

Determination of CERES TOA Fluxes Using Machine-Learning Algorithms

Bijoy Vengasseril Thampi¹, Takmeng Wong²

Constantine Lukashin²

¹Science Systems and Applications, Inc., Hampton, VA

²NASA Langley Research Center, Hampton, VA

CERES Science Team Meeting, 16-18 May 2017



Objective

- It is observed that **5.6%** of all CERES Terra/Aqua data contains missing cloud cover information or insufficient imager data for a reliable scene identification (some times it can reach up to 50% of data for a specific scene type)
- The unavailability of imager data lead to gaps in global radiation budget dataset.
- In this study, our objective is to develop a Machine learning methodology for the improved determination of CERES scene type and subsequent clear-sky TOA flux estimation using standalone CERES TOA radiance measurements (without any MODIS/Imager data).

Machine learning Algorithms

Random Forests (RF)

- is an ensemble learning method for classification and regression.
- Random forests operate by constructing a multitude of decision trees and outputting the class that gets maximum number of votes from the forest.
- Main advantages of RF method are:
 - *they have faster runtimes*
 - *can deal with unbalanced and missing data*
 - *has the ability to handle data without preprocessing or rescaling.*

Artificial Neural networks (ANN)

- ANN is based on a large number of neural units loosely modelling the way a biological brain solves problem.
- They are exceptionally good at performing pattern recognition and other tasks that are very difficult to program using conventional techniques.
- Programs that employ neural nets are also capable of learning on their own and adapting to changing conditions.

Clear-sky Flux estimation



- Scene Classification - **Random Forests (RF)** method
 - Developed by Breiman and Cutler(2000)
- TOA Flux estimation – **Artificial Neural network (ANN)** method
 - ANN methodology outlined in Lukashin and Loeb(2003)

Input data

Input Variables	IGBP surface types
Solar & viewing zenith-angles	Water bodies
Relative azimuth angle	Bright Desert
CERES Shortwave (SW) and Longwave (LW) broadband radiances	Dark Desert
LW surface emissivity	Grasslands
Broadband surface-albedo	Croplands and cities
Surface skin temperature	Evergreen Forests
Precipitable water	Deciduous Forests
Wind speed	Woody Savannas and Shrub lands
	Permanent and Fresh-snow
	Sea Ice

All 10 variables are used in the RF analysis while variables in **bold letters** are only used for ANN analysis.

Using RF, The TOA radiances are classified in to clear and cloudy classes first.

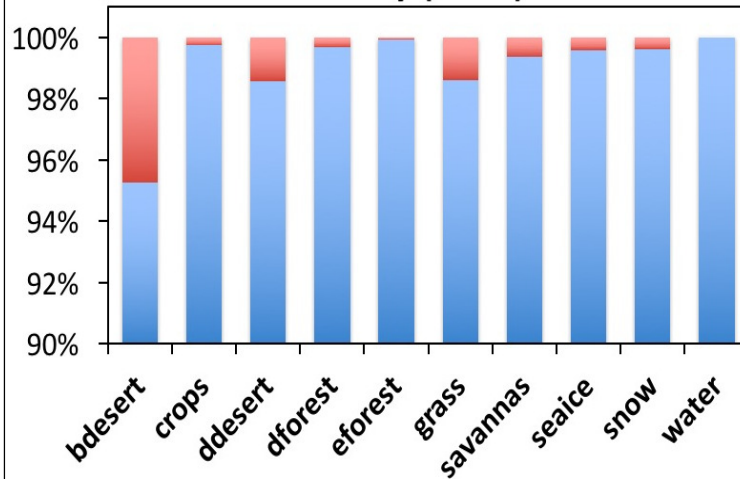
In the second step, the classified radiances are converted in TOA fluxes using the ANN method.

Training data : 2003-2014
Test data : 2015

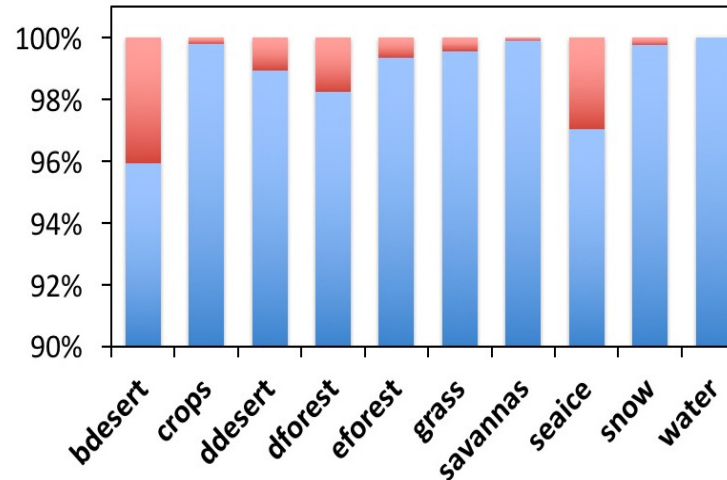
This allows conversion of CERES TOA radiances to TOA fluxes without using any cloud Imager data

RF scene classification Results

January (Clear)



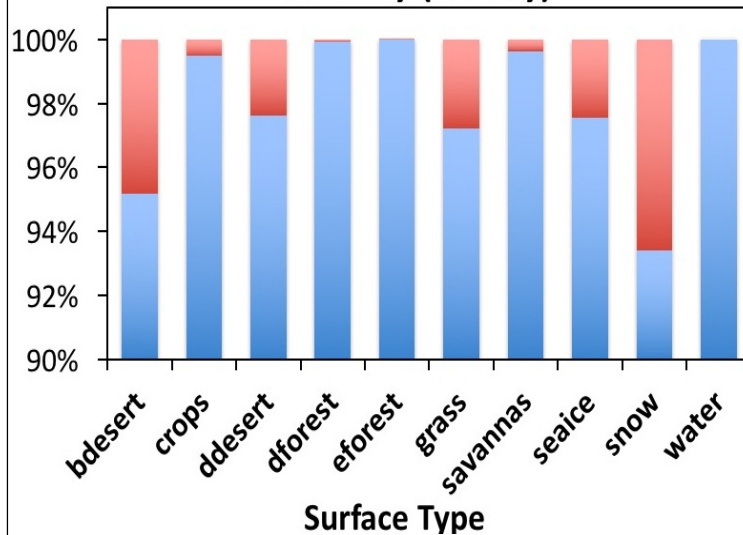
July (Clear)



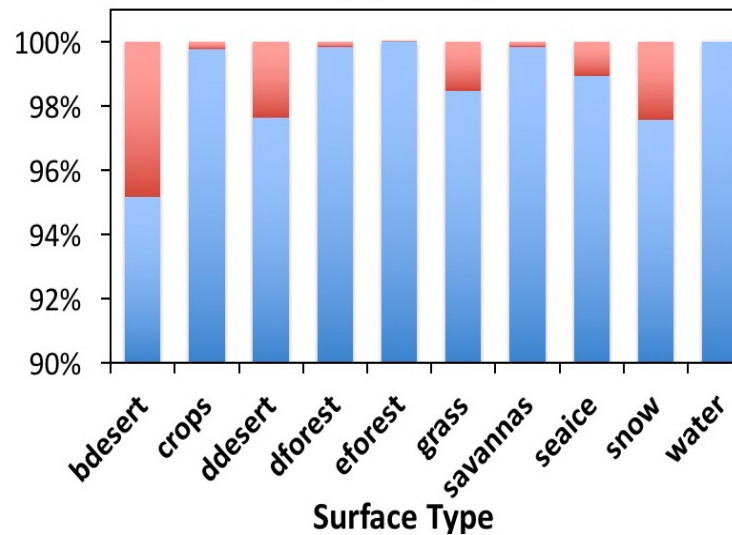
Year : 2015
(Day time)

RED –
misclassified
data points

January (Cloudy)

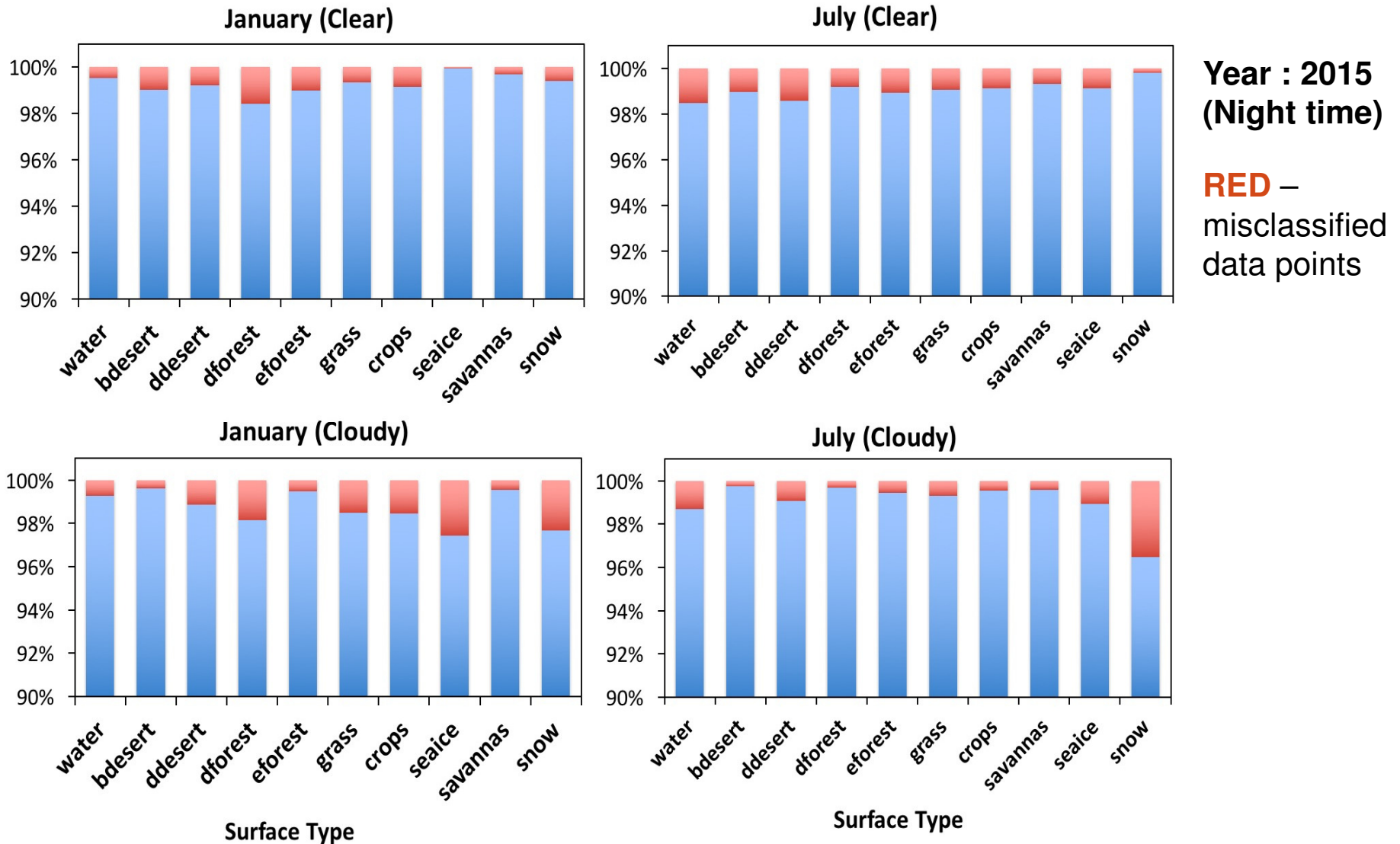


July (Cloudy)



- Scene Classification rate in general is > 98% for most of the surface types
- A misclassification rate of 3-6% is observed for surface types like bright deserts, snow and seaice.

RF scene classification Results



- Both SW radiance and albedo are not included in the night time analysis
- Scene classification rate in general > 98% for most of the surface types
- Misclassification rate is relatively high (>3%) for surface types like snow and seaice.

ANN TOA clear-sky Flux estimation

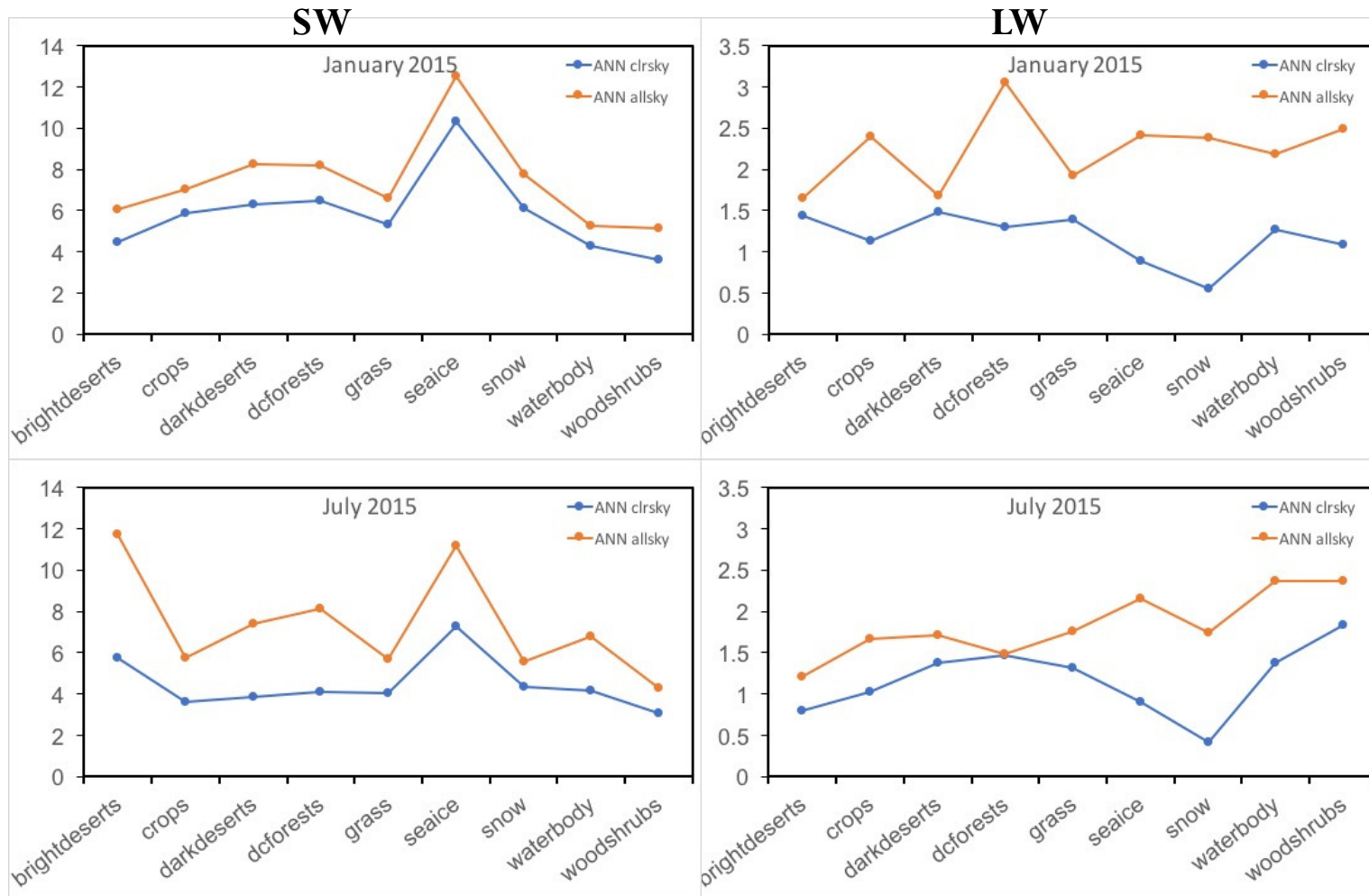
- Once the clear-scene identification is carried out by Random Forest method, CERES radiance to flux conversion is carried out by employing a feed-forward error back-propagation (**FFEB**) artificial neural network (ANN) method (**Loukachine and Loeb, 2003**).
- The technique is then validated by comparing ANN-derived TOA fluxes with CERES (original) TOA fluxes.
- In the ***modified ANN method***, only clear-sky SSF monthly data (2003-2014) is used to train the ANN and results were compared with earlier ANN methodology.

Intercomparison- ANN_{clear} Vs ANN_{allsky}

- ANN radiance to flux conversion of RF classified data (clear) is conducted using both modified ANN and original ANN method
- Analysis of the ANN derived Flux values that closer to actual SSF TOA Flux values show that ANN clear sky method produce better results for all surface types for majority of the instances (>60%) compared to the ANN all sky method.

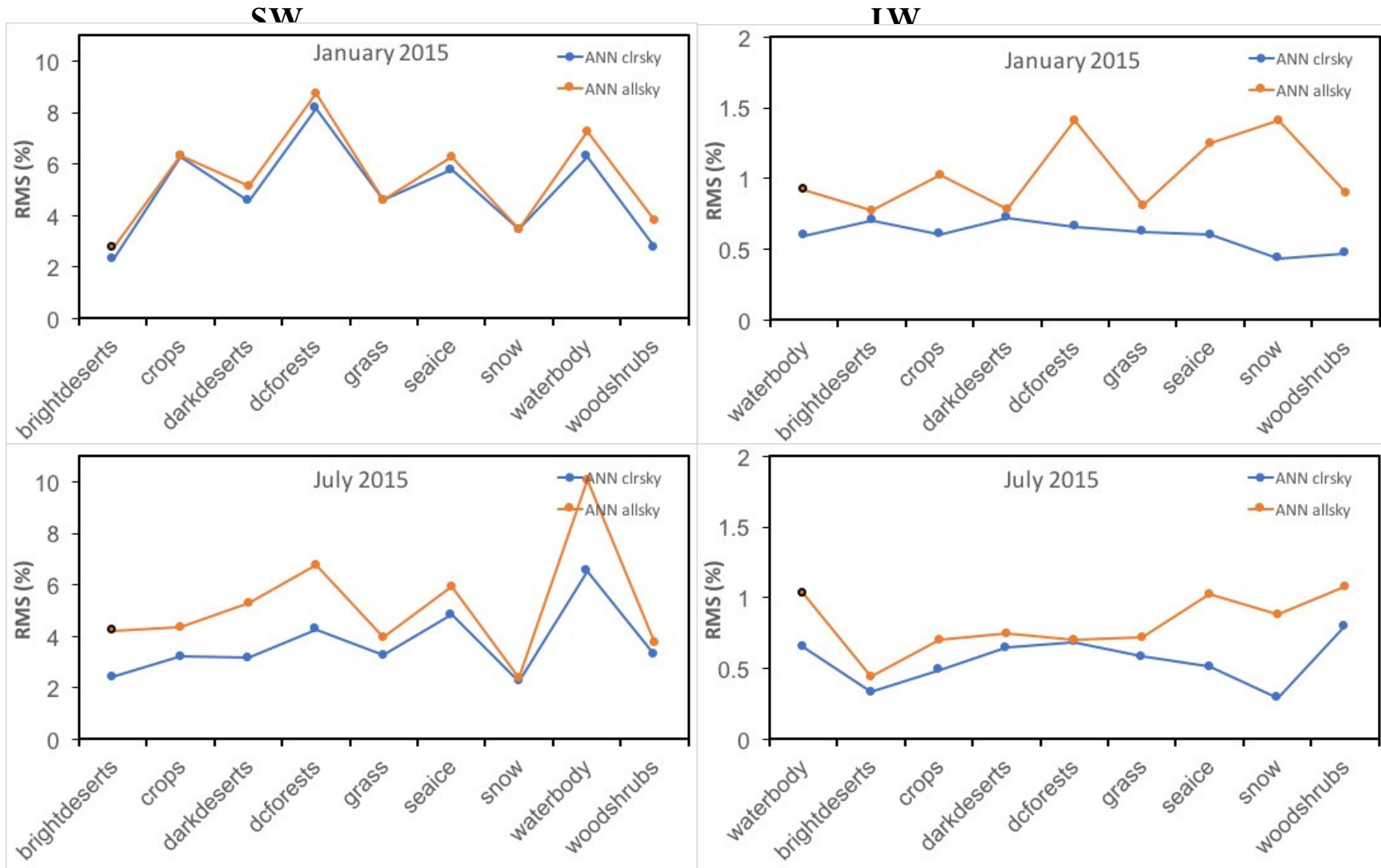
SURFACE TYPE	SW		LW	
	JAN (%)	JUL(%)	JAN(%)	JUL(%)
bdesert	64.5	67.3	84.1	63.7
crops	59.2	63.6	85.4	88.8
ddesert	57.3	64.7	82.8	77.1
dforest	65.0	68.6	63.7	59.8
grass	65.5	73.9	80.4	49.5
savannas	62.2	74.3	59.2	61.8
seaice	62.4	68.6	76.0	68.9
snow	63.5	77.4	60.9	71.2
water	58.1	67.9	67.4	67.0

BIAS



- Mean bias is relatively lower for the ANN clear sky method compared to the all sky method estimated for the Clear-sky SW (left) and LW (right) TOA Fluxes.

Relative RMS (%)



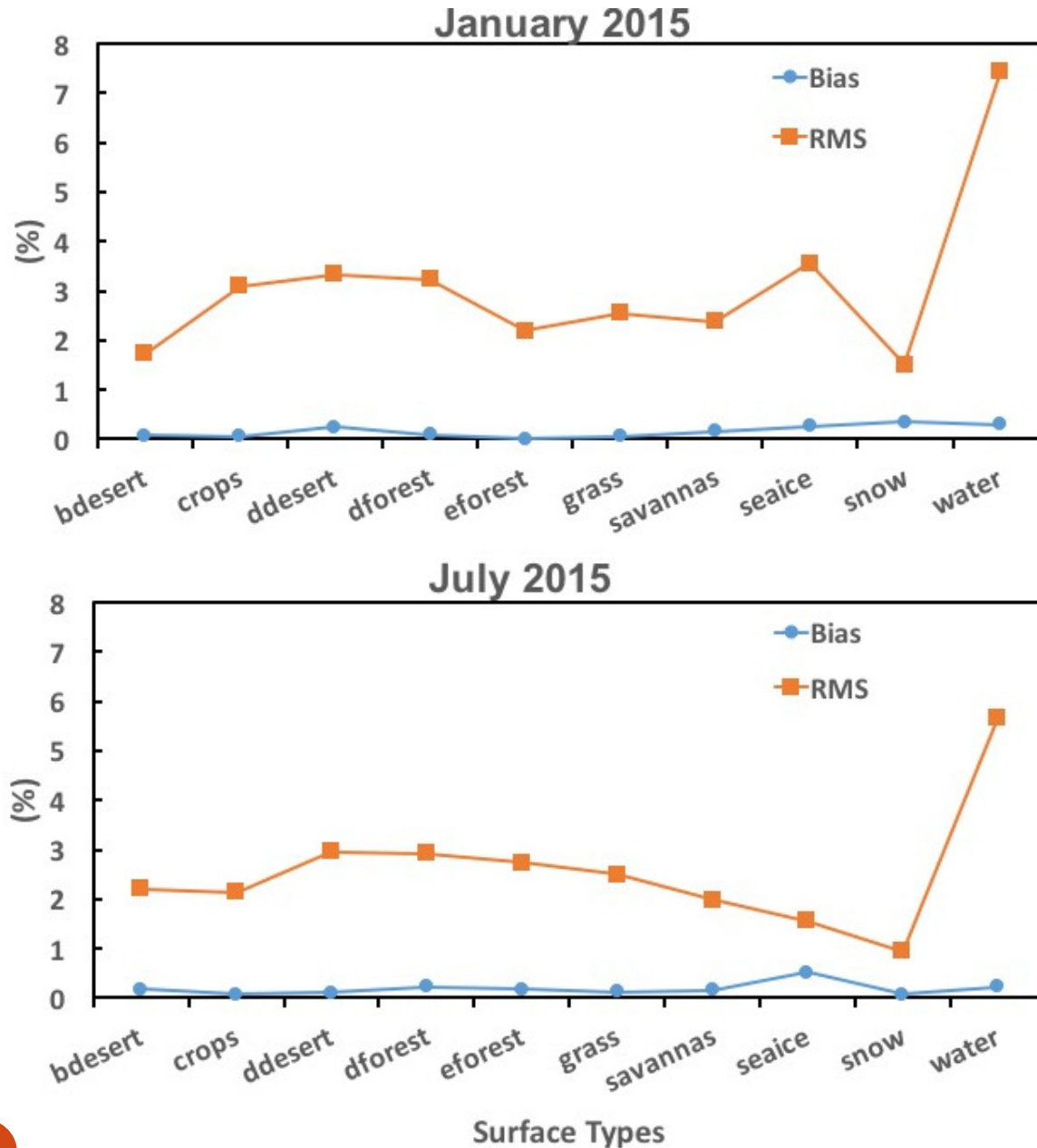
- Mean RMS (%) also show relatively lower for the ANN clear sky method compared to the all sky method estimated for the Clear-sky SW (left) and LW (right) TOA Fluxes.

Summary

- A new methodology for Imager independent CERES TOA clear-sky flux retrieval is developed incorporating **Random Forests** scene classification and **Artificial Neural Network** flux estimation methods.
- RF misclassification rate for (Clear and cloudy, Day time) shows lower values ($< 2\%$) for Water bodies, Crops, Evergreen forest, etc.
- Modified ANN clear-sky method produce more accurate TOA flux values ($>60\%$ of data) compared to earlier ANN method with relatively lower Bias.

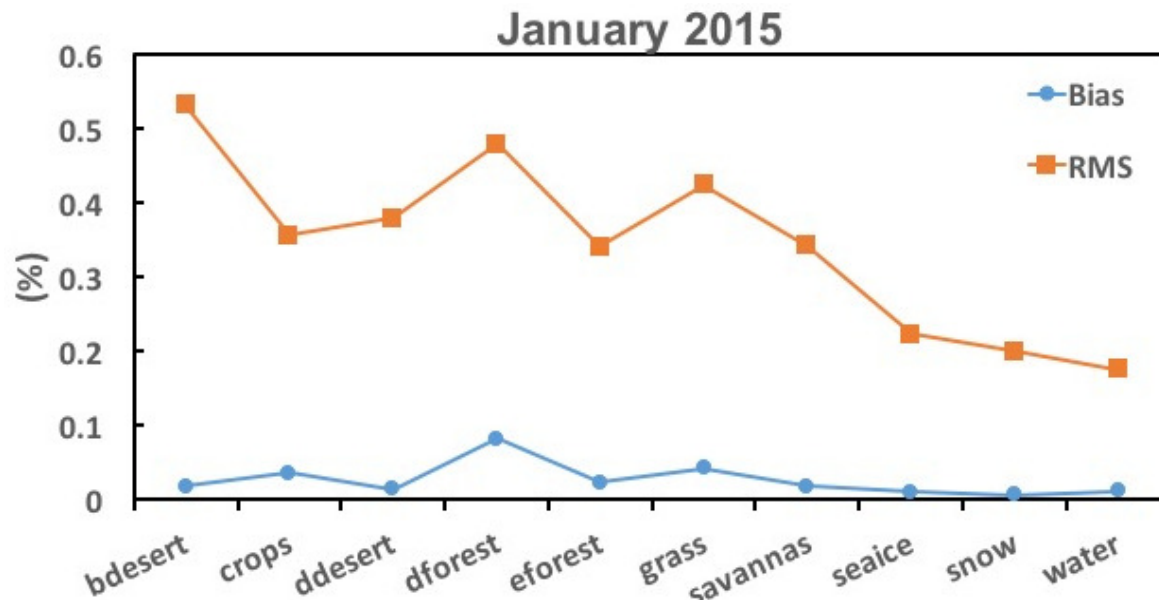
Thank you...

Global mean BIAS and RMS : Shortwave flux

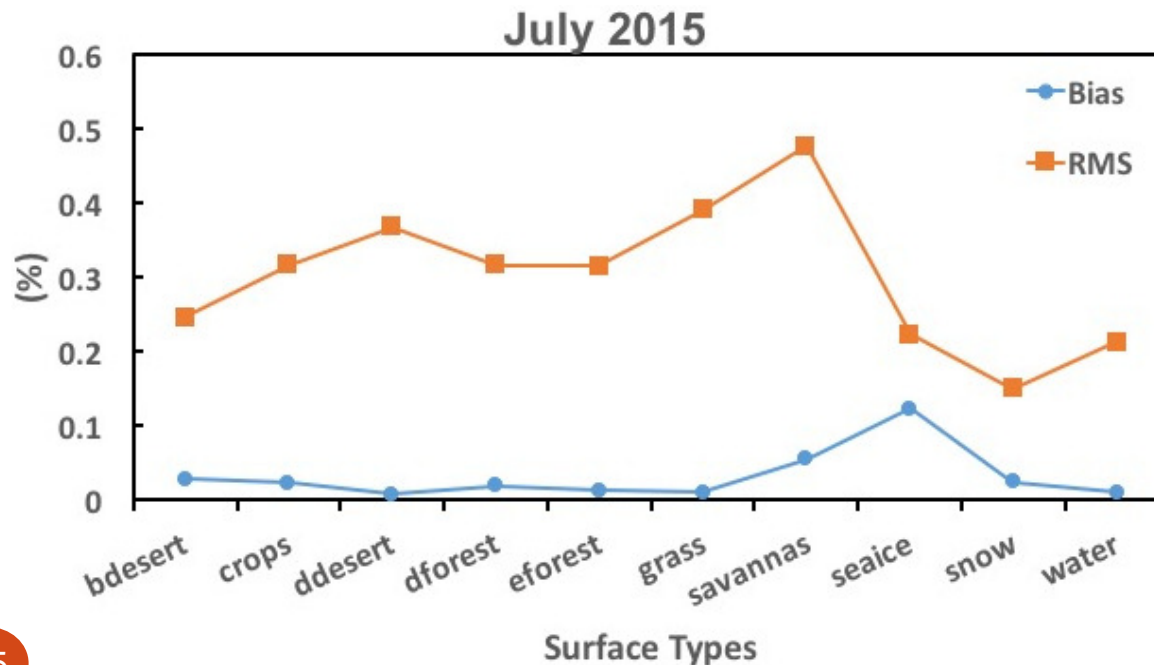


- Global mean Bias in SW flux on the other hand is lower ($< 1\%$) for all the surface types
- Mean RMS in SW Flux (%) for most surface types are below 4% while it is relatively higher (6-8%) for water surface.

Global mean BIAS and RMS : Longwave flux



- Similarly, the global mean Bias for the LW flux is also lower ($< 0.15\%$) for all the surface types compared to the SW flux values.



- Compared to the SW flux, Mean RMS for the LW Fluxes are usually below 1%