Hanzehogeschool Groningen
University of Applied Sciences

Data Fusion Architectures and Models

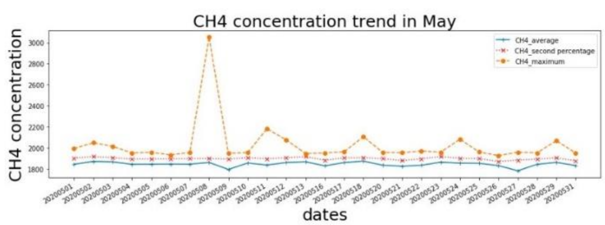
Spaceborne Tropospheric Monitoring Instrument(TROPOMI) Project portfolio

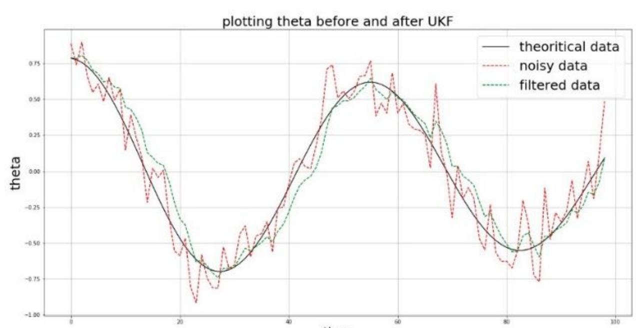
and

Unscented Kalman filter assignment

By
Ing. Walaa Khalili

Lecturer
Pr. Ronald van Elburg





June 17, 2020

Contents

First Part of the portfolio

TROPOMI project

1. TROPOMI Project

1.1 Introduction

Spaceborne Tropospheric Monitoring Instrument(TROPOMI) is a Dutch-made satellite instrument that explore air quality, was launched on October 13, 2017. TROPOMI has four detectors combined together that can detect wavelengths in the ultraviolet UV, ultraviolet-visueel UVIS , shortwave infra-red(SWIR) and near infrared (NIR).[1]

Images of the Earth are produced using these equipment every second, with a resolution of seven by seven kilometers. [1] As shown in figure(1).

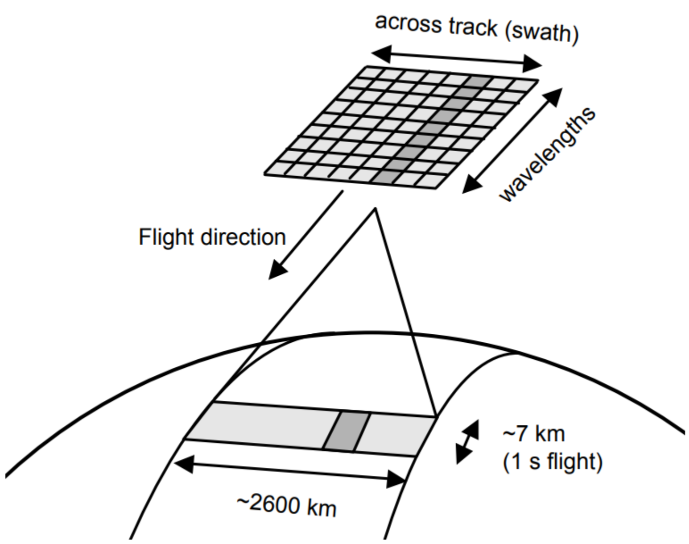


Figure 1:TROPOMI measurement principle[2]

Our main task is to combine multi- sensor data; to define CH4 leakage locations. Data we have downloaded is (L2), has been combined from nine wavelengths ,but to detect CH4 they have processed SWIR(short wave IR) wavelengths. [3] (L1) to (Lb1) data have been processed and the images resolution have been enhanced, in an understandable representational format for CH4 which is parts per billion(ppb).

1.2 Steps done in the jupyter notebook “TROPOMI project-CH4 average”

What I have done in the jupyter notebook(TROPOMI project-CH4 average) is as follows:

1. I downloaded data that cover whole china for one month (May) manually(in section 1.5. Interface design an explanation is given why it is manually downloaded), and share it with my group using a shared file.

2. Three orbits covers whole china. Thus each day has three files.

3. For each day I aligned data based on location(spatial-alignment). A 2D array was initiated, rows are days and columns are orbits.

4. I defined a region of investigation based on center latitude and longitude of china which can be found using google search engine , I tried to cover a maximum region china by adding a margin of 20 km to the maximum and subtracting it from the minimum latitude and longitude.

5. I read the data using netCDF library. [3a]

6. Each orbit has 3000 scanlines[appendix], based on longitude and latitude of the scan area of earth. I flattened the obtained longitude and latitude from each file and searched for the defined region of china. Furthermore, locating the region of interest within the scanlines, then CH4 is extracted for this specific region, and store it in an 2D array so I can then extract my features.

7. I extracted feature from features by taking the average, percentile, I found that the most important is maximum feature, where it indicates which day is CH4 leakage is the highest and action based on that should be taken. As it is illustrated in (Figure2).

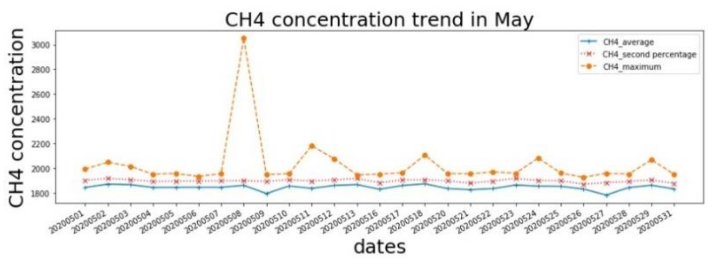


Figure 2:CH4 concentration trend in May

8. The ability to narrow down the area of interest, by knowing each longitude and latitude of each city, this area of interest should not be less than the TROPOMI resolution which is 7X7 km

9. In another file (tropomi project.ipynb),I plotted an image by using basemap library, just for visualizing files downloaded from TROPOMI websites.[3a]

1.3 . Error propagation analyses for the CH4

"Uncertainties for the CH4 due to measurement noise. CH4 data set exhibits a random error of 14.0 ppb (0.8 %) and a systematic error of 4.3 ppb (0.2 %)". [4]

Error propagation analyses is further investigated in section 2 "Unscented Kalman filter assignment".

1.4. System requirement

To analyze the requirement of a system , the system should consist of hardware, software and team, and their soft skills as well. Moreover for a successful and clear system requirements a defined customer requirements are needed. [5]

1.4.1 Customer requirements

My team and me we selected the Chinese professor requirement, to be our goal, which focus to detect CH4 leakage and locate that incident, in an easy and fast way in real time, this is called (the customer need), and based on advice he wanted to investigate the ability to use TROPOMI, to achieve his needs. A great importance for TROPOMI is that "it detect and quantify of an accidental emission from a satellite during routine operations (i.e., without pointing the satellite to a previously known target area), which demonstrates the unique value of satellite remote sensing, and the TROPOMI instrument in particular”.[6]

TROPOMI data is an open source and free to download, in addition the availability of near real time data of CH4.[5a] In the introduction I mentioned what is TROPOMI, and what type of sensors TROPOMI consists of.

1.4.2 Translating a set of requirement into sensor fusion architectures

1.4.2.1 JDL sensor architecture

JDL model has six levels .Implementing JDL model to Interpret these requirements , as follows[7] :

1. “Sub-object refinement(L0)and Object refinement(L1) “: is equivalent to L1 up to L2 TROPOMI data

2. “Situation refinement (assessment) “: processing the data obtained from (L2) , such plotting the image and notice the inside attribute for each image

3. “Impact assessment” : locate leakage and concentration of CH4

4. “Process Resource management”: adaptive data acquisition and processing to support sensing objectives such getting data from databases (Copernicus Open Access Hub or earth engine). [8]

5. Cognitive refinement : Figure(2) shows the recent User interface , previously user interface was defined as visualizing and monitoring without any interaction in the fusion process, that give the user the responsibilities of an observer.

Recently JDL give significant role for the user in the fusion process in such a way that interpret his needs, further discussion can be found in (section user interface design).Figure (3) show the levels of the JDL model.[9]

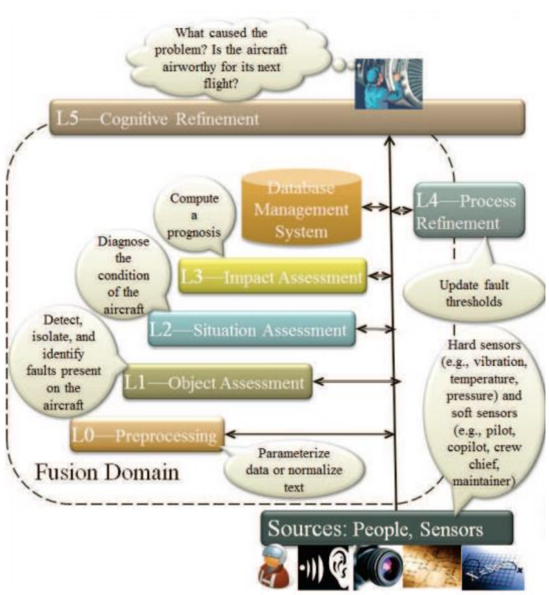


Figure 3:JDL model representations of all the level, specially cognitive refinement level, and present the ability for a user interaction

1.4.2.2 Durrant-Whyte[7]

Durrant-Whyte distinguishes by strategy, implement Durrant-Whyte on this project requirements as follows:

1. Complementary: Four detectors take readings of gasses in air.

2. Competitive : the priority is for the gasses which has a significant impact on the air pollution and weather monitoring.

3. Cooperative: the four sensors can detect different wavelength of the spectrometer of different gasses, each sensor detect specific wavelength and by combining their readings a complementary view can be given for air quality monitoring.

1.4.2.3 Dasarathy model[7]

Dasarathy distinguishes six types of fusion, implement Dasarathy model on this project requirements as follows:

Data In - Data Out: L1(raw data) and Lb1 (processed data)

Data In - Features Out: DaI-FeO: this fit in the Lb1 to L2 produces Features

Features In – Features Out: L2 data contains extracted features from raw data ,from these feature we try to extract features by using mean, percentile, etc.

Features In - Decision Out: DaI-DeO: L2 data contains features extracted from raw data, in this project there are no decision come out from the project. But the user take the decision based on what he receives from the analyzed data..

In my point of view, JDL model is the best model describes the TROPOMI project in an easy ,understandable way. Table 1, demonstrate the nine gasses and specify which detectors used to detect these gasses. [10]

Table 1: demonstrate the nine gasses and specify which detectors used to detect these gasses

Product	Spectrometer	Application
Ozone	UV, UVIS	Ozone layer monitoring, UV-index forecast, Climate monitoring
NO_2	UVIS	Air quality forecast and monitoring
CO	SWIR	Air quality forecast and monitoring
CH_2O	UVIS	Air quality forecast and monitoring
CH_4	SWIR	Climate monitoring
SO_2	UVIS	Air quality forecast and monitoring, Climate monitoring, Volcanic plume detection
Aerosol	UVIS, NIR	Air quality forecast and monitoring, Climate monitoring, Volcanic plume detection
Clouds	UVIS, NIR	Climate monitoring
UV-Index	UVIS	UV index forecast

1.5. Interface design

I selected two papers from seven papers to read about the user interface design. They are as follows :

1. Rethinking Level 5: Distributed Cognition and Information Fusion[11]

2. DFIG Level 5 (User Refinement) issues supporting Situational Assessment Reasoning.[12]

My understanding of it is as follows, Maria Nilsson et al [11] ,summaries different studies one of these study is , Blasch and Plano[12] , where they refer to that humans can have a good role in the information fusion process. Maria Nilsson et al, claims that ,In order to understand actual information fusion process, science should go further than Human computer interface(HCI) and take cognitive science into consideration(e.g. distributed cognition), it can be proved that HCI interface limited the human from using their cognitive skills for detecting patterns and their analytical skills. Moreover, they mention that Dasarathy indicates that there are a number of studies which indicate that the more information is not always better.

It has been shown that the user is no longer act as an observer of information fusion system; rather, the humans are included in the information fusion process. The user can act on this process as a feedback provider, can give responses to the system, and the system process these responses, it was discussed that "without human inputs, the DFS [Data Fusion System] refinements, based only on the data received, may be time consuming and useless" . that means more interaction between the human and the fusion system and therefore, enable rethink the

In addition, Blasch and Plano[12] emphasize on the most important step is defining the user requirement, since a good user interface can interpret these requirement as monitoring, visualization or something matches the user perceive , the user interface should be concise but detailed, showing little information could hide important information, more information is sometimes cause confusion.

The user can be interactive with the fusion systems, by trying to predict ,react to certain alarms, search for similar incident in past data, but in a fully automated system, user is marginalized and that is incorrect, where the user has no role. A good user interface that give the user a role to take decision and actions based on the his needs, e.g such combining readings from three sensors or more, he wanted to predict a certain action after he input a certain parameter, he wanted to find similarity in some of the past data. The user seeks questions and the system translates these requests into actionable items using interface to present or to monitor.

1.5.1 System management interface

Unfortunately, we could not design an interface for streaming management , because data that are no longer online are flagged as “offline”. After a while, the requested search item will be available for download through the original website.[12a] Methane data is available only offline, that is why it is not possible to access data in real time and why I downloaded it manually and stored them locally.[12b] In figure (4), In TROPOMI project actually we have two users, the first user is the ICT developer, the second one is the customer, I included both of them to the fusion process, where the first user can interact with TROPOMI website, by selecting different parameters, and based on the user expertise he make his analysis and outputs , for the second user he can only take an action based on the processed data got from the first user.

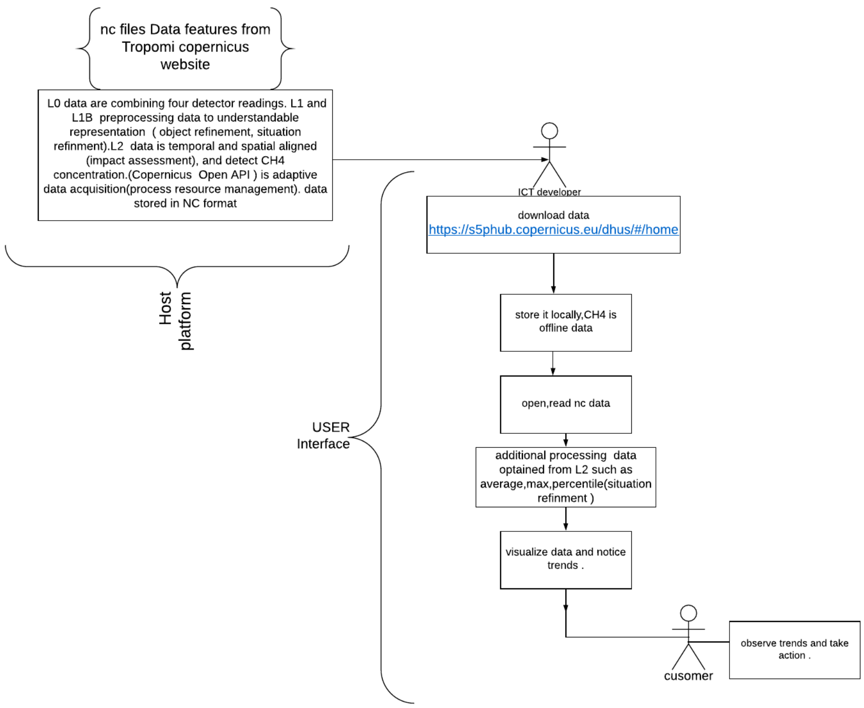


Figure 4: The user role in TROPOMI project, the first user is the ICT developer can select files for specific region or specific days. extract extra features from nc files ,the second user which is the customer can only observe and can take action based on his expertise.

1.6. SOSA standards

What I found on SOSA standards for data Interfaces management and related to (TROPOMI Copernicus open Hub)[12a] , which is the host platform, the following:

“Sensor interface enable a system to read out information from the input signal generated by complex sensors, providing a suitable output signal that is easy for a user to display or process. A host platform interface module is responsible for all communication with the host platform. Its primary function is data translation to/from formats and messages required by the host platform . The SOSA host platform may contain external (to the SOSA sensor) processing capability that could use the data products of the SOSA sensor, and external communications. It will provide the SOSA sensor with its tasking. It is assumed that the host platform provides any User Interface (UI) that may be required; while the SOSA system may provide data and information to be used by a UI, the UI is outside the scope of the SOSA system." [13]

1.7. Virtualization

Virtualization means "the process of creating a software based or virtual version of something whether that can be compute, server, storage networking, servers or applications." [14]

The most beneficial of virtualization, is the ability to failure management and also safety, if for any reason a server fails, another server running the same application takes over, thereby minimizing the interruption in service.[15]

This term can't be implemented in TROPOMI project as a whole, maybe could be implemented to the server where data are stored.

To make a virtual representation of a whole process it is most suitable to the drying process project.

Second Part of the portfolio

Unscented Kalman filter(UKF) assignment

2. Unscented Kalman filter(UKF) assignment

In this jupyter notebook, I only changed the section related to g-h Kalman filter to unscented Kalman filter. which I found an example in this website. [1,2]

The Kalman Filter, the Unscented Kalman Filter, and extended Kalman filter is an unsupervised algorithm for tracking a single target in a continuous state space, the difference is that while the Kalman Filter linearize dynamics to affine functions, This introduces errors in posteriori mean and covariance of the transformed Gaussian Random variable, the Unscented Kalman Filter is designed to operate under arbitrary dynamics by using a deterministic sampling approach without the need of calculating Jacobian matrices which they are expensive. [3]

To implement UKF I needed to define specific functions which are f(x) and h(x). f(x) is a function to return the nonlinear state transition variables (theta, omega), h(x) The update step converts the sigmas into measurement space and return theta and omega.[4,4a]

By implementing UKF library "UKF(dim_x=2, dim_z=2, dt=dt, fx=fx, hx=hx, points=points)",on the pendulum simulation. [1,2]

Definition of these library parameters [1,2] as follows:

dim_x: Number of state variables for the filter, in our case they are theta and omega

dim_z: Number of measurement inputs, if the sensor provides position in (x,y), dim_z would be 2.

dt: Time between steps in seconds.

f(x) and h(x), previously defined.

Points : Sigma points derived from this function(MerweScaledSigmaPoints) I changed the parameters of the function which are as follows:

n: dimensionality of the state(integer number), alpha defined as the distribution of the sigma point around the mean, beta Incorporates prior knowledge of the distribution of the mean ,kappa Secondary scaling parameter either equal to n or 3n. and couldn't notice much difference in the simulation, so I plotted the pendulum process before and after implementing UKF to observe any unusual behavior. The same for function(JulierSigmaPoints).

To notice the behavior of the simulation after implementing Unscented Kalman Filter, I vary parameters to get the optimal behavior of the UKF, such parameters is process noise(Q). when Q is small, then the behavior of UKF will be un-well [5,6], nothing found inform how to estimate that, so I tried different values from 0 up to 10. If it is smaller than 4 the behavior not predicted well, when increasing the value more than 4 I couldn't notice any difference in the behavior of the simulation.

The next different figures show the plot while varying the kappa parameter.

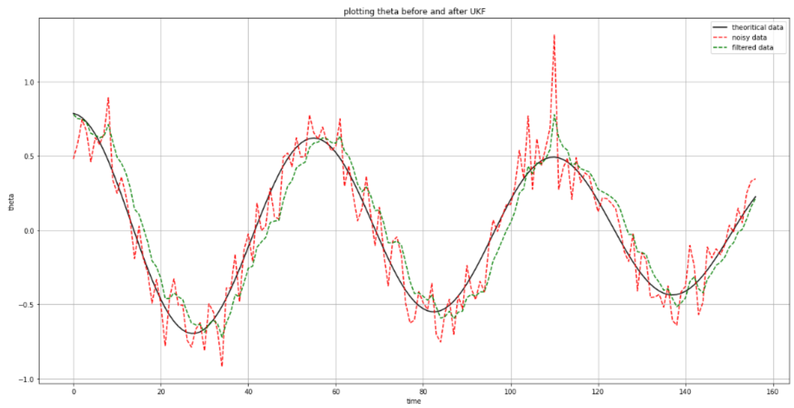


Figure 4:when kappa=0

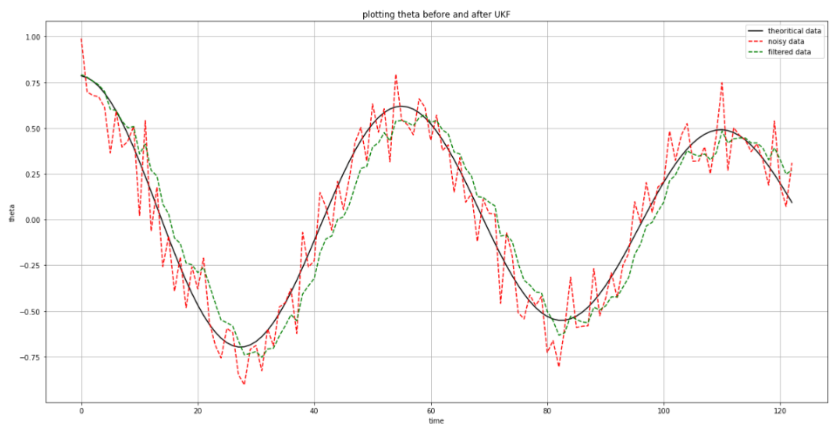


Figure 5:when kappa=1

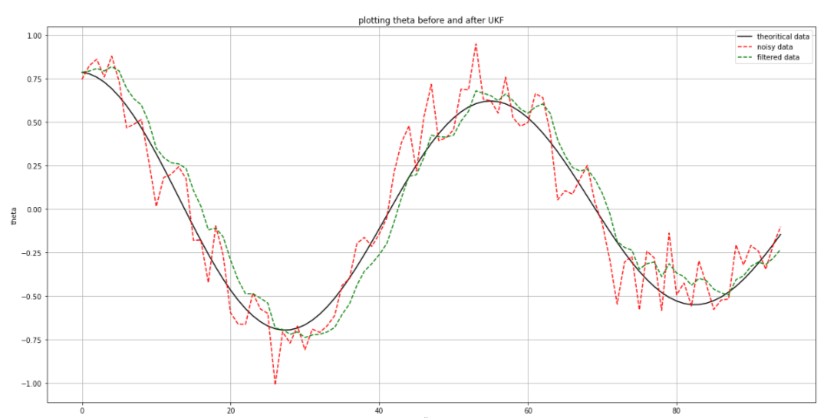


Figure 6:when kappa=4

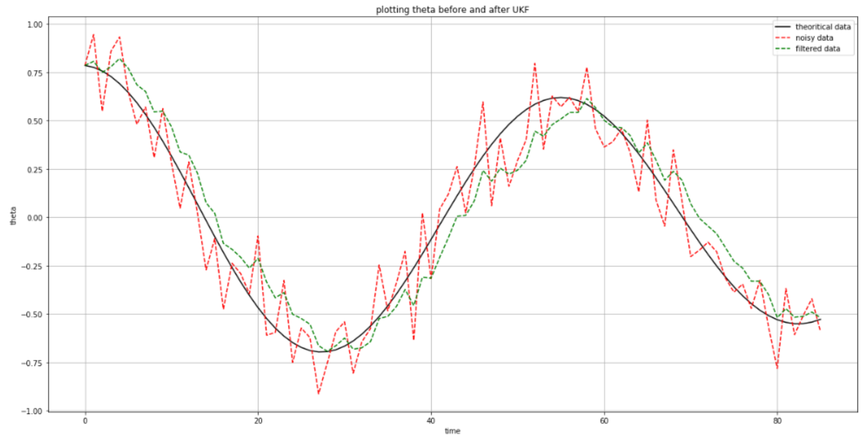


Figure 7:when kappa =5

2.1 . Error Propagation on a pendulum simulation

The sources of errors in a pendulum measurements are the following[7]:

1. Errors while measuring Time between steps in seconds for the pendulum oscillation, to reduce this by taking measurement many times.

2. Instrument errors, depend on the instrument if it is digital or analog, analog instrument introduce more errors than the digital one.

3. The precise measurement of the length of the pendulum is difficult to take by using meter sticks or rulers.

And there are many aspects that produce errors while measuring pivot angle and angular speed.

How to calculate uncertainty.[8]

a) Adding and Subtracting Uncertainties :

(3.4 ± 0.2 cm) + (2.1 ± 0.1 cm) = (3.4 + 2.1) ± (0.2 + 0.1) cm = 5.5 ± 0.3 cm

(3.4 ± 0.2 cm) − (2.1 ± 0.1 cm) = (3.4 − 2.1) ± (0.2 + 0.1) cm = 1.3 ± 0.3 cm

b) Multiplying or Dividing Uncertainties:

(3.4 cm ± 5.9%) × (1.5 cm ± 4.1%) = (3.4 × 1.5) cm2 ± (5.9 + 4.1)% = 5.1 cm2 ± 10%

(3.4 cm ± 5.9%) ÷ (1.7 cm ± 4.1 %) = (3.4 ÷ 1.7) ± (5.9 + 4.1)% = 2.0 ± 10%

There are a library to implement uncertainties(error propagation), a simple example by implementing this library as follows[9]:

```
from uncertainties import ufloat

x = ufloat(2, 0.25)

print(x)

out: 2.0 +/-0.25

square = x**2  # Transparent calculations

print(square)

out: 4.0+/-1.0 # by multiply the output of uncertainty by adding the uncertainty values

square.nominal_value

out: 4.0

print(square - x*x)

out: 0.0+/-0 # correlations taken into account
```

3. Summary

The purpose of this project was to implement data fusion techniques, by investigating the feasibility of TROPOMI to detect CH4 in daily basis, I found that “TROPOMI can detect and quantify of an accidental emission from a satellite during routine operations (i.e., without pointing the satellite to a previously known target area), which demonstrates the unique value of satellite remote sensing, and the TROPOMI instrument in particular”.[6] By specifying the region of interest it is easily to detect the CH4 concentration amounts leaked out that region.

For the best of my understanding by specifying the region of interest , and by specifying date it is easy to capture CH4 concentration amounts leaked out that area ,and when locating the maximum of CH4 concentration it is easy to extract the longitude and latitude of that specific incidents.

As it can be shown in Figure(8), the intensity of CH4 vary from low to high.

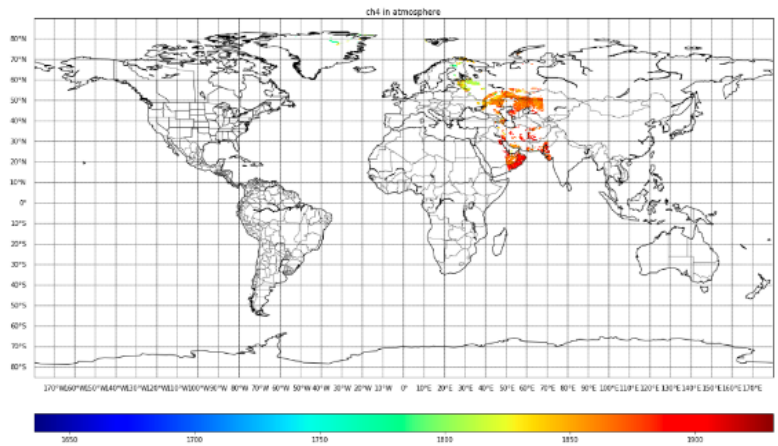


Figure 8: Plot of one nc file(one orbit) illustrates CH4 and longitude and latitude

The difficulty is to know the location of the extracted longitude and latitude since we don't have a labeled places . On the other hand using TROPOMI is not feasible for daily detection since the data released after processing it, I suppose that the analysis takes time before it becomes public. In addition TROPOMI capture CH4 column concentrations when it is close to the Earth's surface, that means the wind has a great impact on the location of CH4. As it can be seen in Figure(9)[appendix]

Figure 9: TROPOMI CH4 detecting methodology

In this analysis a trend was extracted on daily basis for one month for a big area of china, it could be also monthly, such as noticing the difference of air pollutions before Corona crisis and after. Or seasonal to notice a trend in wet months or in dry months, many analysis could be done with TROPOMI data. In a nut shell It is much feasible if TROPOMI used in research purposes and not used for detecting a location of leakages .

In my point of view, JDL model is the best model describes the TROPOMI project in an easy ,understandable way.

For unscented Kalman filter it's unsupervised technique for predicting nonlinear process in more accurate and non-expensive way compared to Kalman filter and extended Kalman filter. Error Propagation give reliability to the readings.

4. References

4.1 TROPOMI Project.

[1] Tropomi. (n.d.). Retrieved May 27, 2020, from https://www.knmi.nl/kennis-en-datacentrum/uitleg/tropomi

[2] Input/output data specification for the TROPOMI L01b data processor, https://earth.esa.int/documents/247904/3119978/Sentinel-5P-Level-01B-input-output-data-specification

[3] High resolution observations from the Sentinel-5P TROPOMI satellite instrument, https://kfolkertboersma.files.wordpress.com/2018/09/bira_press_20180619.pdf

[3a] Acgeospatial. (n.d.). Acgeospatial/Sentinel-5P. Retrieved June 16, 2020,from https://github.com/acgeospatial/Sentinel-5P/blob/master/Sentinel_5P.ipynb

[4] A scientific algorithm to simultaneously retrieve carbon monoxide and methane from TROPOMI onboard Sentinel-5 Precursor, 14 June 2019, from https://pure.mpg.de/rest/items/item_3069359/component/file_3069360/content

[5](n.d.). Retrieved May 27, 2020, from https://www.visual-paradigm.com/guide/requirements-gathering/requirement-analysis-techniques/

[5a] Sentinel-1 Data Offer. (n.d.). Retrieved June 16, 2020, from https://scihub.copernicus.eu/userguide/WebHome

[6]Pandey, S., Gautam, R., Houweling, S., Sadavarte, P., Borsdorff, T., Hasekamp, O., . . . Hees. (2019, December 26). Satellite observations reveal extreme methane leakage from a natural gas well blowout. Retrieved May 27, 2020, from https://www.pnas.org/content/116/52/26376.short

[7] H., & M. (n.d.). Data Fusion: Concepts and Ideas. Retrieved June 08, 2020, from https://link-springer-com.nlhhg.idm.oclc.org/book/10.1007/978-3-642-27222-6

[8] High resolution observations from the Sentinel-5P TROPOMI satellite instrument, 19 June 2018, https://kfolkertboersma.files.wordpress.com/2018/09/bira_press_20180619.pdf

[9] Joseph T. Bernardo, (July 6-9, 2015) ,Cognitive and Functional Frameworks for Hard/Soft Fusion for the Condition Monitoring of Aircraft, https://c4i.gmu.edu/~pcosta/F15/data/fileserver/file/472287/filename/Paper_1570104471.pdf

[10](PDF) Revisiting the JDL Model for Information Exploitation. (n.d.). Retrieved May 27, 2020, from https://www.researchgate.net/publication/260037206_Revisiting_the_JDL_Model_for_Information_Exploitation

[11] M. Nilsson and T. Ziemke, "Rethinking Level 5: Distributed Cognition and Information Fusion," 2006 9th International Conference on Information Fusion, Florence, 2006, pp. 1-8, doi: 10.1109/ICIF.2006.301559.

[12] E. Blasch and S. Plano, "DFIG Level 5 (User Refinement) issues supporting Situational Assessment Reasoning," 2005 7th International Conference on Information Fusion, Philadelphia, PA, 2005, pp. xxxv-xliii, doi: 10.1109/ICIF.2005.1591828.

[12a] (n.d.). Retrieved June 16, 2020, from https://s5phub.copernicus.eu/dhus/

[12b] Sentinel-5P OFFL CH4: Offline Methane | Earth Engine Data Catalog. (n.d.). Retrieved June 16, 2020, from https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S5P_OFFL_L3_CH4

[13] Technical Standard for SOSA™ Reference Architecture, Edition 1.0, Version 2, 2019 7th January, https://publications.opengroup.org/s180

[14] Virtualization Technology & Virtual Machine Software: What is Virtualization? (2020, June 03). Retrieved June 11, 2020, from https://www.vmware.com/solutions/virtualization.html

[15] 1 Introduction to Virtualization. (n.d.). Retrieved June 15, 2020, from https://docs.oracle.com/cd/E20065_01/doc.30/e18549/intro.htm

4.2 Unscented Kalman filter assignment references .

[1] UnscentedKalmanFilter. (n.d.). Retrieved June 12, 2020, from https://filterpy.readthedocs.io/en/latest/kalman/UnscentedKalmanFilter.html

[2] Rlabbe. (n.d.). Rlabbe/Kalman-and-Bayesian-Filters-in-Python. Retrieved June 12, 2020, from https://github.com/rlabbe/Kalman-and-Bayesian-Filters-in-Python/blob/master/10-Unscented-Kalman-Filter.ipynb

[3] M. Rampelli and D. Jena, "Advantage of Unscented Kalman Filter over Extended Kalman Filter in dynamic state estimation of power system network," Michael Faraday IET International Summit 2015, Kolkata, 2015, pp. 278-283, doi: 10.1049/cp.2015.1644.

[4] Kalman-and-Bayesian-Filters-in-Python / 10-Unscented-Kalman-Filter.ipynb, https://share.cocalc.com/share/7557a5ac1c870f1ec8f01271959b16b49df9d087/Kalman-and-Bayesian-Filters-in-Python/10-Unscented-Kalman-Filter.ipynb?viewer=share]

[4a] The Unscented Kalman Filter for Nonlinear Estimation, https://www.seas.harvard.edu/courses/cs281/papers/unscented.pdf

[5] How to set the initial value of x0, P0,R,Q in Unscented Kalman Filter and how to embedded the measurment noise in UKF? ,https://www.researchgate.net/post/How_to_set_the_initial_value_of_x0_P0_R_Q_in_Unscented_Kalman_Filter_and_how_to_embedded_the_measurment_noise_in_UKF

[6] Frankfrank 1, DaveDave 5, SneftelSneftel 30k77 gold badges5353 silver badges8686 bronze badges, & ThomasThomas 1. (1963, June 01). Explain process noise terminology in Kalman Filter. Retrieved June 12, 2020, from https://stackoverflow.com/questions/19537884/explain-process-noise-terminology-in-kalman-filter

[7] What are the sources of errors in a simple pendulum experiment and precautions to minimize them?, https://www.quora.com/What-are-the-sources-of-errors-in-a-simple-pendulum-experiment-and-precautions-to-minimize-them

[7] Johnson, L. (2019, March 02). How to Calculate Uncertainty. Retrieved June 12, 2020, from https://sciencing.com/how-to-calculate-uncertainty-13710219.html

[8] Uncertainties. (n.d.). Retrieved June 12, 2020, from https://pypi.org/project/uncertainties/

4.3 Appendix

1

2