Using Machine Learning Algorithms to Detect And Classify RFI: Illustration Using Soil Moisture Active/Passive (SMAP) Data

A. Bringer[(1)], A. Bete[(1)],
J. T. Johnson[(1)], P. Mohammed [(2)], S. Misra [(3)]

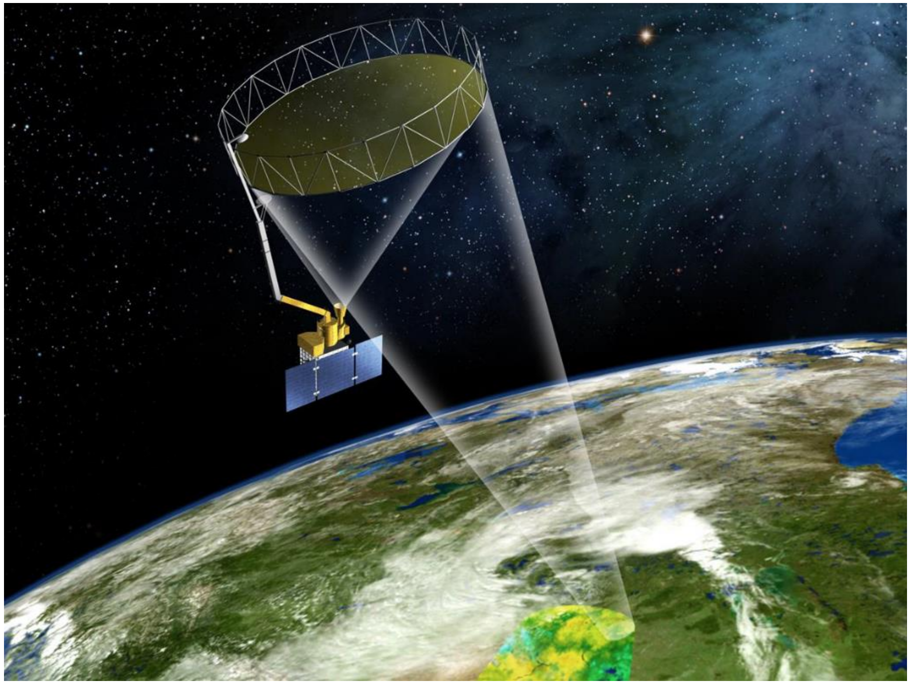
(1) ElectroScience Laboratory,
The Ohio State University Columbus, OH 43212,
(2) Goddard Flight Center,
Greenbelt, Maryland
(3) Jet Propulsion Laboratory,
California Institute of Technology, California

The Ohio State University

RFI 2022

Soil Moisture Active/Passive Mission

- Launched on January 31, 2015:
  - Spatial resolution: 36 km
  - Temporal resolution: 3 days
- Since April 24, 2015:
  - Soil moisture measurement acquired by the L-Band Radiometer
- Radiometer Measurements:
  - Brightness Temperature: H and V Polarization
  - 3^{rd} and 4^{th} Stokes parameters

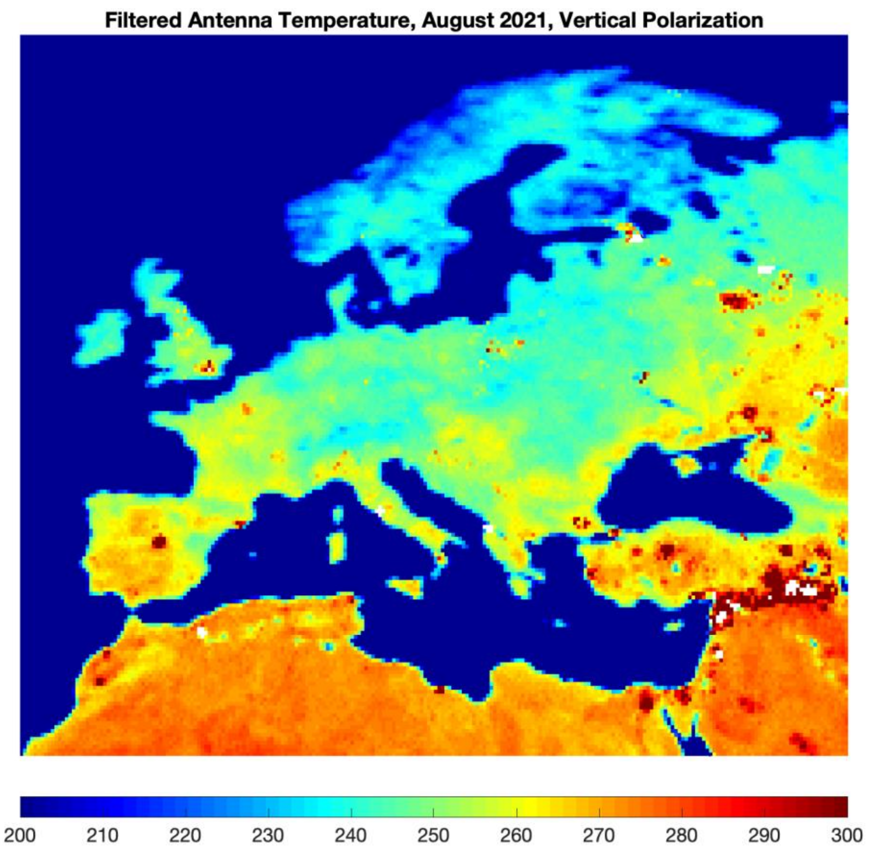


Despite the fact that SMAP radiometer is operating at a protected band for radiometry, SMAP measurements are contaminated by Radio Frequency Interference (RFI)

Soil Moisture Active/Passive Mission

ElectroScience LABORATORY

- However, the SMAP microwave radiometer was specially designed to include a digital back-end that adds additional capabilities in detecting and filtering RFI contributions.
- Despite the overall good performance of the RFI detection algorithms, it has been demonstrated that **continuous moderate wideband RFI** are still an issue as they often remain **undetected**.
- Such **residual RFI** is a **major** concern for radiometry applications as it can cause **biases** in soil moisture retrievals.



This presentation investigates the use of machine learning algorithms to detect and classify RFI using SMAP data.

Objectives

- SMAP Digital backend provides a plethora of information on RFI environment over the 7 years of data
- Many ways to present and analyze this data
- Previous studies analyzed maps and their time history to obtain insight about the changes of the RFI environment
- Those analyses don't provide information on the RFI sources directly (i.e source types, frequency most flagged…)
- Given the large quantity of RFI data available, machine learning algorithms can be used to classify RFI sources and to potentially help detecting the residual RFI sources.
- Previous work from Mohammed et al, 2021 demonstrated the potential of using Machine learning algorithms to detect RFI using a classification algorithm that classifies the spectrogram as: RFI and no RFI.
- In this work, we refine the classification by incorporating more classes in order to get information on the RFI source type

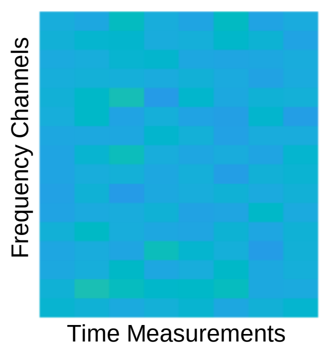
Why Machine learning? Which Algorithms?

- Many machine learning algorithms exist for different applications. Our goal is to classify the RFI sources depending on their types. Deep learning is widely used for classification of object in images for instance.
- In our study, the RFI will be the object to classify in the SMAP individual spectrograms that will be used as images.
- To classify SMAP spectrograms, we used a supervised learning algorithm which relies on a large portion of the data to be labeled. In that case, the labeled data are used as input and the model is trained iteratively until the model correctly identified the RFI type.
- Convolutional Neural Networks (CNN) are commonly used in deep learning for classification purposes. Several preexisting CNNs are now available and can be used for our analysis.
- Transfer learning provides the benefit of taking an existing network and replacing the final layers to be retrained on a new data set. This allows a previously tested CNN to be used for a new application.
- Therefore, our method is based on transfer learning of some of these preexisting networks on the SMAP dataset to classify RFI source types.

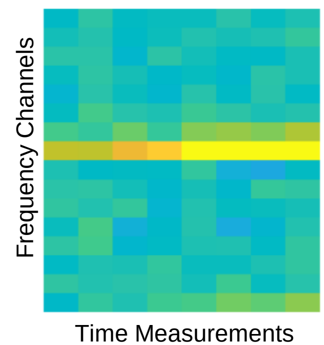
Methodology

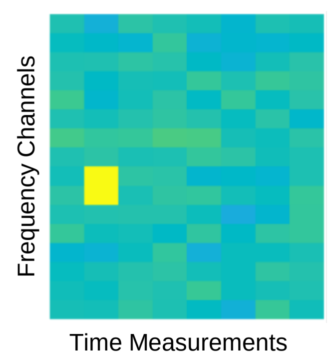
- The first step of this work is to identify the classes and manually populate these classes with examples among SMAP data.
- It was chosen to remove any spectrograms that are located over a coastline as the difference between the antenna temperature acquired over land and over ocean is large and the machine learning algorithm can interpret that difference as RFI. This allows to limit the false alarm detection.
- It was decided in a first analysis to divide the SMAP data into 3 classes:

No RFI

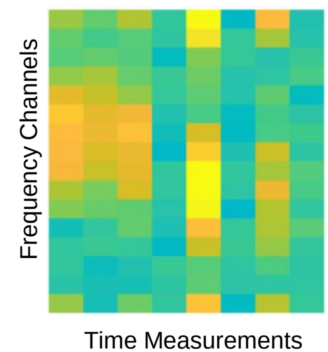


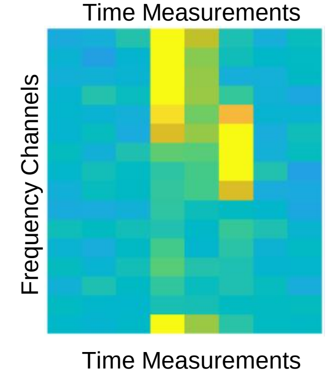
Narrow Band RFI





Wide Band RFI



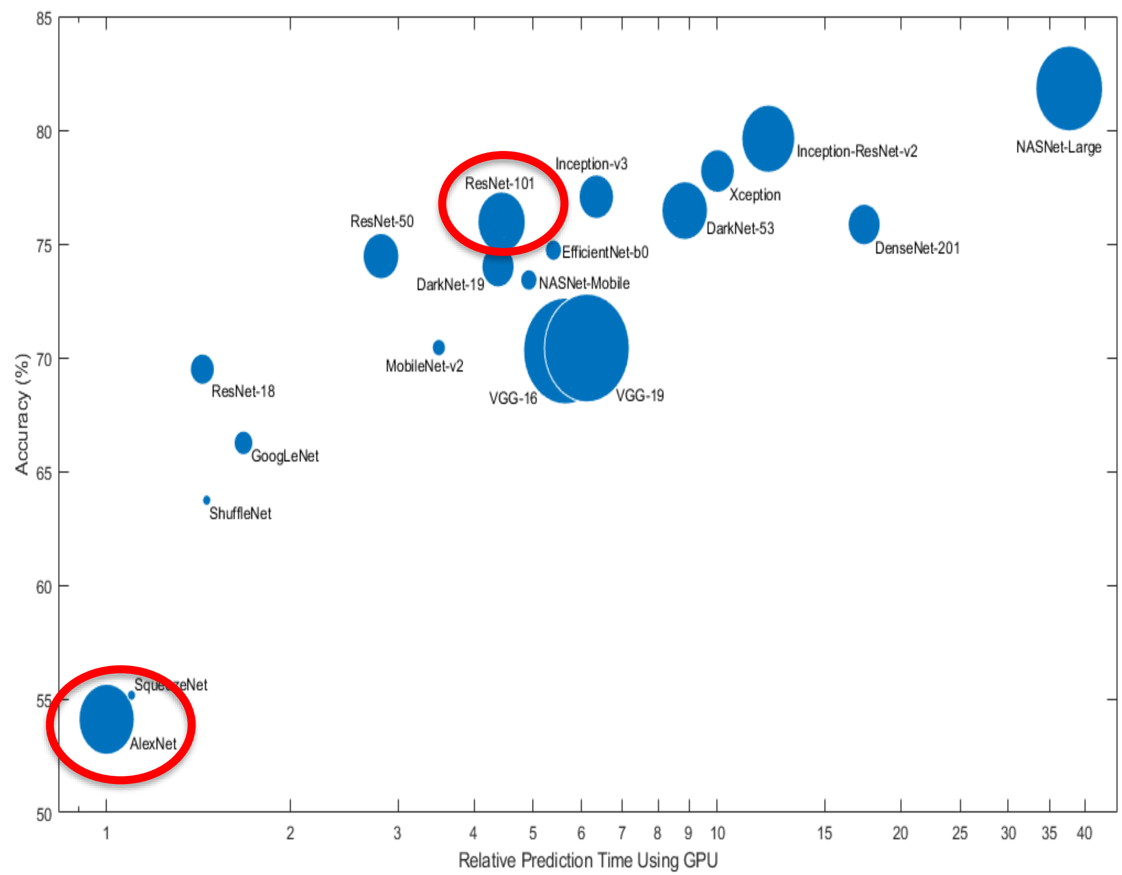


# spectrograms:	No RFI	Narrow Band RFI	Wide Band RFI
	279 (training)	279 (training)	218 (training)
	12 (validation)	12 (validation)	9 (validation)

Methodology

ElectroScience LABORATORY

- For this study, transfer learning is investigated using pre-existing networks. To compare their performance, two different Convolutional Neuronal Networks have been tested: AlexNet and ResNet 101.
- The CNN are usually evaluated on their accuracy, speed, and size.
- AlexNet has a lower accuracy but it is fast to train
- Resnet 101 presents a good compromise between accuracy and training time
- For our classification analysis, AlexNet and Resnet 101 were trained over 800 spectrograms and validated using around 40 spectrograms

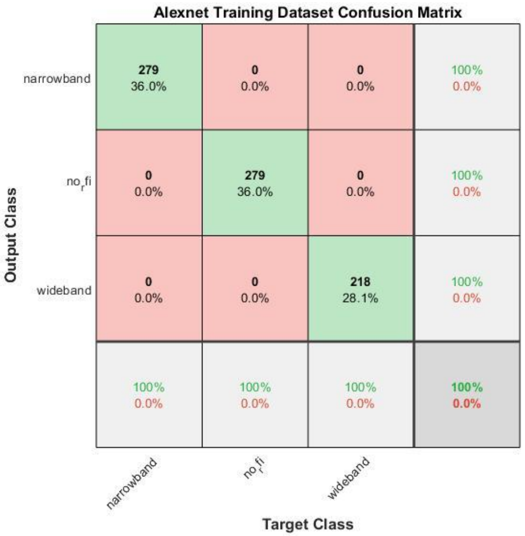


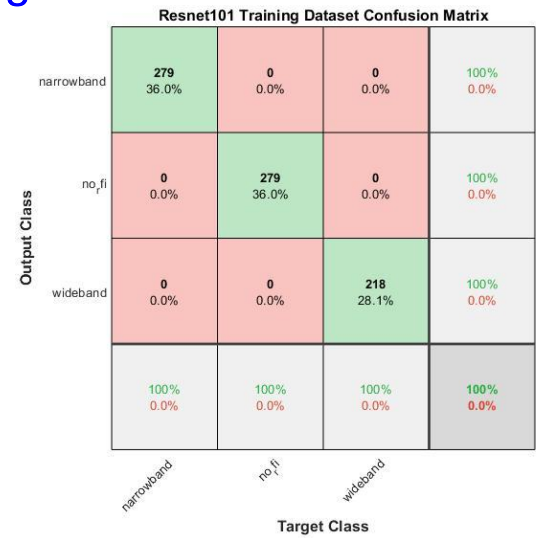
https://www.mathworks.com/help/deeplearning/ug/pretrained-convolutional-neural-networks.html

Results

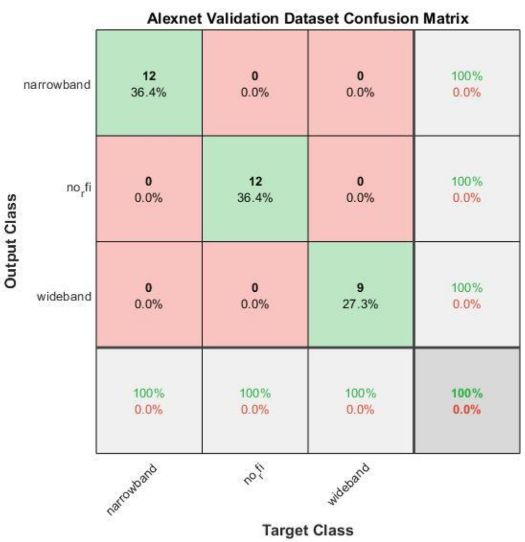
- Both networks were trained using the same parameters (learning rates, iterations…)
- During the training phase, the networks are tested on validation images every 130 iterations during training.
- The confusion matrices for the validation and training phases show the accuracy of the networks after training was completed.
- The accuracy of the training phase of each network reaches 100 % for each spectrogram class as expected demonstrating the overall good training of the two networks.
- While training, the two networks were validated on ~ 30 spectrograms. The confusion matrices for each CNN confirms demonstrates that the networks classify new images well.

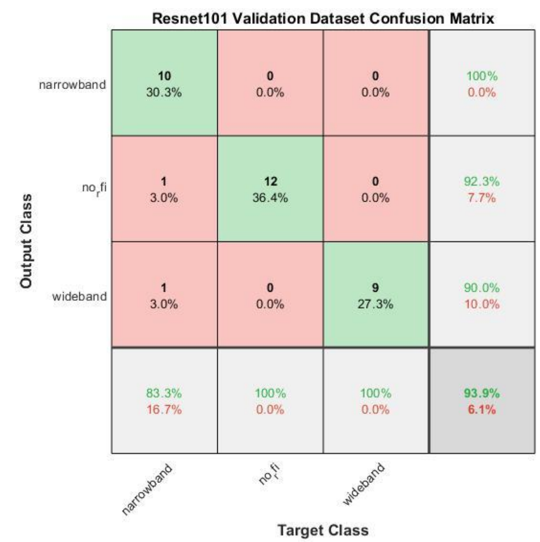
Training Phase





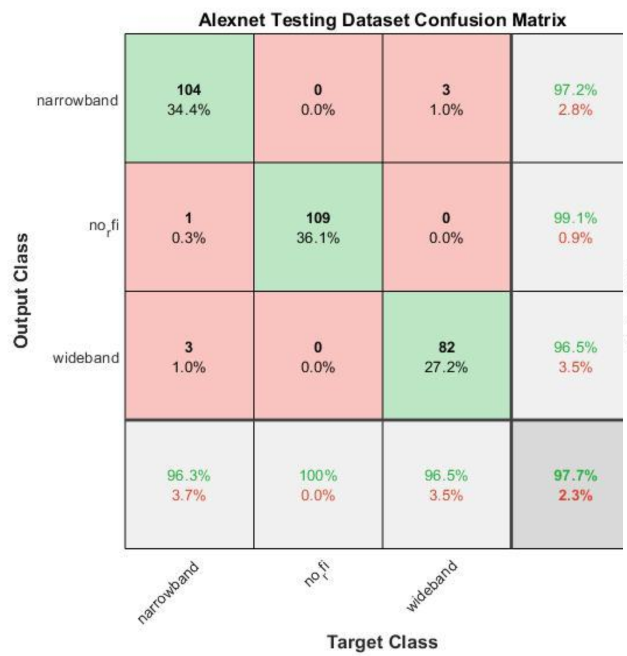
Validation Phase





Results

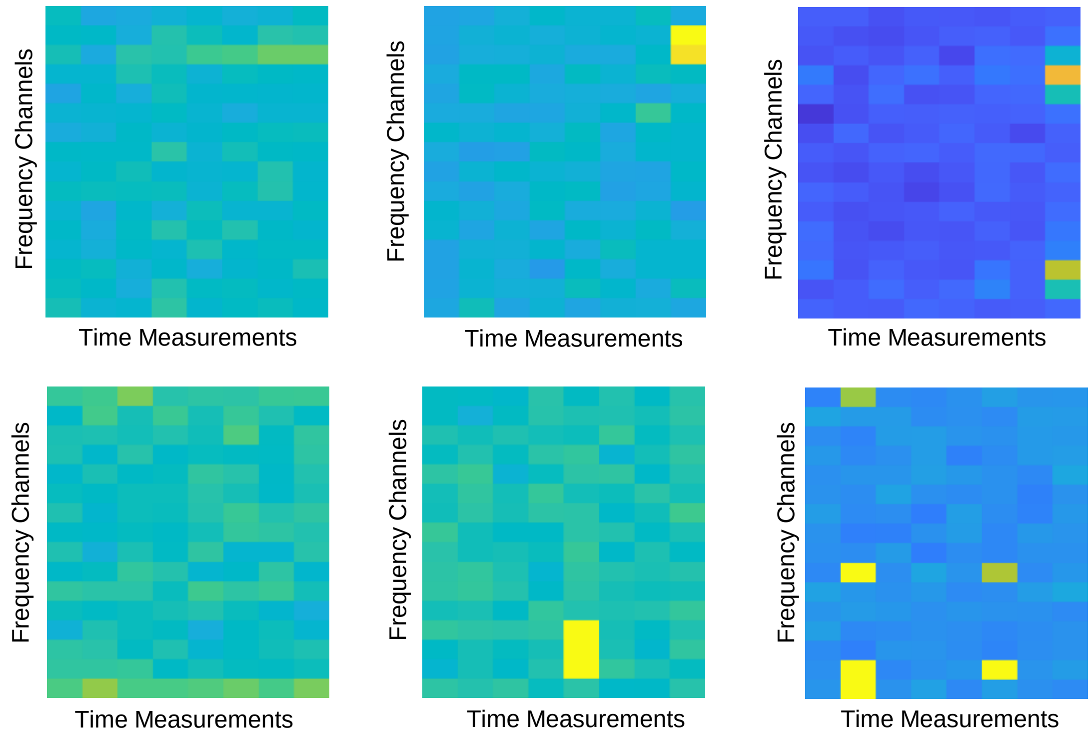
- The next step is to test the two trained networks on an unknown data set
- A new data set was then generated and was composed of 109 spectrograms labeled as No RFI, 108 spectrograms labeled as narrowband RFI and 85 spectrograms identified as wideband RFI.
- The confusion matrices of the two networks show that the Resnet101 has a better overall accuracy and performs better at classifying narrowband and wideband RFI sources.
- However, Alexnet seems more accurate to classify the no-RFI spectrograms.
- These results confirms the possibility of using classification deep learning algorithms to characterize RFI sources



Results

ElectroScience LABORATORY

- These spectrograms present examples where the two CNNs are not able to classify the RFI sources correctly
- The misclassification can be due to several factors:
  - The number of spectrograms considered for the training is too small and all the combinations of narrowband and wideband RFI are not encountered
  - The number of wideband RFI was limited in our training phase
  - All RFI levels have to be equally represented, especially the lower and medium RFI.



Narrowband misclassified as no RFI

Narrowband misclassified as wideband

Wideband misclassified as narrowband

Next Steps

ElectroScience LABORATORY

- This first analysis is promising but suffers from limiting factors.
- We are currently working on expanding the training and validation data sets in order to represent better the RFI environment. The challenges are to find a sufficient number of spectrograms representative of our different classes
- Increase the number of classes to refine the RFI classification in order to get more information about the global RFI environment.
- A new CNN can also be developed in order to identify portions of the spectrogram impacted by RFI. Since the RFI areas would be selected, the network would be able to effectively test for more complicated RFI types.
- Additional CNNs will be tested. We will also consider other types of deep learning (compare supervised and unsupervised learning)

Conclusion

- In this talk, we presented our initial analysis on implementing supervised deep learning algorithms in order to classify RFI sources.
- The potential of using machine learning algorithms for detecting RFI in SMAP spectrograms was previously demonstrated by Mohammed et al. They used transfer learning to classify the spectrograms into RFI/No RFI classes. This study refines the RFI class into Narrowband and Wideband RFI which would provide more information of the RFI environment.
- This analysis illustrated the possibility of using pre-existing CNNs such as Alexnet and Resnet101 to distinguish narrowband and wideband RFIs. This analysis needs to be expanded over a larger data sets.
- We are working on testing additional CNNs. We will also consider other types of deep learning (compare supervised and unsupervised learning) and other kinds of detection algorithms such as clustering algorithms.