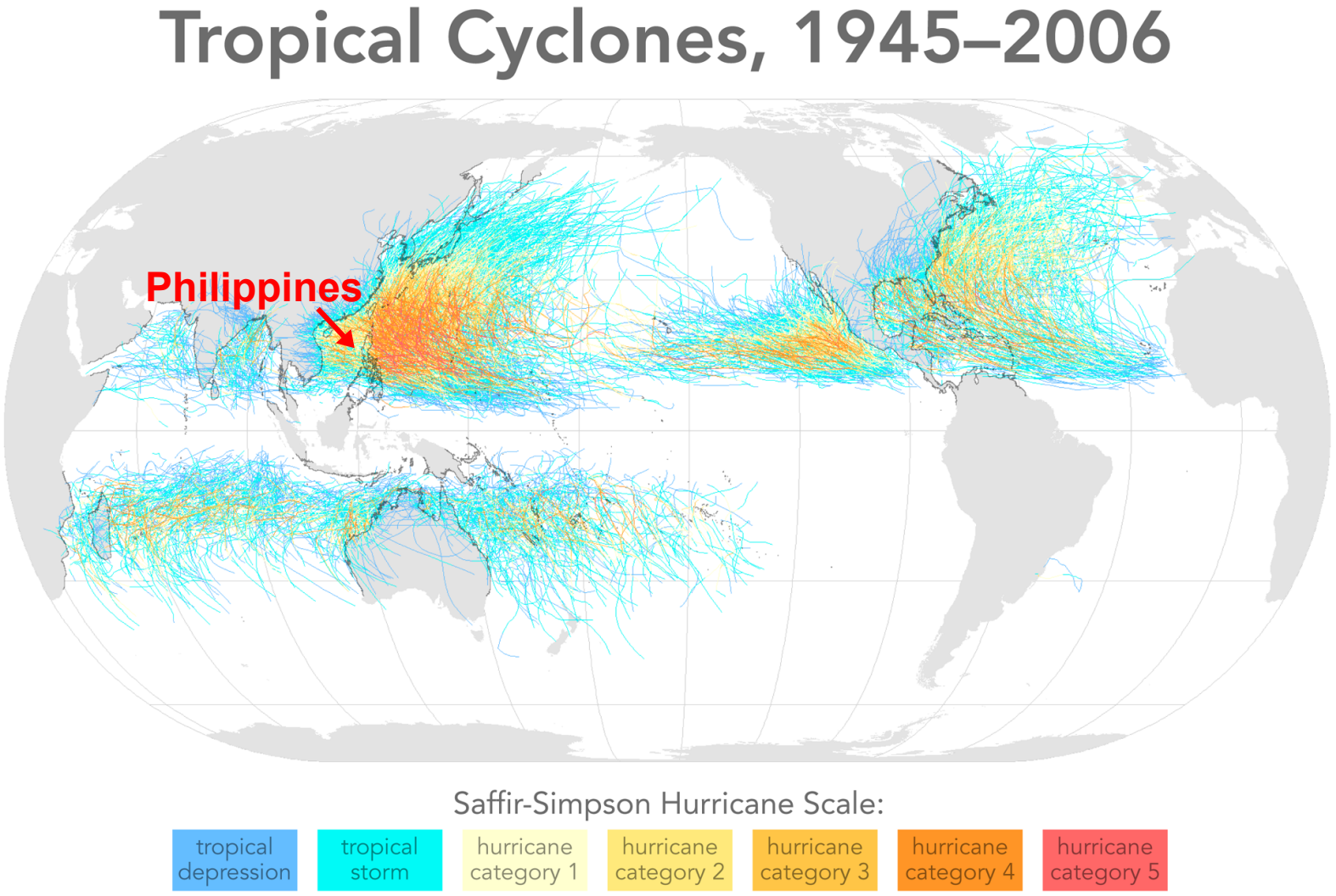
Jay Samuel Combinido
Advanced Science and Technology Institute
Department of Science and Technology
Quezon City, Philippines

APAN43

Meteorological characteristics of the Philippines







https://upload.wikimedia.org/wikipedia/commons/6/6f/Tropical_cyclones_1945_2006_wikicolor.png

Typhoon Ketsana (Ondoy) 2009

http://opinion.inquirer.net/12757/large-areas-of-metro-manila-sinking

TS Lingling (Agaton) 2014

http://www.ibtimes.co.uk/dramatic-photos-floods-landslides-philippines-1433015

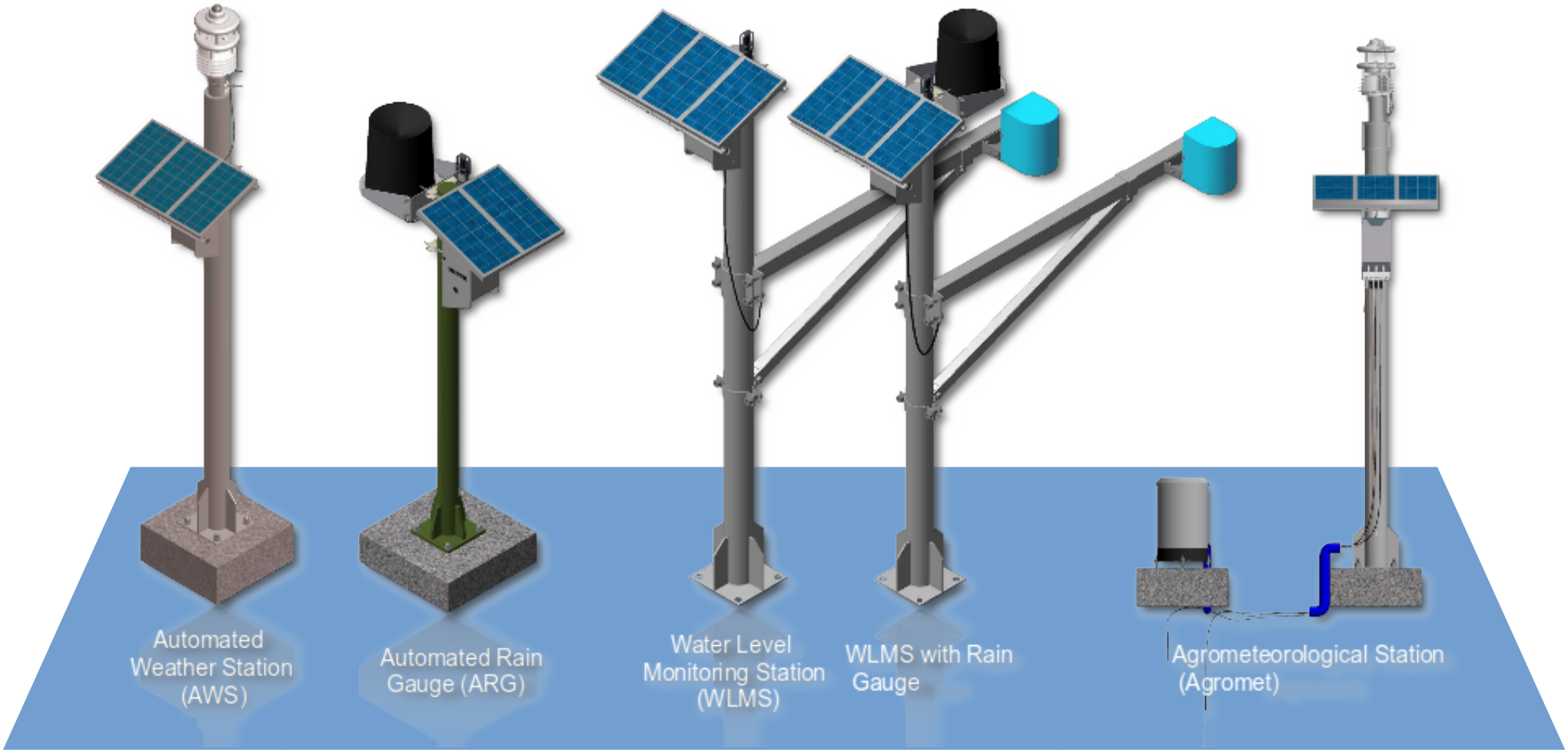
Monsoon rains (2016)

http://www.dailymail.co.uk/news/article-3739923/Young-boys-struggle-swim-deep-floodwaters-monsoon-kills-five-people-leaves-70-000-homeless-Philippines.html

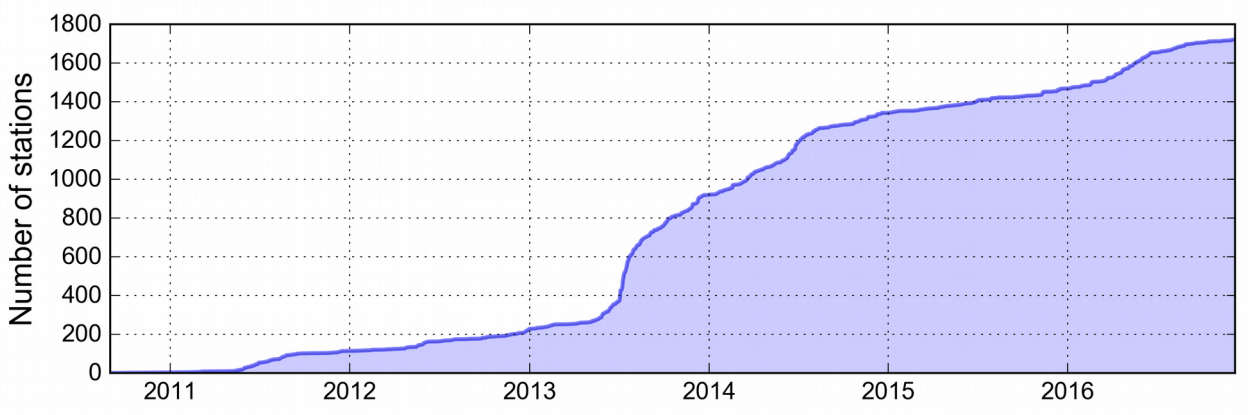
Monsoon rains (2012)

[illegible]srucksack.com/201[illegible]lessons-from-[illegible]/
[illegible]ww.interaksyon.com [illegible] the FB account of [illegible] Ruazol)

ASTI weather stations

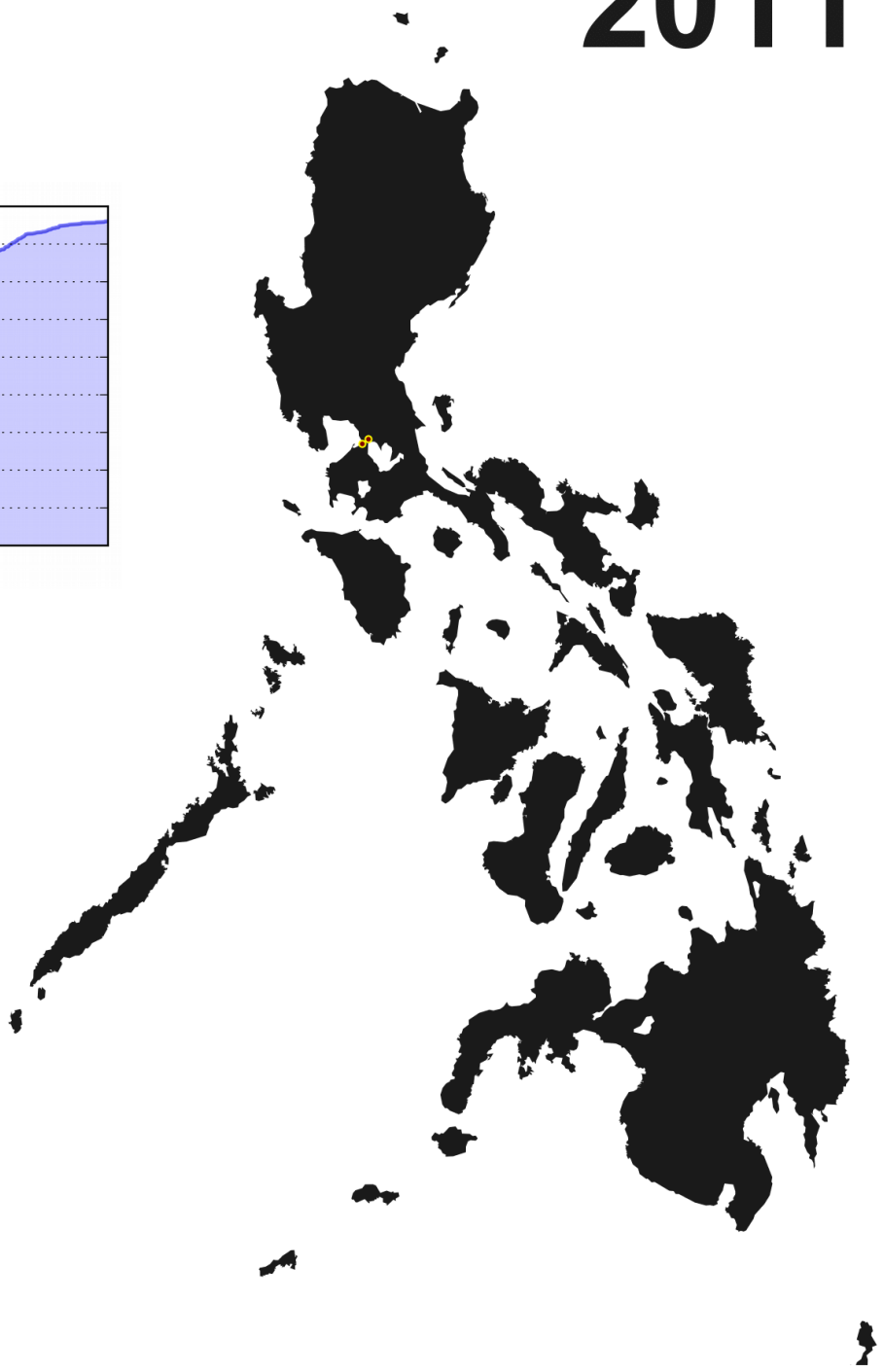


AWS deployment



~1800 stations

0.35 billion records



Applications of weather data

- Weather monitoring
- Early warning or advisories
- Rainfall forecasting
- Flood forecasting

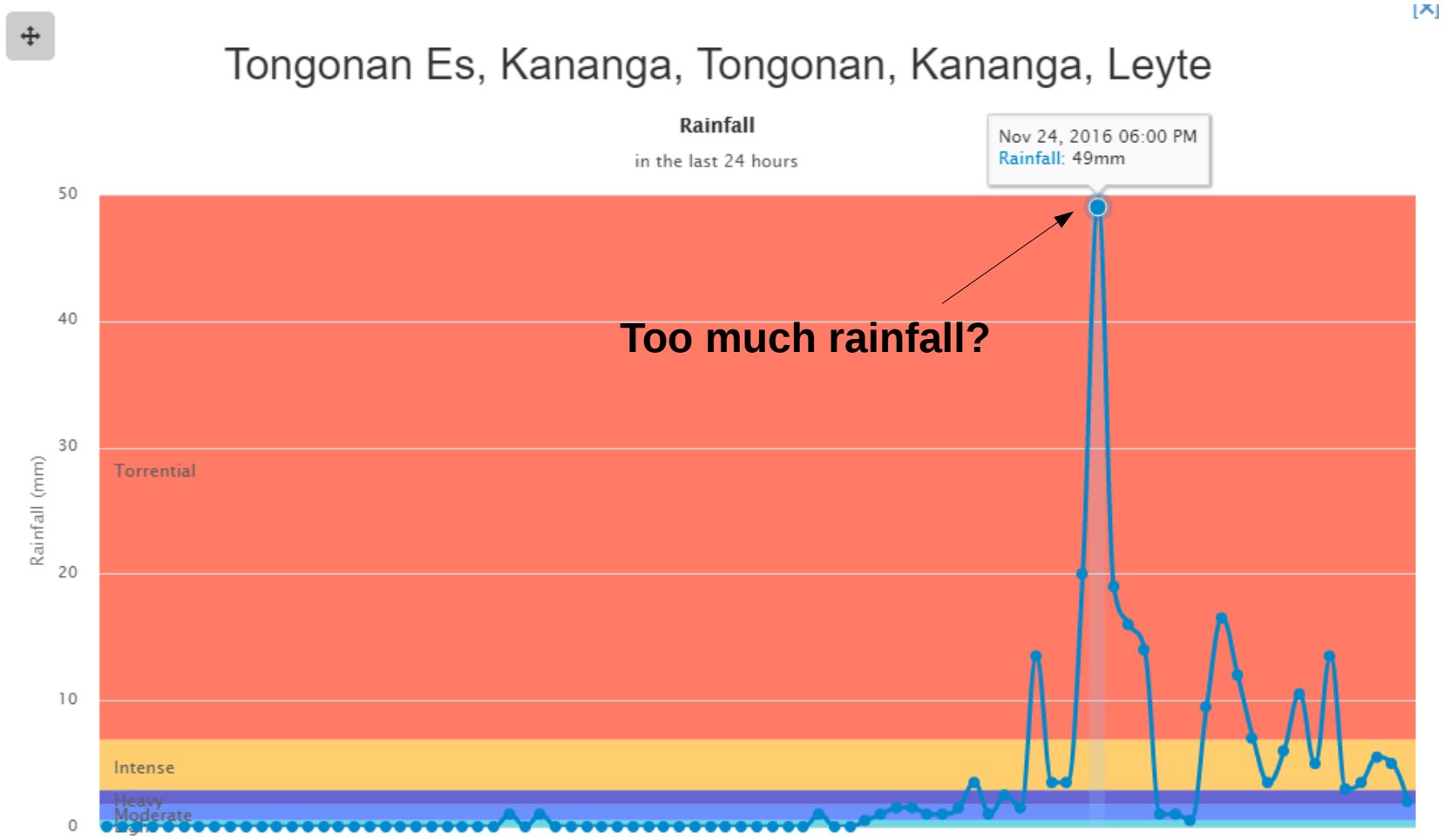
- Research (e.g. NWP, climate, validation)

Motivations of QC

- **Make it easier for data users to identify suspicious and erroneous data, and to highlight corrected values**
- avoid the issuance of warnings or advisories based on anomalous data [1]
- Minimize analysis and weather prediction errors [2, 3]
- Identify calibration, measurement, and communication errors
- Detect deterioration and malfunction of sensors

1. Nagata, K. (2010), The importance of data quality control in disaster prevention and mitigation, JMA/WMO workshop on quality management in surface and upper-air observations in RA II (Asia), Tokyo, Japan.
2. Bertrand, C., Gonzales Sotelino, L., and Journee, M. (2013) Quality control of 10-min air temperature data at RMI, Adv. Sci. Res., 10, 1-5.
3. Qin, Z.K., Zou, X., Li, G., and Ma X.L. (2010) Quality control of surface station temperature data with non-Gaussian observation-minus-background distributions, J. Geophys. Res., 115, D16312.

Incident reported by a user



** NOAH Program*

Quality control checks

Table 1: Quality control procedure applied to ASTI automated weather stations for 15-minute data.

Validation procedures	Air pressure (hPa)	Air temperature (°C)	Relative humidity (%)	Rainfall (mm)	Wind speed (km/h)	...
Timestamp check			$\lvert t - \text{round}\,(t,p)\rvert \leq 2.5$ minutes			...
			$\text{rank}\,(\lvert t - \text{round}\,(t,p)\rvert) = 1$			...
Range check	$640 \leq P_t \leq 1022$ (PAGASA, 2014)	$0 \leq T_t \leq 50$ (PAGASA, 2014)	$0 \leq RH_t \leq 95$	$0 \leq R_t \leq 30$ (PAGASA, 2014)	$0 \leq U_t \leq 270$	...
				$0 \leq R_t^d \leq 600$		...
Step check	$\lvert P_t - P_{t-1}\rvert \leq 0.6$	$\lvert T_t - T_{t-1}\rvert \leq 1.5$	$\lvert RH_t - RH_{t-1}\rvert \leq 30$		$\lvert U_t - U_{t-1}\rvert \leq 75$	...
Persistence check	$\left\langle (P_x - \langle P_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$	$\left\langle (T_x - \langle T_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$	$\left\langle (RH_x - \langle RH_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$		$\left\langle (U_x - \langle U_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.4$	...
Internal consistency check		$T > T_d$ (Zahumensky, 2004)	rain > 0 and value > 60	value $= 0$ and duration $= 0$	speed $= 0$ and direction $= 0$	...
				value > 0 and duration > 0	speed $\neq 0$ and direction $\neq 0$	...
				(Zahumensky, 2004)	(Zahumensky, 2004)	...
				value > 0 and RH > 60		...
					$U_t < U_{\max}$	...
Spatial check			not yet implemented			...

t: time stamp of a measurement record; p: sampling period of the station in minutes; round(t,p): rounded timestamp based on p; P_t: air pressure at time t; P_{t-1}: air pressure 15 minute before time t; P_x: $(P_t, P_{t-1}, P_{t-2}, ..., P_{t-12})$; T_t: air temperature at time t; T_{t-1}: air temperature 15 minute before time t; T_x: $(T_t, T_{t-1}, T_{t-2}, ..., T_{t-12})$; T_d: dew point temperature; RH_t: relative humidity at time t; RH_{t-1}: relative humidity 15 minute before time t; RH_x: $(RH_t, RH_{t-1}, RH_{t-2}, ..., RH_{t-12})$; R_t: rainfall amount at time t; R_t^d: daily accumulated rainfall ending at time t; U_t: wind speed at time t; P_{t-1}: wind speed 15 minute before time t; U_x: $(U_t, U_{t-1}, U_{t-2}, ..., U_{t-12})$; $U_{\max}$: maximum wind speed

implemented using Apache Spark™

Quality control checks

Table 1: Quality control procedure applied to ASTI automated weather stations for 15-minute data.

Validation procedures	Air pressure (hPa)	Air temperature (°C)	Relative humidity (%)	Rainfall (mm)	Wind speed (km/h)	...
Timestamp check			$\lvert t - \text{round}\,(t,p)\rvert \leq 2.5$ minutes $\text{rank}\,(\lvert t - \text{round}\,(t,p)\rvert) = 1$			... ...
Range check	$640 \leq P_t \leq 1022$ (PAGASA, 2014)	$0 \leq T_t \leq 50$ (PAGASA, 2014)	$0 \leq RH_t \leq 95$	$0 \leq R_t \leq 30$ (PAGASA, 2014) $0 \leq R_t^d \leq 600$	$0 \leq U_t \leq 270$	... ...
Step check	$\lvert P_t - P_{t-1}\rvert \leq 0.6$	$\lvert T_t - T_{t-1}\rvert \leq 1.5$	$\lvert RH_t - RH_{t-1}\rvert \leq 30$		$\lvert U_t - U_{t-1}\rvert \leq 75$	...
Persistence check	$\left\langle (P_x - \langle P_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$	$\left\langle (T_x - \langle T_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$	$\left\langle (RH_x - \langle RH_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$		$\left\langle (U_x - \langle U_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.4$	...
Internal consistency check		$T > T_d$ (Zahumensky, 2004)	rain > 0 and value > 60	value $= 0$ and duration $= 0$ value > 0 and duration > 0 (Zahumensky, 2004) value > 0 and RH > 60	speed $= 0$ and direction $= 0$ speed $\neq 0$ and direction $\neq 0$ (Zahumensky, 2004) $U_t < U_{\max}$	... ...
Spatial check			not yet implemented			...

t: time stamp of a measurement record; p: sampling period of the station in minutes; round(t,p): rounded timestamp based on p; P_t: air pressure at time t; P_{t-1}: air pressure 15 minute before time t; P_x: $(P_t, P_{t-1}, P_{t-2}, ..., P_{t-12})$; T_t: air temperature at time t; T_{t-1}: air temperature 15 minute before time t; T_x: $(T_t, T_{t-1}, T_{t-2}, ..., T_{t-12})$; T_d: dew point temperature; RH_t: relative humidity at time t; RH_{t-1}: relative humidity 15 minute before time t; RH_x: $(RH_t, RH_{t-1}, RH_{t-2}, ..., RH_{t-12})$; R_t: rainfall amount at time t; R_t^d: daily accumulated rainfall ending at time t; U_t: wind speed at time t; P_{t-1}: wind speed 15 minute before time t; U_x: $(U_t, U_{t-1}, U_{t-2}, ..., U_{t-12})$; $U_{\max}$: maximum wind speed

implemented using Apache Spark™

Quality control checks

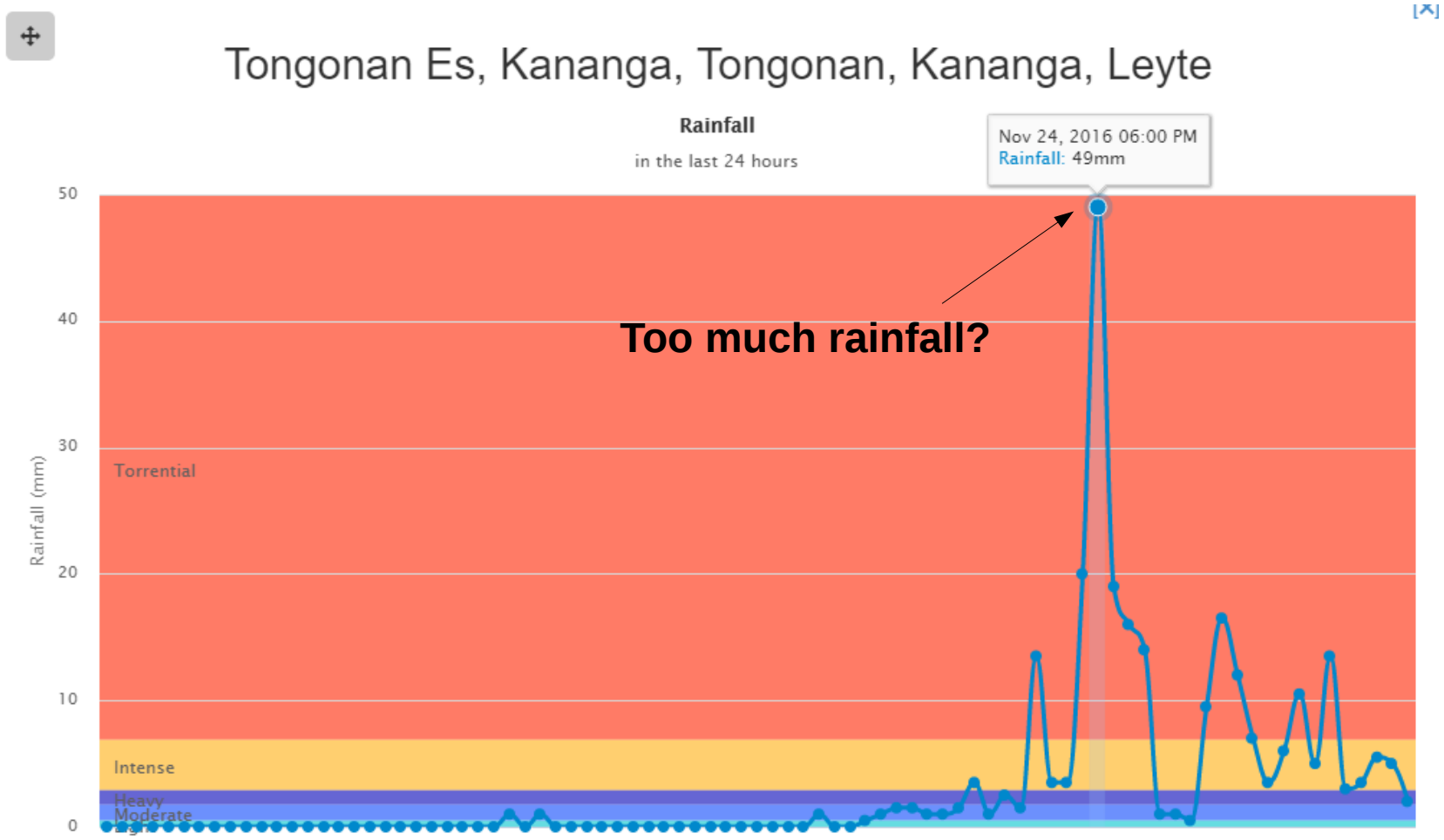
Table 1: Quality control procedure applied to ASTI automated weather stations for 15-minute data.

Validation procedures	Air pressure (hPa)	Air temperature (°C)	Relative humidity (%)	Rainfall (mm)	Wind speed (km/h)	...
Timestamp check			$\lvert t - \text{round}\,(t,p)\rvert \leq 2.5$ minutes			...
			$\text{rank}\,(\lvert t - \text{round}\,(t,p)\rvert) = 1$			...
Range check	$640 \leq P_t \leq 1022$ (PAGASA, 2014)	$0 \leq T_t \leq 50$ (PAGASA, 2014)	$0 \leq RH_t \leq 95$	$0 \leq R_t \leq 30$ (PAGASA, 2014)	$0 \leq U_t \leq 270$	...
				$0 \leq R_t^d \leq 600$		...
Step check	$\lvert P_t - P_{t-1}\rvert \leq 0.6$	$\lvert T_t - T_{t-1}\rvert \leq 1.5$	$\lvert RH_t - RH_{t-1}\rvert \leq 30$		$\lvert U_t - U_{t-1}\rvert \leq 75$	...
Persistence check	$\left\langle (P_x - \langle P_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$	$\left\langle (T_x - \langle T_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$	$\left\langle (RH_x - \langle RH_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$		$\left\langle (U_x - \langle U_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.4$	...
Internal consistency check		$T > T_d$ (Zahumensky, 2004)	rain > 0 and value > 60	value $= 0$ and duration $= 0$	speed $= 0$ and direction $= 0$	...
				value > 0 and duration > 0	speed $\neq 0$ and direction $\neq 0$	...
				(Zahumensky, 2004)	(Zahumensky, 2004)	...
				value > 0 and RH > 60		...
					$U_t < U_{\max}$	...
Spatial check			not yet implemented			...

t: time stamp of a measurement record; p: sampling period of the station in minutes; round(t,p): rounded timestamp based on p; P_t: air pressure at time t; P_{t-1}: air pressure 15 minute before time t; P_x: $(P_t, P_{t-1}, P_{t-2}, ..., P_{t-12})$; T_t: air temperature at time t; T_{t-1}: air temperature 15 minute before time t; T_x: $(T_t, T_{t-1}, T_{t-2}, ..., T_{t-12})$; T_d: dew point temperature; RH_t: relative humidity at time t; RH_{t-1}: relative humidity 15 minute before time t; RH_x: $(RH_t, RH_{t-1}, RH_{t-2}, ..., RH_{t-12})$; R_t: rainfall amount at time t; R_t^d: daily accumulated rainfall ending at time t; U_t: wind speed at time t; P_{t-1}: wind speed 15 minute before time t; U_x: $(U_t, U_{t-1}, U_{t-2}, ..., U_{t-12})$; $U_{\max}$: maximum wind speed

implemented using Apache Spark™

Incident reported by a user



** NOAH Program*

Quality control checks

Table 1: Quality control procedure applied to ASTI automated weather stations for 15-minute data.

Validation procedures	Air pressure (hPa)	Air temperature (°C)	Relative humidity (%)	Rainfall (mm)	Wind speed (km/h)	...
Timestamp check			$\lvert t - \text{round}\,(t,p)\rvert \leq 2.5$ minutes			...
			$\text{rank}\,(\lvert t - \text{round}\,(t,p)\rvert) = 1$			...
Range check	$640 \leq P_t \leq 1022$ (PAGASA, 2014)	$0 \leq T_t \leq 50$ (PAGASA, 2014)	$0 \leq RH_t \leq 95$	$0 \leq R_t \leq 30$ (PAGASA, 2014)	$0 \leq U_t \leq 270$	...
				$0 \leq R_t^d \leq 600$		...
Step check	$\lvert P_t - P_{t-1}\rvert \leq 0.6$	$\lvert T_t - T_{t-1}\rvert \leq 1.5$	$\lvert RH_t - RH_{t-1}\rvert \leq 30$		$\lvert U_t - U_{t-1}\rvert \leq 75$	...
Persistence check	$\left\langle (P_x - \langle P_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$	$\left\langle (T_x - \langle T_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$	$\left\langle (RH_x - \langle RH_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$		$\left\langle (U_x - \langle U_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.4$	...
Internal consistency check		$T > T_d$ (Zahumensky, 2004)	rain > 0 and value > 60	value $= 0$ and duration $= 0$	speed $= 0$ and direction $= 0$	...
				value > 0 and duration > 0	speed $\neq 0$ and direction $\neq 0$	...
				(Zahumensky, 2004)	(Zahumensky, 2004)	...
				value > 0 and RH > 60		...
					$U_t < U_{\max}$	...
Spatial check			not yet implemented			...

t: time stamp of a measurement record; p: sampling period of the station in minutes; round(t,p): rounded timestamp based on p; P_t: air pressure at time t; P_{t-1}: air pressure 15 minute before time t; P_x: $(P_t, P_{t-1}, P_{t-2}, ..., P_{t-12})$; T_t: air temperature at time t; T_{t-1}: air temperature 15 minute before time t; T_x: $(T_t, T_{t-1}, T_{t-2}, ..., T_{t-12})$; T_d: dew point temperature; RH_t: relative humidity at time t; RH_{t-1}: relative humidity 15 minute before time t; RH_x: $(RH_t, RH_{t-1}, RH_{t-2}, ..., RH_{t-12})$; R_t: rainfall amount at time t; R_t^d: daily accumulated rainfall ending at time t; U_t: wind speed at time t; P_{t-1}: wind speed 15 minute before time t; U_x: $(U_t, U_{t-1}, U_{t-2}, ..., U_{t-12})$; $U_{\max}$: maximum wind speed

implemented using Apache Spark™

Quality control checks

Table 1: Quality control procedure applied to ASTI automated weather stations for 15-minute data.

Validation procedures	Air pressure (hPa)	Air temperature (°C)	Relative humidity (%)	Rainfall (mm)	Wind speed (km/h)	...
Timestamp check			$\lvert t - \text{round}\,(t,p)\rvert \leq 2.5$ minutes			...
			$\text{rank}\,(\lvert t - \text{round}\,(t,p)\rvert) = 1$			...
Range check	$640 \leq P_t \leq 1022$ (PAGASA, 2014)	$0 \leq T_t \leq 50$ (PAGASA, 2014)	$0 \leq RH_t \leq 95$	$0 \leq R_t \leq 30$ (PAGASA, 2014)	$0 \leq U_t \leq 270$	...
				$0 \leq R_t^d \leq 600$		...
Step check	$\lvert P_t - P_{t-1}\rvert \leq 0.6$	$\lvert T_t - T_{t-1}\rvert \leq 1.5$	$\lvert RH_t - RH_{t-1}\rvert \leq 30$		$\lvert U_t - U_{t-1}\rvert \leq 75$	...
Persistence check	$\left\langle (P_x - \langle P_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$	$\left\langle (T_x - \langle T_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$	$\left\langle (RH_x - \langle RH_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$		$\left\langle (U_x - \langle U_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.4$	...
Internal consistency check		$T > T_d$ (Zahumensky, 2004)	rain > 0 and value > 60	value $= 0$ and duration $= 0$	speed $= 0$ and direction $= 0$	...
				value > 0 and duration > 0	speed $\neq 0$ and direction $\neq 0$	...
				(Zahumensky, 2004)	(Zahumensky, 2004)	...
				value > 0 and RH > 60		...
					$U_t < U_{\max}$	...
Spatial check			not yet implemented			...

t: time stamp of a measurement record; p: sampling period of the station in minutes; round(t,p): rounded timestamp based on p; P_t: air pressure at time t; P_{t-1}: air pressure 15 minute before time t; P_x: $(P_t, P_{t-1}, P_{t-2}, ..., P_{t-12})$; T_t: air temperature at time t; T_{t-1}: air temperature 15 minute before time t; T_x: $(T_t, T_{t-1}, T_{t-2}, ..., T_{t-12})$; T_d: dew point temperature; RH_t: relative humidity at time t; RH_{t-1}: relative humidity 15 minute before time t; RH_x: $(RH_t, RH_{t-1}, RH_{t-2}, ..., RH_{t-12})$; R_t: rainfall amount at time t; R_t^d: daily accumulated rainfall ending at time t; U_t: wind speed at time t; P_{t-1}: wind speed 15 minute before time t; U_x: $(U_t, U_{t-1}, U_{t-2}, ..., U_{t-12})$; $U_{\max}$: maximum wind speed

implemented using Apache Spark™

Quality control checks

Table 1: Quality control procedure applied to ASTI automated weather stations for 15-minute data.

Validation procedures	Air pressure (hPa)	Air temperature (°C)	Relative humidity (%)	Rainfall (mm)	Wind speed (km/h)	...
Timestamp check			$\lvert t - \text{round}\,(t,p)\rvert \leq 2.5$ minutes			...
			$\text{rank}\,(\lvert t - \text{round}\,(t,p)\rvert) = 1$			...
Range check	$640 \leq P_t \leq 1022$ (PAGASA, 2014)	$0 \leq T_t \leq 50$ (PAGASA, 2014)	$0 \leq RH_t \leq 95$	$0 \leq R_t \leq 30$ (PAGASA, 2014)	$0 \leq U_t \leq 270$	...
				$0 \leq R_t^d \leq 600$		...
Step check	$\lvert P_t - P_{t-1}\rvert \leq 0.6$	$\lvert T_t - T_{t-1}\rvert \leq 1.5$	$\lvert RH_t - RH_{t-1}\rvert \leq 30$		$\lvert U_t - U_{t-1}\rvert \leq 75$	...
Persistence check	$\left\langle (P_x - \langle P_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$	$\left\langle (T_x - \langle T_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$	$\left\langle (RH_x - \langle RH_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$		$\left\langle (U_x - \langle U_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.4$	...
Internal consistency check		$T > T_d$ (Zahumensky, 2004)	rain > 0 and value > 60	value $= 0$ and duration $= 0$	speed $= 0$ and direction $= 0$	...
				value > 0 and duration > 0	speed $\neq 0$ and direction $\neq 0$	...
				(Zahumensky, 2004)	(Zahumensky, 2004)	...
				value > 0 and RH > 60		...
					$U_t < U_{\max}$	...
Spatial check			not yet implemented			...

t: time stamp of a measurement record; p: sampling period of the station in minutes; round(t,p): rounded timestamp based on p; P_t: air pressure at time t; P_{t-1}: air pressure 15 minute before time t; P_x: $(P_t, P_{t-1}, P_{t-2}, ..., P_{t-12})$; T_t: air temperature at time t; T_{t-1}: air temperature 15 minute before time t; T_x: $(T_t, T_{t-1}, T_{t-2}, ..., T_{t-12})$; T_d: dew point temperature; RH_t: relative humidity at time t; RH_{t-1}: relative humidity 15 minute before time t; RH_x: $(RH_t, RH_{t-1}, RH_{t-2}, ..., RH_{t-12})$; R_t: rainfall amount at time t; R_t^d: daily accumulated rainfall ending at time t; U_t: wind speed at time t; P_{t-1}: wind speed 15 minute before time t; U_x: $(U_t, U_{t-1}, U_{t-2}, ..., U_{t-12})$; $U_{\max}$: maximum wind speed

implemented using Apache Spark™

Quality control checks

Table 1: Quality control procedure applied to ASTI automated weather stations for 15-minute data.

Validation procedures	Air pressure (hPa)	Air temperature (°C)	Relative humidity (%)	Rainfall (mm)	Wind speed (km/h)	...
Timestamp check			$\lvert t - \text{round}\,(t,p)\rvert \leq 2.5$ minutes			...
			$\text{rank}\,(\lvert t - \text{round}\,(t,p)\rvert) = 1$			...
Range check	$640 \leq P_t \leq 1022$ (PAGASA, 2014)	$0 \leq T_t \leq 50$ (PAGASA, 2014)	$0 \leq RH_t \leq 95$	$0 \leq R_t \leq 30$ (PAGASA, 2014)	$0 \leq U_t \leq 270$	...
				$0 \leq R_t^d \leq 600$		...
Step check	$\lvert P_t - P_{t-1}\rvert \leq 0.6$	$\lvert T_t - T_{t-1}\rvert \leq 1.5$	$\lvert RH_t - RH_{t-1}\rvert \leq 30$		$\lvert U_t - U_{t-1}\rvert \leq 75$	...
Persistence check	$\left\langle (P_x - \langle P_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$	$\left\langle (T_x - \langle T_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$	$\left\langle (RH_x - \langle RH_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$		$\left\langle (U_x - \langle U_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.4$	...
Internal consistency check		$T > T_d$ (Zahumensky, 2004)	rain > 0 and value > 60	value $= 0$ and duration $= 0$	speed $= 0$ and direction $= 0$	...
				value > 0 and duration > 0	speed $\neq 0$ and direction $\neq 0$	...
				(Zahumensky, 2004)	(Zahumensky, 2004)	...
				value > 0 and RH > 60		...
					$U_t < U_{\max}$	...
Spatial check	not yet implemented					...

t: time stamp of a measurement record; p: sampling period of the station in minutes; round(t,p): rounded timestamp based on p; P_t: air pressure at time t; P_{t-1}: air pressure 15 minute before time t; P_x: $(P_t, P_{t-1}, P_{t-2}, ..., P_{t-12})$; T_t: air temperature at time t; T_{t-1}: air temperature 15 minute before time t; T_x: $(T_t, T_{t-1}, T_{t-2}, ..., T_{t-12})$; T_d: dew point temperature; RH_t: relative humidity at time t; RH_{t-1}: relative humidity 15 minute before time t; RH_x: $(RH_t, RH_{t-1}, RH_{t-2}, ..., RH_{t-12})$; R_t: rainfall amount at time t; R_t^d: daily accumulated rainfall ending at time t; U_t: wind speed at time t; P_{t-1}: wind speed 15 minute before time t; U_x: $(U_t, U_{t-1}, U_{t-2}, ..., U_{t-12})$; $U_{\max}$: maximum wind speed

implemented using Apache Spark™

Quality control checks

Table 1: Quality control procedure applied to ASTI automated weather stations for 15-minute data.

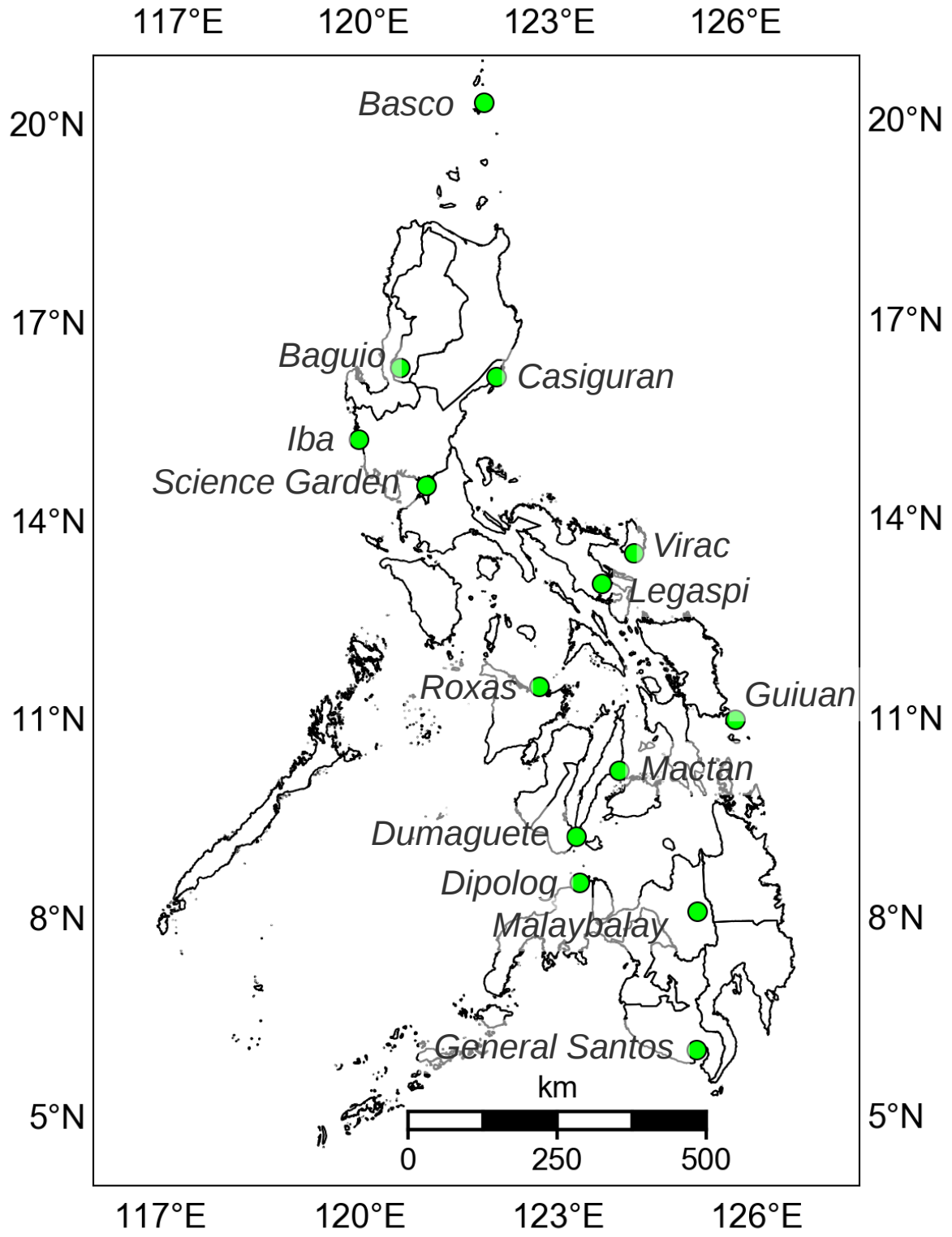
Validation procedures	Air pressure (hPa)	Air temperature (°C)	Relative humidity (%)	Rainfall (mm)	Wind speed (km/h)	...
Timestamp check			$\lvert t - \text{round}\,(t,p)\rvert \leq 2.5$ minutes			...
			$\text{rank}\,(\lvert t - \text{round}\,(t,p)\rvert) = 1$			...
Range check	$640 \leq P_t \leq 1022$ (PAGASA, 2014)	$0 \leq T_t \leq 50$ (PAGASA, 2014)	$0 \leq RH_t \leq 95$	$0 \leq R_t \leq 30$ (PAGASA, 2014)	$0 \leq U_t \leq 270$	...
				$0 \leq R_t^d \leq 600$		...
Step check	$\lvert P_t - P_{t-1}\rvert \leq 0.6$	$\lvert T_t - T_{t-1}\rvert \leq 1.5$	$\lvert RH_t - RH_{t-1}\rvert \leq 30$		$\lvert U_t - U_{t-1}\rvert \leq 75$	...
Persistence check	$\left\langle (P_x - \langle P_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$	$\left\langle (T_x - \langle T_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$	$\left\langle (RH_x - \langle RH_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.1$		$\left\langle (U_x - \langle U_x\rangle)^2 \right\rangle^{\frac{1}{2}} \geq 0.4$	...
Internal consistency check		$T > T_d$ (Zahumensky, 2004)	rain > 0 and value > 60	value $= 0$ and duration $= 0$	speed $= 0$ and direction $= 0$	...
				value > 0 and duration > 0	speed $\neq 0$ and direction $\neq 0$	...
				(Zahumensky, 2004)	(Zahumensky, 2004)	...
				value > 0 and RH > 60		...
					$U_t < U_{\max}$	...
Spatial check	not yet implemented					...

t: time stamp of a measurement record; p: sampling period of the station in minutes; round(t,p): rounded timestamp based on p; P_t: air pressure at time t; P_{t-1}: air pressure 15 minute before time t; P_x: $(P_t, P_{t-1}, P_{t-2}, ..., P_{t-12})$; T_t: air temperature at time t; T_{t-1}: air temperature 15 minute before time t; T_x: $(T_t, T_{t-1}, T_{t-2}, ..., T_{t-12})$; T_d: dew point temperature; RH_t: relative humidity at time t; RH_{t-1}: relative humidity 15 minute before time t; RH_x: $(RH_t, RH_{t-1}, RH_{t-2}, ..., RH_{t-12})$; R_t: rainfall amount at time t; R_t^d: daily accumulated rainfall ending at time t; U_t: wind speed at time t; P_{t-1}: wind speed 15 minute before time t; U_x: $(U_t, U_{t-1}, U_{t-2}, ..., U_{t-12})$; $U_{\max}$: maximum wind speed

implemented using Apache Spark™

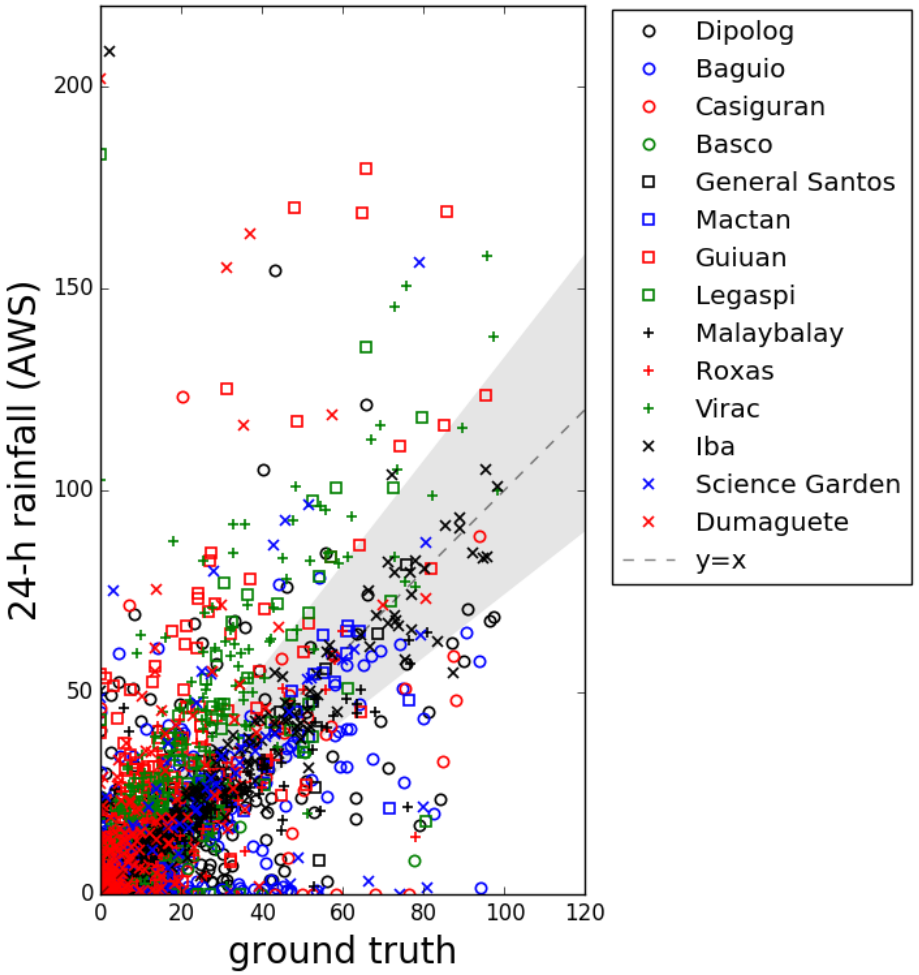
Experiment

- 14 selected stations
- Ground truth obtained from PAGASA synoptic station measurements
- Period: 2012 – 2015
- Variables: rainfall and temperature



** Using 14 test automated weather stations (AWS)*

	RMSE (before QC)	RMSE (after QC)	RMSE difference
Rainfall (mm/d)	14.36	12.04	**2.32** ↓
Temperature (°C)	1.77	1.01	**0.76** ↓



But there's still work to be done.

QC statistics

Table 3: Percentage of flagged meteorological data using the quality control procedure. Data were collected from all automated weather stations from October 2012 to December 2015.

Variables	Total obs	Tests				
		Timestamp	Range	Step	Persistence	Internal consistency
Air pressure	58,228,395	6.29%	1.21%	1.59%	0.12%	—
Air temperature	4,390,234	0.18%	1.41%	0.53%	0.27%	0.00%
Relative humidity	4,390,234	0.18%	3.36%	0.09%	1.79%	0.38%
Rainfall	59,825,488	6.89%	0.03%	—	—	0.41%
Wind speed	4,390,234	0.18%	0.00%	0.00%	65.67%	5.07%
Total	131,224,585					

Future work

- continuous improvements to the algorithm (e.g. *spatial consistency check*)

- algorithm optimization

- **more importantly, to complete the quality control system pipeline, distribution of the quality-controlled data to end-users.**

Summary

- Data quality control is important
- Correct observation data is crucial to forecasters and decision makers
- Garbage in, garbage out

Thank you for listening. :)