Statistical Forecasting of Rainfall from Radar Reflectivity in Singapore

Xiao Liu[1], Vikneswaran Gopal[2] and Reza Hosseini[1]

[1] IBM Research Collaboratory Singapore
[2] Department of Applied Probability and Statistics, National University of Singapore

22nd Conference on Probability and Statistics
in the Atmospheric Sciences
AMS Annual Meeting, Atlanta, 2014
Session: *Statistical model development, statistical forecasting approaches, and ensemble forecasting part 1*

Agenda

- **Introduction**
  - Background and Objectives
  - Data
- **Marshall-Palmer relationship**
- **Recalibration of the Z-R relationship using least squares**
- **Recalibration of the Z-R relationship using quantile regression**
- **Optimization of the Z-R relationship for Singapore**
- **Conclusions**

Other Presentations of Related Work

<u>Conference on Numerical Weather Prediction:</u>

- 9.3. Ensemble Kalman Filter (EnKF) Assimilating the Dropsonde Observations to Reduce the Forecast Track Error of Typhoon Soulik (2013) Based On the Cloud-resolving Model
- 13.2. Recent Advances in High-Resolution Operational NWP, Utilizing WRF-ARW

<u>Conference on Artificial and Computational Intelligence and its Applications to the Environmental Sciences</u>:

- J3.2. A multi-scale solar energy forecast platform based on machine-learned adaptive combination of expert systems

<u>Conference on Climate Variability and Change</u>:

- 8C.4. Simulation of the temporal and spatial characteristics of diurnal rainfall cycle over Borneo

<u>Conference on Weather, Climate, and the New Energy Economy</u>:

- 6.3. Enabling Advanced Weather Modelling and Data Assimilation for Utility Distribution Operations
- 8.1. Outage Prediction and Response Optimization (OPRO)
- 9.1. Very High Resolution Coupled Weather and Wind Power Modeling
- 10.1. Improvements in short-term solar energy forecasting
- 10.2. Two methods in improving onshore wind forecast

<u>Symposium on Advances in Modeling and Analysis Using Python</u>:

- 3.5. A Python-Based Automatic Data Aggregation Framework for Hydrology Models

<u>Superstorm Sandy and the Built Environment: New Perspectives, Opportunities, and Tools</u>:

- 873. Forecast Performance of an Operational Mesoscale Modeling System for Post- Tropical Storm Sandy in the New York City Metropolitan Region

<u>Conference on Probability and Statistics in the Atmospheric Sciences</u>

- 4.2. Customized Verification Applied to High-Resolution WRF-ARW Forecasts for Rio de Janeiro (Thursday)

<u>Symposium on the Urban Environment</u>

- J12.2. High-Resolution, Coupled Hydro-Meteorological Modelling for Operational Forecasting of Severe Flooding Events in Rio de Janeiro

Background

- A joint development between IBM Research and the National Environmental Agency (NEA), Singapore

- Objectives:

1. Recalibrate the Z-R relationship for Singapore
2. Compare different methods that convert radar reflectivity factor (Z) to rainfall intensity (R)
3. Optimize model parameters of the Z-R relationship for Singapore

- <u>Our research is still on-going, and preliminary findings are presented</u>

About Singapore

- Singapore is located at the southern tip of the Malay Peninsula and is 137 kilometres (85 mi) north of the equator.

- Average annual rainfall ~ 2,342mm (92 inches)
  - Rio de Janeiro ~46 inches
  - New York: ~50 inches (total precipitation)
  - Seattle: ~38 inches (total precipitation)

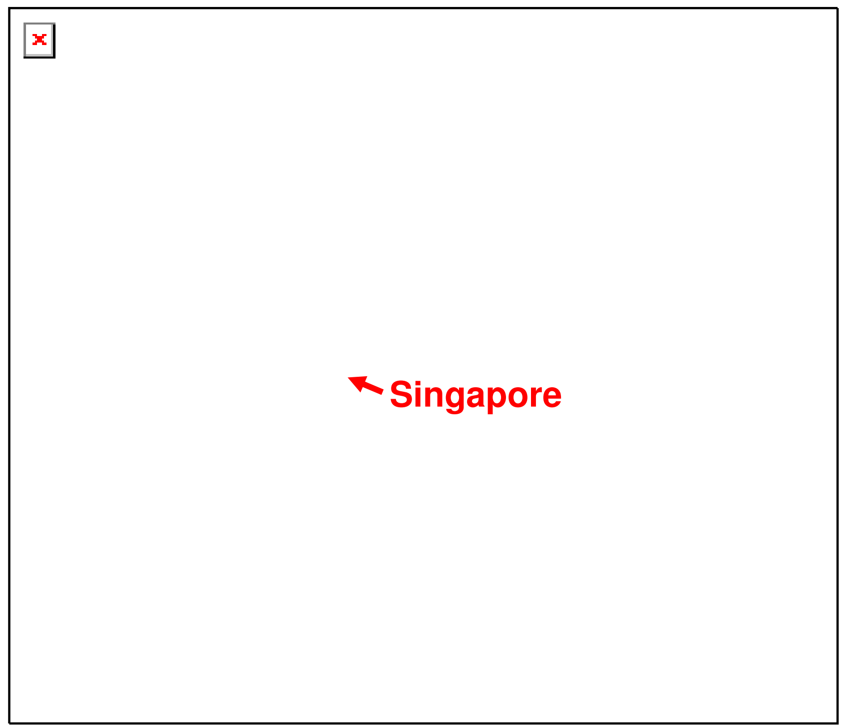
- The “Four Seasons” in Singapore
  - North-East Monsoon Season (Dec to Mar)
  - Inter-Monsoon Season (Apr and May)
  - South-West Monsoon Season (Jun and Sep)
  - Inter-Monsoon Season (Oct and Nov)

- This study focuses on the relationship between (radar) reflectivity and rainfall rate (i.e., the Z-R relationship) for Singapore during the two inter-monsoon seasons

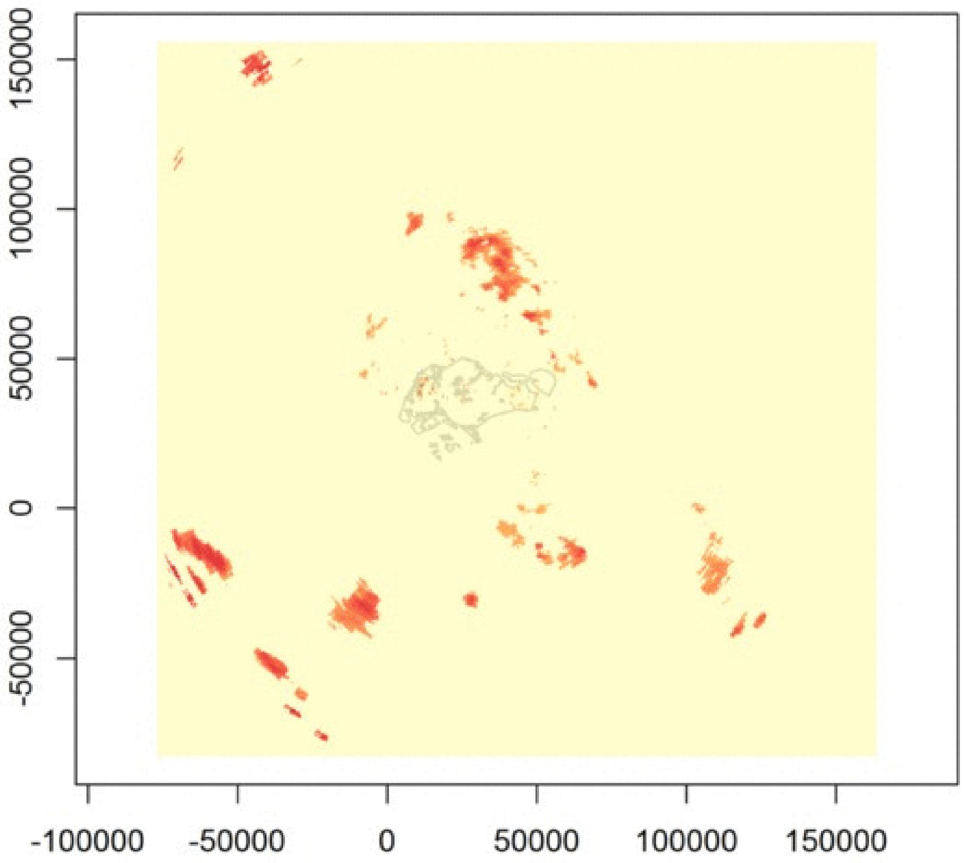
Data

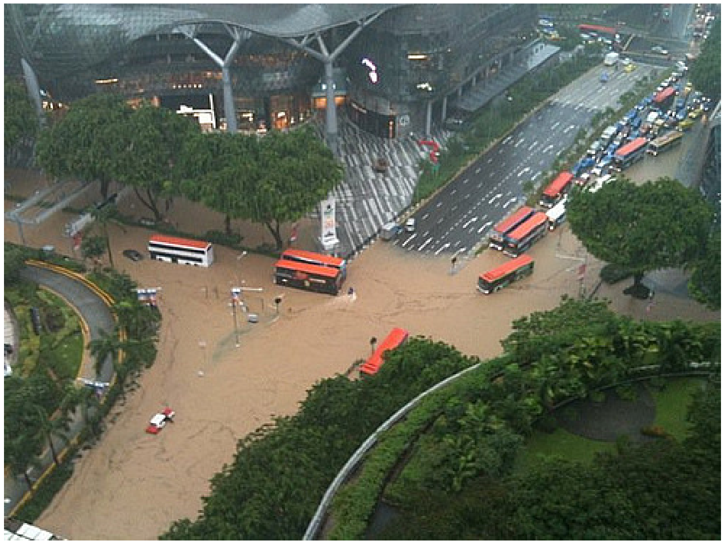
- **Weather radar reflectivity data:**
  - **30 heavy rain events in 2010 and 46 heavy rain events in 2011**
  - **Cartesian grid of 480 by 480 pixels**
  - **Top left corner: E102.892, N2.42799**
  - **Lower right: E105.052, N0.269748**
  - **Spatial resolution: 0.5 by 0.5 kilometres**
  - **Sampling frequency: 5min**

- **Illustration:**



Example: A typical inter-monsoon season convective storm



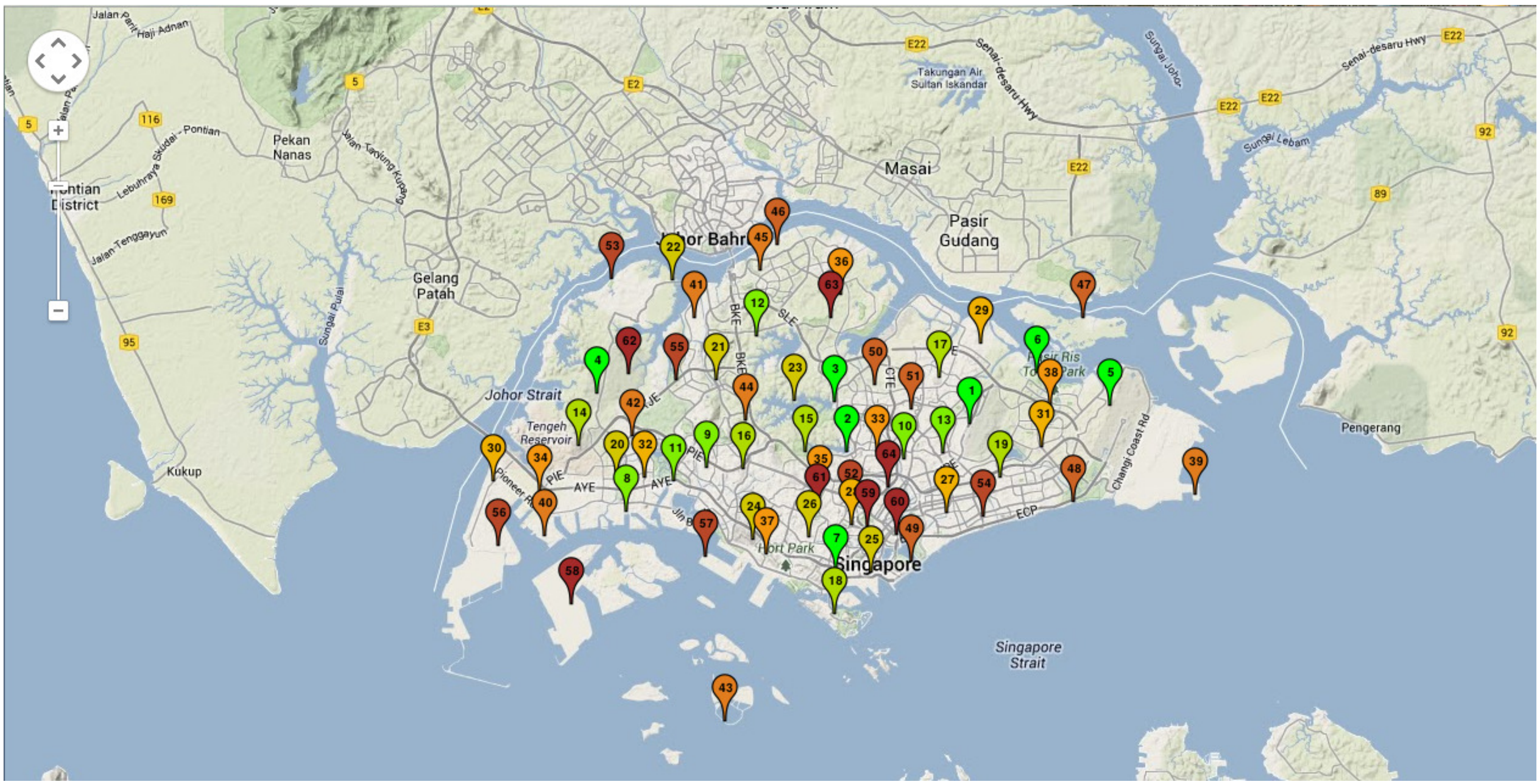


Key features:

- **Developed quickly – challenge for prediction**
- **Lifetime: 1 to 3 hours**
- **Heavy rainfall**

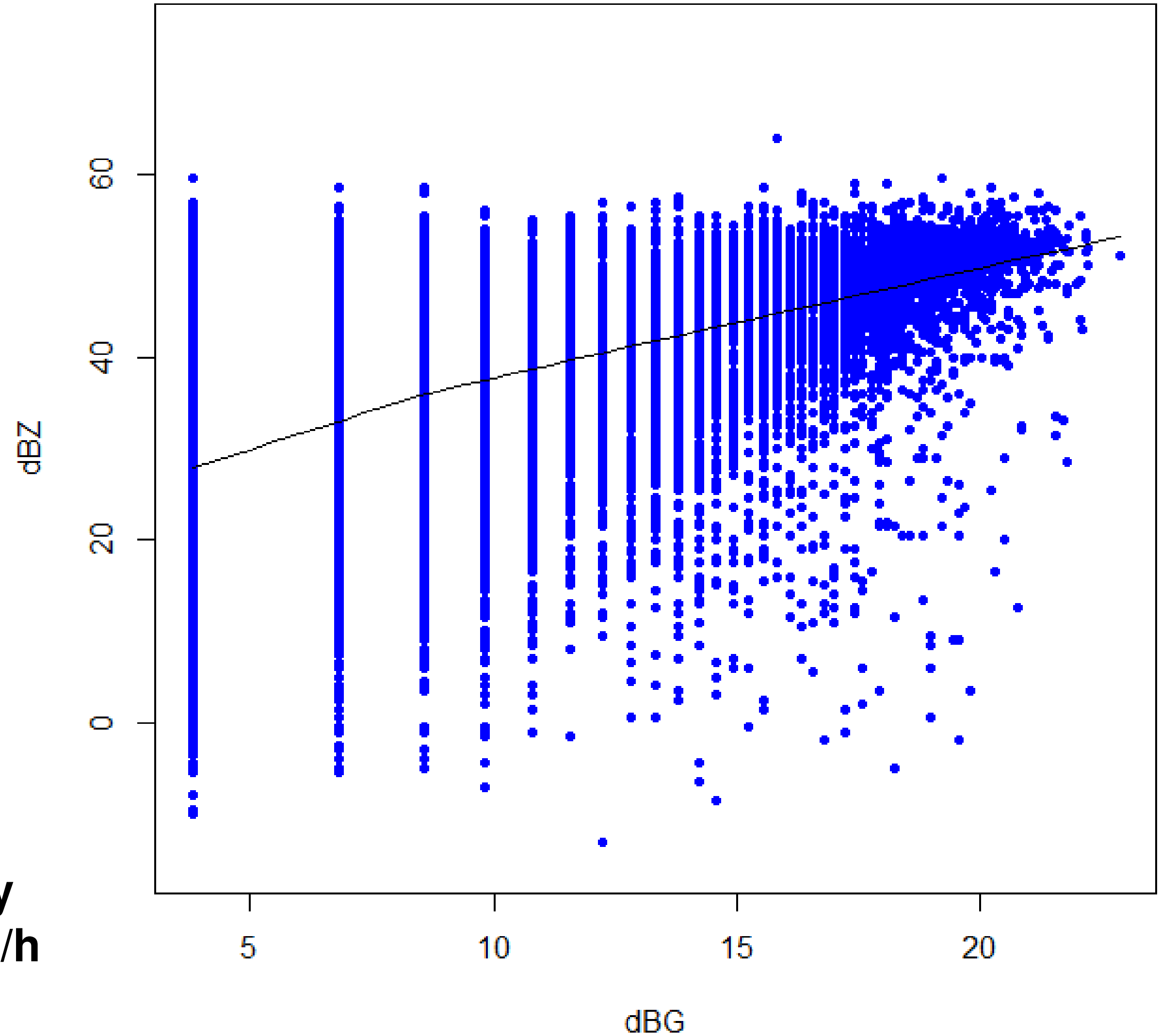
Precipitation Data

- **Rain Gauge Data:**
  - Number of stations: **64**
  - Sampling frequency: **5 minutes**
- **Illustration:**



Precipication and Radar Data

- **Look at dBZ-dBG pairs**
  - **46 inter-monsoon heavy rains in 2011**
  - **88274 pairs**
  - **70008 pairs with zero rainfall**



dBG is based on hourly rainfall intensity in mm/h

Marshall-Palmer Relationship

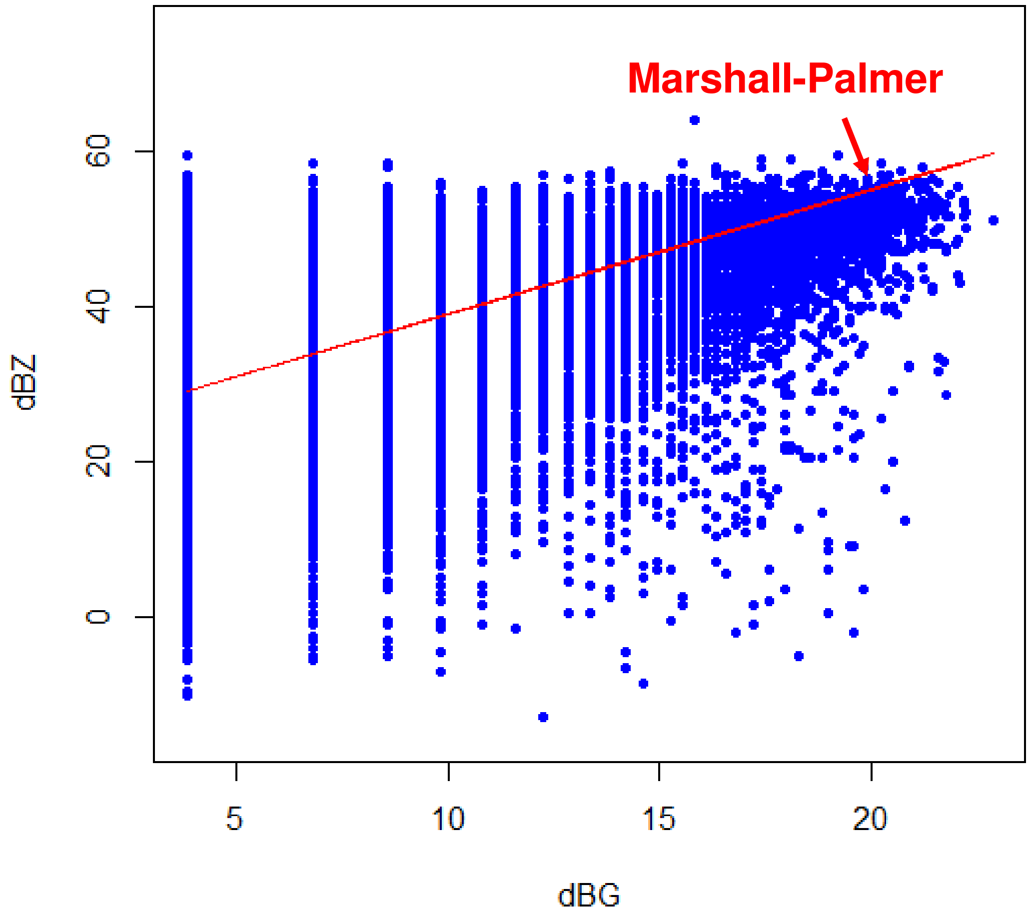
- **MP relationship (Marshall-Palmer, 1948)**

$$Z = aR^b$$

where $a = 200, b = 1.6$

R : rainfall intensity, mm/h; Z : reflectivity factor, $mm^{-6}m^3$

- **Fitting result:**



Performance Assessment of the MP Relationship

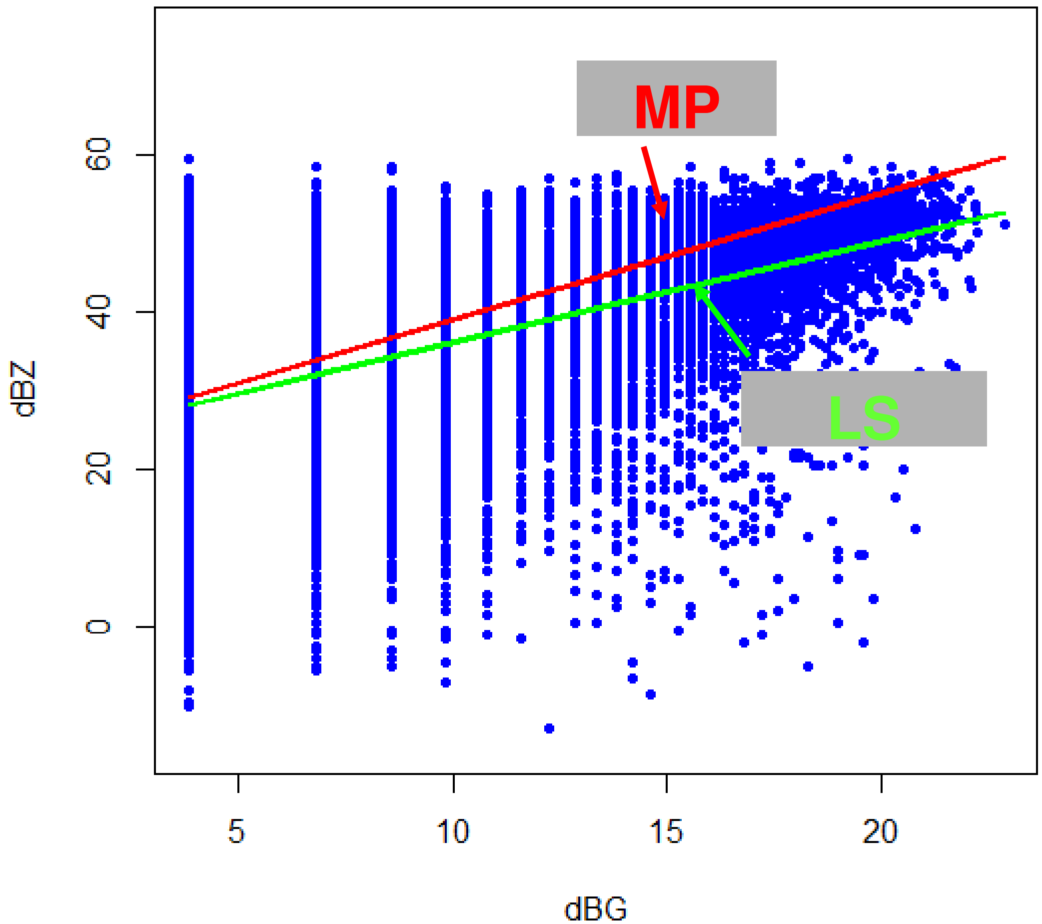
	Marshall-Palmer (tested on 2010 data)
POD (20mm/h ~ 30mm/h)	32%
FAR (20mm/h ~ 30mm/h)	69%
POD (30mm/h ~ 50mm/h)	34%
FAR (30mm/h ~ 50mm/h)	53%
POD (50mm/h ~ 70mm/h)	9%
FAR (50mm/h ~ 70mm/h)	66%

POD: Probability of Detection **FAR: False Alarm Rate**

Least Squares

- **Recalibrate the Z-R relationship for Singapore**
- **Method 1: Least squares (LS)**
  - **Advantages:** simple and commonly used
  - **Disadvantages:** the normality assumption of residuals is violated; sensitive to outliers.
- **Fitting Results:**

$$Z = 214R^{1.28}$$



Performance Assessment and Comparison

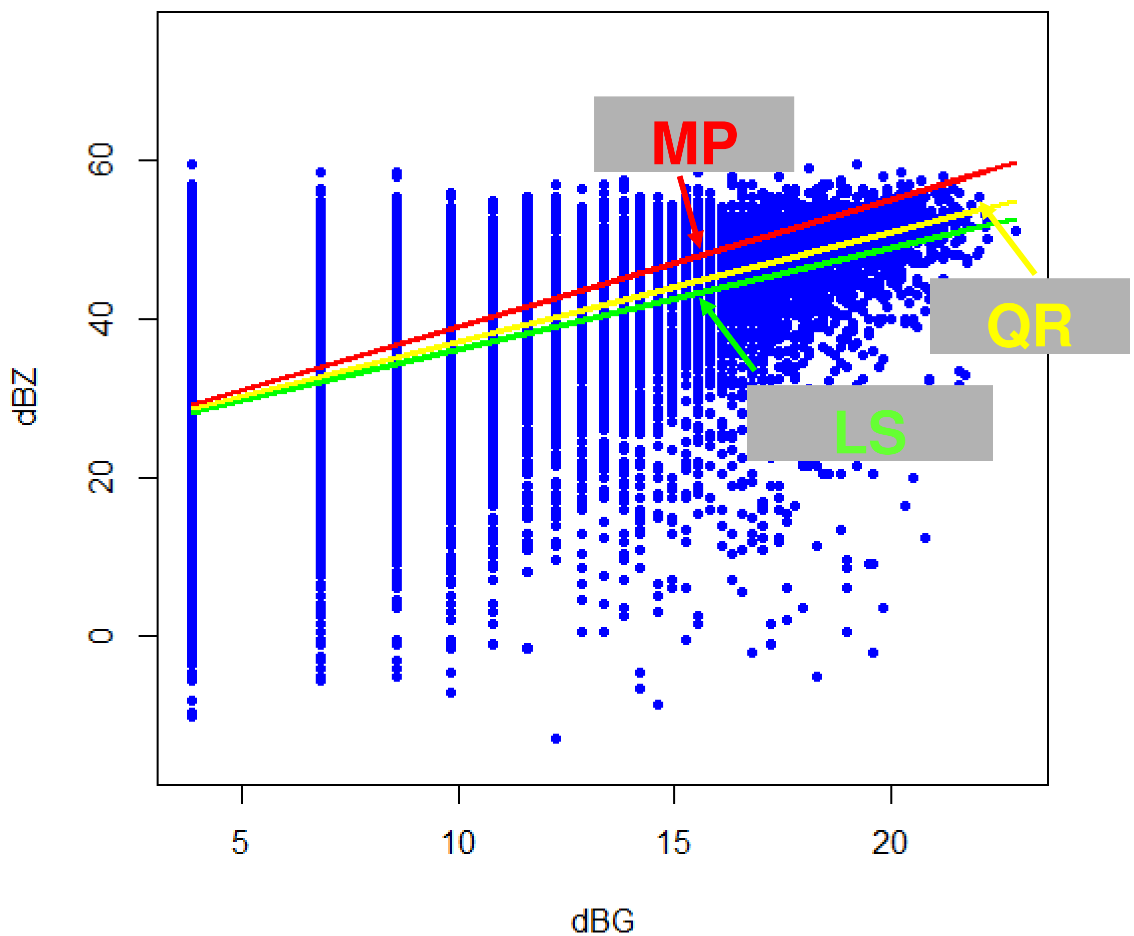
	Marshall-Palmer (tested on 2010 data)	Least Squares (fitted using 2011 data; tested on 2010 data)
POD (20mm/h ~ 30mm/h)	32%	20%
FAR (20mm/h ~ 30mm/h)	69%	89%
POD (30mm/h ~ 50mm/h)	34%	26%
FAR (30mm/h ~ 50mm/h)	53%	90%
POD (50mm/h ~ 70mm/h)	9%	20%
FAR (50mm/h ~ 70mm/h)	66%	95%

- Although the LS method minimizes the sum of squared error, it is apparently NOT a good choice if the goal is to estimate the rainfall intensity from reflectivity.

Quantile Regression

- **Recalibrate the Z-R relationship for Singapore**
- **Method 2: Quantile Regression (QR)**
  - **Advantages:** Robust against outliers; Outperforms least squares when the normality assumption is violated
- **Fitting Results:**

$$Z = 211R^{1.38}$$



Performance Assessment and Comparison

	Marshall-Palmer (tested on 2010 data)	Least Squares (fitted using 2011 data; tested on 2010 data)	Quantile Regression (fitted using 2011 data; tested on 2010 data)
POD (20mm/h ~ 30mm/h)	32%	20%	28%
FAR (20mm/h ~ 30mm/h)	69%	89%	79%
POD (30mm/h ~ 50mm/h)	34%	26%	41%
FAR (30mm/h ~ 50mm/h)	53%	90%	75%
POD (50mm/h ~ 70mm/h)	9%	20%	48%
FAR (50mm/h ~ 70mm/h)	66%	95%	85%

- In terms of POD, the Quantile Regression outperforms the other two for heavy rainfall prediction, especially when the intensity is larger than 50mm/h
- In terms of FAR, the Quantile Regression is not as good as the default MP relationship.
- In general, considering the significant improvement of POD by the quantile regression, we still think the quantile regression outperforms despite the relatively larger FAR.

Optimize Model Parameters

- **Both POD and FAR are determined by the values of a and b**
- **Why not to find the optimum values of a and b that maximize POD subject to the maximum FAR constraint?**
- **In this study, we are particularly interested in predicting extremely heavy rainfall events with intensity within 50~70 mm/h. This leads to the following optimization problem.**

POD for rainfall intensity within 50~70mm/h (to be maximized)

$$\begin{aligned} &\max \quad \mathrm{POD}(a,b) \\ &\text{s.t.} \quad \mathrm{FAR}(a,b) \leq 0.66 \end{aligned}$$

FAR of MP relationship for rainfall intensity within 50~70mm/h

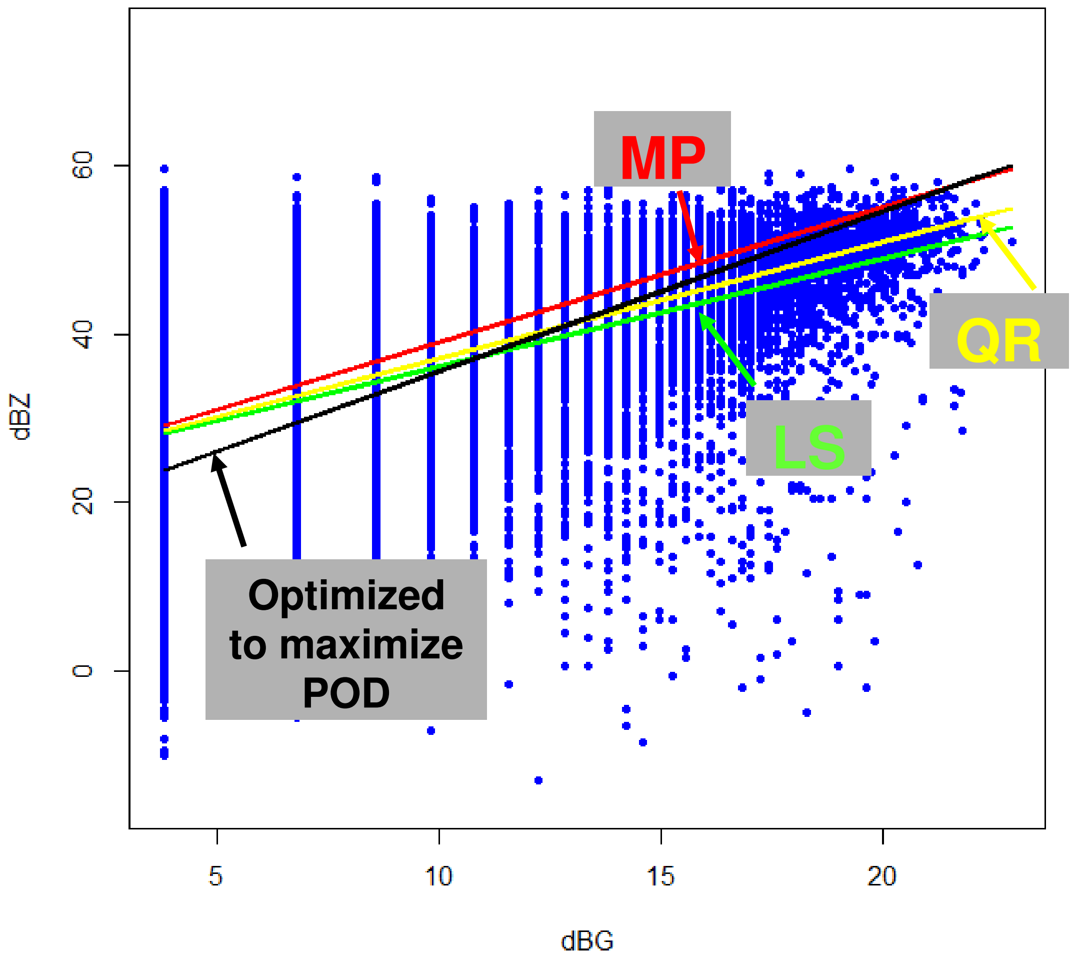
We are searching for a and b that 1) maximize the POD for rainfall intensity within 50~70mm/h, and 2) with FAR not greater than that of the default MP relationship

Optimize Model Parameters

- Recalibrate the Z-R relationship for Singapore
- Method 3: Maximization of POD for Rainfall Intensity within 50-70mm/h

- Results:

$$Z = 45R^{1.9}$$



Performance Assessment and Comparison

	Marshall-Palmer (tested on 2010 data)	Least Squares (fitted using 2011 data; tested on 2010 data)	Quantile Regression (fitted using 2011 data; tested on 2010 data)	With Optimized a and b (fitted using 2011 data; tested on 2010 data)
POD (20~30mm/h)	32%	20%	28%	44%
FAR (20~30mm/h)	69%	89%	79%	70%
POD (30~50mm/h)	34%	26%	41%	55%
FAR (30~50mm/h)	53%	90%	75%	59%
POD (50~70mm/h)	9%	20%	48%	25%
FAR (50~70mm/h)	66%	95%	85%	59%

Conclusions:

1. The Z-R relationship in Singapore can be significantly improved over the default Marshall-Palmer relationship;
2. When the goal is to predict the rainfall intensity from reflectivity, the optimum values of a and b are those that maximize the POD subject to the maximum FAR constraint;
3. It is interesting to see that a good fitting of the dBZ-dBG pairs (such as least squares or quantile regression) does not necessarily imply high accuracy in rainfall prediction based on reflectivity.

Next Steps

- Continue the refinement of the optimization
- Evaluate against additional events
- Apply categorical metrics based upon rainfall intensity
- Experiment with quasi-operational deployment