A Machine Learning Framework for ARM Data Quality Analysis

Application: MWR Rain Contamination Detection

SHUAIQI TANG

Lawrence Livermore National Laboratory

2019 ARM/ASR PI meeting, Rockville, MD, 10-13 June 2019

This work was performed under the auspices of the U. S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Release Number: LLNL-PRES-776810

Participants of the LLNL ML project

Dr. Shaocheng Xie
PI of the project
Research Scientist
GL at **LLNL**

Dr. Charan Thimmisetty
Data Scientist
at **CISCO**, **Xiao**'s past postdoc

Dr. Shuaiqi Tang
Research Scientist
at **LLNL**

Dr. Xiao Chen
Co-PI
Data Scientist
at **LLNL**

A ML Framework for ARM Data Quality Analysis

Dr. Dave Turner
PM at **NOAA**

Dr. Maria Cadeddu
Principal Atmospheric Engineer
at **ANL**

Rain and post-rain contamination of MWR retrieved LWP

3-Channel MWR (MWR3C)
2011 - current

MWR Radome in cloudy sky condition[1]

MWR Radome during and after Rain[1]

- MWR measures brightness temperature (Tb) and uses them to retrieve cloud liquid water path (LWP) and other variables.
- LWP retrieved from MWR is important for cloud parameterization development and validation.
- MWR retrieved LWP is contaminated by rain water on the radome.
- Water may be present on the radome even after the rain stops
- Rain flag is not enough to identify the rain contamination period

[1]Ada Vittoria Bosisio and Maria P. Cadeddu ,“lensContamination”.

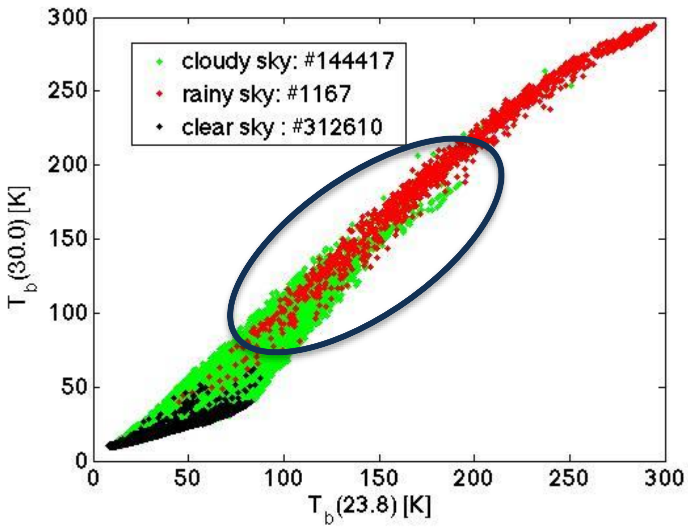
Possible methods to detect rain contamination

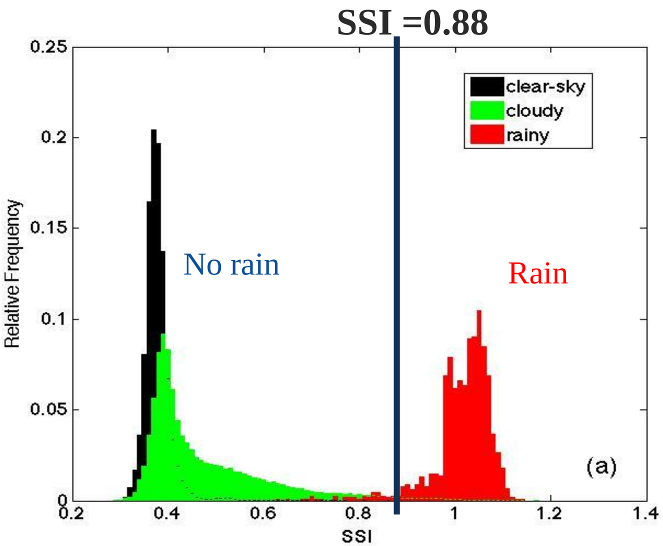
- Method 1: Tb_{23} < 100K and Tb_{30} < 100K (applied in mwrret.c2)
- Method 2: SSI < 0.88 (from *Maria P. Cadeddu (ANL) et al.)*

Tb_{89} is not used

$$SSI = \frac{Tb_{30} - c_0}{Tb_{23}}$$

*c*0 is the intercept of the straight line *Tb*(*30*) = *c*0 + c1**Tb*(*23*) that relates the two values under **clear sky conditions**

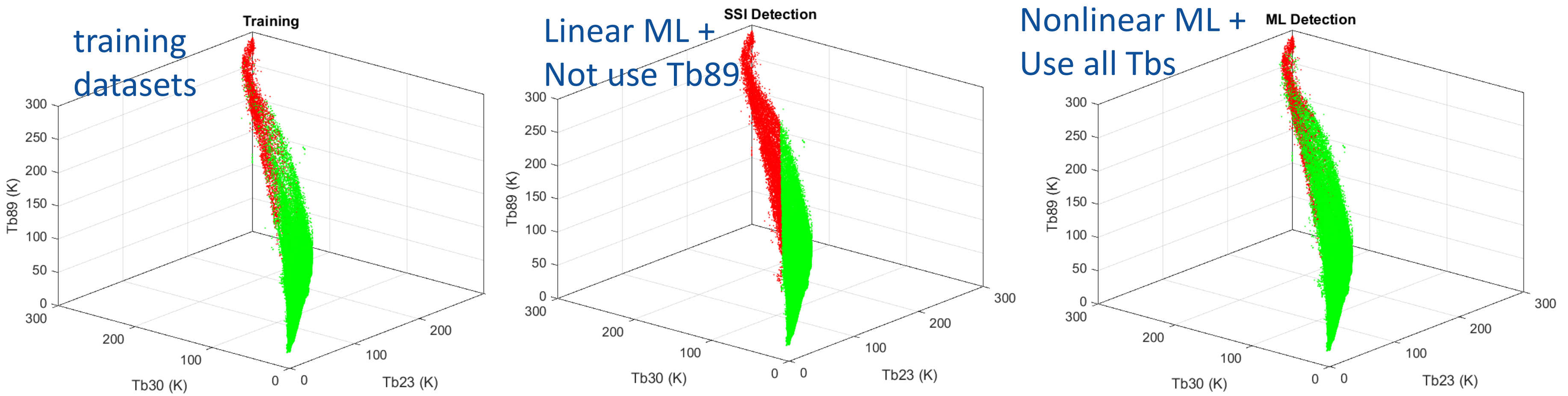




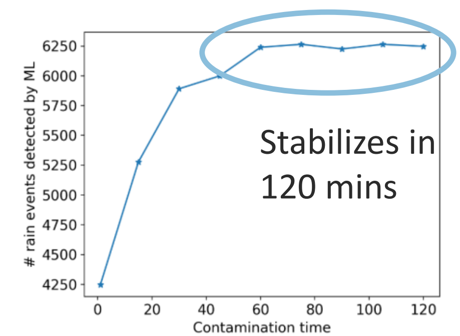
Linear classification, not well separated

A non-linear machine learning framework to detect contamination of water on the radome

- **support vector machine (SVM):** nonlinear kernel transformation.
- Use all three channels (Tb_{23}, Tb_{30}, Tb_{89}).



Rain & contamin NoContamin



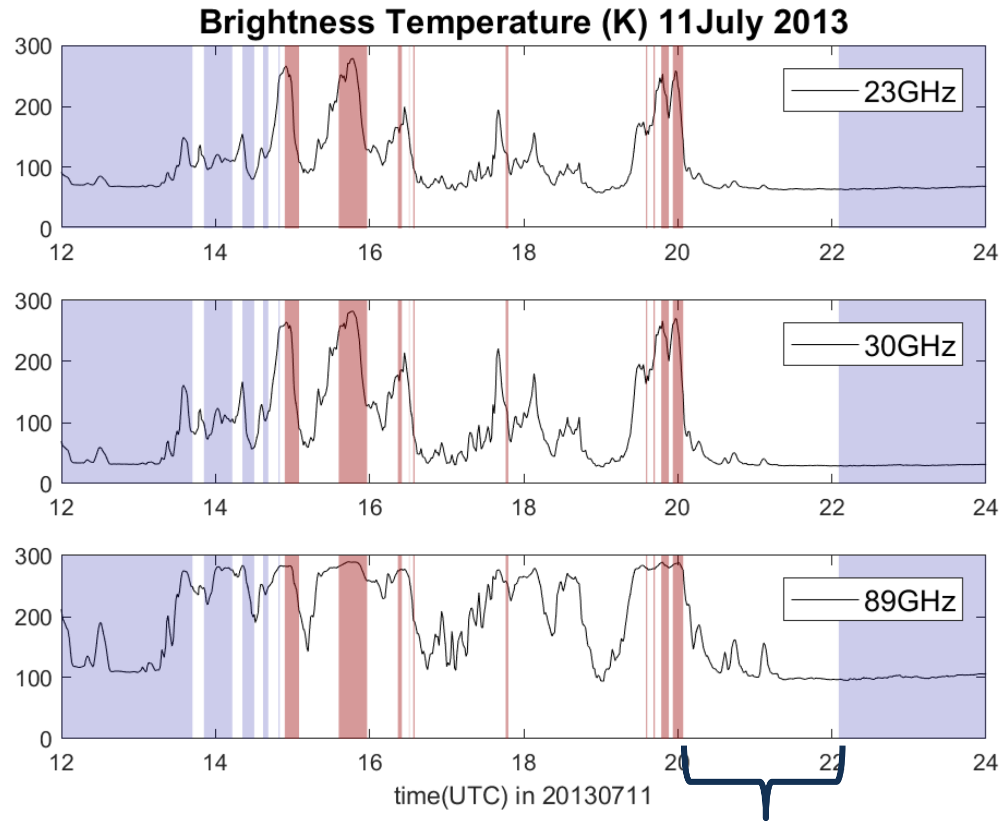
- Clean training datasets (excluding 2-hr data after a rain event)

U.S. DEPARTMENT OF ENERGY

Improved training data by excluding potential contamination data

Training datasets are separated into two categories:

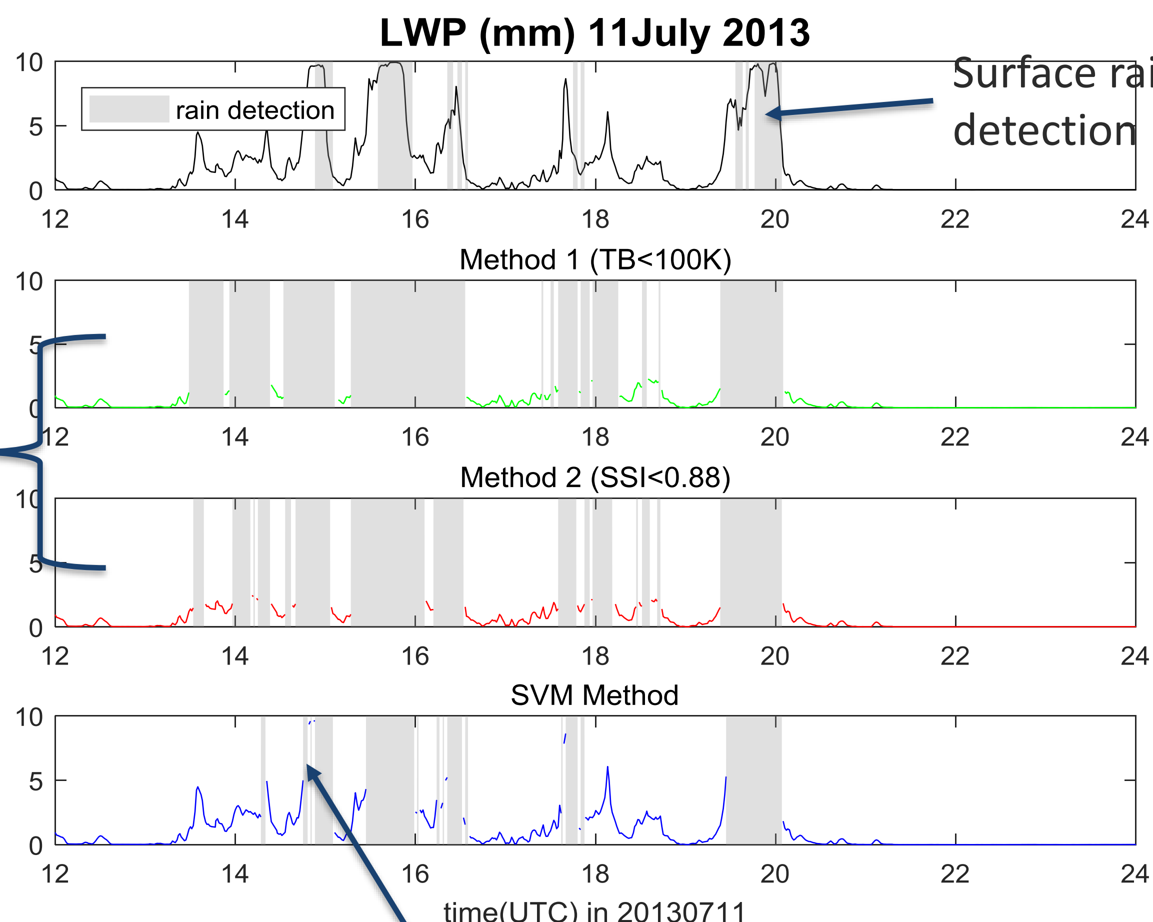
1. rainy: RR>2mm/hr (red)
2. Norain: RR=0 & rain_gauge=0 (blue)



2-hour data after rain is excluded

1-year (2013) Tb data are used to train the ML model.

Test results in 11 July, 2013



Detect long contamination period, remove all large LWP values

Less contamination detection, cover all rain periods, keep some large LWP values

Test results in the full year 2013

		True Positive	False Positive
	% of time detected	% of RR>2mm/hr detected	% of LWP<0.1 mm detected
RR>2mm/hr	0.85	-	0.02
Method 1	1.60	83.1	0.00
Method 2	1.86	88.4	0.00
SVM ML	1.35	94.5	0.24

- The SVM method identify **much less contamination time** than the other two methods.
- All three methods have **high true positive rate**. SVM method performs the best.
- there are **a few false positives** when LWP is low.

Summary

ARM

By closely working with instrumental mentors and retrieval experts, we

- Developed a support vector machine (SVM) method with nonlinear kernel transformation to address MWR rain contamination problem;
- Better cleaned the training data;
- Compared the SVM results with other detection methods for MWR rain contamination problem.
    - **much less contamination time**
    - **better identify rain contamination**

- This framework is easy to be implemented for other applications.

Issues and Suggestions

- A **ground truth** is needed to verify and evaluate the results.
  - We tried to use TSI, but it does not represent the real condition of MWR radome
  - A camera is set up to look at the MWR radome at SGP
- More variables from other surface measurements may be added to improve the reliability of the ML algorithm.
  - E.g., relative humidity