Environment. Technology. Resources. Rezekne, Latvia
Proceedings of the 16th International Scientific and Practical Conference. 2025. Volume II, 383-388

Pollution Detection Model Utilizing Anomaly Detection Techniques Performing Analysis of Satellite Data

Roberts Volkovičs
Modelling of sociotechnical systems
Vidzeme University of Applied Sciences
Valmiera, Latvia
roberts.volkovics@va.lv

***Abstract*—In this article review was created of nowadays popular way in science of working on climate change topics by processing and performing analysis of open data from Earth orbiting satellites. Open data from satellites are provided by The European Space Agency or ESA, National Aeronautics and Space Administration or NASA via portals and cloud services under their management and by other organizations via such cloud service as Creodias, for example. Open data from satellites are available for research for other topics as well, but in the scope of this article the topic of climate change is in the focus by achieving research goal and proving hypothesis that it is possible to build the pollution detection model and pollution location detection model by utilizing anomaly detection techniques and performing analysis of satellite data.**

Keywords — Analysis, Anomaly Detection (AD), Location Detection, Model, Pollution Location, Satellite Data.

I. INTRODUCTION

Climate change and related topics are very popular in science nowadays and many researches are working in this field with the aim to improve the overall situation that is constantly worthening regarding climate on the Earth. As the evidence to situation worthening such phenomena could be mentioned as increasing global air temperature, increasing global water temperature in the oceans, weather anomalies like heavy storms and other.

Author of the article works on different topic related to anomaly detection and has found that open data from satellites provide good basis for experiments with anomaly detection techniques in the area of big and dense data. This research in the field of climate change was done performing analysis of satellite data with the aim to create the model capable to detect pollution or source of pollution using anomaly detection techniques. The hypothesis of the research: "it is possible to build a model for analysis of satellite sensor data for pollution detection or pollution location detection utilizing anomaly detection techniques."

II. MATERIALS AND METHODS

During this research an open data from The European Space Agency or ESA "Copernicus" [1] project was used to perform the analysis to achieve the research goal and prove the hypothesis, that it is possible to construct the model based on anomaly detection techniques and algorithms, that could be utilized for pollution detection and pollution location detection on Earth.

National Aeronautics and Space Administration or NASA, has prepared a guide on how to access "Copernicus" project open data [2]. Besides the option to download "Copernicus" project's data from several portals, NASA provides an option to download "Copernicus" project's data as well and there are options to request data in different formats. For that purpose, there is "Earth data Search" tool available online [3]. Additional information on "Sentinel" project satellites, for example, products, naming conventions could be found in "Senti Wiki" portal [4].

"Sentinel" project has 6 types of satellites with several satellites in for each type going around the Earth by different orbits as demonstrated in "Fig. 1".

All the satellites of "Sentinel" project provide following "Copernicus" project services:

Online ISSN 2256-070X
https://doi.org/10.17770/etr2025vol2.8566

- The Copernicus Atmosphere Monitoring Service (CAMS) provides continuous data and information on atmospheric composition. The service describes the current situation, forecasts the situation a few days ahead, and analyses consistently retrospective data records for recent years. The Copernicus Atmosphere Monitoring Service supports many applications in a variety of domains including health, environmental monitoring, renewable energies, meteorology, and climatology [5];

Fig. 1. Sentinel satellites

- The Copernicus Marine Service (or Copernicus Marine Environment Monitoring Service) is the marine element of Copernicus, the Earth observation component of the European Union's Space programme. It provides free, regular, and systematic authoritative information on the state of the Blue (physical), White (sea ice) and Green (biogeochemical) ocean, on a global and regional scale [6];
- The Copernicus Land Monitoring Service (CLMS) provides geospatial information on land cover and its changes, land use, vegetation state, water cycle and Earth's surface energy variables to a broad range of users in Europe and across the World in the field of environmental terrestrial applications [7];
- The Copernicus Climate Change Service (C3S) supports society by providing authoritative information about the past, present and future climate in Europe and the rest of the World [8];
- The Copernicus service for Security applications aims to support European Union policies by providing information in response to Europe’s security challenges [9];
- The Copernicus Emergency Management Service (Copernicus EMS) provides all actors involved in the management of natural disasters, man-made emergency situations, and humanitarian crises with timely and accurate geo-spatial information derived from satellite remote sensing and completed by available in situ or open data sources [10].

In the scope of this article, the main attention is paid to “Sentinel-5P” mission which provides open data for “The Copernicus Climate Change Service (C3S)” service.

Sentinel-5p TROPOMI sensor applications are encompassing the air quality and pollution, the ozone-layer monitoring, the climate change, and the aviation safety. All these applications are presented in depth, with the specification of the corresponding product types [8].

“The Copernicus data space ecosystem” [11] is one of the portals where open data from different satellites are made available; similar services for different organization satellites are provided by “Creodias” [12], which is cloud service provider for Earth observation data.

In case to retrieve “raw” data from satellites the Copernicus service or the Creodias service could be utilized. To avoid computation on mapping “raw” satellite data to Earth area, in scope of the research, data were requested and downloaded locally from NASA “Earth Search Data” service [3]. These data are mapped to the Earth area and instead of representing exact results of satellite sensor scan over particular Earth area at specific moment in time, the data represent (in one NCDF4 file) one week of moving average for each day of observations over the Earth area. To prove the hypothesis of the research NCDF4 files over five-year long period are processed.

Central and South America was chosen as the region of choice for the research, because it is known to the science that winds in stratosphere usually are bidirectional in this region: with east-wind or west-wind component of direction.

The winds in the stratosphere and mesosphere are usually estimated from temperature data collected by satellites. The winds at these high levels are assumed to be geostrophic. Overall, in the midlatitudes, they have a westerly component in the winter and an easterly component in the summer [13].

The phenomenon of winds in the stratosphere and mesosphere having east-wind or west-wind component of direction has been so stable in long term that even some changes in this behaviour are considered as anomaly by some researchers who are monitoring this phenomenon closely.

In article [14], the graphical data representation of described phenomenon could be found as demonstrated in “Fig. 2”.

Based on phenomenon described and rules of logic and probability, it could be assumed, that any kind of pollution or pollution location, if it exists at all, must show up after analysis of satellite data when performing data representation in graphical diagram as horizontally (Easterly-Westerly) oriented regions of lower/higher probability areas or polygons of research metric or variable visual representation. For this research, visual representation of research metric or variable serves as the main method for proving hypothesis.

As research metrics or variables density of ozone in stratosphere from two different altitudes were used.

As a pollution factor freon gas will be considered, which reacts with ozone molecules repeatedly turning ozone in oxygen and causing free atom of oxygen to construct other chemical elements in different chemical reactions. Such a pollution must be visible as anomalies in long term geospatial data, especially very close to pollution location, in case properly visualized.

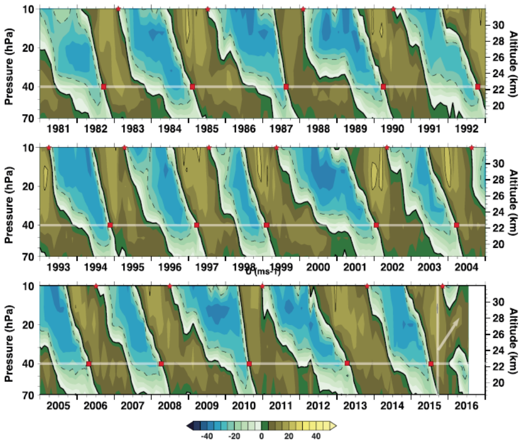

Fig. 2. Phenomenon East and West component of wind alternation [14].

All the mentioned factors combined were joined in the plan of the research to process 5 years of "Sentinel-5" data utilizing the model for pollution detection based on anomaly detection techniques. The model was created as program written in MATLAB using MATLAB modules for NCDF4 file processing, geospatial data processing, anomaly detection techniques, visual representation and regular functionality for file and variable processing in software program cycles the same way as in programs written in other programming languages.

Calculations are done considering rules of statistics - utilizing variable scaling techniques and anomaly scoring, and considering that in normal conditions density of ozone in stratosphere has normal probability distribution.

In "Fig. 3" the web interface of portal Creodias is visible, allowing to choose the region of interest for which to request and download satellite data. To calculate the distribution of the sample of ozone density the distant region in the ocean was chosen assuming there is the minimum of impact of human caused pollution.

In "Fig. 4" the web interface of portal Creodias allowing to pick the data from satellite is demonstrated.

In the "Fig. 5" the histogram of variable of ozone density is visible.

The histogram visible in the "Fig. 5" is in almost perfect form of bell, which suggests that the density of the ozone in normal circumstances has normal probability distribution.

Normal probability distribution is the most frequently occurring distribution for most measurable metrics in nature, like air temperature in a certain season, a time of a day and an area.

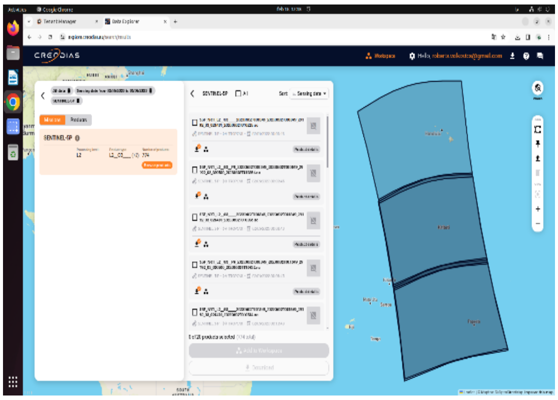

Fig. 3. Creodias, choosing the region of interest

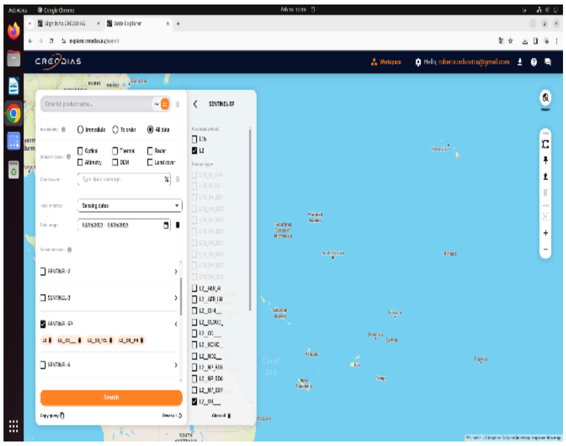

Fig. 4. Creodias choosing the product of interest

That allows to perform statistical calculations and anomaly detection assuming that ozone density during normal conditions has normal probability distribution.

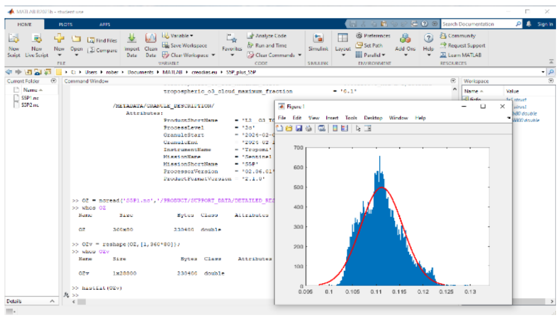

Fig. 5. Histogram of ozone density

In the "Fig. 6" the screenshot of MATLAB displaying the structure of NCDF4 file is visible.

Statistical operations with variables having normal distributions allow to perform such calculations as:

- anomaly detection based on z-score methodology;
- addition of average values of the same variable at two different altitudes;
- scaling for scoring processing of anomaly detection.

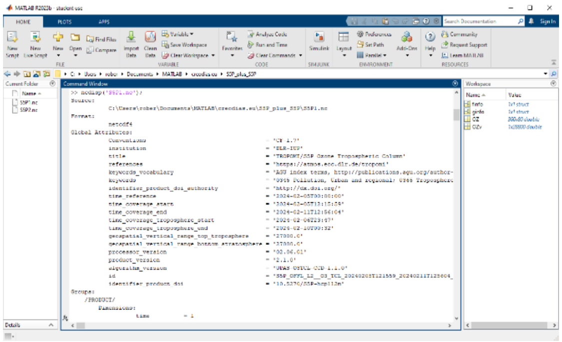

Fig. 6.. NCDF4 structure from "Sentinel5"

Normal distribution is one of the most widely spread distributions in nature and many statistics methods were created based on this type of distribution and are only available for it.

III. Results and discussion

After processing the data, the list of graphical visualizations was created showing the overall anomaly level over region of interest for the years 2019, 2020, 2021, 2022, 2023.

More than 330 files for each year were processed by anomaly detection techniques calculating the minimum level of ozone density, the maximum level of ozone density and the normal level of ozone density.

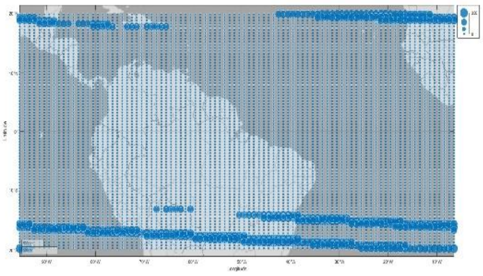

Fig. 7. 2019. Anomaly scoring category minimums ozone density

The minimum level of ozone density for the year 2019 is visible in “Fig. 7”

As visible in “Fig. 7”, for the year 2019 the minimums of ozone density tend to draw horizontal line at the top and at the bottom of the graph.

In Fig. 8, the maximum ozone density is displayed for the year 2019.

As demonstrated in “Fig. 8”, the maximums tend to create horizontal areas in the middle of the geographical region.

In the “Fig. 9”, overall results are displayed for all three scoring categories: minimums, maximums, and normal level of anomalies of ozone density for the year 2019.

Light blue colour represents anomaly scoring level for minimums, green for maximums and regular blue for both minimums and maximums having normal level.

In the “Fig. 10”, the minimums of ozone density are displayed for the year 2020.

In the “Fig. 11”, the minimums of ozone density are displayed for the year 2021.

In the “Fig. 12”, the minimums of ozone density are displayed for the year 2022.

In the “Fig. 13”, the minimums of ozone density are displayed for the year 2023.

In the “Fig. 14” the map of populated places from “Latin America and the Caribbean (LAC) Population Database” is visible [15.]

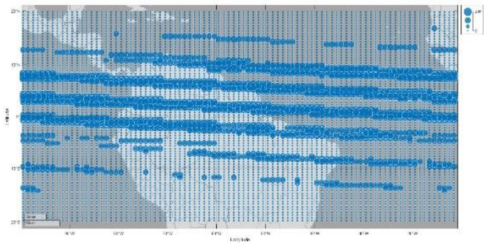

Fig. 8. 2019. anomaly scoring category maximums ozone density

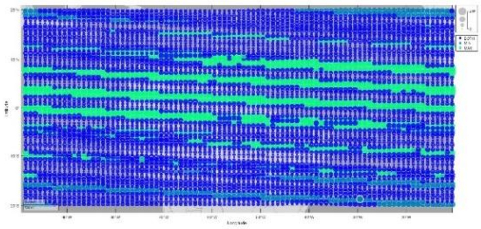

Fig. 9. 2019. anomaly scoring all categories ozone densitv

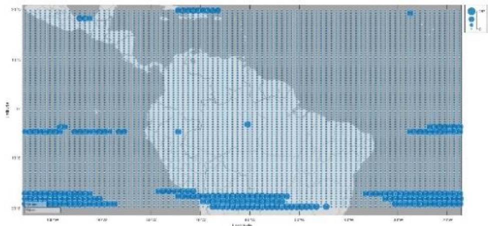

Fig. 10. 2020. anomaly scoring category minimums ozone density

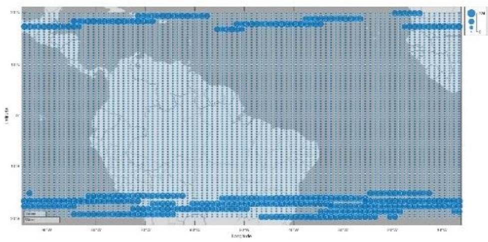

Fig. 11. 2021. anomaly scoring category minimums ozone density

At some extent, figures: “Fig. 7”, “Fig. 10”, “Fig. 11”, “Fig. 12”, “Fig. 13” are following the same pattern and clearly demonstrate horizontally oriented areas of ozone minimum very close to most populated places visible in map from "Fig. 14".

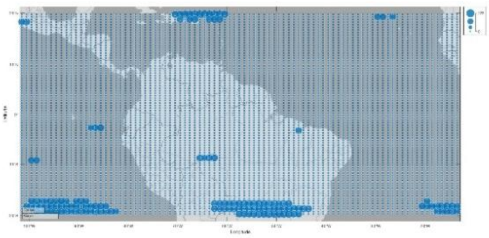

Fig. 12. 2022. anomaly scoring category minimums ozone density

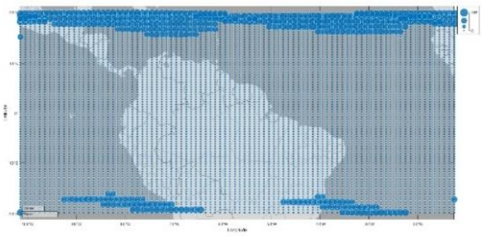

Fig. 13. 2023. anomaly scoring category minimums ozone density

As previously discussed, such minimums of ozone density could be caused by pollution of freon gas and the close location to populated places could be direct cause of the pollution.

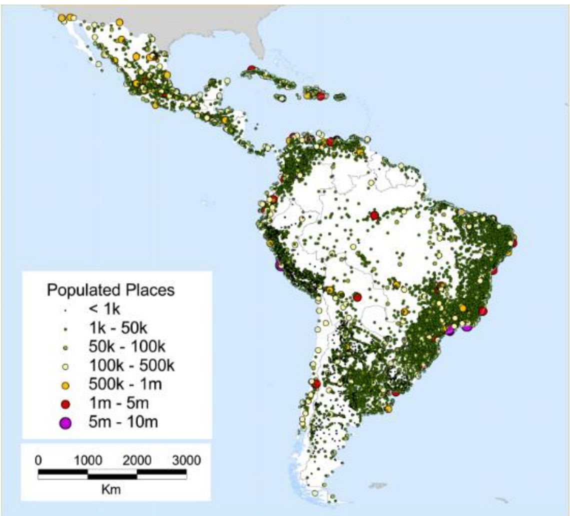

Fig. 14. Central and South America populated places [15]

It should be considered that winds near earth could very well affect direction of how freon gas lifts until stratosphere, but still it could be assumed that the location of pollution could be detected by this approach to certain precision of geographical area.

IV. Conclusions

After analysis and processing of open data from "Sentinel-5" project data sets [16] and [17], it was possible to generate graphical representation of anomaly scoring level our outlierness for specific geographic region for 5 years long period. Since geographic region is in tropic area of Earth where wind directions in stratosphere mostly have only east and west components, it is possible to conclude that horizontally oriented lines, areas, or polygons representing anomaly scoring category minimums in particular geographical areas and those are most likely very close to freon gas pollution sources.

Areas with anomaly scoring category minimums are very close in geographic map with those areas of highest density of human population like large cities with many million inhabitants which most likely is the cause of freon gas pollution as it was used some time ago for such devices as refrigerators, conditioners etc.

Luckily humanity has agreed to not produce anymore such type of devices with freon gas components and situation with ozone layer density started to improve recently.

The goal of the research is achieved by proving hypothesis that it is possible to build the model with capability of pollution detection performing analysis of satellite data and proving hypothesis by graphical representation of anomaly scoring results.

Acknowledgments

This research was funded by the European Commission, Research Executive Agency grant number 101079206, "Twinning in Environmental Data and Dynamical Systems Modelling for Latvia" (TED4LAT).

References

[1] Copernicus, "Europe's eyes on Earth." [Online]. Available: https://www.copernicus.eu/en [Accessed: Sep. 19, 2024].
[2] National Aeronautics and Space Administration, "TUTORIAL FOR DATA DOWNLOAD FROM THE COPERNICUS OPEN ACCESS HUB," *National Aeronautics and Space Administration*, 2024. [Online]. Available: https://appliedsciences.nasa.gov/sites/default/files/2021-10/Part2_Data-Download_EN.pdf [Accessed: Sep. 19, 2024].
[3] National Aeronautics and Space Administration, "Earth data Search," *National Aeronautics and Space Administration*, 2024. [Online]. Available: https://search.earthdata.nasa.gov/search [Accessed: Sep. 19, 2024].
[4] Copernicus, "Senti Wiki," [Online]. Available: https://sentiwiki.copernicus.eu/web/sentiwiki [Accessed: Sep. 19, 2024].
[5] Copernicus, "The Copernicus Atmosphere Monitoring Service (CAMS)," [Online]. Available: https://www.copernicus.eu/en/copernicus-services/atmosphere [Accessed: Sep. 19, 2024].
[6] Copernicus, "The Copernicus Marine Service," [Online]. Available: https://www.copernicus.eu/en/copernicus-services/marine [Accessed: Sep. 19, 2024].
[7] Copernicus, "The Copernicus Land Monitoring Service (CLMS)," [Online]. Available: https://www.copernicus.eu/en/copernicus-services/land [Accessed: Sep. 19, 2024].
[8] Copernicus, "The Copernicus Climate Change Service (C3S)," [Online]. Available: https://www.copernicus.eu/en/copernicus-services/climate-change [Accessed: Sep. 19, 2024].
[9] Copernicus, "The Copernicus service for Security applications," [Online]. Available: https://www.copernicus.eu/en/copernicus-services/security [Accessed: Sep. 19, 2024].

[10] Copernicus, "The Copernicus Emergency Management Service (Copernicus EMS)," [Online]. Available: https://www.copernicus.eu/en/copernicus-services/emergency [Accessed: Sep. 19, 2024].
[11] Copernicus, "The Copernicus data space ecosystem," [Online]. Available: https://dataspace.copernicus.eu/ [Accessed: Sep. 19, 2024].
[12] Creodias, "CREODIAS offers a cloud computing solution for EO data processing," [Online]. Available: https://creodias.eu/ [Accessed: Sep. 19, 2024].
[13] Britannica, "Winds in the stratosphere and mesosphere," [Online]. Available: https://www.britannica.com/science/climate-meteorology/Winds-in-the-stratosphere-and-mesosphere [Accessed: Sep. 24, 2024].
[14] P. A. Newman, L. Coy, S. Pawson, and L. R. Lait, "The anomalous change in the QBO in 2015–2016," *Geophys. Res. Lett., 43, 8791–8797*, 2016. [Online serial]. Available: https://dx.doi:10.1002/2016GL070373 [Accessed: Sep. 24, 2024].
[15] Centro Internacional de Agricultura Tropical (CIAT), United Nations Environment Program (UNEP), Center for International Earth Science Information Network (CIESIN), Columbia University, and the World Bank (2005), "Latin America and the Caribbean (LAC) Population Database," [Online]. Available: https://gisweb.ciat.cgiar.org/population/report.htm [Accessed: Sep. 24, 2024].
[16] "2018-04-30 to 2020-07-07 Sentinel-5P TROPOMI Tropospheric Ozone Column V1 (S5P_L2__O3_TCL)", [Online]. Available: https://dx.doi.org/10.5270/S5P-8aqg6um [Accessed: Sep. 24, 2024].
[17] "2018-04-30 ongoing Sentinel-5P TROPOMI Tropospheric Ozone Column V2 (S5P_L2__O3_TCL)", [Online]. Available: https://dx.doi.org/10.5270/S5P-hcp1l2m [Accessed: Sep. 24, 2024].