International Journal of Technology and Engineering Studies
volume 5 issue 2 pp. 30-39 doi: https://dx.doi.org/10.20469/ijtes.5.40001-2

Adopting the Hadoop Architecture to Process Satellite Pollution Big Data

Badr-Eddine Boudriki Semlali*
Faculty of Sciences and Techniques,
Abdelmalek Essaadi University,
Tetouan, Morocco

Chaker El Amrani
Faculty of Sciences and Techniques,
Abdelmalek Essaadi University,
Tetouan, Morocco

Guadalupe Ortiz
Escuela Superior de Ingenierìa,
University of Cadiz,
Cadiz, Spain

Abstract: This research aims to monitor abnormal climate changes and supervise Air Quality (AQ), especially in Morocco. This study aims to contribute to finding a solution to the AQ degradation and climate change issues by using Remote Sensing (RS) techniques. RSBD in NRT is collected from six sources: the MDEO ground station of EUMETSAT data, the EOSDIS data of NASA, the NESDIS data of NOAA, and the Copernicus platform, some MGS data, and the Raspberry PI sensors data. The current manuscript explains the different aspects of the used satellite data, proving that satellite data could be regarded as Big Data (BD). Accordingly, this research has proposed a Hadoop BD architecture and explained how to efficiently process RS environmental data. This architecture comprises six main layers: the data sources, data ingestion, data storage, data processing, data visualization, and the monitoring layer. The aforementioned architecture automatically collects filters, extracts, and stores data into the HDFS. This proposed model would be beneficial in managing adverse climate conditions and prevent natural disasters.

Keywords: *RS, satellite sensors data, BD architecture*

Received: 03 February 2019; ***Accepted:*** 20 March 2019; ***Published:*** 26 April 2019

I. INTRODUCTION

In this last decade, the world suffers from various environmental problems, and several natural disasters, including air pollution, abnormal climate change, earthquakes, and so on [1, 2]. At this stage, it is important to precisely supervise the climatic and pollution data in atmospheric layers in the troposphere such as the temperature, the humidity, the wind speed and the concentration of trace gases. Recently, satellite data are become widely employed in many potential applications such as tracking pollutant plumes, AQ monitoring support and weather forecasting [3, 4]. In this investigation we have used both polar and geostationary satellites; the main difference between the two is the nature of the orbit. The polar satellite passes always by a Sun Synchronous Orbit (SSO) scanning the north and south poles of the earth [5]. The geostationary satellite passes by a unique orbit over the equator within the same speed of earths rotation making twenty-four hours per orbit [5].

RS techniques refer to the use of the technologies, measuring the specifications of the earth's surface, ocean and atmosphere components without making physical contact with it through Electromagnetic Energy (EME) [6]. This technique employs plenty of types of sensors. Generally, satellite sensors are divided into two main categories: active instruments such as Light Detection

*Correspondence concerning this article should be addressed to Badr-Eddine Boudriki Semlali, Faculty of Sciences and Techniques, Abdelmalek Essaadi University, Tetouan, Morocco. E-mail: Badreddine.boudrikisemlali@uae.ac.ma

and Ranging (LIDAR) and Radar, which illuminate the scanned object with their energy, emit radiation over the target, and then detect the reflected or backscattered radiation [7]. The second type is passive instruments, which detect natural radiation emitted from the target, especially sunlight and photon [8].

In this research, we collect satellite data from the Mediterranean Dialogue Earth Observatory (MDEO) terrestrial station installed in Abdelmalek Essaadi University of Tangier [9], the National Aeronautics and Space Administration (NASA) [10], the National Oceanic and Atmospheric Administration (NOAA), the European Space Agency (ESA), some Meteorological Ground Station (MGS) [11] and from a Raspberry PI ground sensor in Near Real-Time (NRT). These satellites produce daily a large number of datasets, coming from various sources and diverse sensors within different spatial, temporal and spectral resolutions, the daily amount of collected data reaches one hundred Gigabit (GB). These data have also different file formats including the Binary Universal Form for the Representation of meteorological data (BUFR) assumed by the World Meteorological Organization (WMO) [12], the Hierarchical Data Format (HDF5) developed by the National Center for Supercomputing Applications (NCSA) [13], the Network Common Data Form (NetCDF) created by University Corporation for Atmospheric Research (UCAR) [14], and so on. These formats have various physical structures. Accordingly, decrypting these kinds of files demand some specific libraries and interfaces, that make the read of RS data more difficult for non-expert end-users and are continuously increasing storage spaces [15].

The velocity of RS data is so high; it is meaning the rate of which data come there are an important number of files coming per day with a big size. For instance, the Suomi National Polar-orbiting Partnership (NPP) satellite provides data with a rate of two thousand files per day. EPS-Africa channel provides also about three thousand daily files. The Sentinel 5P affords also a huge daily data encompassing forty-eight GB [16]. Similarly, NASA and NOAA data have also an important velocity. Accordingly, RS data are regarded as BD according to the attribute definition of BD based in the 4Vs (Volume, Variety, Velocity, and Veracity), thus the processing of RS BD includes several challenges in term of data collection, storage and handling; as a result, it is necessary to develop a BD platform enabling data collection, sort, categorizing, analyze and storage. Designing a BD architecture is the best way to split the problem of RSBD. We have to design and make in place well all the essential components of BD where each layer platforms have a specific function. These architectures help to trace the data pipeline, either batch, and streaming processing. Generally, we find six layers in BD architecture: data sources, ingestion layer, Hadoop storage platform, Hadoop management layer, Hadoop infrastructure layer, security layer, and monitoring layer. In this study, we will cover the basics of RS technique, and its use for satellites and sensors. Moreover, we will prove that the received data are massive and complex and can be considered as BD. We also proposed an innovative architecture to process this kind of data. The aforementioned architecture will automatically collect remote data, filter subsets and extract and store data into HDFS; moreover, the processing layer will include some Artificial Intelligence (AI) algorithms increasing data accuracy. Finally, the visualization layer will show data in NRT into interactives maps and charts helping in the decision-makers. This architecture should be supplied with another Cloud Computing architecture that may optimize the time execution; this task would be an interesting work to be conducted in the future.

II. BACKGROUND

Generally, BD refers to diverse and complex data with a huge volume, which go beyond the management is the ability of current architectures and platform. There are three definitions of BD which are: the attribute definition based in the four salient (Volume, Velocity, Variety, and Veracity) [17]. The second one is the comparative definition that refers to datasets that go beyond the ability of storage, managing, and processing. The last denotation is the architectural definition that suggests that BD limits the ability to perform effective processing within the traditional relational approaches, or requires the use of significant horizontal scaling for efficient processing. Satellite data become widely used in many applications such as tracking pollutant plumes, supporting AQ providing input for AQ models [18], monitoring Aerosol Optical Depth (AOD) [19] and estimating the Ozone precursor [20]. In our study, we apply RS technique to supervise AQ of Morocco [21], and monitor climate changes in NRT.

We acquired data from four organizations which are the MDEO providing NRT data from multiple EUMETSAT satellites [22]. These data support scientific researches such as pollution monitoring, early warning against disasters, particularly, storm and flood, and it supervises the climate changes, We have gathered also from the NASA, the NOAA, the ESA and some MGS through their websites in order to apply a data validation [23]. In addition, we have developed a Raspberry PI robot, equipped with MQ sensors in order to support RS data validation. Fig. 1 shows the main sources of the acquired satellite data, the satellites, the channel and their products. We collect data from ten polar satellites including the MetOp [24], Terra AURA Sentinel and so on, except for the Meteosat, that is second generation (MSG) and is geostationary [25]. In addition, we exploited many active and passive satellite such as the Tropospheric Monitoring Instrument (TROPOMI), Fourier Transform Spectrometer (TANSO-FTS), Moderate Resolution Imagin Spectrometer (MODIS), the Global Ozone Monitoring Experiment2 (GOME-2) [26] and so on. Table 1 shows the used channel with their extension, file and size per day. In our case study, RS data come with a high velocity reaching forty thousand files per day with an average latency of thirty minutes. These data increase continuously the storage space with 56 GB per day. The collected data are stored in scientific files format, particularly the NetCDF, HDF5 and BUFR [12]. Accordingly, and according to attribute definition (Venue, Volume, Variety, Veracity, Velocity and so on) of BD [12] we can affirm that satellites data are BD.

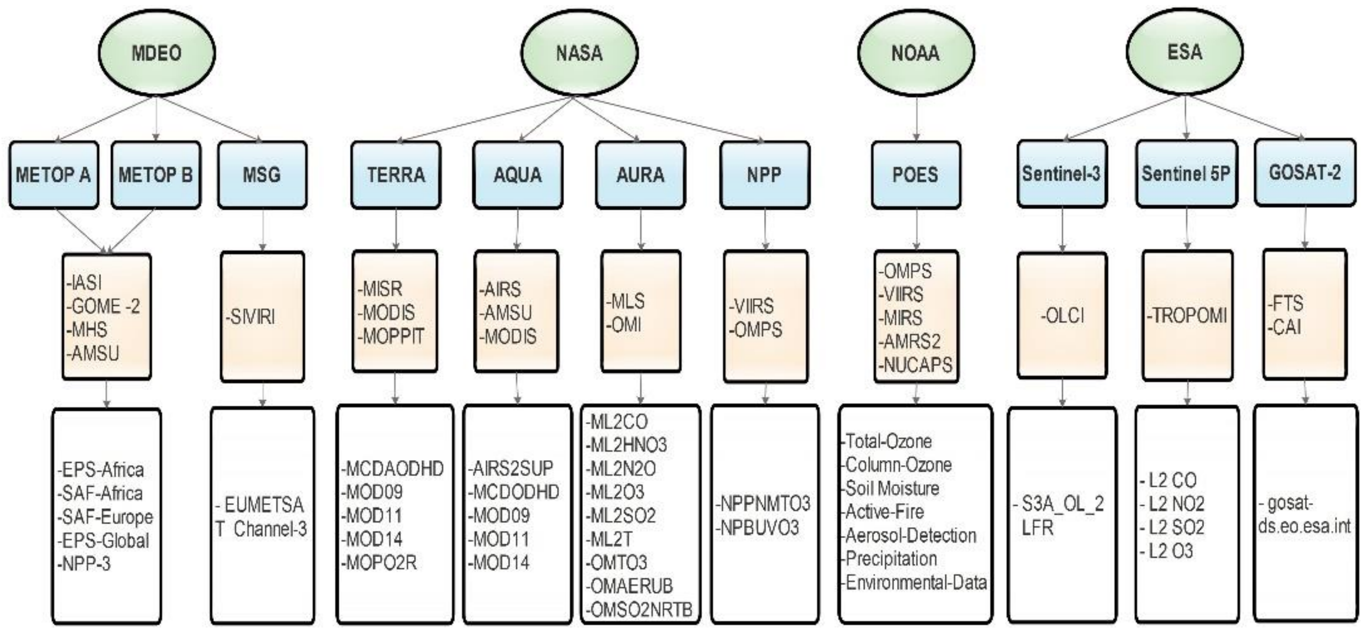


Fig. 1. Sources channel of the used satellite data

TABLE 1
USED SATELLITE DATA WITH THEIR SPECIFICATIONS

Product Name	File is Format	(Files/Day)	Data Amount (Mb/Day)
Sentinel-3	NetCDF	41	14000
Sentinel-5P	NetCDF	85	4400
EPS-Africa	BUFR, Bin	9000	2200
EPS-Global	Bin	1000	180
Data_Channel_3	GRIB, HDF5	300	240
NPP-3	NetCDF, Bin	1000	1100
SAF-Africa	BUFR, HDF5	2000	700
SAF-Europe	BUFR, Bin HDF5	5000	3800
AIRS2SUP_NRT.006	HDF5	640	5400
MCDAODHD	HDF5	4	4
ML2CO_NRT.004	HDF5	90	25
ML2H2O_NRT.004	HDF5	90	25
ML2HNO3_NRT.004	HDF5	90	25

TABLE 1
CONTINUE

Product Name	File is Format	(Files/Day)	Data Amount (Mb/Day)
ML2N2O_NRT.004	HDF5	90	25
ML2O3_NRT.004	HDF5	90	30
ML2SO2_NRT.004	HDF5	90	25
ML2T_NRT.004	HDF5	90	25
MOD11_L2	HDF5	250	590
MOD14	HDF5	280	4900
MOP02R_NRT	HDF5	30	660
NMTO3NRT	HDF5	14	123
NPBUVO3-L2-NRT	HDF5	14	7
OMAERUV	HDF5	15	86
OMSO2NRTb	HDF5	15	361
OMTO3	HDF5	15	530
Precipitation	NetCDF	15	410
Active-Fire-EDR	NetCDF	1100	90
Aerosol-Optical-Depth	NetCDF	600	20000
Nadir-Profile-Ozone	NetCDF	1100	14
Precipitation-and-Surface	NetCDF	2500	3300

III. METHODS AND MATERIALS

Designing a BD architecture is the best way to split the problems of BD processing. We have to design and make in place well all the essential components of BD, where each layer has a particular function. There are seven layers as illustrated in Fig. 2, which are: the data sources, the ingestion layer, the Hadoop storage, management, infrastructure, security and the monitoring layer, as explained below:

- The RS data sources layer designates the different RSBD sources that need an efficient treatment in the ingestion layer. In this paper, our principal RS data sources are the MDEO ground station of EUMETSAT data, EOSDIS data access of NASA, the NESDIS [27] data access of NOAA and the Copernicus platform of ESA.
- The next layer is the ingestion layer since it involves connection to plenty of data sources, filters, subsets, extracts and removes noised and inaccurate datasets. This ingestion layer decompresses and filters RS BD of the selected countries using their satellite orbits cross times, and bounding rectangle of longitude and latitude. Then, many environmental and pollutants variables such as temperature, humidity and the concentration of several trace gazes particularly CO, CO_2, NO_x, CH_4 are extracted from plenty converted files. The final step in this layer is to subset and remove inaccurate values before they are stored inside the distributed files system especially the Hadoop HDFS, Amazon S3, or GlusterFS.
- The storage layer is dedicated to save and store the output data from the ingestion layer into distributed databases, particularly the Hadoop HDFS. HDFS facilitates parallel data processing mainly by using a master/slaves topology [28].
- The management layer includes all necessary tools for RSBD batch processing; in our research MapReduce and Dryad are the programming models that will be applied, the PIG language that would be an interesting language to use in order to develop some smarts algorithms based in AI, Deep Learning (DL) and Data Mining (DM) helping in the decision-makers.
- The final step of the BD processing is the visualization, it helps to show final results into charts, maps, tables, and reports.

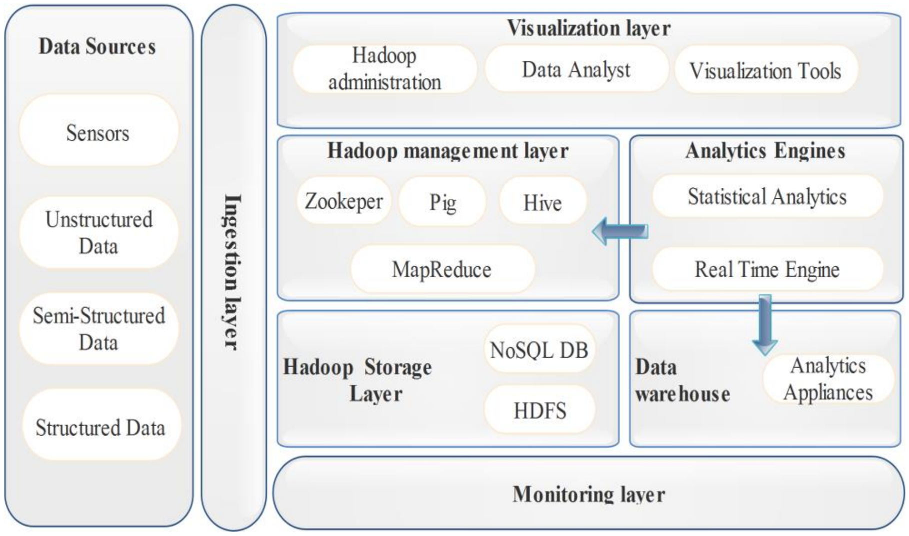


Fig. 2. The proposed Hadoop BD architecture

IV. RESULTS

The ingestion layer permits the selection of the interesting pollution or climate variables from a wide range of the stored Raster datasets or from the converted files. In this investigation, we were interested on twelve variables including temperature, humidity, pressure, wind speed, AOD, VCD of trace gases and so on [29]. The output of the extraction script is saved into a CSV file (see Fig. 4 and 5). The final plots are the refined datasets that could be integrated and used in the forthcoming as input of AI algorithms. In addition, the final output has a unified structure and format. It can be integrated into an HDFS to apply some environmental models (CALMET and CALPUFF) and algorithms of interpolation fusion and validation based on DL from Fig. 3, we can notice that the number of plots is also different among countries. Big countries such as China, USA and Australia reach more than forty million plots per day, however for moderate countries surface as, for instance, Morocco and Spain, there are only seven million plots per day. We remark also that channels using a sensor with a high resolution including (MOD14, MOD11, and AOD) occur a huge number of the measured values. Fig. 6 and 7 show the Vertical Mixing Ration (VMR) with the Part per Million or Billion unit (PPMv-PPBv) of the Ozone (O_3) and Carbone Monoxide (CO) of Morocco in 11/06/2018, respectively. The altitude of this measurement is between 0-200 meters. These maps are an example of the visualization tools showing the low tropospheric air pollution of Morocco helping decision makers. We notice that the density of the trace gases is significant near industrials zones located in Casablanca, Safi and Tangier. More over the concentration of CO and SO_2 are high in the coastal areas near ports due to the high activities of the maritime transport.

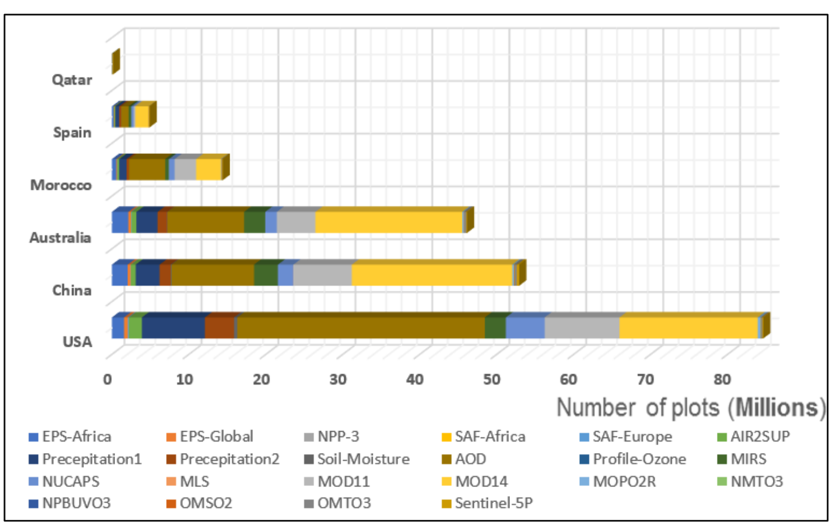
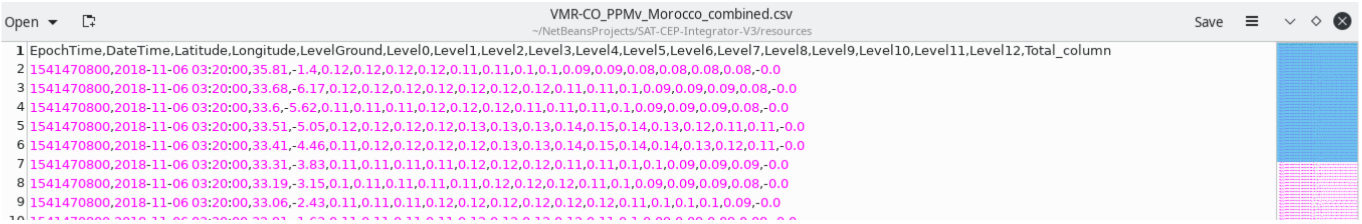


Fig. 3. The number of plots after the processing

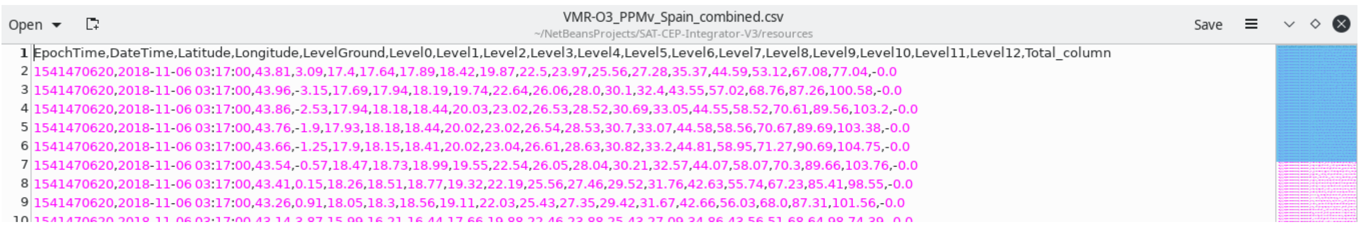


Open | Save

VMR-CO_PPMv_Morocco_combined.csv
~/NetBeansProjects/SAT-CEP-Integrator-V3/resources

```
1 EpochTime,DateTime,Latitude,Longitude,LevelGround,Level0,Level1,Level2,Level3,Level4,Level5,Level6,Level7,Level8,Level9,Level10,Level11,Level12,Total_column
2 1541470800,2018-11-06 03:20:00,35.81,-1.4,0.12,0.12,0.12,0.12,0.11,0.11,0.1,0.1,0.09,0.09,0.08,0.08,0.08,0.08,-0.0
3 1541470800,2018-11-06 03:20:00,33.68,-6.17,0.12,0.12,0.12,0.12,0.12,0.12,0.12,0.11,0.11,0.1,0.09,0.09,0.09,0.08,-0.0
4 1541470800,2018-11-06 03:20:00,33.6,-5.62,0.11,0.11,0.11,0.12,0.12,0.12,0.11,0.11,0.11,0.1,0.09,0.09,0.09,0.08,-0.0
5 1541470800,2018-11-06 03:20:00,33.51,-5.05,0.12,0.12,0.12,0.12,0.13,0.13,0.13,0.14,0.15,0.14,0.13,0.12,0.11,0.11,-0.0
6 1541470800,2018-11-06 03:20:00,33.41,-4.46,0.11,0.12,0.12,0.12,0.12,0.13,0.13,0.14,0.15,0.14,0.14,0.13,0.12,0.11,-0.0
7 1541470800,2018-11-06 03:20:00,33.31,-3.83,0.11,0.11,0.11,0.11,0.12,0.12,0.12,0.11,0.11,0.1,0.1,0.09,0.09,0.09,-0.0
8 1541470800,2018-11-06 03:20:00,33.19,-3.15,0.1,0.11,0.11,0.11,0.11,0.12,0.12,0.12,0.11,0.1,0.09,0.09,0.09,0.08,-0.0
9 1541470800,2018-11-06 03:20:00,33.06,-2.43,0.11,0.11,0.11,0.12,0.12,0.12,0.12,0.12,0.12,0.11,0.1,0.1,0.1,0.09,-0.0
```

Fig. 4. The CSV output containing the VMR of CO in Morocco (PPMv)



Open | Save

VMR-O3_PPMv_Spain_combined.csv
~/NetBeansProjects/SAT-CEP-Integrator-V3/resources

```
1 EpochTime,DateTime,Latitude,Longitude,LevelGround,Level0,Level1,Level2,Level3,Level4,Level5,Level6,Level7,Level8,Level9,Level10,Level11,Level12,Total_column
2 1541470620,2018-11-06 03:17:00,43.81,3.09,17.4,17.64,17.89,18.42,19.87,22.5,23.97,25.56,27.28,35.37,44.59,53.12,67.08,77.04,-0.0
3 1541470620,2018-11-06 03:17:00,43.96,-3.15,17.69,17.94,18.19,19.74,22.64,26.06,28.0,30.1,32.4,43.55,57.02,68.76,87.26,100.58,-0.0
4 1541470620,2018-11-06 03:17:00,43.86,-2.53,17.94,18.18,18.44,20.03,23.02,26.53,28.52,30.69,33.05,44.55,58.52,70.61,89.56,103.2,-0.0
5 1541470620,2018-11-06 03:17:00,43.76,-1.9,17.93,18.18,18.44,20.02,23.02,26.54,28.53,30.7,33.07,44.58,58.56,70.67,89.69,103.38,-0.0
6 1541470620,2018-11-06 03:17:00,43.66,-1.25,17.9,18.15,18.41,20.02,23.04,26.61,28.63,30.82,33.2,44.81,58.95,71.27,90.69,104.75,-0.0
7 1541470620,2018-11-06 03:17:00,43.54,-0.57,18.47,18.73,18.99,19.55,22.54,26.05,28.04,30.21,32.57,44.07,58.07,70.3,89.66,103.76,-0.0
8 1541470620,2018-11-06 03:17:00,43.41,0.15,18.26,18.51,18.77,19.32,22.19,25.56,27.46,29.52,31.76,42.63,55.74,67.23,85.41,98.55,-0.0
9 1541470620,2018-11-06 03:17:00,43.26,0.91,18.05,18.3,18.56,19.11,22.03,25.43,27.35,29.42,31.67,42.66,56.03,68.0,87.31,101.56,-0.0
```

Fig. 5. The CSV output containing the VMR of O_3 in Morocco (PPMv)

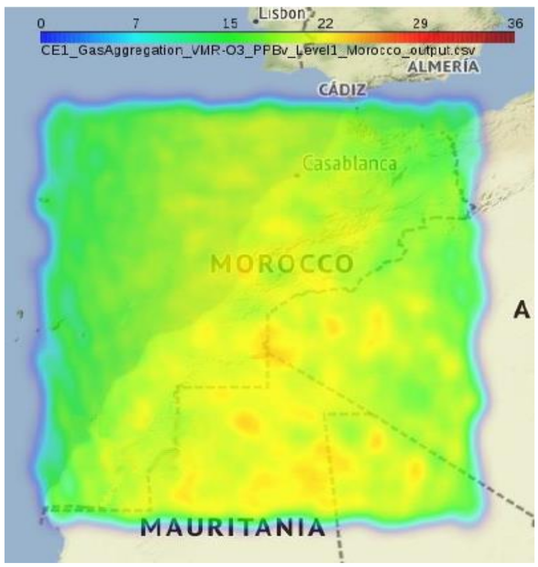


Fig. 6. VMR of O_3 (PPBv)

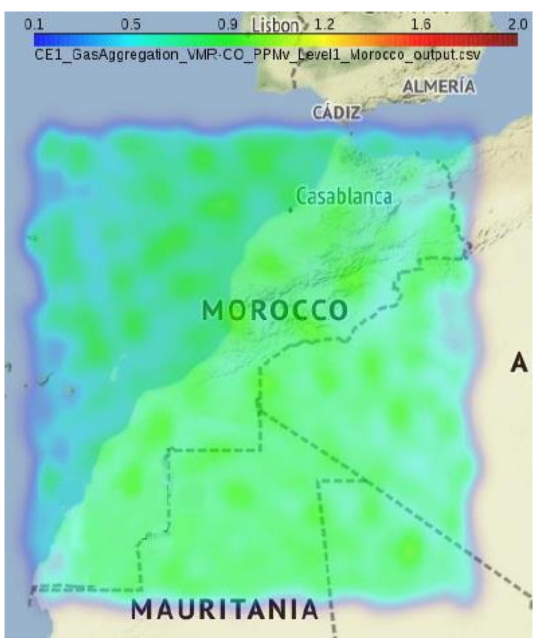


Fig. 7. VMR of CO (PPMv)

V. COMPARISON

In this section, we will show and compare other studies witch process environmental RS data.

There is some other investigations witch:

- Monitor the atmospheric pollution using the IASI data collected from the MetOp satellites. This study has also developed an algorithm validating and optimizing the quality of the IASI datasets [30].
- Depict the changes in the surface temperature using the MGS data gathered from the SEVIRI sensor. A developed algorithm used these variations as input to predict earthquakes in Algeria [31].
- Supervise continuously the sea, land, ice, and forest by processing the ENVISAT data with a novel parallel architecture based on Hadoop and MapReduce [32].
- Develop a java-based application processing the MDEO data in NRT to monitor and correlate the air pollution in Morocco with their emitting sources such as factories, power thermal plants and so on [21].

TABLE 2
COMPARISON AMONG RELATED WORKS

Study	RS Data					RS Data Processing		RS Application
	Sources	Satellites	Sensors	Size/day	Velocity/day	Architecture	Tools	
This study	4	11	16	More than 56 Gb	More than 36000 Files	Single & distributed (Hadoop)	Java Python Bash	AQ Forest fire Climate changes
[30]	1	1	1	About 566 Mb	About 2000 Files	Single	Python	Air Pollution
[31]	1	1	1	About 10 Mb	About 8 Files	Single	AEPA	Earthquake prediction
[32]	1	1	1	About 500 Mb	About 5 Files	Distributed (Hadoop)	Pig	Climate changes
[21]	2	2	4	More than 1 Gb	About 2200 Files	Single	Java	Air pollution

From Table 2, we notice that the comparison focuses mainly on the features and the characteristics of the processed RS data, particularly: the RS data sources, size, velocity the processing paradigm, the language of development and the field of application.

We remark also that all the aforementioned studies RS data for an environmental application, particularly: AQ monitoring, forest fire detection, climate changes supervision and earthquakes prediction.

The majority of these studies used only one RS data source, one satellite, and a maximum of 4 satellite sensors. The total size treated per day does not exceed 1 GB and 2200 files. However, our study uses four data sources especially: The MDEO, the NASA, the NOAA and the Copernicus data and processes daily 56 GB stored in more than 36.000 files.

Accordingly, our input of RS data is very consistent and will help to occur accurate results after. In addition, our proposed architecture of processing could be executed in a single architecture executed on a standard computer or distributed using a Hadoop or a Cloud cluster.

VI. CONCLUSION AND RECOMMENDATIONS

During this last decade, air pollution and climate change have been two phenomena that affect the environments safety and human is health. This is due to the emission of pollutant gases particularly, CO, CO_2, NO_x and so on, from industrials, transport and agriculture activities. Thus, the continuous monitoring of atmospheric composition in NRT become highly significant. The key solution is to employ RS techniques that provide global scale satellite data in NRT. Our research aims to monitor abnormal climate changes and supervise AQ especially in Morocco. We collected RSBD in NRT from six sources which are the MDEO ground station of EUMETSAT data, the EOSDIS data of NASA, the NESDIS data of NOAA, the Copernicus platform, some MGS data, and the Raspberry PI sensors data. The handicap is that these datasets have not only a huge volume and velocity but also come with different file formats since they cover a global scale, and they come from different satellite sensors with a wide spectral resolution. Accordingly, we have proved that RS data are BD according to the four salient: volume, variety, velocity, and veracity. In our case, RSBD is heavy in terms of size, reaching 56 GB per day, stored in different files extensions including (BIN, BUFR, NetCDF, HDF5, etc.) and have a high velocity averaged within 36.000 daily files. There is no doubt that the existing systems and architectures are so limited to handle the NRT RSBD. As a result, we have adopted the Hadoop architecture to RSBD to process efficiently this kind of data. This architecture is composed of six main layers as follow: the data sources, data ingestion, data storage, data processing, data visualization, and the monitoring layer. The aforementioned architecture automatically collects filters, extracts and stores data into the HDFS. Moreover, the processing layer will include some AI algorithms increasing data accuracy. Finally, the visualization layer will show data in NRT into interactives maps and charts helping decision-makers. This architecture should be supplied with another cloud computing architecture that may optimize the time execution. This task would be an interesting work to be conducted in the future.

Declaration of Conflicting Interests

There are no competing interests in this research.

Acknowledgment

The corresponding author of this manuscript is thankful to the CAMPUS MARE NOSTRUM for the grant and the formation about the techniques of the English redaction of the scientific articles which was in the university of Murcia in Spain between 16-18 July 2019. https://bit.ly/2Rba9M4

REFERENCES

[1] K. R. Smith, N. Bruce, K. Balakrishnan, H. Adair-Rohani, J. Balmes, Z. Chafe, M. Dherani, H. D. Hosgood, S. Mehta, D. Pope *et al.*, "Millions dead: How do we know and what does it mean? Methods used in the comparative risk assessment of household air pollution," *Annual Review of Public Health*, vol. 35, pp. 185–206, 2014. doi: https://doi.org/10.1146/annurev-publhealth-032013-182356

[2] Z.-X. Wang and J.-M. Lin, "Application of the greens function in attenuation factor to solve the status of human health on the earth," *International Journal of Technology and Engineering Studies*, vol. 4, no. 1, pp. 32–41, 2018. doi: https://doi.org/10.20469/ijtes.4.10005-1

[3] E. Chuvieco and R. G. Congalton, "Application of remote sensing and geographic information systems to forest fire hazard mapping," *Remote Sensing of Environment*, vol. 29, no. 2, pp. 147–159, 1989. doi: https://doi.org/10.1016/0034-4257(89)90023-0

[4] Y. X. Lee and Z. W. Zhong, "A study of the relationship between adverse weather conditions and flight delay," *Journal of Advances in Technology and Engineering Research*, vol. 2, no. 4, pp. 113–117, 2016. doi: https://doi.org/10.20474/jater-2.4.2

[5] R. J. Boain, *AB-Cs of sun-synchronous orbit mission design*. Pasadena, CA: Jet Propulsion Laboratory, National Aeronautics and Space., 2004.

[6] G. Chijioke, "Satellite remote sensing technology in spatial modeling process: Technique and procedures," *International Journal of Science and Technology*, vol. 2, no. 5, pp. 309–315, 2012.

[7] S. Gupta. (2018) Active and passive remote sensing. [Online]. Available: https://bit.ly/2GfIFi6

[8] L. Zhu, J. Suomalainen, J. Liu, J. Hyyppa, H. Kaartinen, and H. Haggren, "A review: Remote sensing sensors," in *Multi-purposeful Application of Geospatial Data*, R. B. Rustamov, S. Hasanova, and M. H. Zeynalova, Eds. London, UK: IntechOpen, 2018.

[9] C. El Amrani, G. L. Rochon, T. El-Ghazawi, G. Altay, and T. Rachidi, "Development of a real-time urban remote sensing initiative in the mediterranean region for early warning and mitigation of disasters," in *IEEE International Geoscience and Remote Sensing Symposium*, Munich, Germany, 2012, pp. 2782–2785.

[10] C. L. Parkinson, A. Ward, and M. D. King, *Earth science reference handbook: A guide to NASAs earth science program and earth observing satellite missions*. Washigton, DC, WA: National Aeronautics and Space Administration, 2006.

[11] A. Garcia-de Prado, G. Ortiz, J. Boubeta-Puig, and D. Corral-Plaza, "Air4people: A smart air quality monitoring and context-aware notification system," *Journal of Universal Computer Science*, vol. 24, no. 7, pp. 846–863, 2018.

[12] V. Karhila and A. Hernandez-Carrascal, "BUFR: A meteorological code for the 21st century," in *EUMETSAT Meteorological Satellite Conference*, Cordoba, Spain, 2010.

[13] L. Gosink, J. Shalf, K. Stockinger, K. Wu, and W. Bethel, "HDF5-fastquery: Accelerating complex queries on HDF datasets using fast bitmap indices,"

in *18th International Conference on Scientific and Statistical Database Management (SSDBM'06)*, Vienna, Austria, 2006, pp. 149–158.

[14] R. Rew and G. Davis, "NetCDF: An interface for scientific data access," *IEEE Computer Graphics and Applications*, vol. 10, no. 4, pp. 76–82, 1990. doi: https://doi.org/10.1109/38.56302

[15] Y. Ma, H. Wu, L. Wang, B. Huang, R. Ranjan, A. Zomaya, and W. Jie, "Remote sensing big data computing: Challenges and opportunities," *Future Generation Computer Systems*, vol. 51, pp. 47–60, 2015. doi: https://doi.org/10.1016/j.future.2014.10.029

[16] R. Showstack, "Sentinel satellites initiate new era in earth observation," *Eos, Transactions American Geophysical Union*, vol. 95, no. 26, pp. 239–240, 2014. doi: https://doi.org/10.1002/2014eo260003

[17] R. Youssra and R. Sara, "Big data and big data analytics: Concepts, types and technologies," *International Journal of Research and Engineering*, vol. 5, no. 9, pp. 524–528, 2018. doi: https://doi.org/10.21276/ijre.2018.5.9.5

[18] T. Lyons, J. Kenworthy, C. Moy, and F. Dos Santos, "An international urban air pollution model for the transportation sector," *Transportation Research Part D: Transport and Environment*, vol. 8, no. 3, pp. 159–167, 2003. doi: https://doi.org/10.1016/s1361-9209(02)00047-0

[19] H.-D. Lee, J.-W. Yoo, M.-K. Kang, J.-S. Kang, J.-H. Jung, and K.-J. Oh, "Evaluation of concentrations and source contribution of PM10 and SO_2 emitted from industrial complexes in Ulsan, Korea: Interfacing of the WRF–CALPUFF modeling tools," *Atmospheric Pollution Research*, vol. 5, no. 4, pp. 664–676, 2014. doi: https://doi.org/10.5094/apr.2014.076

[20] M. A. Zaytar and C. El Amrani, "Ozonet: Atmospheric ozone interpolation with deep convolutional neural networks," *Geoscientific Instrumentation, Methods and Data Systems Discussions*, pp. 1–16, 2019. doi: https://doi.org/10.5194/gi-2018-53

[21] B.-R. B. Semlali, C. El Amrani, and S. Denys, "Development of a java-based application for environmental remote sensing data processing," *International Journal of Electrical and Computer Engineering*, vol. 9, no. 3, pp. 1978–1986, 2019. doi: https://doi.org/10.11591/ijece.v9i3.pp1978-1986

[22] J. Schulz, P. Albert, H.-D. Behr, D. Caprion, H. Deneke, S. Dewitte, B. Durr, P. Fuchs, A. Gratzki, P. Hechler *et al.*, "Operational climate monitoring from space: The EUMETSAT satellite application facility on climate monitoring (CM-SAF)," *Atmospheric Chemistry and Physics*, vol. 9, no. 5, pp. 1687–1709, 2009. doi: https://doi.org/10.5194/acp-9-1687-2009

[23] B. B. Semlali and E. A. C., "Towards remote sensing datasets collection and processing," *International Journal of Embedded and Real-Time Communication Systems (IJERTCS)*, vol. 10, no. 3, pp. 49–67, 2019. doi: https://doi.org/10.4018/ijertcs.2019070104

[24] M. A. Zaytar and C. E. Amrani, "Metop satellites data processing for air pollution monitoring in morocco," *International Journal of Electrical & Computer Engineering*, vol. 8, no. 6, pp. 4584–4592, 2018. doi: https://doi.org/10.11591/ijece.v8i6.pp4584-4592

[25] J. Schmetz, P. Pili, S. Tjemkes, D. Just, J. Kerkmann, S. Rota, and A. Ratier, "An introduction to Meteosat Second Generation (MSG)," *Bulletin of the American Meteorological Society*, vol. 83, no. 7, pp. 977–992, 2002. doi: https://doi.org/10.1175/1520-0477(2002)083<0992:staitm>2.3.co;2

[26] R. Munro, M. Eisinger, C. Anderson, J. Callies, E. Corpaccioli, R. Lang, A. Lefebvre, Y. Livschitz, and A. P. Albinana, "GOME-2 on MetOp," in *Proceeding of The EUMETSAT Meteorological Satellite Conference*, Helsinki, Finland, 2006, pp. 12–16.

[27] National Oceanic and Atmospheric Administration. (2019) NOAA National Environmental Satellite, Data, and Information Service (NESDIS). [Online]. Available: https://bit.ly/310aTXX

[28] C. Wang, F. Hu, X. Hu, S. Zhao, W. Wen, and C. Yang, "A hadoop-based distributed framework for efficient managing and processing big remote sensing images," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 2, no. 4, p. 63, 2015. doi: https://doi.org/10.5194/isprsannals-ii-4-w2-63-2015

[29] B. S. B. E. and E. A. C., "Towards remote sensing datasets collection and processing," in *Transactions on Large-Scale Data- and Knowledge-Centered Systems XLI*, H. A, W. R, and D. TK, Eds. Berlin, Heidelberg: Springer, 2019.

[30] C. Clerbaux, A. Boynard, L. Clarisse, M. George, J. Hadji-Lazaro, H. Herbin, D. J. Hurtmans, M. Pommier, A. Razavi, S. Turquety *et al.*, "Monitoring of atmospheric composition using the thermal infrared iasi/metop sounder," vol. 9, no. 16, pp. 6041–6054, 2009. doi: https://doi.org/10.5194/acp-9-6041-2009

[31] A. Hassini and A. H. Belbachir, "Thermal infrared geostationary satellite sensor data application for

prediction and monitoring earthquake in Algeria," *International Journal of Information Technology and Management*, vol. 15, no. 4, pp. 293–312, 2016. doi: https://doi.org/10.1504/ijitm.2016.079600

[32] M. M. U. Rathore, A. Paul, A. Ahmad, B.-W. Chen, B. Huang, and W. Ji, "Real-time big data analytical architecture for remote sensing application," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 10, pp. 4610–4621, 2015. doi: https://doi.org/10.1109/jstars.2015.2424683