FLOOD FORECASTING AND MONITORING USING SENTINEL-1 AND SMOS SATELLITES: A SUPERVISED PREDICTIVE MODELING

David Bastide [1], Jérôme Gasperi [2], Yann Kerr [3], Richard Moreno [2], Pierre Villard [1]

[1] Capgemini Aerospace and Defense, Toulouse, France – pierre.villard@capgemini.com, david.bastide@capgemini.com
[2] CNES, Toulouse, France – richard.moreno@cnes.fr, jerome.gasperi@cnes.fr
[3] CESBIO, Toulouse, France – yann.kerr@cesbio.cnes.fr

ABSTRACT

Increasingly intense downpours driven by climate change will see flood damages rise from €4.5bn to 23€bn a year by 2050 according to a research published in the journal Nature Climate Change [1]. At this time, a lot of papers are dealing about the possibilities to forecast flood events and to monitor such events as well. Our objective is to combine the two approaches in a learning system in order to have a supervised predictive modeling which is automatically improving itself over time. This paper describes our current work regarding the use of Sentinel-1 and SMOS data to build such a system using the Data Cube developed by Geoscience Australia.

***Index Terms*—** Flood, forecasting, monitoring, Sentinel-1, SMOS, machine learning, data cube

1. INTRODUCTION

Nowadays EO data is easily and quickly accessible from a growing number of satellites. The question is how to deal with an increasing amount of data in an efficient way and to produce some interesting results that can help people taking decisions. In these proceedings we are focusing on some ideas and concepts that can be united to fulfill one objective: efficiently forecasting flood events using an algorithm that is able to learn from its mistakes thanks to a parallel flood monitoring system.

2. THE AUSTRALIAN GEOSCIENCE DATA CUBE

Geoscience Australia (GA) is leading the development of the Australian Geoscience Data Cube (AGDC) to support the management and quantitative analysis of massive volumes of Earth observation (EO) and other geo-scientific data. One of the features of the AGDC approach is that all of the original observations (pixels) are retained for analysis; the data are not mosaicked, binned, or filtered in any way and the source data for each pixel can be traced through the metadata.

The AGDC approach allows us to store heterogeneous data from multiple sensors and sources in one single place and to stack this data over time. It eases the access to the data when focusing on a specific region of interest and frees researchers so they can focus on the direct scientific application and exploitation of the data [2].

3. FLOOD RISK FORECASTING USING SMOS

On one hand, two years ago, Capgemini initiated the implementation of an algorithm developed by the CESBIO (center for the study of the biosphere from space) used to evaluate the probability of a flood event based on SMOS data and weather forecasting [3]. The idea is to compute a flood risk map based on precipitation predictions for the next 5 days merged with information on actual soil moisture conditions. The precipitation forecasts are obtained from the NCEP Forecast data and the soil moisture measurements are acquired from SMOS CATDS L3 Soil Moisture data produced at Ifremer for CNES.

The algorithm behind the risk map is purely statistical: the precipitation predictions are compared to precipitation percentiles from 40 years daily reanalysis data from NCEP. The soil moisture is used to increase or decrease the flood risk from precipitation maps based on initial soil moisture conditions. Wet to Dry initial soil moisture conditions are defined based on the comparison of actual soil moisture from SMOS to the SMOS 4-years dataset archive. So far the algorithm was validated against flood events archives from the Dartmouth Flood Observatory.

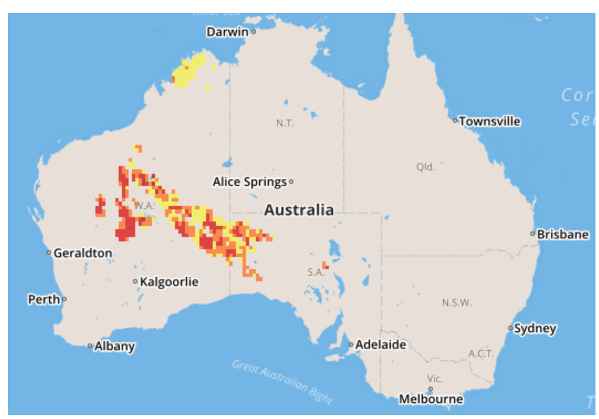

Fig. 1: Flood risk forecasting map based on SMOS data

4. FLOOD MONITORING USING SENTINEL-1

On the other hand, SAR data from Sentinel-1 satellite is easily available from multiple web portals including the French Sentinel data portal (PEPS) and is an excellent product to perform flood mapping operations.

The products are first calibrated, filtered using speckle filtering, and orthorectified. Then a split-based automatic thresholding procedure is used to classify pixels in a "water" class or in a "non-water" class [4]. At the end, an additional segmentation analysis is performed to remove potential false-positive water detections.

Eventually, we identify a new class "flooded area" by comparing the obtained results to the permanent water identified on previous acquisitions. This temporal approach provides a way to produce maps of flooded areas. It also allows the exploitation of the value of the dense Sentinel-1 time series (there will be four Sentinel-1 satellites in a close future).

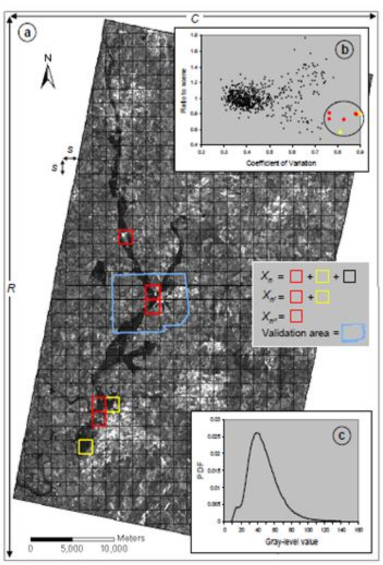

Fig. 2: NRT flood detection using a split-based automatic thresholding procedure based on Sentinel-1 data

5. APACHE SPARK

Apache Spark is an open source cluster computing framework originally developed in the AMPLab at University of California, Berkeley but was later donated to the Apache Software Foundation where it remains today.

In contrast to Hadoop's two-stage disk-based MapReduce paradigm, Spark's multi-stage in-memory primitives provides performance up to 100 times faster for certain applications. By allowing user programs to load data into a cluster's memory and query it repeatedly, Spark is well-suited to machine learning algorithms.

MLlib is Spark's scalable machine learning library consisting of common learning algorithms and utilities, including classification, regression, clustering, collaborative filtering, dimensionality reduction, as well as underlying optimization primitives.

6. SUPERVISED PREDICTIVE MODELING

So far the two systems were running operationally in a Hadoop-based infrastructure to produce both flood risk events maps and flooded areas maps. Recently, we moved from this infrastructure to the Data Cube approach to ease data access and to facilitate the interactions between the two systems.

The underlying idea is quite simple (some of our explanations are illustrated on the next pages): we developed two automatic processing chains to compute both the global flood risk forecasting map and the flood detection map; then the results are stored in a data lake which is the input of our in-memory cluster computing based on Apache Spark to run machine learning algorithms.

In this way, we are able to generate a model which gets better over time and which is able to predict flood events at a Sentinel-1 resolution instead of the SMOS resolution.

On one hand we have flood risk forecasting maps that are computed daily, each time a new L3 SMOS product is made available. On the other hand we have an automatic processing chain downloading all Sentinel-1 products related to our ROI and computing flood detection maps. The combination of the two sources allows detecting false-positive events and false-negative events in order to improve the computed model. Besides, this approach deals with all the Sentinel-1 products and, consequently, exploits the value of the dense Sentinel-1 time series.

Moreover, the use of Sentinel-1 data is a perfect solution to get a flood risk forecasting map at Sentinel-1 resolution instead of the original SMOS resolution. This way, the combination of the products gives the opportunity to have a model taking into account local characteristics of the ground, of the relief, of the land use and so on.

Consequently, this two-side system is an automatically supervised predictive modeling system which is giving more accurate results over time. The accuracy of the predictions is a key factor in decision-making: when a high risk of a flood event is detected, it gives the assurance to correctly take actions in the field, to warn people and to anticipate search and rescue operations.

This on-going project is currently supervised by CNES and the first conclusive results will be published by the end of March 2016.

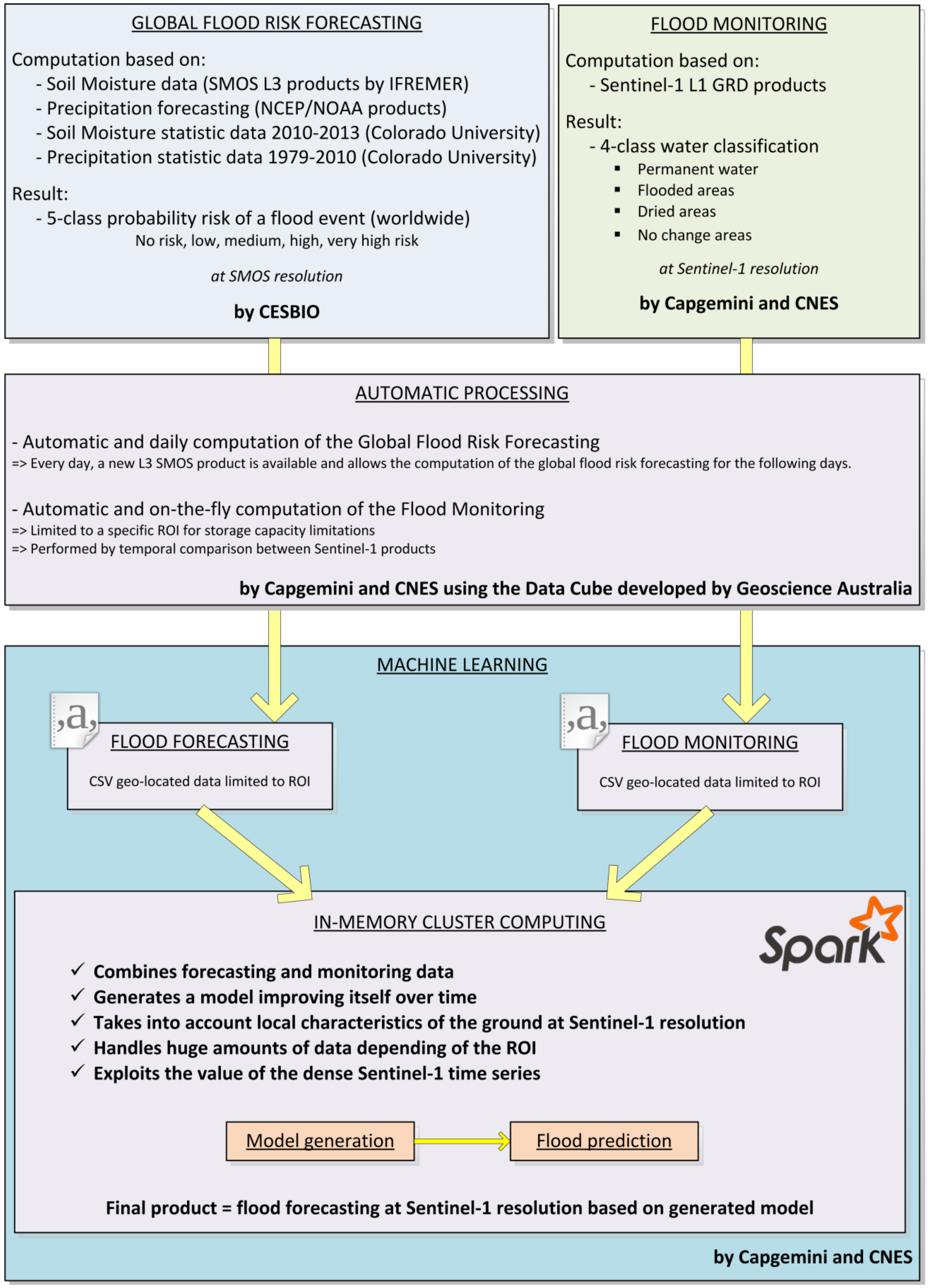

Fig. 3: Flood forecasting and monitoring using Sentinel-1 and SMOS satellites: a supervised predictive modeling

6. PRELIMINARY RESULTS

The main reasons we moved to the Data Cube is the ability to access the data very easily (pictures projected on the same grid and decomposed into smaller tiles), and the possibility to combine it with the Jupyter Notebook. It is a web application that allows you to create and share documents that contain live code (for over 40 programming languages, including those popular in Data Science such as Python, R, Julia and Scala), equations, visualizations and explanatory text. A notebook may also be used to interact with a Spark cluster. This way it allows, very easily, accessing all the data stored in the cube, to perform all the work we want and also to interact with our Spark cluster, all in one place, just using a web browser. To go further, we used Docker instances to allow the deployment of a Data Cube cluster embedding Spark and Jupyter in few minutes.

On one hand, here is an example of the detection of flooded area in India during the monsoon a few months ago:

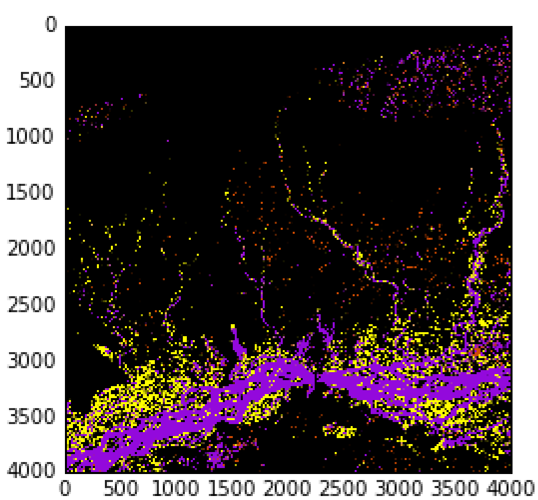

Fig. 4: Flood monitoring using Sentinel-1 data

On the other hand, here is an example of a flood risk forecasting map at global scale using the algorithm developed by CESBIO:

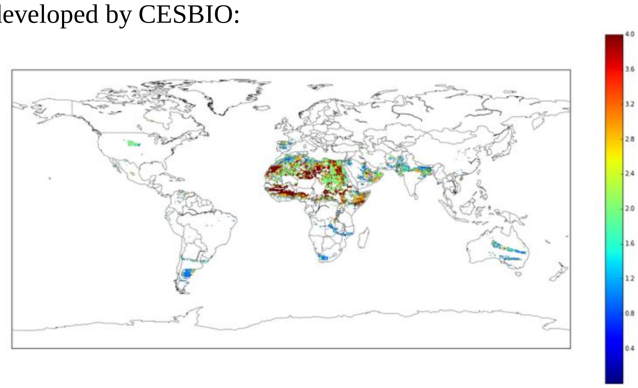

Fig. 5: Flood forecasting using SMOS data

At the moment, the process has been limited to a specific region of interest. The use of this process at a global scale is only a question of infrastructure and available resources in terms of storage, CPU and memory.

The current ROI is limited to the region of Assam in India/Bangladesh. Besides, the quality of the generated model is directly linked to the quantity of data available at the moment the model is trained. Our preliminary results and conclusions will be presented during the conference on Big Data from Space in March.

7. REFERENCES

[1] B. Jongman, S. Hochrainer-Stigler, L. Feyen, J. Aerts, R. Mechler, W.J. Botzen, L. Bouwer, G. Pflug, R. Rojas, P. Ward, "Increasing stress on disaster-risk finance due to large floods", Nature Climate Change, March 2014, http://bit.ly/208CfoA

[2] A. Lewis, S. Oliver, A. Ip, N. Mueller, S. Ring, D. Roberts, M. Thankappan, M. Purss, "Iterating petabyte-scale earth observation processes in the Australian Geoscience Data Cube", Proceedings of the 2014 conference on Big Data from Space (BiDS'14), ESA, http://bit.ly/1H7zLLq

[3] A. Al Bitar, A. Chone, S.K. Tomer, F. Cabot, Y. Kerr, SMOS CATDS Flood Risk, Internal report, CESBIO, 2014

[4] S. Martini, "Towards operational near real-time flood detection using a split-based automatic thresholding procedure on high resolutionTerraSAR-X data", Nat. Hazards Earth Syst. Sci., 9, 303–314, DLR, 2009