This project has received funding from the European's Union Horizon 2020 research innovation program under grant agreement No. 769355

D4.3 Predictive Algorithms v1

Deliverable No.	D4.3	**Due Date**	30-Apr-2019
Type	Report	**Dissemination Level**	Public
Version	1.0	**Status**	Release 1
Description	Deliverable document contains the development of new prediction and forecasting algorithms in the ports where the need is greatest: to allocate energy in more efficient way and to forecast the infrastructure needs and trend of environmental pollution in the ports. For different sectors within ports there are different requirements and thus various predictive models/algorithms must be constructed to include different conditions. The proposed algorithms will be applied using available data measurements to test their performance and forecasting quality. Associated task: T4.5.		
Work Package	WP4		

Authors

Name	Partner	e-mail
Benjamín Molina	P01 UPV	benmomo@upvnet.upv.es
Ignacio Lacalle	P01 UPV	iglaub@upv.es
Carlos Enrique Palau Salvador	P01 UPV	cpalau@dcom.upv.es
Miguel Ángel Llorente	P02 PRODEVELOP	mllorente@prodevelop.es
José Antonio Clemente	P02 PRODEVELOP	jclemente@prodevelop.es
Dejan Štepec	P03 XLAB	dejan.stepec@xlab.si
Tomaž Martinčič	P03 XLAB	tomaz.martincic@xlab.si
Flavio Fuart	P03 XLAB	flavio.fuart@xlab.si
Manuel Devescovi	P04 INSIEL	manuel.devescovi@insiel.it
Charles Garnier	P05 CATIE	c.garnier@catie.fr
Erwan Simon	P05 CATIE	e.simon@catie.fr
Tamara Corsano	P09 SDAG	t.corsano@sdag.it
Andrea Chinese	P09 SDAG	a.chinese@sdag.it
Grigoris Dimitriadis	P10 ThPA SA	gdimitriadis@thpa.gr
Eirini Tserga	P10 ThPA SA	etserga@thpa.gr
Dimitrios Spyrou	P11 PPA SA	dspyrou@olp.gr
Athanasios Chaldeakis	P11 PPA SA	ahaldek@gmail.com
Stefano Bevilacqua	P12 ASPM	s.bevilacqua@monfalconeport.it
Thibault Guillon	P13 GPMB	t-guillon@bordeaux-port.fr

History

Date	Version	Change
14-Mar-2019	0.1	ToC and task assignments
24-Mar-2019	0.2	Final draft section 4; First draft section 7;
31-Mar-2019	0.3	Final draft section 3; First draft sections 2, 5-9;
7-Apr-2019	0.4	Sections 2-9 finalised; Sections 1, 10 final draft;
12-Apr-2019	0.5	Final refinements, overall structure, ready for internal review.
30-Apr-2019	1.0	Official release

Key Data

Keywords	Predictive algorithms, data sources, port operations, machine learning, software, maritime traffic, road traffic, photovoltaic energy
Lead Editor	Dejan Štepec, P03 XLAB
Internal Reviewer(s)	Gilda De Marco, P04 INSIEL Manuel Devescovi, P04 INSIEL Paolo Casoto, P04 INSIEL Eirini Tserga, P10 ThPA SA Innovation Review, P03 XLAB, Joao Pita Costa

Abstract

Artificial Intelligence (AI) is becoming one of the main factors in a successful digital transformation of the ports. Larger ports are increasingly becoming aware of the value that is present in a daily collected operational data. The ability to harness operational insights from vast amounts of data that is collected in the ports, will be one of the main advantages of future ports to make them more efficient in terms of energy efficiency, hinterland multimodal transport needs and to better anticipate harmful actions.

This deliverable presents the first version of predictive algorithms task (T4.5) in WP4 due in M12. We outline the tasks that were identified in detail, along with the methodology that will be used to tackle the proposed tasks. The most important ingredient for this task is the data that is captured in the ports or external data that can be used to complement internal data. This document presents data sources that were identified and will be used to successfully implement proposed tasks.

The tasks were identified based on the existing documentation about requirements and use cases, as well as based on the review of the state-of-the-art in the literature, existing trends and examples from the maritime industry, our AI expertise and available internal and external data.

- **Prediction of vessel call data from FAL forms and other sources:** In this task internal data about vessel calls is utilized to predict vessel calls and their durations. General statistical analysis and visualizations are also performed. Vessel call data is available in every port as is obtained from FAL forms which are legally required, thus making this task generally applicable to every port at a low cost. Internal vessel call data that was identified is presented along with the detailed descriptions of the subtasks, methodology and initial results.
- **Use of AIS data:** AIS data is widely used in maritime domain and is becoming extremely useful for data analytics tasks, especially because of its quantity. In this task we visualize and analyse the data around the ports, provide port congestion indicators out of AIS data and to some extent ETA prediction for the incoming ships. AIS data sources that were identified are presented, ranging from different national authorities in EU and USA, internal AIS receivers in the ports as well as open-source providers. A literature review of challenges of AIS data usage has been performed as part of this task.
- **Use of satellite imagery:** Obtaining operational insights from remote sensing imagery presents an emerging field, offering the ports increased situational awareness by giving them the ability to monitor their port from Space and compare it in a global perspective to understand their unique differentiators in the global market. Different cost-effective data sources were identified, and initial results are presented. This task will also uniquely utilize AIS data offering additional capabilities, as well as reuse of the methods applied in AIS data tasks.
- **Analysis and prediction of road traffic conditions with connection to port operations:** In this task hinterland multimodal transport requirements in the Port of Monfalcone, Port of Piraeus and Port of Thessaloniki are addressed. A common task of short-term traffic volume prediction has been identified. Predictions will be correlated with port operations to provide estimates on the impact that congestions have on them. Different data sources were identified, mainly from regional road operators or openly available data.
- **Prediction of renewable energy production:** In this task ports are provided with the ability to estimate the potential of renewable energy production for different time resolutions. Task is focused on Port of Bordeaux use case, but the developed methods are general and applicable for any port. Different open data sources are identified about the weather and measured photovoltaic power. Live, as well as historical data sources are presented, with initial results based on this data.

Statement of originality

Table of contents

List of tables

List of figures

List of acronyms

Acronym	Explanation
AIS	Automatic Identification System
ANN	Artificial Neural Network
API	Application Programming Interface
ARIMA	Auto-Regressive Integrated Moving Average
BOEM	Bureau of Ocean Energy Management
CEP	Complex Event Processing
CMS	Copernicus Maritime Surveillance
CMSAF	Climate Monitoring Satellite Application Facility
COLREG	Convention on the International Regulations for Preventing Collisions at Sea
DEBS	ACM International Conference on Distributed and Event-based Systems
DMA	Danish Maritime Authority
DOTA	Large-scale Dataset for Object DeTection in Aerial Images
DPO	Data Protection Officer
EDA	Exploratory Data Analysis
ETD	Estimated Time of Departure
EMSA	European Maritime Safety Agency
EO	Earth Observation
ERT	Extremely Boosting Decision Trees
ESA	European Space Agency
ETA	Estimated Time of Arrival
FAL	Convention on Facilitation of Maritime Traffic
FVG	Friuli Venezia Giulia
GBDT	Gradient Boosting Decision Trees
GDPR	General Data Protection Regulation
GPS	Global Positioning System
GTS	Global Telecommunication System
GUI	Graphical User Interface
HIT	Hellenic Institute of Transport
HRSC	High Resolution Ship Collections
IDE	Integrated Development Environment
IMO	International Maritime Organization
IT	Information Technology
JRC	Joint Research Center
LAI	Leaf Area Index
LULC	Land Use and Land Cover
MAE	Mean Average Error
ML	Machine Learning
MMSI	Maritime Mobile Service Identity
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NIR	Near-Infrared imaging

Acronym	Explanation
NMEA	National Marine Electronics Association
NOAA	National Oceanic and Atmospheric Administration
NSRD	National Solar Radiation Database
NWP	Numerical Weather Prediction
PCI	Port Congestion Index
PEI	Port Environmental Index
PMIS	Port Management Information System
PVGIS	PhotoVoltaic Geographical Information System
RCNN	Region-based Convolutional Neural Network
RFID	Radio-Frequency Identification
RGB	Read Green Blue
RMSE	Root Mean Square Error
ROI	Return on Investment
SAR	Satellites with Aperture Radar imaging
SEC	Safe and Efficient Cargo
SOLAS	International Convention for the Safety of Life at Sea
SUMO	Simulation of Urban MObility
SVM	Support Vector Machine
SYNOP	Surface Synoptic Observations
TCP	Transmission Control Protocol
VEINS	Vehicles in Network Simulation
VHF	Very High Frequency
VHR	Very High Resolution
WMO	World Meteorological Organization

1. About this document

This document aims to present the initial work that has been done regarding predictive algorithms by defining tasks that were identified, along with data sources and methods that will be used to implement the proposed tasks. The main idea was to find tasks that are applicable especially to small and medium sized ports and can be addressed with the data that is captured internally by ports, or accessible as open data to provide cost efficient solutions, that are at the same time scalable to be used also in larger ports. One of the main drivers behind the definitions of the tasks and their scope, was the data that can be obtained by ports as stakeholders and can be used in the project. The proposed tasks were defined, along the requirements, as well as, state-of-the-art in literature and industry trends.

1.1. Deliverable context

Keywords	Lead Editor
Objectives	**Objective 5 Develop predictive algorithms** This deliverable provides a selection of predictive algorithms that will be developed. In addition, initial results are presented for each section. Main results of objective 5 are being achieved through development of predictive algorithms that have the potential of significantly increasing the efficiency in one or more of the following areas: **1. Energy demand:** achieved through prediction of renewable energy prediction (section 7). **2. Hinterland multimodal transport needs:** achieved as explained in section 3 (predicting vessel calls duration, type and quantity of cargo), section 4 (predicting ETA and port events), section 5 (ship traffic analytics through remote observation) and section 6 (traffic peaks, congestion). **3. Anticipation of environmentally harmful actions:** section 4 proposes implementation of a Port Congestion Index that will be used to anticipate peaks in vessel emissions in ports, ETA (Estimated Time of Arrival) estimation could be used to trigger slow steaming and schedule port arrivals. Detection of additional events in the port area can help to prevent accidents involving harmful substances. This can be supported using remote sensing solutions proposed in section 5, which will help to increase situational awareness of the ports.
Exploitable results	The main exploitable result arising from this deliverable are the predictive algorithms and, to some extent, statistical analysis.
Work plan	This deliverable is the result of work performed from M7 to M12 in Task 4.5 - Predictive algorithms. Content of this deliverable will be used to develop further predictive algorithms in T4.5. Results (software) will also be used, as input to all modelling tasks of WP4 and as codebase to delivering Operational Tools in WP6.
Milestones	This is the first deliverable for verification of MS5 Predictive models/algorithms established (M24). The second deliverable, D4.5 will be released in M24.
Deliverables	This deliverable is an elaboration of the initial plan for the development of predictive models drafted in D4.1. Scenarios and use cases defined in D3.3 and D3.4 have been used to identify relevant predictive-related tasks. In addition, the draft version of D3.2 has been used, to relate predictive algorithms to specific user requirements. This deliverable is the first in the series of two, with the final version D4.4 Predictive Algorithms v2 scheduled for M24.
Risks	**WT5#6 Technical activities are not completed on time, are not aligned with the main objective, are not accurate or present a lack of consistency.** This deliverable shows that technical activities related to T4.5 have been executed in a timely fashion, as well as provides a plan ("future work") to execute successfully the task. **WT5#9 Some processes cannot be modelled as they depend on too many factors or they are overmuch random.** Initial work shows promising results, thus minimising the risk of unfeasible models.

1.2. The rationale behind the structure

This report describes the work performed in T4.5 of the PIXEL project and plan for future work. Except the introduction and conclusion, each section describes a specific subtask related to PIXEL data analytics.

Those sections are organised as follows:

1. **Introduction:** Problem statement, state-of-the-art overview, relation to PIXEL requirements.
2. **Data sources**: A section describing available data sources.
3. **Predictions and analytics for PIXEL**: Description of work planned for PIXEL.
4. **Initial work:** Reporting work and initial results at M12 of the project.
5. **Plan and future work:** Plan until the end of the project and, in some cases, an outline of work to be performed beyond PIXEL.

Appendix 1 lists data sources that will be used for data analytics in PIXEL. For each data source, a list of fields has been provided: Dataset name, Data Source, Description, Usage in PIXEL, Algorithms, Sharing of results, License and terms of use, Comments, DPO (Data Protection Officer) assessment.

1.3. Version-specific notes

This is the first in a series of two deliverables. This report provides an overview of work performed in the first six months of the task execution, as well as an outline of pending actions. The final version of this deliverable, due in M24 of the project, will provide a full report about applied analytics, and achieved results.

2. Introduction to PIXEL data analytics

The maritime sector and port operations are undergoing a process of digital transformation. It usually starts by digitalisation of document exchange for compliance purposes, operations management and monitoring. While this helps speeding up day-to-day operations, it also creates data repositories rarely used afterwards. The opportunity hidden in these massive data repositories, has triggered the emergence of the data economy[1], a global digital ecosystem where data is used to create value from accumulated information. One of the key ingredients of the data economy is the application of data analytics and artificial intelligence methods to extract and interpret the information.

The overall goal of PIXEL, according to the GA, *"is to close the gap between small and large ports by providing an easy-to-use open source smart platform for operational data interchange."* Large ports have the capacity to invest resources in data science teams that work on the development of smart platforms to assist port operations[2]. For example, the port of Rotterdam has been investing in development of a "Self-learning computer that predicts vessel arrival"[3].

In line with these global trends Task 4.5 Predictive algorithms *"aims at identification of predictive algorithms in ports that will support achievement of PIXEL objectives and will relate to all tasks in WP4, where is need for predictive algorithms. We will select the predictive tasks which will contribute most in the optimization of port-operation processes and resources."* The task description further lists possible candidates for implementation: *"forecasting of infrastructure needs (ship arrivals, congestions, rail transport demand, flows of containers and freights for hauliers, scheduling of transportation), the predicted energy demand (peaks, energy distribution) and the forecasting directions (trend) of environmental pollution"*.

According to the above definition, as a preliminary step, predictive tasks to support solving real-life scenarios in PIXEL must be identified, prioritized and selected for implementation. The task definition only provides a list of tentative candidates/examples.

PIXEL stakeholders have expressed their need for implementation of predictive algorithms, as well as address digital transformation through PIXEL Requirements analysis[4] and Use cases and scenarios manual[5]. In addition to the explicitly expressed stakeholder requirements, there is a clear need stated in the PIXEL Models[6] report, for forecasting information to be used as input for modelling (thus achieving "*connection with tasks in WP4"*).

All these documents point to a crucial need to forecast inbound and outbound traffic flows. Traffic flows are related to the maritime traffic, which includes ETA, cargo type and quantity, vessel call duration, seasonality and other related data. These can be estimated from several standard data sources, such as FAL forms, AIS data and satellite imagery. FAL form data is mostly owned and processed by ports, satellite imagery is an external data source, while AIS is in between. Some ports collect AIS from their own receiving network, while AIS data providers, provide commercial access to AIS data. On the other hand, road traffic, parking slots and congestion predictions, may contribute significantly to optimisation of port operations, as well as lowering the environmental impact especially on port-city surroundings, like those in Monfalcone, Piraeus and Thessaloniki.

For this reason, four chapters of this deliverable are concerned with traffic flows prediction.

The last chapter is addressing the issue of the prediction of renewable energy creation. Increased share of renewable electrical energy has many advantages but comes at cost, as an unstable source that must be complemented by traditional sources. To decrease this dependency, some sort of energy forecasting is needed. The forecast can help budgeting of other power sources, but most importantly, it could assist in optimising port

[1] https://en.wikipedia.org/wiki/Data_economy
[2] https://www.escolaeuropea.eu/odiseo/issue-29-winter-2018/artificial-intelligence-ports-are-beginning-to-take-up-positions/
[3] https://www.portofrotterdam.com/en/news-and-press-releases/self-learning-computer-predicts-vessel-arrival
[4] D3.2– PIXEL Requirements Analysis [M12] – This deliverable includes the analysis of the requirements, transform them into pilots, as well as the dependencies and interrelationships among them.
[5] D3.3 and D3.4 – Use cases and scenarios manual [M4, M9] – Definition of use cases and different scenarios to deploy the PIXEL proposal.
[6] D4.1 – PIXEL Models [M9] – This report provides an overall description of the methodology for the environmental management of the port and of the selected models, algorithms and numerical methods.

operations in such a way that the most energy-consuming tasks are aligned, as much as possible, with periods when energy from alternative sources abounds.

Data generated through the application of predictive algorithms in those three areas (maritime traffic predictions, road traffic prediction, energy production prediction) will be used to achieve the overall goal of PIXEL *"to use this data to reduce environmental impact in Port Cities and surrounding areas by improving the knowledge and control of the port operations, optimizing processes and improving management."*

We efficiently address the overall goal of PIXEL, **closing the gap between small and large ports**, only if we consider the cost of development and execution of predictive algorithms. The main components that affect cost are (1) cost of data access, (2) cost of IT (Information Technology) resources (storage, computational power), (3) cost of software development and (4) integration with existing ICT systems in ports. Small ports cannot afford, for example, paying access to global AIS data sources or access to commercial satellite images, nor can they have access to huge computational resources needed, for some of the analytical methods.

We propose a cost-effective approach, that considers data available in ports (FAL forms, vessel calls, limited AIS data, gate access data, traffic data), open or semi-open data (AISHub, road traffic) and emerging technologies with open data access (like ESA Sentinel satellite imagery). The aim is to offer a set of extremely cost-effective approaches that maximise benefits for small and medium ports. As we base our approach on standard and widely available data sources, they are scalable and applicable to ports of any size.

While section *"1.4.1 Predictive algorithms"* of the Grant Agreement focuses on State of the Art of energy efficiency, it is clear that the main ambition of that section is to "*adapt predictive algorithms to different port scenarios"* and test and validate them *"in real situations"* by "*applying novel forecasting methods to forecast the infrastructure needs, predict the energy flows, optimize the environmental effects in the port areas".* As the definition of T4.5 states as well, the ambition is to go beyond energy-related use cases.

The second part of T4.5 definition describes, in general terms, techniques used for the implementation of predictive algorithms: *"The input for the predictive models will be operational data produced by devices, sensors and will be used to develop and test forecasting algorithms. There could be unneeded, irrelevant and redundant attributes, which may decrease the accuracy of predictive model. The attribute/feature selection, will be performed as initial step for predictive modelling, which will select attributes that will give better outcome accuracy. After the initial step, several predictive models/algorithms will be developed and adapted to different scenarios. The classification or regression models can be used, it depend on the required output (outputs can be continuous or divided into classes). During the construction of the predictive models, the number of cases of training and testing will also be determined, where training dataset is a set of cases, used to fit the parameters (machine learning methods have parameters, that need to be tuned), and test set is therefore a set of cases used to assess the performance. The forecast frequency also must be determined whether the predictions will be for one day in advance, or one hour in advance. To sum up, there will be constructed different predictive models, each with its specific input data (available attributes), purpose (ship arrivals, energy demand), the length of training/testing data, forecasting frequency and the selected machine learning method (SVM (Support Vector Machine), ANN (Artificial Neural Network), etc.)."* These techniques are further elaborated, to a certain degree of detail, in sections 3 through 7 of this document. Section 3 is the most detailed in this sense, as it is also the most advanced at time of writing this report. The next version of this deliverable, will report these models in full detail, including code snippets used.

The main ingredient for the implementation of predictive algorithms is the data that needs to be collected. Authors of this document went to great lengths into identification and analysis of existing data sources, those owned and managed by ports, but also those available publicly or commercially outside of the Port Community system itself. In addition to preliminarily exploratory data analysis, initial legal (license and usage) and privacy-related assessment of the data sources has been performed. As already stated in previous reports, data (non)availability is the main risk for the successful selection "*of the predictive tasks which will contribute most in the optimization of port-operation processes and resources*" and successful execution of this task.

2.1. Relation with PIXEL objectives and use cases

All work described in D4.3 and performed as part of T4.5 aims at fulfilment of **objective 5: "*Develop predictive algorithms*"**.

The development of predictive algorithms for vessel calls predictions (FAL forms, AIS data), maritime traffic predictions and ETA (AIS, Satellite) and event detection in ports (AIS data) has the potential to significantly increase the efficiency in two areas: *hinterland multimodal transport* by predicting cargo flows through the port and *anticipation of environmentally harmful actions*. Environmentally harmful actions are predicted either indirectly through calculation of PCI or directly, by *predicting* potentially harmful events like collisions or anchor dragging, or by providing the ports with increased situational awareness using predictive algorithms developed on satellite imagery. The development of predictive algorithms for road traffic has the potential of significantly increasing the efficiency in two of the three areas: *hinterland multimodal transport* (related with T4.3) and *anticipation of environmentally harmful actions* (related with T4.1 and T4.4). Furthermore, this prediction will serve as an input for achieving objective 7 in "*mitigating possible environmental and health effects of port activities*", as the means of verification of its success is to empirically test and verify them in real scenarios. The development of the methods for prediction of renewable energy production, has the potential to increase the use of renewable energy and thus contribute to more sustainable development, by predicting the *energy demand* (related with T4.2) that can be supplemented with renewable sources.

The following **use cases** have directly expressed their needs in the areas of these types of predictions through D3.4 (Use cases and scenarios).

2.1.1.Use case of GPMB

GPMB use case is mostly connected to energy management. One of the challenges of GPMB is to become a greener more cost-attractive port. To reach this, GPMB aims to leverage on the potential of photovoltaic electricity and on self-consumption. This is directly connected with the task of renewable energy production prediction.

As expressed in D3.4 (Use cases and scenarios manual v2): "*GPMB is convinced that the electricity price will be an advantage to attract new activities in the port area. Instead of depending on the market price of electricity, GPMB aims to leverage on the potential of photovoltaic electricity and on self-consumption. But the investments are high, the return on investment (ROI) is difficult to achieve and the capability of the GPMB electric network must be assessed*." Moreover, according to D3.4, in GPMB a first approximation of 30 000 m² of rooftops could be used to install solar panels.

Some of the proposed tasks are also connected with objectives (2) – advanced port statistics analysis and (3b) - estimating port traffic evolution.[7] This can be directly connected with tasks regarding vessel call data where GPMB has provided us not only with historical data but also with API to access live information. Additionally, GPMB has also access to AIS receiver which data will be similarly provided through API and utilized as presented in section about AIS data. Similarly, satellite imagery can improve situational awareness and provide operational insights about port operations and development.

As expressed in D3.4 (Use cases and scenarios manual v2): "*To raise awareness and to provide decision tools about port operations and port development thanks to the integration of PIXEL components (advanced port statistics analysis) to PCS (Port Community Systems) and to "big data calculations"."*

2.1.2.Use case of Port of Monfalcone

Port of Monfalcone use case relates to monitoring regional road network to predict short-term traffic volume and to detect and forecast traffic peaks.

As expressed in D3.4 (Use cases and scenarios manual v2): *"A PIXEL predictive model will be able to detect and forecast peak traffic that could create congestion in the port and urban areas."*

[7] As presented in section 4.3.1 of D3.4 – Use cases and scenarios manual v2

2.1.3.Use case of Port of Piraeus

Use case of Port of Piraeus does not explicitly mention predictive algorithms, but the use case is concerned with measuring the pollution that is caused by traffic towards the port. One of the objectives to be achieved, is to improve the access to the port which can be leveraged by better understanding of traffic volumes around the port area and is addressed in our traffic use case regarding Port of Piraeus.

As expressed in D3.4 (Use cases and scenarios manual v2): *"The Port of Piraeus is confronted with accessibility and connection problems, both between the port area and the greater Athens and Piraeus cities. Moreover, the development of economic activity, the growth of tourism and the rise in movements of goods and passengers puts a high level of pressure on both the coastal and urban area and on the main transport corridors. PPA will benefit from PIXEL to improve the access to the seaport so that it can ensure sustainable economic growth in the port city of Piraeus, leveraging enabled communications at data level proportioned by PIXEL."*

2.1.4.Use case of Port of Thessaloniki

Port of Thessaloniki use case relates to our traffic use case regarding ThPA by predicting traffic volume around the port area and roads leading to the port. As expressed in their objectives, their main objective (1) is to optimize the traffic between the city and the port area.

As expressed in D3.4 (Use cases and scenarios manual v2):

"To optimize the traffic between the city and the port area and thus alleviate congestion or bottlenecks, caused by its operations."

"ThPA SA will use the outcomes of the PIXEL project to boost its knowledge of ports operations and act proactively, in order to minimize all impacts (congestion/bottlenecks, air quality)."

2.2. Relation to requirements

Tasks related to predicting vessel calls are proposed based on information provided in requirements[8] **Monitor expected port calls [11]**, **Estimate air pollution impact of handling cargo [56]**, **Estimate noise pollution impact of handling cargo [57]** and **Estimate air pollution impact of bulk cargo operations [58]**, as well as or identified as useful for other modelling tasks in WP4.

The road traffic predictions have a wide coverage among requirements for three use cases: ASPM/SDAG provided 5 requirements, ThPA provided 6 requirements and PPA provided one requirement.

Some requirements dealing with forecasting photovoltaic energy production have already been expressed by GPMB: **Modelling and analysis of energy consumption during ship handling procedures [15]**, **Support weather sensor/service [17]** and **Optimization of photovoltaic energy production and consumption [19]**.

Table 1 provides a full list of requirements directly or indirectly related to the implementation of predictive algorithms.

[8] All requirements are numbered according to identifiers provided in D3.2– PIXEL Requirements Analysis.

Table 1. Relation to PIXEL requirements

Requirement	Addressed in sections
Common functional requirements	
Interaction with models [41]	These requirements are initially addressed in all subtasks of task T4.5. All research results of T4.5 will follow conventions defined in WP6 to allow fulfilment of those requirements in WP6.
Catalogue of models [62]	
Operational Interface [71]	
Analyse historical data [81]	
Port of Bordeaux – Energy Management Use Case	
Monitor expected port calls [11]	3. Predicting vessel call data 4. Use of AIS data
Modelling and analysis of energy consumption during ship handling procedures [15]	3. Predicting vessel call data 4. Use of AIS data 7. Prediction of renewable energy production
Support weather sensor/service [17]	7. Prediction of renewable energy production
Optimization of photovoltaic energy production and consumption [19]	7. Prediction of renewable energy production
Expose data to VIGIEsip system [82]	3. Predicting vessel call data 4. Use of AIS data
Port of Monfalcone – SDAG – Intermodal Transport Use Case	
Integration with the SILI Information System [23]	6. Analysis and prediction of road traffic conditions
Integration with ASPM video monitoring system [25]	6. Analysis and prediction of road traffic conditions
Traffic peak and congestion monitoring at the port facility [26]	6. Analysis and prediction of road traffic conditions
Port congestion forecasting [34]	3. Predicting vessel call data 4. Use of AIS data 5. Use of satellite imagery 6. Analysis and prediction of road traffic conditions
Port - SDAG highway congestion forecasting [37]	6. Analysis and prediction of road traffic conditions
Port of Thessaloniki – Port City Integration Use Case	
Support real-time gate surveillance sensors [52]	6. Analysis and prediction of road traffic conditions
Support traffic data provided by third party [55]	6. Analysis and prediction of road traffic conditions
Estimate air pollution impact of handling cargo [56]	3. Predicting vessel call data 4. Use of AIS data 6. Analysis and prediction of road traffic conditions
Estimate noise pollution impact of handling cargo [57]	3. Predicting vessel call data 4. Use of AIS data 6. Analysis and prediction of road traffic conditions
Estimate air pollution impact of bulk cargo operations [58]	3. Predicting vessel call data 4. Use of AIS data 6. Analysis and prediction of road traffic conditions
Visualize the traffic status [106]	6. Analysis and prediction of road traffic conditions
Port of Piraeus – Port City Integration Use Case	
Port - City road congestion forecasting [91]	6. Analysis and prediction of road traffic conditions

2.3. Data sources

Identifying data sources to be used for prediction algorithms, is a crucial task for a successful fulfilment of needs and requirements put forward by ports, involved in PIXEL. Data sources are described in detail in each chapter (3 - 7) dealing with the implementation of predictive algorithms. Furthermore, each data source has a summary table provided in Annex 1. Table 2 provides an overview of proposed data sources.

Table 2. List of data sources

Dataset name	Short description	Usage in PIXEL
GPMB vessel call data	Vessel call data for GPMB.	Analysis and prediction of vessel calls. Correlation of port operations with road traffic.
PPA Vessel call data	Vessel call data for PPA.	
ASPM Vessel call data	Vessel call data for ASPM.	
ThPA Vessel call data	Vessel call data for ThPA.	
DEBS (ACM International Conference on Distributed and Event-based Systems) 2018 Challenge dataset	Data used for DEBS 2018 Challenge that was prepared from AIS data for the Mediterranean Sea.	Implementing long-term ETA and destination port prediction.
Danish Maritime Authority AIS data	Historical AIS data for Danish waters.	ETA prediction, AIS data analytics around port area, satellite imagery data fusion.
U.S. Coast Guard AIS data	Historical AIS data for USA coastal area.	
AISHub data	Live AIS data from AISHub	
Thessaloniki car fleet data	Car fleet equipped with GPS providing location and speed information.	Correlation of port operations and regional road network for analysis and prediction purposes.
Traffic data form the Thessaloniki stationary sensor network	Traffic data form the Thessaloniki stationary sensor network: traffic counters and speed sensors.	
Thessaloniki traffic data for vehicles entering/leaving the port	Traffic data for vehicles entering and leaving the port by gate and timestamp	
Traffic data from SILI system	Traffic information from cameras and gates connected to SILI system. Direction of the traffic, lane number, exact times, license plates, nationality and type of vehicle	
Live data report from the police (traffic department) for Piraeus data traffic.	The data provides the traffic situation of the main Athens - Piraeus road arteries.	
Traffic reports from the Region of Attica observatory on traffic congestions.	These data reports are published every 6 months. These data are concerning to the traffic on the main streets of Attica.	
Data obtained from a third-party subscription database in csv format.	These data are being prepared by the third party so that PPA can do their tests and evaluate the results. These will occur within May 2019.	

Dataset name	Short description	Usage in PIXEL
ESA Sentinel satellite imagery.	Optical and SAR satellite imagery provided by Sentinel-1 and Sentinel-2 satellites.	To develop methods for ship detection and classification which will be further used for traffic analysis and predictions of port operations and intermodal transport forecasts.
Planet Labs satellite imagery	Optical satellite imagery over California (OpenCalifornia) provided by PlanetScope (Dove) satellite constellation.	
Kaggle Airbus ship detection	Satellite imagery with annotated ships for ship detection from optical imagery.	
HRSC 2016	HRSC (High Resolution Ship Collections) 2016 satellite imagery with annotated ships for ship detection and classification from optical imagery.	
xView dataset	Satellite imagery with annotated ships (and other classes) for object detection from optical imagery.	
DOTA	DOTA (Large-scale Dataset for Object DeTection in Aerial Images) satellite imagery with annotated ships (and other classes) for object detection from optical imagery.	
Kaggle Planet Labs	Satellite imagery from Planet Labs with ship/not ship imagery and annotations.	
PVGIS Data	Web applications to browse and query GIS databases of solar radiation and other climatic parameters.	To develop methods for forecasting solar energy production based on weather conditions using past data of production.
PVoutput data set	Free online service for sharing and comparing photovoltaic solar panel output data.	
OpenWeatherMap	Current weather, daily forecast for 16 days, and 3-hourly forecast 5 days.	
OpenDataSoft data	Fields of analysis and forecasts in grid points, resulting from the atmospheric model Arome on the metropolis.	

3. Predicting vessel calls data from FAL forms and other sources

3.1. Introduction

Ports internal data can provide plenty of information to analyse and optimise operations in the port. One of the richest and most general information that all the ports store is that of vessel calls. Vessels are usually announced at least 24 hours prior to their arrival. This is not only due to port authority regulations, but also due to national maritime agencies, that require this kind of information to be submitted before arrival to their territorial waters. The forms that need to be submitted, are standardized and were agreed in FAL (Convention on Facilitation of Maritime Traffic)[9]. One of the main objectives was to standardize the procedures and declarations that need to be submitted to the public authorities, to prevent unnecessary delays in maritime traffic.

The documents that need to be provided to the public authorities have been standardized by IMO (International Maritime Organization) in a form of seven documents (FAL 1-7) which are listed below:

- IMO General Declaration
- Cargo Declaration
- Ship's Stores Declaration
- Crew's Effects Declaration
- Crew List· Passenger List
- Dangerous Goods

These documents are usually provided to the public authorities through "single window" solutions. This was recently recognized and mandated by IMO, through the adoption of mandatory requirements, for the electronic exchange of these documents[10]. Most of the ports and countries already have this in place, with ports using their PMIS (Port Management Information System).

For T4.5 task, the most important is FAL Form 1 (IMO General Declaration), which captures most of the relevant information about the ship, cargo and voyage. Below some of the most important information that is captured in FAL Form 1 is provided:

- Name of the ship
- Type of the ship
- IMO number
- Port of arrival/departure
- Flag State of ship
- Gross & Net tonnage
- Last port of call/Next port of call
- Cargo declaration
- Date and time of arrival/departure
- Last & next port of call

Ports usually have historical data available and this kind of data can be used for different predictive analytics tasks that could benefit the ports in optimising their operations and at the same time, reduce environmental impacts. Some of the tasks that can be addressed with this kind of data are presented below:

- General statistical analysis and visualization of the vessel call data: The amount of historical data that can be made available, can provide useful insights for the port authorities. Patterns, such as seasonality

[9] http://www.imo.org/en/OurWork/Facilitation/ConventionsCodesGuidelines/Pages/Default.aspx
[10] https://worldmaritimenews.com/archives/188573/imo-electronic-data-exchange-in-a-single-window/

of the cargo can be discovered, processing time of each type of cargo, influence of weekends, holidays, weather or other external factors. This can help port authorities better understand their operations, types of cargo that they are processing, quantities and time intervals. Equipped with such information, they can make more informative decisions that are data driven.

- Prediction of vessel call duration: Vessel call data usually contains the exact time of entry to the port and departure. Given historical data, one could learn to predict on port entry, what will be the duration of the vessel call, given the amount of cargo that will be processed, type of cargo, time of the year, day of the week, etc. More accurate prediction can help ports optimize their operations.
- Prediction of a vessel call: For each of the ports, there is probably a lot of traffic patterns that can be discovered out of historical data, such as regular shipping lines, seasonality of the cargo etc. This can be used to predict the amount and type of cargo that will arrive to the port in the next days or weeks. This can help ports, at planning resources more efficiently in advance.

In the next sections different data sources are described, tasks that will be addressed in PIXEL project, initial results and plan for future work.

3.2. Vessel call data sources

In this section vessel call data or FAL Form data sources are described, that were made available by PIXEL ports. Most of the ports, do not store directly the FAL Forms but selection of data from the FAL forms. The good thing that came from standardization efforts is that similar data is stored by port PCS systems, but problems usually arise with attributes that are maybe stored differently or at a different resolution (e.g. arrival/departure times).

3.2.1.Port of Bordeaux

Port of Bordeaux provided us with historical vessel call data from 2010 to 2017 (extracted from VIGIEsip database), that is presented in Figure 1. There are almost 4500 vessels calls captured during those years that have entries for both arrival and departure. Besides basic information about the ship such as ship name, there is also information about the cargo, amount of cargo and actual arrival and departure times. Type of cargo is captured in two levels, there is a high-level cargo description and a more detailed one. Amount of cargo is presented as tonnage of the cargo that was loaded and unloaded.

Année Statis	Numéro d'es	Sens	Poste à quai	Nom Navire	Libellé Type	Tonnage dét	Date de poste à quai	Libellé du type de construction	Code type fi	Libellé type
2017	0004	Entrée	436	GRANDE RIVIERE	I.METHANOI	5.993	1 Jan 17 08:45:00	PRODUITS CHIMIQUES (CHEMI(	05	LIQ.V
2017	0004	Sortie	436	GRANDE RIVIERE		0	2 Jan 17 20:45:00	PRODUITS CHIMIQUES (CHEMI(	05	LIQ.V
2017	0005	Entrée	432	LIBERTAS-H	I.CONTENEU	2.669	2 Jan 17 10:45:00	PORTE-CONTENEURS INTEGRA	09	P.C.
2017	0005	Sortie	432	LIBERTAS-H	E.CONTENE	5.403	4 Jan 17 08:15:00	PORTE-CONTENEURS INTEGRA	09	P.C.
2017	0006	Entrée	433	INCE EVRENYE	I.UREE VRA	5.000	2 Jan 17 11:05:00	VRAQUIER (BULK)	06	SOL.V
2017	0006	Sortie	433	INCE EVRENYE		0	3 Jan 17 21:50:00	VRAQUIER (BULK)	06	SOL.V
2017	0010	Entrée	436	BAYAMO	I.BUTADIENI	1.002	4 Jan 17 09:35:00	GAZ LIQUEFIES (GAZ)	04	GAZ.L
2017	0010	Sortie	436	BAYAMO		0	4 Jan 17 22:25:00	GAZ LIQUEFIES (GAZ)	04	GAZ.L
2017	0011	Entrée	449	ARKLOW VALE		0	4 Jan 17 20:30:00	CARGO DE PLUS DE 500 TJB	13	CEREAL
2017	0011	Sortie	449	ARKLOW VALE	E.MAIS VRA	4.628	5 Jan 17 22:30:00	CARGO DE PLUS DE 500 TJB	13	CEREAL
2017	0013	Entrée	436	CORAL OBELIA	I.BUTADIENI	2.100	4 Jan 17 23:55:00	GAZ LIQUEFIES (GAZ)	04	GAZ.L
2017	0013	Sortie	436	CORAL OBELIA		0	5 Jan 17 12:40:00	GAZ LIQUEFIES (GAZ)	04	GAZ.L
2017	0014	Entrée	449	ALECTO		0	5 Jan 17 22:05:00	CARGO DE PLUS DE 500 TJB	13	CEREAL
2017	0014	Sortie	449	ALECTO	E.CEREALE:	3.363	7 Jan 17 00:25:00	CARGO DE PLUS DE 500 TJB	13	CEREAL
2017	0015	Entrée	417	ELBETOR		0	5 Jan 17 22:50:00	CARGO DE PLUS DE 500 TJB	13	CEREAL
2017	0015	Sortie	417	ELBETOR	E.MAIS VRA	3.200	7 Jan 17 00:50:00	CARGO DE PLUS DE 500 TJB	13	CEREAL
2017	0017	Entrée	432	LAURA ANN	I.CONTENEU	1.961	6 Jan 17 22:48:00	PORTE-CONTENEURS INTEGRA	09	P.C.

Figure 1. Vessel call data from Port of Bordeaux

Besides historical AIS data, Port of Bordeaux has also made available some of the FAL Forms information through their VIGIEsip API (Application Programming Interface), so that live information can be gathered. With this kind of data, predictions can be made on real time basis, as well as the provided dataset can be further extended with new vessel calls.

3.2.1.Port of Monfalcone

Port of Monfalcone presents the smallest port in our consortium and provided us with vessel call data for the year 2018. This data consists of almost 700 vessel calls, with various attributes that were captured. An example of this data is provided in Figure 2. In terms of properties that are captured, it represents the most complete data. Besides accurate arrival and exit times, expected departure date is also captured. Unique to other vessel call data, this data also captures moored time, flag of the ship, draught, length, last port of call, next port of call and the shipping agent. Type and amount of cargo that was processed is also captured, though the amount of cargo is missing in some entries. Similar data for additional years will be provided.

ARR. DATE	DEP. DATE	TIME MOOREI	NR. NAVI ARRIVATE NELL'ANN	SHIP'S NAME	IMO NO.	FLAG	G.T.	DRAUGHT I	LENGHT N	CARGO DESCRIPT.	TYPE	TONN	WHARF	LAST PORT OF CALL	NEXT PORT	EXPECTED DEPARTURE (ETD)	SHIPPING AGENT
29/12/17 12.18	2/1/18 10.05	93:47:00	649	SOCAR 5	//	ITA	1789	5,00	65	CARBONE	SBARCO	5300	A2A	KOPER	KOPER	Tuesday, January 2, 2018	CATTARUZZA
29/12/17 11.10	3/1/18 16.35	125:25:00	650	RECEP KURU	9040948	TR	3229	6,00	99	VERGELLA	IMBARCO	3200	7	RAVENNA	HAIFA	Wednesday, January 3, 2018	MARLINES
3/1/18 7.17	3/1/18 22.30	15:13:00	1	CALYPSO	9370305	NLD	2545	6,00	80	CAOLINO	SBARCO	1507	6	RAVENNA	BARLETTA	Wednesday, January 3, 2018	TSL
2/1/18 7.10	3/1/18 23.40	40:30:00	2	EGE BEY	9103013	CK	5966	6,60	131	BILLETTE	SBARCO	7705	6_7	NOVOROSSIJSK	ISTANBUL	Wednesday, January 3, 2018	MARLINES
20/12/17 17.32	4/1/18 8.20	350:48:00	653	ATALANTE	9363168	MLT	14501	8,50	150	CEREALI	SBARCO	11405	De Franc	KOPER	RIJEKA	Thursday, January 4, 2018	TSL
31/12/17 12.25	4/1/18 14.30	98:05:00	654	SOCAR 6	//	ITA	1789	5,00	65	CARBONE	SBARCO	5300	A2A	KOPER	KOPER	Thursday, January 4, 2018	CATTARUZZA
3/1/18 8.55	4/1/18 16.50	31:55:00	3	FIDES	9030852	ITA	33825	6,80	178	ROTABILI	SBA_IMB		3	ORTONA	RAVENNA	Thursday, January 4, 2018	CATTARUZZA
4/1/18 7.20	4/1/18 17.30	10:10:00	4	KARL-JAKOB K	9344409	NLD	3057	5,90	90	CELLULOSA	SBARCO	3966	5	CEUTA	MILOS	Thursday, January 4, 2018	TSL
3/1/18 6.50	4/1/18 21.05	38:15:00	5	NAZMI C	9577769	PA	24210	8,00	180	BRAMME	SBARCO	23299	9	MARIUPOL	ISTANBUL	Thursday, January 4, 2018	NOGAROSPED

Figure 2. Vessel call data from Port of Monfalcone

3.2.1.Port of Piraeus

Port of Piraeus provided sample data for the last 3 months of 2018 (around 5000 vessel calls), that is presented in Figure 3. Compared to other ports, this data contains IMO numbering, which automatically identifies the ship and can also be used to gather additional data about the ship from vessel databases that are available, such as Lloyd's register[11], though such data is not available for free. Port of Piraeus captures both, arrival and departure times but the problem is that for cargo ships, only the date is provided, without the exact time. This limits the usability of such data for vessel call duration prediction and road traffic volume prediction and correlation. Information about the amount and type of cargo that was processed is also provided, which will be used for statistical analysis and different operational insights will be obtained from such data. Some of the issues, such as more accurate entry and exit times, cannot be resolved, as there is no data available (as per PPA response). Despite that the data for almost 5000 vessel calls was provided, most of the entries have missing entries, especially for type and the amount of cargo that was handled. The data for at least a couple of years was also requested to perform historical analysis and to train predictive algorithms.

ID	VESSEL NAME	VESSEL IMO	DATE	FLAG	PORT LOCATION	PORT DESCRIPTION	STATUS	CARGO VOLUMES	CARGO TONES	TYPE OF CARGO	TRANSIT VOLUMES	TRANSIT TONES	TYPE OF TRANSIT	CARGO TONES
1	KRITI SFAKIA	9187227	01-Oct-18	1	ZZ888	PIRAEUS	3	0	0		0	0		0
2	BOTAFOGO	9395329	01-Oct-18	1	CYLMS	LIMASSOL	3	0	0		0	0		0
3	EMERALD LEADER	9361811	01-Oct-18	1	GBPRU	PORTBURY	1	112	159	ΑΥΤΟΚΙΝΗΤΟ	0	0		0
3	EMERALD LEADER	9361811	01-Oct-18	1	ESVGO	VIGO	1	30	35	ΑΥΤΟΚΙΝΗΤΟ	0	0		0
3	EMERALD LEADER	9361811	01-Oct-18	1	GBPRU	PORTBURY	1	0	0		1	52	ΕΡΠΥΣΤΡΙΟΦΟΡΑ (	0
3	EMERALD LEADER	9361811	01-Oct-18	1	DEBRV	BREMERHAVEN	1	0	129	ΚΟΛΛΑ	0	0		0
3	EMERALD LEADER	9361811	01-Oct-18	1	DEBRV	BREMERHAVEN	1	0	0		4	32	MAFI ΜΕ ΦΟΡΤΙΟ	0
3	EMERALD LEADER	9361811	01-Oct-18	1	ESTAR	TARRAGONA	1	51	67	ΑΥΤΟΚΙΝΗΤΟ	0	0		0
3	EMERALD LEADER	9361811	01-Oct-18	1	DEBRV	BREMERHAVEN	1	2	4	ΜΕΤΑΧΕΙΡΙΣΜΕΝΑ Ο)	0	0		0
3	EMERALD LEADER	9361811	01-Oct-18	1	BEZEE	ZEEBRUGGE	1	1	25	TRAILER	0	0		0
3	EMERALD LEADER	9361811	01-Oct-18	1	GBPRU	PORTBURY	1	1	7	ΤΡΟΧΟΣΠΙΤΟ	0	0		0
3	EMERALD LEADER	9361811	01-Oct-18	1	GBPRU	PORTBURY	1	1	8	ΦΟΡΤΗΓΑ	0	0		0
3	EMERALD LEADER	9361811	01-Oct-18	1	DEBRV	BREMERHAVEN	1	0	0		0	38	ΚΟΛΛΑ	0
3	EMERALD LEADER	9361811	01-Oct-18	1	BEZEE	ZEEBRUGGE	1	192	219	ΑΥΤΟΚΙΝΗΤΟ	0	0		0
4	HOEGH INCHON	9088249	01-Oct-18	1	ITLIV	LIVORNO	1	2	3	ΑΥΤΟΚΙΝΗΤΟ	0	0		0
4	HOEGH INCHON	9088249	01-Oct-18	1	MATNG	TANGIER	1	1	2	ΑΥΤΟΚΙΝΗΤΟ	0	0		0
4	HOEGH INCHON	9088249	01-Oct-18	1	BEZEE	ZEEBRUGGE	1	1	22	ΕΡΠΥΣΤΡΙΟΦΟΡΑ ΟΧΙ	0	0		0
4	HOEGH INCHON	9088249	01-Oct-18	1	GBTYN	TYNE	1	0	0		144	199	ΑΥΤΟΚΙΝΗΤΟ	0
4	HOEGH INCHON	9088249	01-Oct-18	1	ESBCN	BARCELONA	1	5	9	ΑΥΤΟΚΙΝΗΤΟ	0	0		0
1	EMERALD LEADER	9361811	01-Oct-18	2	TRAUT	AUTOPORT	1	0	0		12	22	ΑΥΤΟΚΙΝΗΤΟ	0
1	EMERALD LEADER	9361811	01-Oct-18	2	TRAUT	AUTOPORT	1	0	0		25	38	ΑΥΤΟΚΙΝΗΤΟ	0
1	EMERALD LEADER	9361811	01-Oct-18	2	TRAUT	AUTOPORT	1	0	0		14	26	ΑΥΤΟΚΙΝΗΤΟ	0
2	HOEGH INCHON	9088249	01-Oct-18	2	TRDRC	DERINCE	1	0	0		24	41	ΑΥΤΟΚΙΝΗΤΟ	0
5	AEGEAN HIGHWAY	9464388	02-Oct-18	1	JPYOK	YOKOHAMA	1	0	0		4	7	ΑΥΤΟΚΙΝΗΤΟ	0
5	AEGEAN HIGHWAY	9464388	02-Oct-18	1	CNSHA	SHANGHAI	1	0	0		0	1	ΚΟΛΛΑ	0
5	AEGEAN HIGHWAY	9464388	02-Oct-18	1	JPKOB	KOBE	1	0	0		2	4	ΑΥΤΟΚΙΝΗΤΟ	0
5	AEGEAN HIGHWAY	9464388	02-Oct-18	1	SGSIN	SINGAPORE	1	0	0		3	6	ΑΥΤΟΚΙΝΗΤΟ	0
5	AEGEAN HIGHWAY	9464388	02-Oct-18	1	JPNGO	NAGOYA	1	0	0		70	122	ΑΥΤΟΚΙΝΗΤΟ	0
5	AEGEAN HIGHWAY	9464388	02-Oct-18	1	KR888	MASAN	1	0	0		390	444	ΑΥΤΟΚΙΝΗΤΟ	0

Figure 3. Vessel call data from Port of Piraeus

Port of Piraeus additionally provided the data for ferries in January 2019, but the usage is very limited as only the dates are provided, gross tonnage, vessel name and length. In task T4.5 we will focus on cargo ships only.

[11] https://ihsmarkit.com/products/maritime-ships-register.html

3.2.2. Port of Thessaloniki

Port of Thessaloniki provided us with sample data from the container terminal, presented in Figure 4 and conventional cargo terminal, presented in Figure 5. Sample data consists of 100 vessel calls. The attributes provided by the Port of Thessaloniki are similar to the ones provided by Port of Bordeaux, except that there is no information about the amount of cargo that was processed - in the case of container terminal, the number of containers that were processed, while in the case of conventional cargo terminal, information about the tonnage of the cargo that was processed.

For the conventional cargo terminal, cargo type is missing. This information, besides the amount of cargo is crucial for the statistical analysis, as well as for the prediction of vessel call duration as this not only relies on the amount of cargo, but also on the type of cargo and plenty of other factors, that will be presented in later sections.

Additional information, such as the amount of cargo (i.e. tonnage or number of the containers), type of cargo, as well as historical data for at least couple of years, was requested from Port of Thessaloniki and will be provided.

VO_CODE	VESSEL_NAME	ORIGIN_PORT_ID	DESTIN_PORT_ID	ARRIVAL_DATE	DEPARTURE_DATE	ARRIVAL_DATE_SCHED	DEPARTURE_DATE_SCHED
MXPRIDE1908	MAX PRIDE	GRSKG	CYLMS	2019-02-24 17:24:01.697	2019-02-24 17:24:01.697	2019-02-24 00:00:00.000	2019-02-25 00:00:00.000
TGS006W19	AURETTE A	GRPIR	GRPIR	2019-02-23 14:55:41.877	2019-02-23 14:55:41.877	2019-02-22 07:00:00.000	2019-02-25 06:00:00.000
11577DE908	MERITO	GRVOL	GRPIR	2019-02-23 13:41:38.377	2019-02-24 15:47:42.697	2019-02-22 00:00:00.000	2019-02-23 00:00:00.000
12318AD907	MED TEKIRDAG	TRTEK	EGDAM	2019-02-23 02:47:44.383	2019-02-24 15:44:49.830	2019-02-22 00:00:00.000	2019-02-24 00:00:00.000
OTF4RS1MA	LION	NULL	NULL	2019-02-22 08:14:38.383	2019-02-22 08:14:38.383	2019-02-18 10:00:00.000	NULL
TGS007E19	GABRIEL A	MTMAR	GRPIR	2019-02-21 23:40:07.973	2019-02-23 04:31:13.813	2019-02-21 02:00:00.000	2019-02-23 06:00:00.000
0TF4RS1MA	LION	GRPIR	TRIST	2019-02-21 14:54:11.373	2019-02-23 00:29:25.840	2019-02-21 07:06:00.000	2019-02-23 13:00:00.000
04607NC908	MSC LISA	GRPIR	TRIZM	2019-02-20 22:03:02.223	2019-02-21 21:56:00.983	2019-02-20 00:00:00.000	2019-02-22 00:00:00.000
VOASPR517W	UNI-ASPIRE	GRSKG	GRPIR	2019-02-19 14:52:36.290	2019-02-20 14:44:28.690	2019-02-19 16:30:07.000	2019-02-21 16:30:07.000
CD016N19	CAFER DEDE	TRGEM	TRIZM	2019-02-18 18:48:45.180	2019-02-19 23:20:46.160	2019-02-18 13:16:50.000	2019-02-19 13:16:50.000
12590AD906	CONTSHIP TOP	TRTEK	EGDAM	2019-02-18 07:54:49.883	2019-02-18 16:48:43.770	2019-02-18 00:00:00.000	2019-02-20 00:00:00.000
MXPRIDE1907	MAX PRIDE	GRSKG	CYLMS	2019-02-17 07:50:06.377	2019-02-18 15:42:10.947	2019-02-16 00:00:00.000	2019-02-17 00:00:00.000
TGS006E19	AURETTE A	MTMAR	GRPIR	2019-02-17 02:12:36.843	2019-02-18 23:54:25.410	2019-02-15 00:00:00.000	2019-02-17 00:00:00.000
11577DE907	MERITO	GRVOL	GRPIR	2019-02-16 20:06:05.567	2019-02-17 19:37:47.700	2019-02-15 00:00:00.000	2019-02-16 00:00:00.000
04217NC907	MSC FABIENNE	GRPIR	TRIZM	2019-02-15 19:13:42.037	2019-02-17 00:09:03.057	2019-02-14 00:00:00.000	2019-02-15 00:00:00.000
TGS005W19	GABRIEL A	GRPIR	MTMAR	2019-02-15 18:12:35.033	2019-02-16 18:10:32.220	2019-02-13 00:00:00.000	2019-02-14 00:00:00.000
VOWNBL024W	WARNOW BELUGA	GRSKG	GRPIR	2019-02-13 20:35:23.510	2019-02-15 17:06:37.203	2019-02-12 23:03:18.000	2019-02-14 09:03:18.000
12122NC906	MSC SARAH	TRIZM	GRPIR	2019-02-12 10:47:12.940	2019-02-15 16:08:49.630	2019-02-12 00:00:00.000	2019-02-13 00:00:00.000
AX001W19	ASTERIX	TRGEM	TRIZM	2019-02-11 01:51:23.927	2019-02-12 11:28:38.737	2019-02-10 00:00:00.000	2019-02-11 00:00:00.000
MXPRIDE1906	MAX PRIDE	GRSKG	CYLMS	2019-02-10 09:48:07.290	2019-02-12 01:05:06.043	2019-02-10 00:00:00.000	2019-02-11 00:00:00.000

Figure 4. Vessel call data from Port of Thessaloniki container terminal

date_katapl	departure_date	from_limani	to_limani
2018-12-21 17:00:00.000	2019-01-10 17:10:00.000	009EEU	005PVE
2018-12-26 15:00:00.000	2019-01-03 09:40:00.000	528SPD	052ALI
2018-12-27 01:20:00.000	2019-01-11 08:30:00.000	075YEY	066CND
2018-12-27 14:20:00.000	2019-01-02 13:50:00.000	600VAS	009TSI
2018-12-28 21:48:00.000	2019-01-03 07:50:00.000	009EEU	009EEU
2018-12-29 10:45:00.000	2019-01-01 09:00:00.000	009APY	009APY
2018-12-30 07:30:00.000	2019-01-01 13:00:00.000	075NVS	068BOJ
2018-12-30 08:05:00.000	2019-01-05 11:50:00.000	052NEM	052NEM
2018-12-30 15:10:00.000	2019-01-02 08:05:00.000	009EEU	009EEU
2018-12-31 12:00:00.000	2019-01-04 09:25:00.000	009APY	009APY
2018-12-31 14:00:00.000	2019-01-03 18:00:00.000	052IZM	009PIR
2019-01-01 03:15:00.000	2019-01-05 13:50:00.000	072KHE	068BOJ
2019-01-02 07:50:00.000	2019-01-07 19:00:00.000	072MRU	009EEU
2019-01-02 14:36:00.000	2019-01-04 14:06:00.000	046MAR	009PIR

Figure 5. Vessel call data from Port of Thessaloniki conventional cargo terminal

3.3. Predictions and analytics for PIXEL

In PIXEL project focus will be on the tasks that were presented in the introduction. In this section we give more details about each of the tasks. Vessel call data represents the main input for these tasks, as this data is the most general source of information that can be obtained by the ports and usually, if available / stored, is very similar, because it is obtained from the FAL Forms, which are standardized by the IMO. The differences arise from the fact that, usually the ports store only selection of the data that is present on the FAL Forms, something that can easily be resolved, as is just the matter of what is stored; the data is available. In the following subsections problems that will be addressed, are described.

3.3.1. General statistical analysis and visualization of the vessel call data

The amount of data that can be gathered about the vessel calls, offers possibilities to gather operational insights, to boost efficiency of port operations and at the same time reduce environmental impacts. As presented, this data is usually gathered from FAL Forms which are in place for quite a few years. Some of the ports, as presented above, store selection of the data that is available in FAL Forms, for different periods of time. One of the goals of this task is to show the ports as stakeholders, that the data captured in FAL Forms, offers additional capabilities, which can lead to useful insights. Thus, it is worth storing it in a proper manner, with all the attributes that are relevant for statistical analysis, for at least a few years back.

On a first look, this kind of data does not offer much but as will be presented in the initial results section, a lot of useful insights can be obtained. Based on the initial data, which was obtained from the ports, Port of Bordeaux stores most of the data that is captured in FAL Forms and the data for a few years back was provided for now (2010-2017). This data contains the exact entry and departure times, cargo types and the amount of cargo that was processed, along with other data, which is not so relevant. From such data, useful statistics can be gathered. Below, some of them are listed.

- Distribution of vessel call duration.
- Vessel call duration based on entry day/holidays.
- Number of entry and exists based on a day of a week/time of year etc.
- Vessel call duration given cargo type, specific ships, etc.
- Statistics about cargo/ship types in general.
- Analysis of type and amount of cargo that is imported/exported given time of the year – seasonality.
- How regular are specific shipping lines, can we predict them in advance more precisely?
- What is the influence of weather on port operations?
- How efficient is the port at processing different cargo (tonnage, hours known)?

Above, few of the statistics that will be computed and/or visualized, are presented. This kind of data will be visualized for the port operators and updated with live data, such as provided by the Port of Bordeaux through their VIGIEsip API. Different popular Python Data Science tools will be used, such as Pandas[12], to process the input data and different visualization libraries to present the results, such as Matplotlib[13] and Seaborn[14]. These insights will not only help the ports, but they are also helpful for other tasks that are presented in the next sections, as this exploratory data analysis is useful, to better understand the data to be selected, with the right features for predictive tasks (e.g. vessel call duration prediction) and to understand their influence on the problem.

[12] https://pandas.pydata.org/
[13] https://matplotlib.org/
[14] https://seaborn.pydata.org/

3.3.2.Prediction of vessel call duration

Port of Bordeaux and Port of Thessaloniki store the exact time of ship arrival and departure. This kind of data can be used to predict the duration of a vessel call on entry to the port. Such information can greatly boost port scheduling and operations of the port. For more precise vessel call duration, different attributes are needed, either provided by the ports in vessel call data, other internal port data or provided by external providers.

Different attributes have different impact on the duration of a vessel call. EDA (Exploratory Data Analysis) presented in the next section, will provide answers to some of the questions. The most obvious attribute that is captured in FAL Forms and is stored as internal data in most of the ports, is the amount of cargo that was processed during a vessel call. The amount of cargo has some correlations with the vessel call duration, as the more cargo that needs to be processed, the longer the ship will probably stay in the port. But the strength of this correlation depends on plenty of other factors, which have influence on the duration of a vessel call. The processing capacity for diverse cargo types is different and has maybe even bigger influence on the duration of a call than the amount of cargo itself. Another factor that influences the duration of a call, is weather, as if the winds are too strong, cranes cannot operate, or operate at a reduced capacity. Rain can also have influence, especially on the processing of certain types of bulk cargo. Duration of the vessel call also depends on the day of arrival at the port, as processing capacity during the weekends or holidays, is probably reduced.

This task mainly focuses on predicting a vessel call duration. The first main task will be feature engineering, which means finding useful attributes in the data or producing new ones from the data available and domain expertise, which can be used to boost prediction accuracy. Different regression methods will be utilized to predict vessel call duration. In this deliverable, we present results obtained with a simple linear based method, to see the amount of linear correlation that is present in the data. In the next deliverable, more advanced methods that can model nonlinear dependencies, will follow. Besides general approaches, such as Regression Decision Trees, Support Vector Machines, Neural Networks, ensemble-based methods will also be tested out, such as Random Forest method and state of the art approaches such as XGBoost[15]. Different automatic feature selection methods will be tested out and different protocols for splitting the data to training, validation and testing sets to evaluate the performance. Different error metrics will be computed, such as MAE (Mean Average Error) and RMSE (Root Mean Square Error)[16], compared to the actual departure times, as well as, estimated departure times on time of entry (where available), in order for our methods to be compared, not only to the actual departure times but also to the departure times, that were estimated by the port authorities on time of entry.

3.3.3.Prediction of a vessel call

Some of the ports store vessel call data for years back, which offers possibility to discover patterns that could help predict future trends in the port, in terms of cargo type, volumes or even a specific ship, especially in cases of a regular line, well in advance, out of historical data.

Figure 6 visualizes the difference in terms of days that have passed, since the last visit of a ship, which has visited Port of Bordeaux a lot of times. From the results it is concluded that is probably a regular shipping line, because in 40% of the time, 7 days have passed, since the last visit, while in almost 80% of the time, 6-9 days have passed. Besides that, from historical data, it is concluded that this ship comes weekly and that it is a container ship, which usually brings about the same amount of cargo to be processed - about 3000 tonnes.

[15] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). ACM, New York, NY, USA, 785-794. https://arxiv.org/pdf/1603.02754.pdf

[16] Chai, T. and Draxler, R. R.: Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature, Geosci. Model Dev., 7, 1247-1250. https://www.researchgate.net/publication/262980567_Root_mean_square_error_RMSE_or_mean_absolute_error_MAE

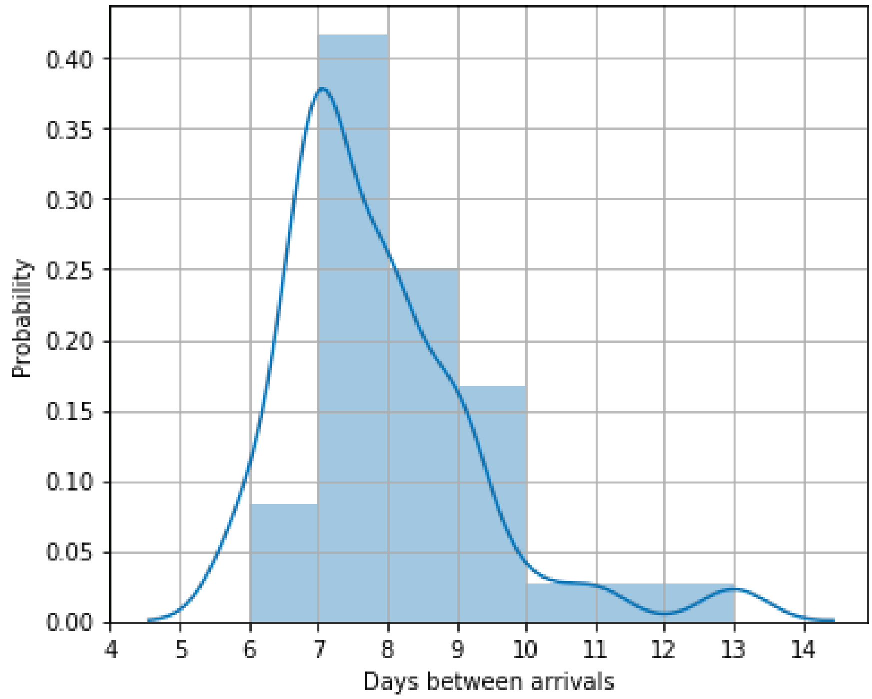


Figure 6. A typical graph for a regular shipping line.

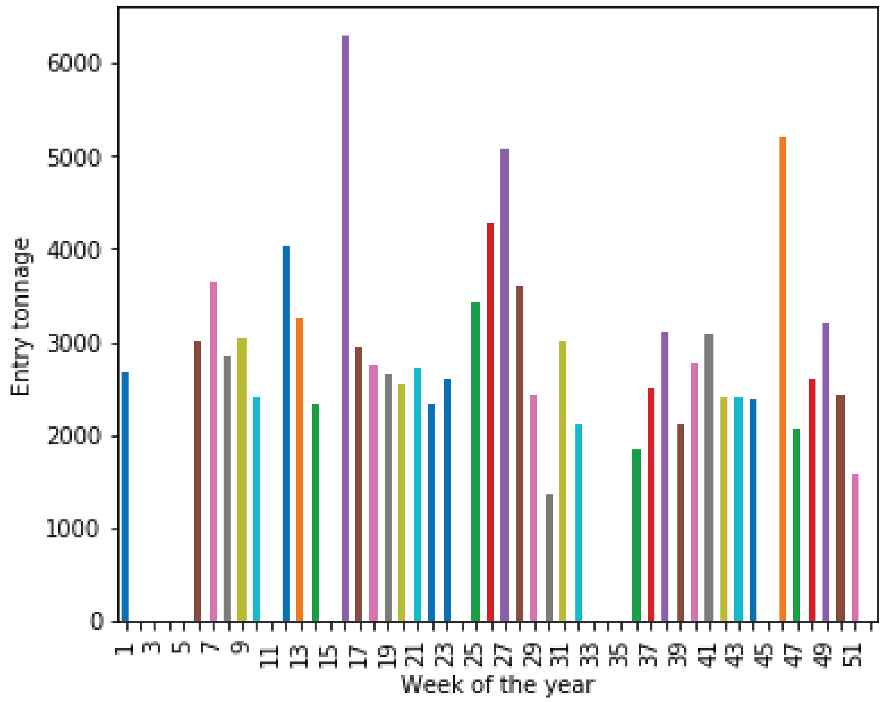


Figure 7. The amount of cargo that is processed per week for the ship presented in Figure 6.

This is just one example that clearly shows, that vessel call data can be used to discover patterns, which can help at predicting future cargo volumes, more accurately and far in advance. Moreover, seasonality of the cargo, collective vacations, days of the week, etc., are observed. Some of the initial results of EDA, are presented in the next section. Predicting future volumes and type of cargo, will be the main goal of this task. Crucial for the success of this task, is the amount of data. Currently, there is enough data only for the Port of Bordeaux. Different regression and classification methods will be used, including state-of-the-art methods, some of which were mentioned in the preceding subsection.

3.4. Initial results

In this section initial results, based on Port of Bordeaux historical vessel call data, are presented. The focus will be on general statistical analysis and visualizations of vessel call data, which also serves as exploratory data analysis (EDA), to address other tasks that were presented in the preceding section. Based on this data analysis, initial results for vessel call duration prediction are also presented.

To begin with, some basic statistical analysis is conducted, that will give insights on the type of cargo that is processed in the Port of Bordeaux. Figure 8 presents different types of cargo and the number of ships that have unloaded or loaded that specific type of cargo. From this bar plot, it is concluded that corn is the main export product and that is only exported. It is also noticeable that containers are always loaded and unloaded, while there are certain types of cargo that are only imported. Bar plot shows only the most frequent types of cargo.

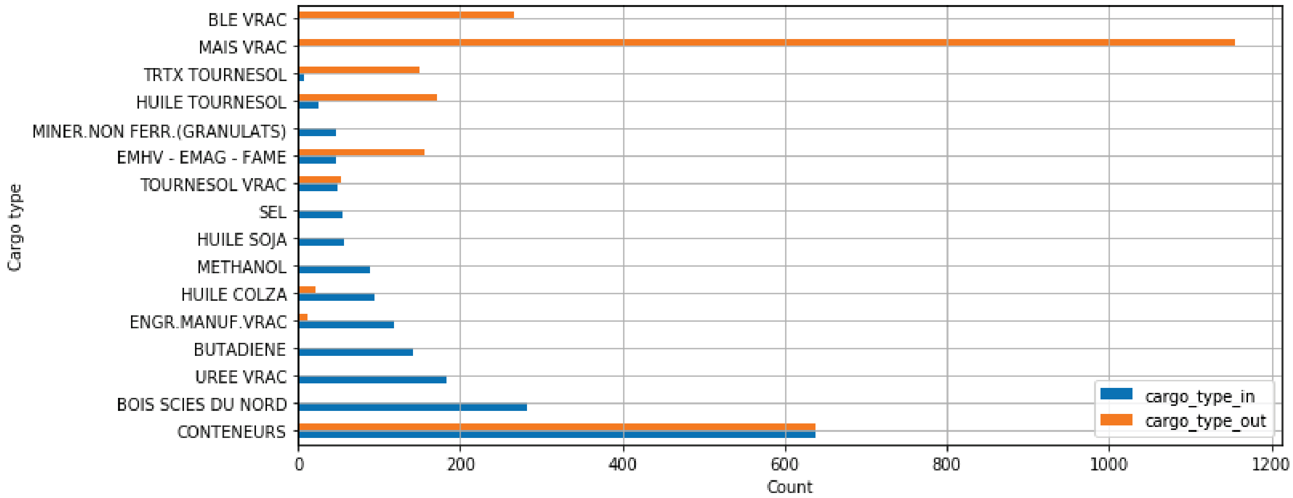


Figure 8. Cargo types and the number of ships that have loaded or unloaded cargo.

Figure 9 represents this in more detail, with the high-level classification of the cargo types, compared to the previous example. There are examples of vessel calls, where no cargo was either loaded or unloaded, probably of ships arriving for the maintenance, something that rarely happens.

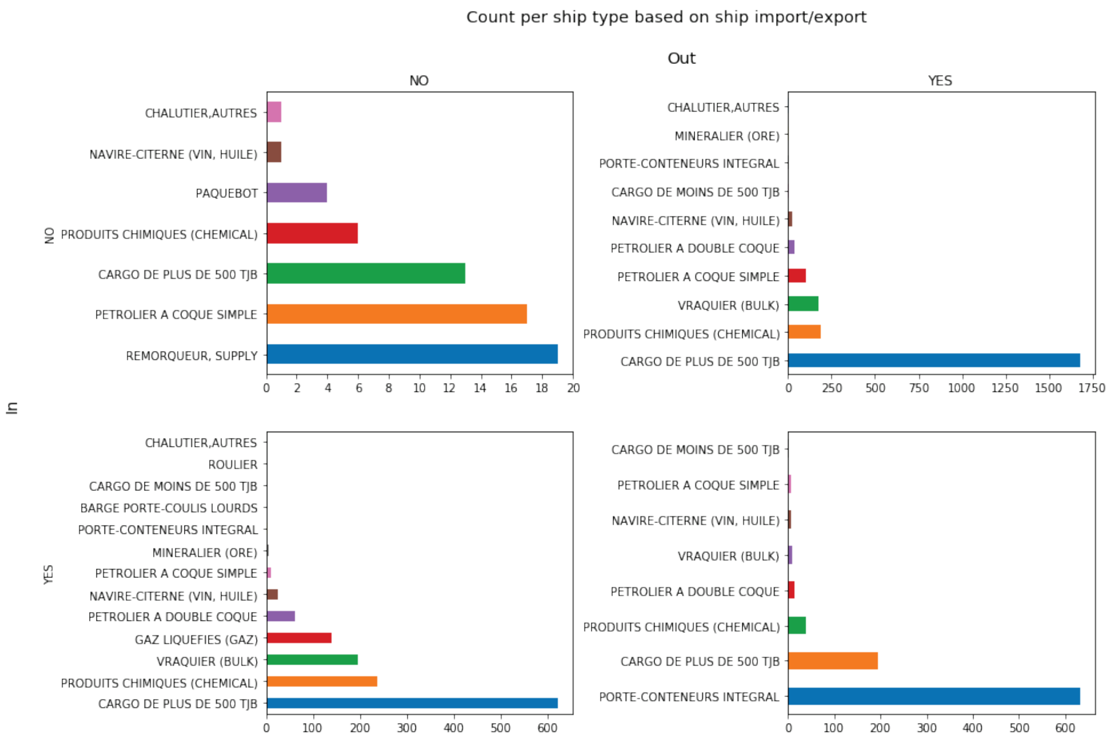


Figure 9. Import/Export, number of ships given high-level classification of cargo type in Port of Bordeaux

Heatmap in Figure 10, presents mapping between import and export cargo (i.e. if specific ship arrived with cargo X to be unloaded, what was cargo Y, that was loaded). It is concluded, that majority of the ships that came empty (or no cargo was unloaded), left with corn and other agricultural products.

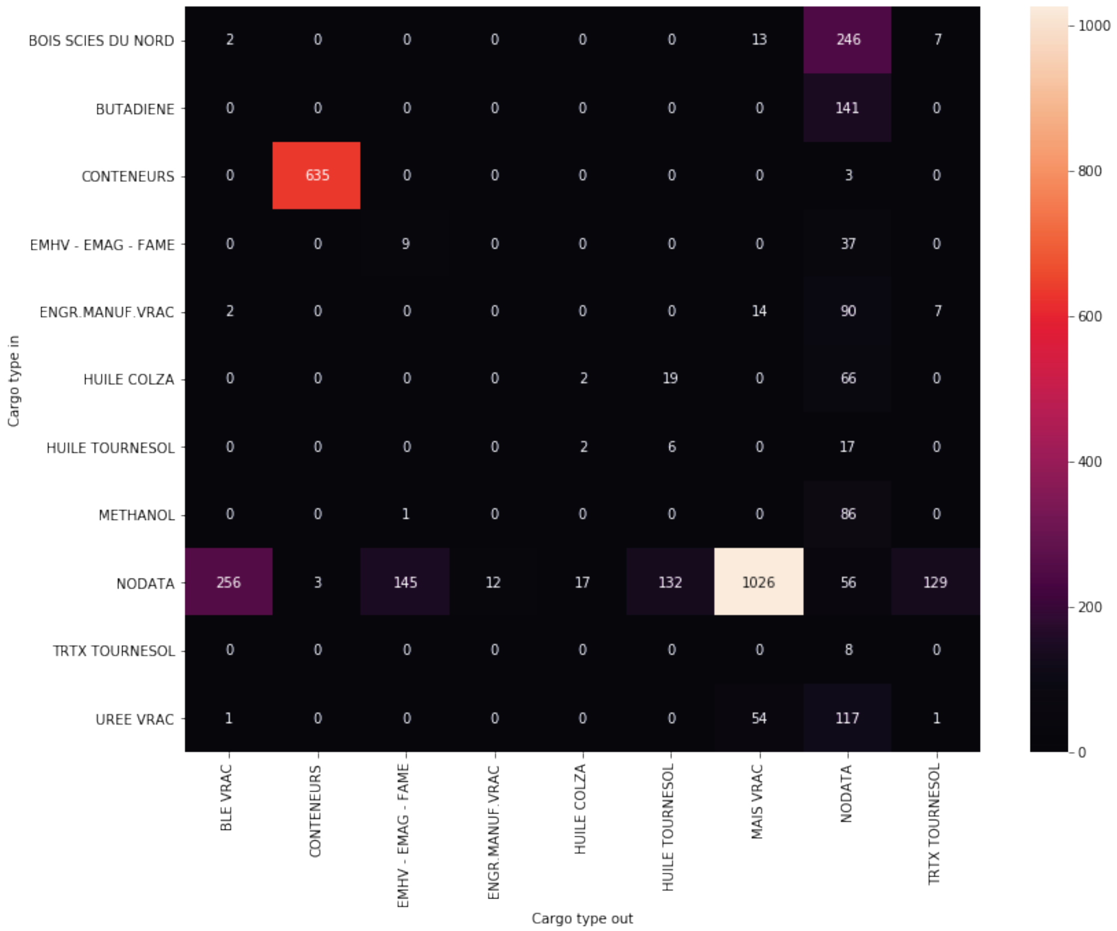


Figure 10. Heatmap representing the map between unloaded and loaded cargo

From historical data that was made available by the Port of Bordeaux from 2010 to 2017, seasonality of the cargo, can also be analysed. Figure 11 presents distribution of the cargo per week, for the whole historical data that was made available, normalized by the maximum amount for each cargo type per week. A first observation is that containers are unloaded at a steady rate through the whole year, but this is not the case with some of the products that are seasonal. An example of this, is fertilizer (UREE VRAC). Peaks of imports are observed between week 7 to week 20 and between weeks 38 to 47. These are exactly the parts of the year, which fertilizer is needed in agriculture. Similarly, seasonality of some of the agricultural products that are imported is observed, such as sunflower seeds (TOURNESOL VRAC), which are harvested in autumn. Moreover, besides containers and fertilizers, wood from the north is also imported a lot, especially between week 10 and week 30, probably due to winter conditions before and after. Interesting example is also cement, which is imported mostly between May and September, which coincides with the peak of construction.

Figure 12 presents the same heatmap, but for cargo that was loaded (i.e. exported), where some of the peaks are clearly seen. Corn (MAIS VRAC), which presents the main agricultural product that is exported, has its peak in the end of September/start of October, which correlates with the harvesting season. Of course, there are some delays between harvesting and the actual export. Similar peaks can be seen with other agricultural products, such as wheat (BLE VRAC), rapeseed (COLZA VRAC) and others. This kind of data is useful for vessel call/cargo volume prediction well in advance.

Figure 11. Heatmap representing seasonality of the imported cargo.

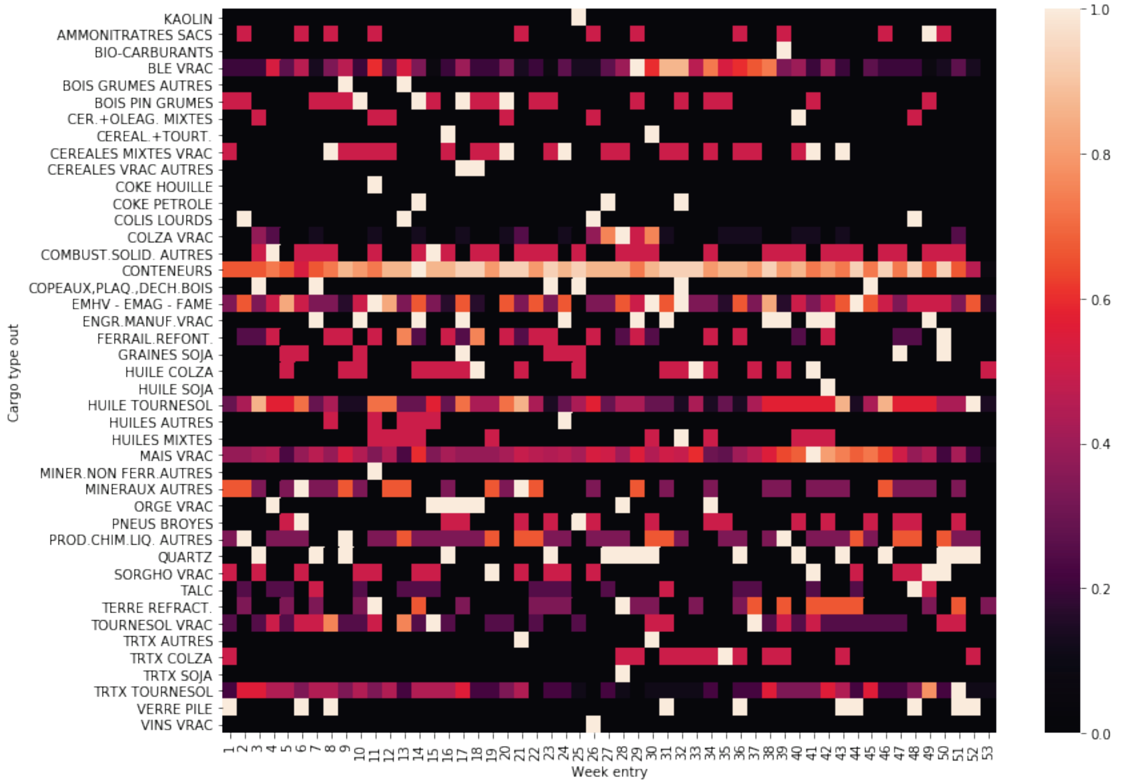


Figure 12. Heatmap representing seasonality of the exported cargo.

One of the tasks to be tackled, is to predict the duration of a vessel call on arrival. Figure 13 presents the distribution of vessel call duration. Average duration of a vessel call is 53.5 hours. The most obvious factor that one would say has the most influence on the duration of a vessel call, is that of the amount of cargo that needs to be processed. Figure 14 visualizes correlations between cargo type and hours spent in the port. For some of the cargo types, there is some linear correlation, while for others, there is almost no correlation. For containers, the number of containers that were processed would be a more proper figure as input data for the processing time, since tonnage does not correlate well, as it depends on what is in the containers.

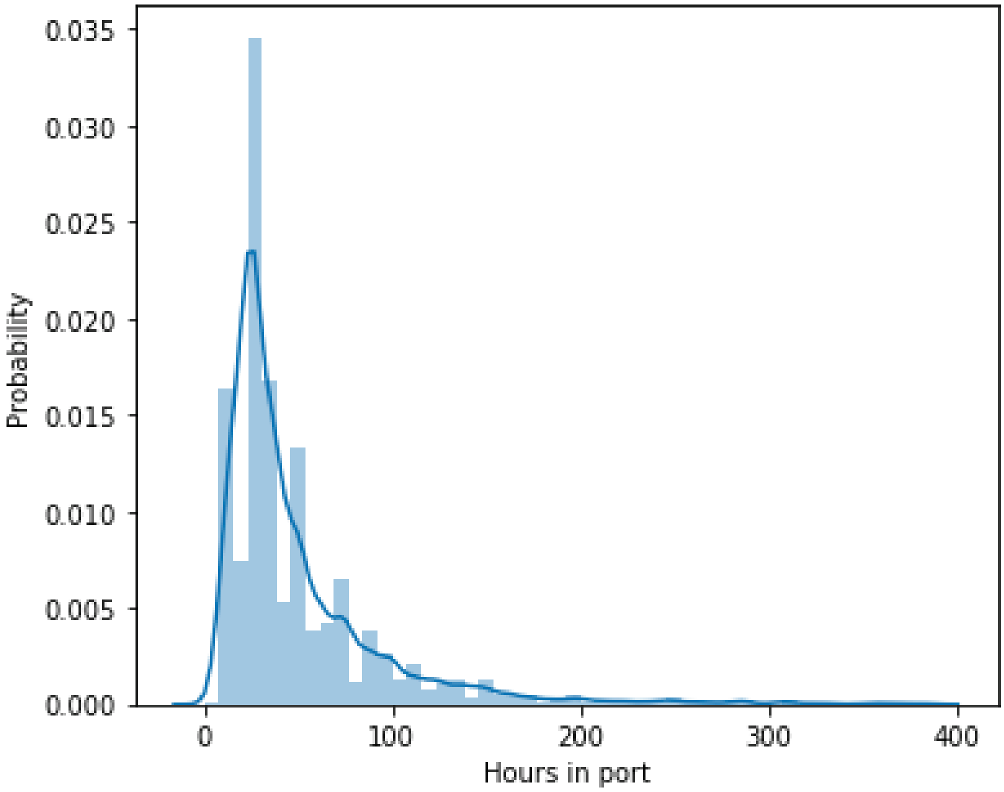


Figure 13. Distribution of vessel call duration in Port of Bordeaux between 2010 and 2017.

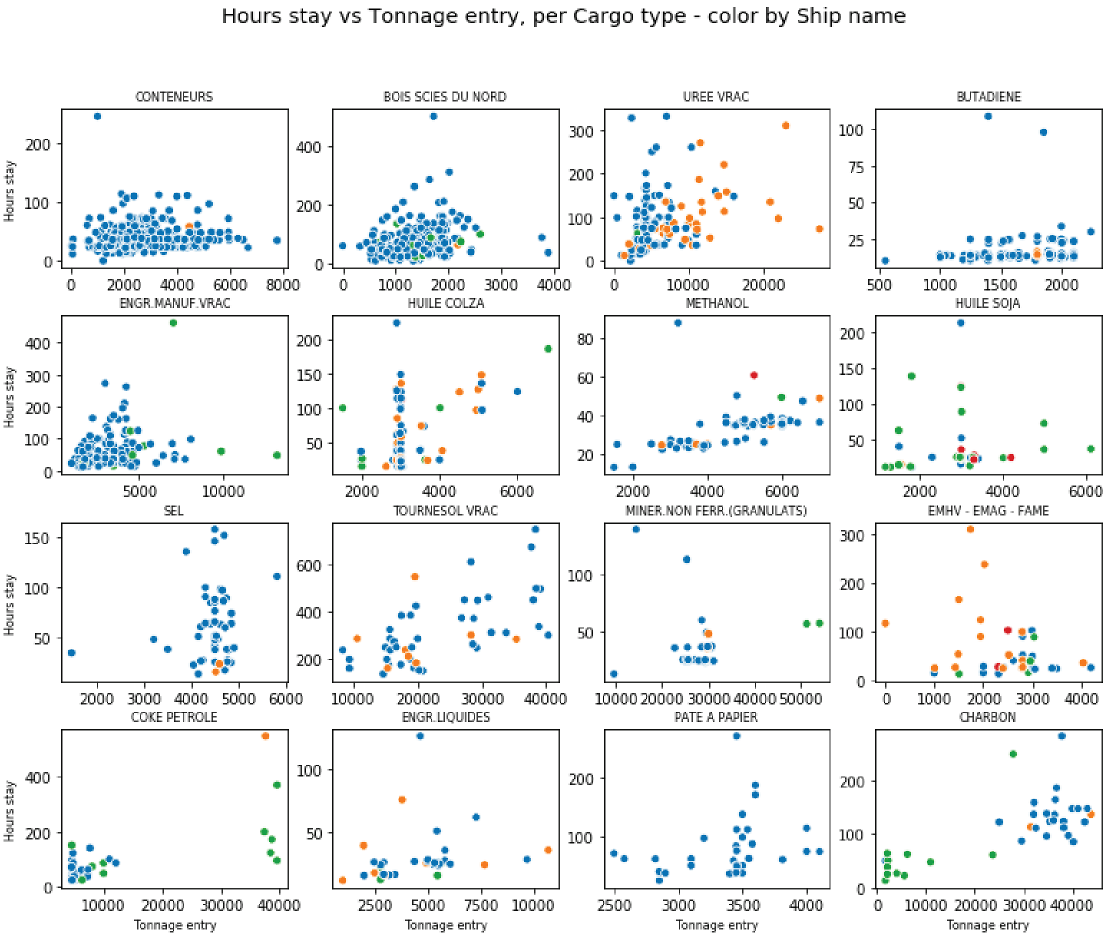


Figure 14. Correlation between the cargo tonnage to be unloaded and vessel call duration.

Besides amount of cargo that needs to be processed, there is plenty of other factors that have influence on the duration of a vessel call. Figure 15 presents how the day of the arrival, on a specific day of the week, influences the duration of a vessel call, as well as holidays, in a form of a boxplot[17]. As it is shown in the picture below, ships that arrive on weekends stay generally longer in the port. This is probably due to reduced manpower during the weekend. Similar can be seen for the ships that arrive on Fridays, especially if cargo needs more than 1 day to be processed. Holidays have quite an influence as it can be seen, while if the following day is a holiday, ships in general stay longer in the port. A longer stay is also especially noticeable around the weekends.

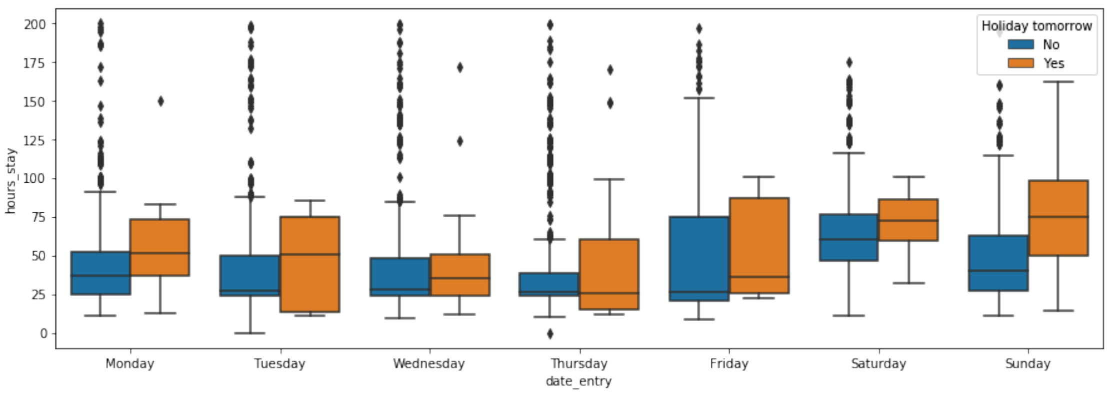


Figure 15. What influence has day of the week and holidays on vessel call duration.

Different cargo types have different processing times. Figure 16 presents what is the processing capacity in terms of tonnage per hours, for different cargo types. We can see that that some of the cargo has more variability

[17] https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51

compared to others such as containers, where the process is more efficient and repeatable. Though, ship that came in with cargo X to unload, might load cargo Y instead, while hours of stay that are available, does not capture that. However, through the graph, it is inferred that type of cargo matters.

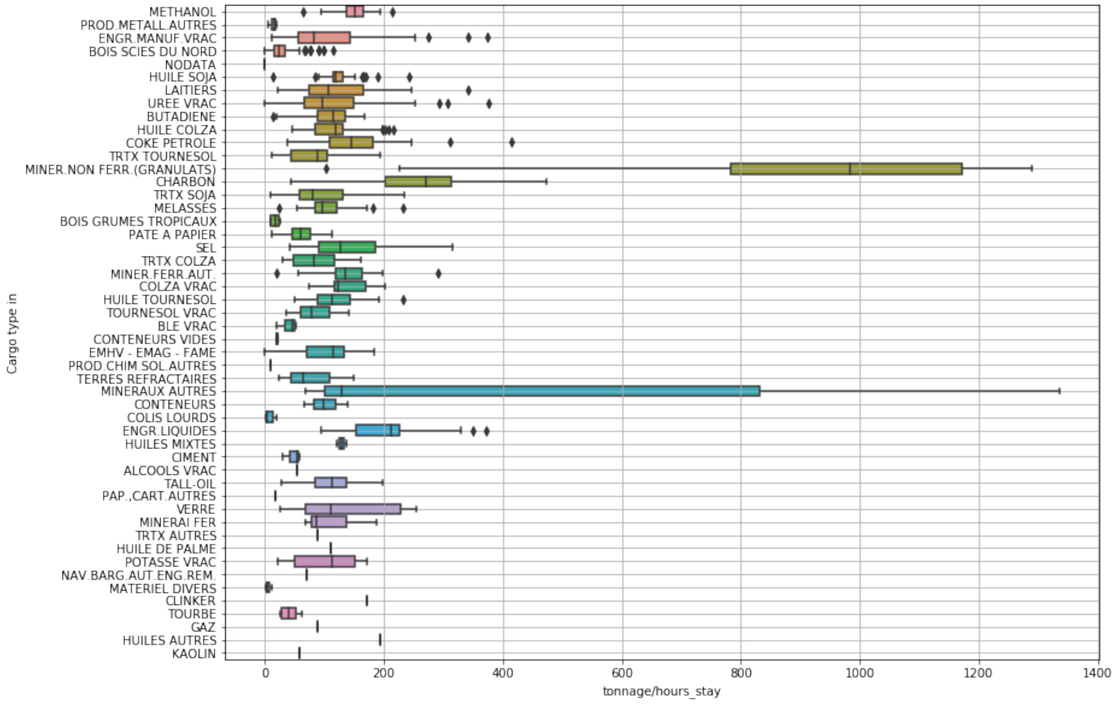


Figure 16. Processing capacity in terms of tons per hour for different cargo types.

In the last few examples, different attributes that have different influence on the duration of a vessel call were demonstrated. Below, there is a list of some of the attributes, in a high-level way, which have been used in this initial stage, to develop prototypes to predict vessel call duration.

- Ship name
- Ship type
- Cargo type in & out
- Tonnage entry & exit
- Berth
- Number of ships in the port on entry
- Number of same ships (cargo in & out) in the last few days
- Entry day of the week
- Weekend on entry (± 3 days)
- Holiday on entry (± 3 days)
- Entry hour
- Weather data

For weather data, data from NASA (National Aeronautics and Space Administration)[18], was used, such as precipitation, temperature, wind speed and direction at 10m and 50m. Data was aggregated with different moving averages etc., but no direct correlation was found so far. This data is modelled and does not represent real measurements. Currently, solutions on getting weather data directly from the port are searched, in the cases that sensors are installed on their equipment.

[18] https://power.larc.nasa.gov/data-access-viewer

Linear Regression with regularization[19] was used for our initial prototypes, to get a feeling of the level of performance, with such simple linear methods. Two different learning strategies were used. First is to use a variant of a cross-validation[20] called "leave one year out"; basically, leaving out one year, is for testing the performance, while the rest, for training the method. Another strategy, called "forward chaining", only uses training data from the years before the year, which was selected to test the performance. For example, for the year 2014, only the data between 2011 and 2013 is used, to train the methods that are later tested on the data for the year 2014. Same process is performed for all years.

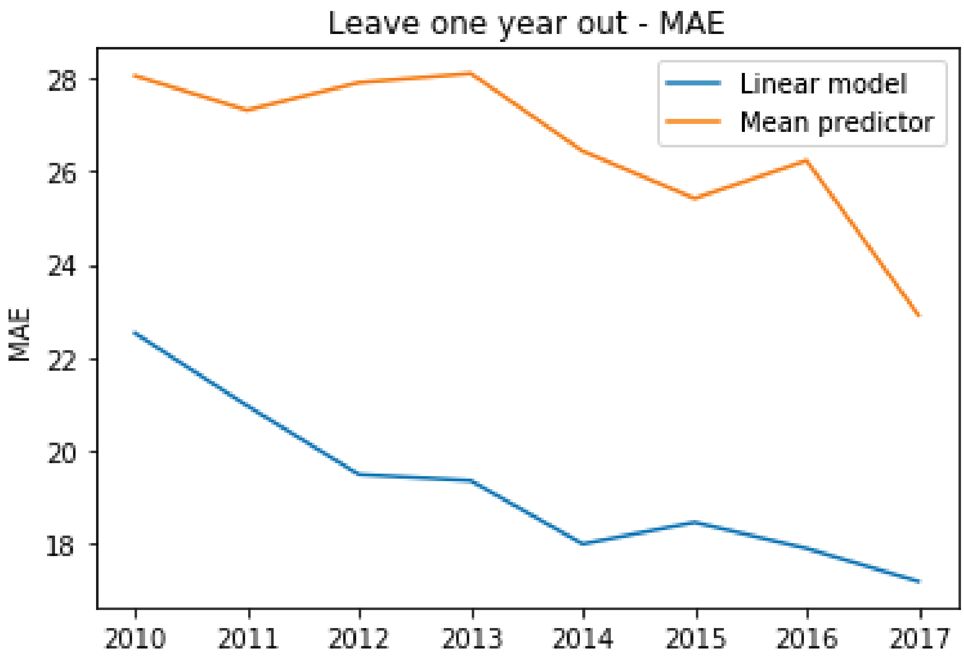


Figure 17. Results (MAE) of the leave one year out strategy of a Linear Regression method and a baseline average predictor.

Figure 17 presents the results for the "leave one year out" strategy. As it is clearly seen, this simple method already outperforms baseline average predictor, by a large margin. Baseline average predictor always predicts precomputed average duration of a vessel call, which was calculated on training data. Average MAE for Linear Regression method was 19.2 hours, while for the baseline predictor was 26.5 hours.

[19] https://scikit-learn.org/stable/modules/linear_model.html
[20] https://en.wikipedia.org/wiki/Cross-validation_(statistics)

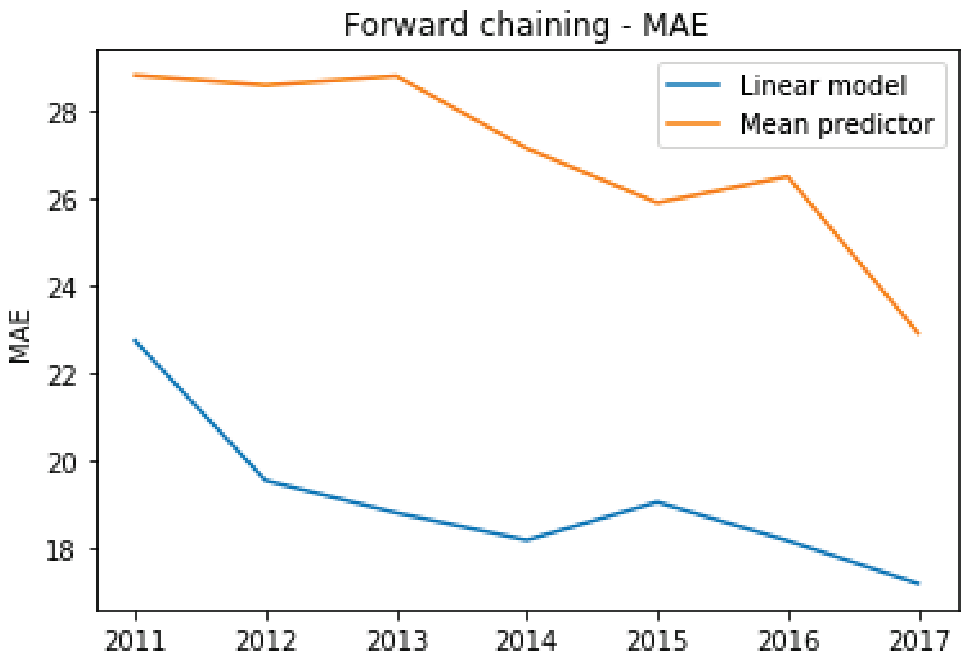


Figure 18. Results (MAE) of the forward chaining strategy of a Linear Regression method and a baseline average predictor.

Figure 18 presents the results for the "forward chaining" strategy, depicting the influence of the amount of data, as the MAE decreases while moving to recent years, mostly due to more data that is used for training the method. The best MAE that was achieved overall for both approaches, was for the last year of our data (i.e. 2017) and was 17.2 hours, compared to 22.9 hours with the average predictor.

3.5. Plan and future work

In this section ports internal data about vessel calls are presented, as well as some of the tasks that can be performed with such data. Vessel call data presents the most general data source that all the ports store, or have information available to store, due to internationally standardized procedures that are in place, to capture this kind of information (FAL Forms). One of the main goals of the tasks is to show ports the importance of this data for their operational use. Currently, only data from Port of Bordeaux was received, that is of enough quality and quantity to be used for all the tasks that were mentioned.

Additional data is expected from other ports, because some of the tasks, will be performed using all the data from the ports; in some cases, for example Port of Piraeus, accurate enough data on arrival and departure times are not properly stored and is thus, not applicable to predict vessel call duration. Some of the demo samples that were sent, do not contain information about the type of cargo and the amount, which was handled. Nevertheless, the developed methods will be general and applicable to any port, which stores enough data, present in FAL Forms. Some of the ports, such as Port of Bordeaux, will provide API access to FAL Forms data, so that predictions will be made on live data.

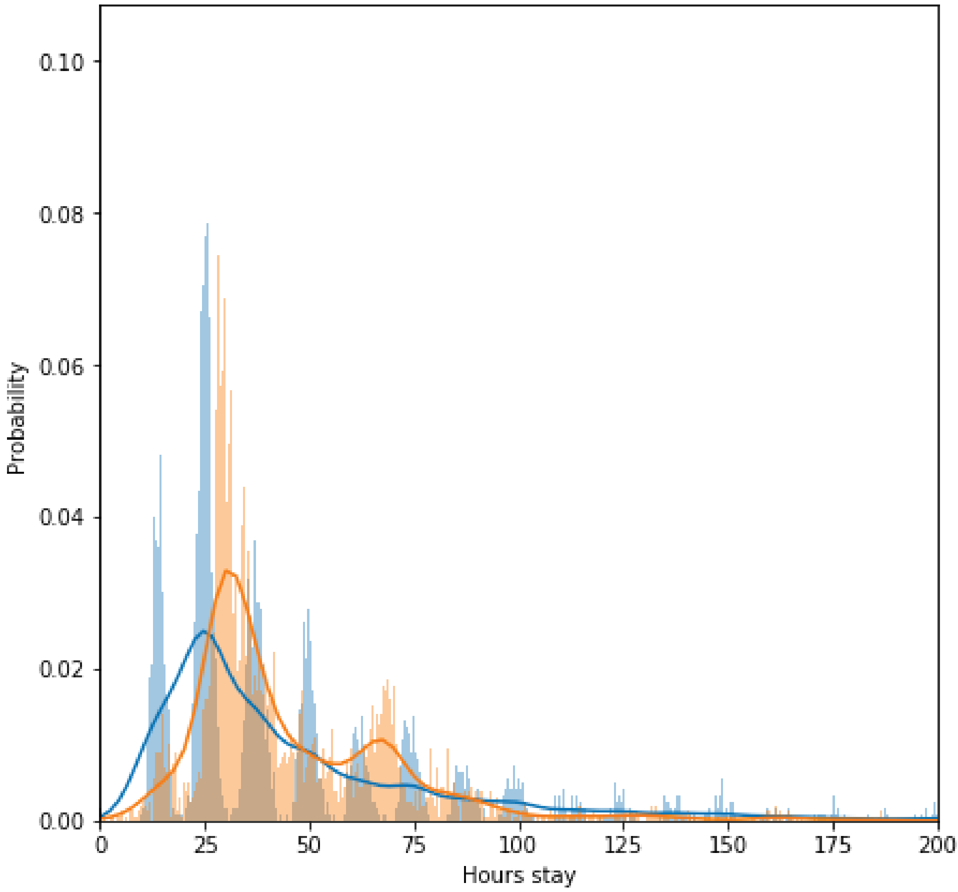


Figure 19. Distribution of ground truth duration of vessel calls and predicted values only on the test sets that were hold out in leave one year out strategy.

Currently, only simple linear based methods were tested out. Advanced Machine Learning methods will be tested out, including state-of-the-art approaches. Figure 19 shows the distribution of predicted values and ground truth values, for all the years in a "leave one year out strategy" (test sets only). In ground truth data where there are groups of vessel calls duration values, it is observed that are discriminative and not detected by simple linear based methods. The goal is to improve the accuracy of current predictions, by utilizing methods that can learn that kind of patterns.

4. Use of AIS data

4.1. Introduction

Automatic Identification System (AIS) was proposed and mandated by IMO (International Maritime Organization) by 2002 SOLAS (International Convention for the Safety of Life at Sea) agreement[21] and its main intention was to prevent collisions on sea, by supplementing additional information, however not replacing existing solutions on board, such as radar and other means that are regulated in COLREG (Convention on the International Regulations for Preventing Collisions at Sea)[22]. Since December 31th, 2004 all vessels exceeding 300GT, are obligated to have an AIS transceiver installed and operational, while each has assigned a unique identifier, MMSI (Maritime Mobile Service Identity). Data such as navigational data, information about the ship and voyage related data, is transmitted via VHF (Very High Frequency) radio between ships and shore stations. The range is limited to the VHF range, which is about 10-20 nautical miles but S-AIS (Satellite-based AIS) is available, which can track ships on open seas. Besides collision avoidance, AIS data is used for many other applications in maritime domain, such as fishing fleet monitoring, maritime security, search and rescue, accident investigation, fleet and cargo tracking and many others.

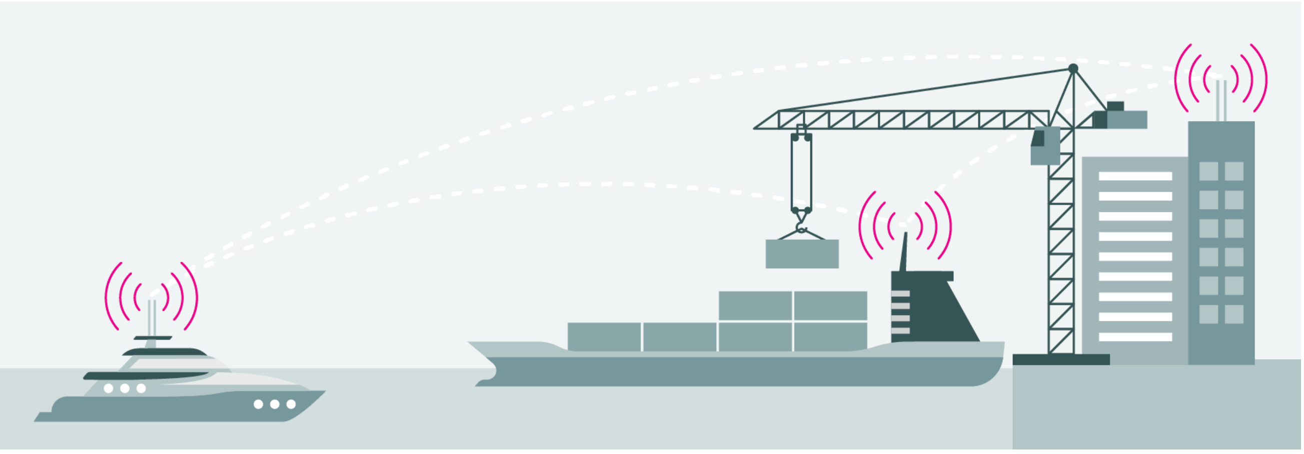


Figure 20. Automatic Identification System (AIS)

Besides natural use cases described above, AIS data also offers opportunities to develop intelligent maritime systems, that could automate some of the tasks mentioned above and provide interesting new use cases. The quantity of this data (spatially and temporally) is vast and could help optimise port operations and maritime traffic in general. Kinematic information (like ship location, speed, course, and heading) and some static information (like MMSI, ship type, ship size) are provided every couple of seconds when ship is underway and every couple of minutes when the ship is anchored. This data is available in almost real-time, while historical data is also available. This data is in the research community exploited in different domains, some of which are listed below.

- **Anomaly detection:** Through historical data, certain motion patterns can be observed, and deviations can be detected on live streaming data. Ships that deviate from standard paths, are of interest as they represent much higher collision risk. Besides historical motion data, ship type, length, draught, etc. can be incorporated, to boost anomaly detection accuracy even more.
- **Route estimation:** Through live and historical AIS data, future trajectory can be predicted for the ship; either short term or long term. This kind of information can be used in tasks, such as destination port prediction and collision prediction, presented below.

[21] http://solasv.mcga.gov.uk/

[22] http://www.imo.org/en/About/Conventions/ListOfConventions/Pages/COLREG.aspx

- **Destination port and ETA prediction:** Route estimation can be used to predict destination port and estimated time of arrival. ETA prediction is particularly useful, to run everyday port operations efficiently, which consequently helps at reducing environmental impacts.
- **Collision prediction:** By predicting future trajectories for multiple ships in the vicinity, collision risks can be estimated for each of the ships.
- **Traffic analysis:** AIS data can be used to analyse traffic patterns in an area of interest. AIS data provides data about ship types, sizes, draught, navigational status and other data presented in the following section that can be used to analyse traffic. Particularly useful can be port congestion indicators **(PCI),** which can be computed out of AIS data. This kind of data can also help at modelling pollution (T4.4), especially the one caused by exhaust fumes that have big influence on nearby living population.
- **Situation detection, classification and prediction:** Apart from unexpected or abnormal behaviours, AIS messages, especially when these are from nearby ports, can be used to detect normal situations in port operations. This information is very useful to expand the knowledge over the port operations, since it provides an automatic recognition that can be more accurate than traditional management systems, based on human inputs. With this data, a better optimization of the resources and better use of the port area can be promoted, with a clear impact on the environment. Some examples of events that can be detected are: anchor dragging, pilotage of a ship, towage, fuel bunkering, excess of speed in port area, etc.

In the following sections the information that is transmitted in AIS messages and the format of AIS messages are described. Different data sources that can be used to gather AIS data are presented, as well as use cases that are relevant for PIXEL project, which will be explored and some results of the initial work. Finally, the plan and future work regarding the use of AIS data in PIXEL project is described.

4.2. Data sources

Raw AIS messages are transmitted in NMEA (National Marine Electronics Association) 0183 protocol[23]. An example of raw AIS stream is presented in Listing 1.

```
!AIVDM,1,1,,B,19NWrrP02sbuuuuhM86hA0=n2<0:,0*12
!AIVDM,1,1,,B,1018lEPP00:tVVIhBeFh0?wj08Rp,0*5A
!AIVDM,1,1,,B,404k0uAu`T`G9bvnQqhCK4o0085o,0*26
!AIVDM,1,1,,B,19NWrrP02sbuv0KhM:Rh@@<F251L,0*2A
!AIVDM,1,1,,B,18M2Et0003:u9hEhETTaM10>051L,0*7C
!AIVDM,2,1,4,B,58M2Et02>2P<Q<;?C39QEA8608QDn22222222216Bpe@@6iF0KDnA3QF,0*74
!AIVDM,2,2,4,B,H88888888888880,2*53
!AIVDM,1,1,,B,17OfpH0P01:u5i7hDQUN2wvN0L0B,0*62
!AIVDM,1,1,,B,6=G<wGSEk?m@01Pf<u1mGij@onqH,0*14
!AIVDM,1,1,,B,17OpSQ0P1l:u8;KhEOj19wvT0@:p,0*4B
```

Listing 1. Raw AIS messages example.

Information that is transmitted can be roughly divided into two categories. Dynamic information, such as speed and positional information, captured from internal sensors installed on board of the ship and static information which is provided either at the installation of the AIS unit (e.g. MMSI number, ship name, size of the ship) or manually entered by the crew for each voyage (e.g. destination port, ETA, draught). Dynamic information is transmitted every 2-10 seconds while under way and every 3 minutes while at anchor. Static information is transmitted every 6 minutes. Some of the most informative fields transmitted in AIS messages are presented in Table 3.

[23] https://www.tronico.fi/OH6NT/docs/NMEA0183.pdf

Table 3. AIS message fields

Dynamic information	Static information
navigational status	MMSI number
rate of turn	IMO number
speed over ground	call sign
longitude & latitude	name of the ship
course over ground	type of ship & cargo type
true heading	length of the ship
timestamp	destination port
	ETA
	draught

Detailed description of each of the fields and possible values are described on the U.S. Coast Guard Navigation Centre website[24, 25]. Some of the interesting fields that have different possible predefined values to be defined, are navigational status, which provides information about the status of the ship, such as under way using engines, moored, at anchor etc., while among static information, ship/cargo type is very useful. This kind of data can be also obtained with IMO number in respective ship registers[26], which provides even more accurate data about the ship.

In the following sections, different sources where AIS data can be obtained, are briefly described. AIS data is collected on an EU wide level by EMSA (European Maritime Safety Agency). This data is collected from national authorities and commercial providers. Besides EU agencies and national authorities, there is also a plethora of commercial providers, where AIS data can be bought from. Some of the AIS sources that are available outside EU are described, as well as a source where one can obtain AIS data freely from amateur receiving stations, from all over the world.

4.2.1.European Maritime Safety Agency (EMSA)

The European Maritime Safety Agency is one of the EU agencies that provides technical assistance and support to the European Commission and Member States, in the development and implementation of EU legislation on maritime safety, pollution by ships and maritime security.[27] Besides that, it is awarded with operational tasks, such as monitoring ship traffic around EU and detecting oil pollution.

One of the systems that EMSA has in place is SafeSeaNet[28], which is a vessel traffic monitoring system put in place, to enhance maritime safety and to protect the environment. EU countries, together with Iceland and Norway. They exchange the data through this system, which enables them to provide and receive information about the ships, ship movements and hazardous cargo. This data includes AIS live information, as well as historical AIS data. Figure 21 presents the web interface used to show AIS data.

[24] https://www.navcen.uscg.gov/?pageName=AISMessagesA
[25] https://www.navcen.uscg.gov/?pageName=AISMessagesAStatic
[26] https://ihsmarkit.com/products/maritime-world-ship-register.html
[27] http://www.emsa.europa.eu/about.html
[28] http://www.emsa.europa.eu/ssn-main.html

Figure 21. SafeSeaNet web interface

This data is mainly available to the national authorities and national agencies, while requests can be made by European Union (EU) Member State national government authorities, EU institutions and bodies, and by projects or programmes established by these parties, working on issues of public interest.[29] The process to obtain the data through EMSA is as reported by other EU projects[30] very long, while sharing limitations are strict, limiting the use of this kind of data in PIXEL project.

4.2.2.National authorities

National authorities gather AIS data and exchange it through SafeSeaNet, which means that most of the national authorities, also store historical AIS data independently of EMSA, some of which also provide access to real-time AIS data. Some of the countries, provide this data (especially historical data) free of charge. Two notable examples are the Danish Maritime authority, which provides historical AIS data for Danish waters from 2006 onwards[31] and historical AIS data that is provided for USA coastline and inland waters by BOEM (Bureau of Ocean Energy Management) and NOAA (National Oceanic and Atmospheric administration)[32] from 2009 to 2017.

[29] http://www.emsa.europa.eu/emsa-documents/data-request-procedure.html
[30] https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php?title=Special:Pdfprint&page=WP4_Deliverable_1
[31] https://www.dma.dk/SikkerhedTilSoes/Sejladsinformation/AIS/Sider/default.aspx
[32] https://marinecadastre.gov/ais/

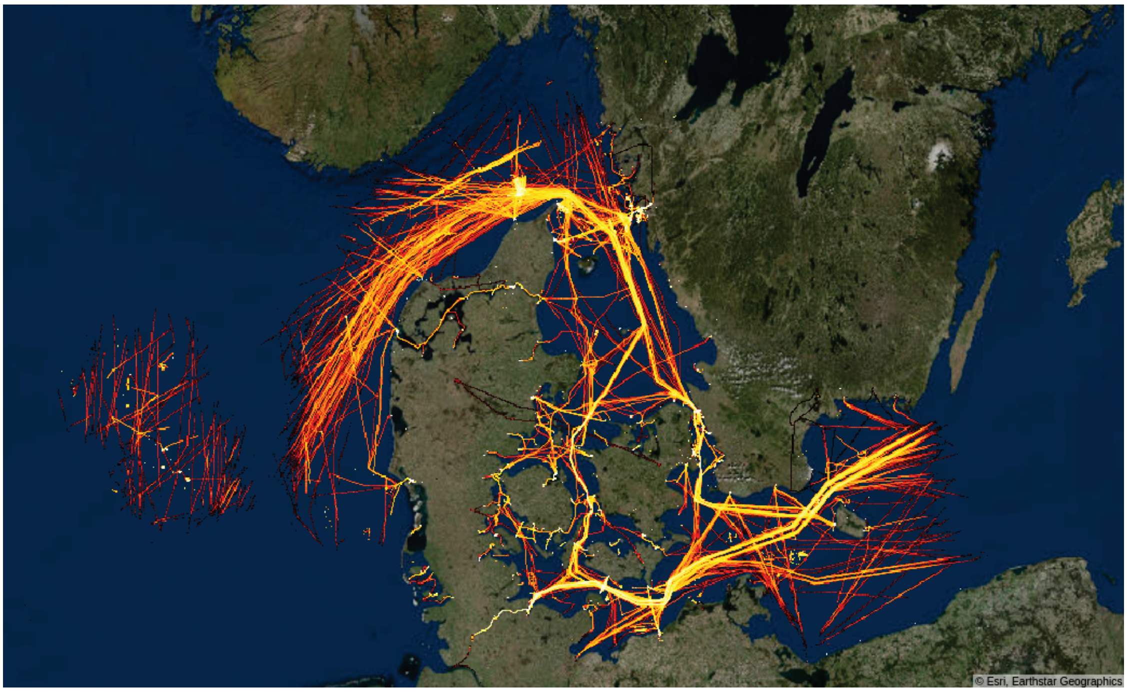


Figure 22. Danish Maritime Authority historical AIS data for 1 day.

AIS data from March 18th, 2019 from Danish waters is visualized in Figure 22. Figure 23 presents AIS data from USA coastal waters, visualized for the first 14 days in August 2017.

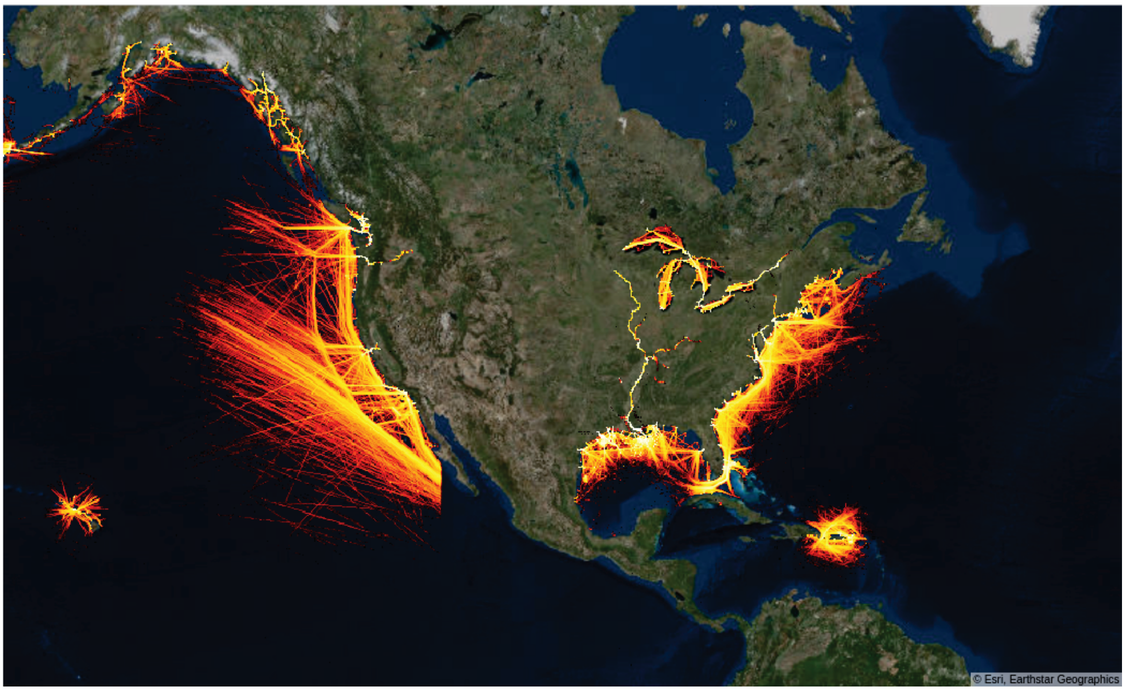


Figure 23. Marine Cadastre USA historical AIS data for 14 days in August 2017.

As reported by ESSNet Big Data project[33], Hellenic Coast Guard also has historical AIS data available from mid-2015 onwards. Requests have been made to obtain some of this data for Greek waters and a response is expected.

[33] https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php?title=Special:Pdfprint&page=WP4_Deliverable_1

4.2.3.Commercial providers

There is plenty of commercial providers for AIS and S-AIS live and historical data. Below some of the commercial providers are listed.

- **MarineTraffic:** MarineTraffic is one of the world's leading providers of ship tracking and maritime intelligence. Besides providing live and historical AIS data, they also provide products derived out of AIS data, such as ETA prediction, congestion rate at the ports, etc.
- **Dirkzwager:** Similar to MarineTraffic, Dirkzwager provides AIS and S-AIS live and historical data. Dirkzwager was chosen by ESS Net Big Data project as their data source, for EU wide AIS data.
- **VesselFinder:** Similar service to Dirkzwager. Historical data available from 2009 onwards but only for terrestrial receivers.
- **FleetMon:** Similar commercial service to MarineTraffic, that provides historical data from 2007 onwards, for terrestrial receivers and from 2013 onwards, for satellite receivers.

AIS data from commercial providers is expensive. ESS Net Big data project, made enquires at Dirkzwager for EU wide data and got a quote for 42000€ for 6 months of data, out of terrestrial receivers only. Later, a special price of 2000€[34] was achieved. We made enquiries at MarineTraffic for getting live AIS data in a small area in front of the Port of Piraeus and got a quote of 250€ per month. Gathering AIS data from roughly the Mediterranean Sea area, would cost around 6000€ per month for terrestrial and satellite data combined. Historical AIS data is usually priced per number of AIS messages/positions, which is the case with VesselFinder and MarineTraffic.

4.2.4.Port management Information System (PMIS)

Some of the ports have AIS receivers installed in their premises and connected to their PMIS, to be used for their internal operations. Port of Bordeaux (GPMB) has its own AIS receiver installed and data is visualized through VIGIEsip[35] as presented in Figure 24 and Figure 25.

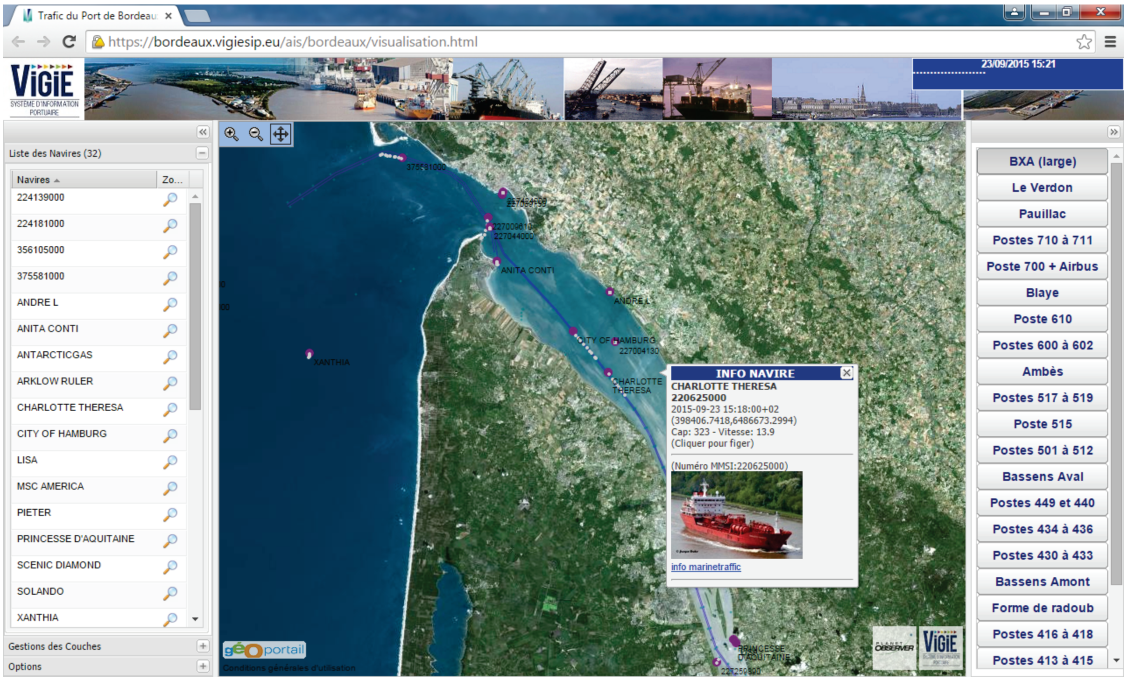


Figure 24. VIGIEsip AIS visualization

Vessels in Figure 25 are drawn in real size, thanks to official dimensions of the vessel recorded in VIGIEsip and controlled by the port officers. The range of AIS receiver in GPMB, is about 30 nautical miles. GPMB will prepare a direct access to the AIS data through VIGIEsip API.

[34] https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php?title=Special:Pdfprint&page=WP4_Deliverable_1
[35] https://bordeaux.vigiesip.eu/ais/html/bordeaux/visualisation.html

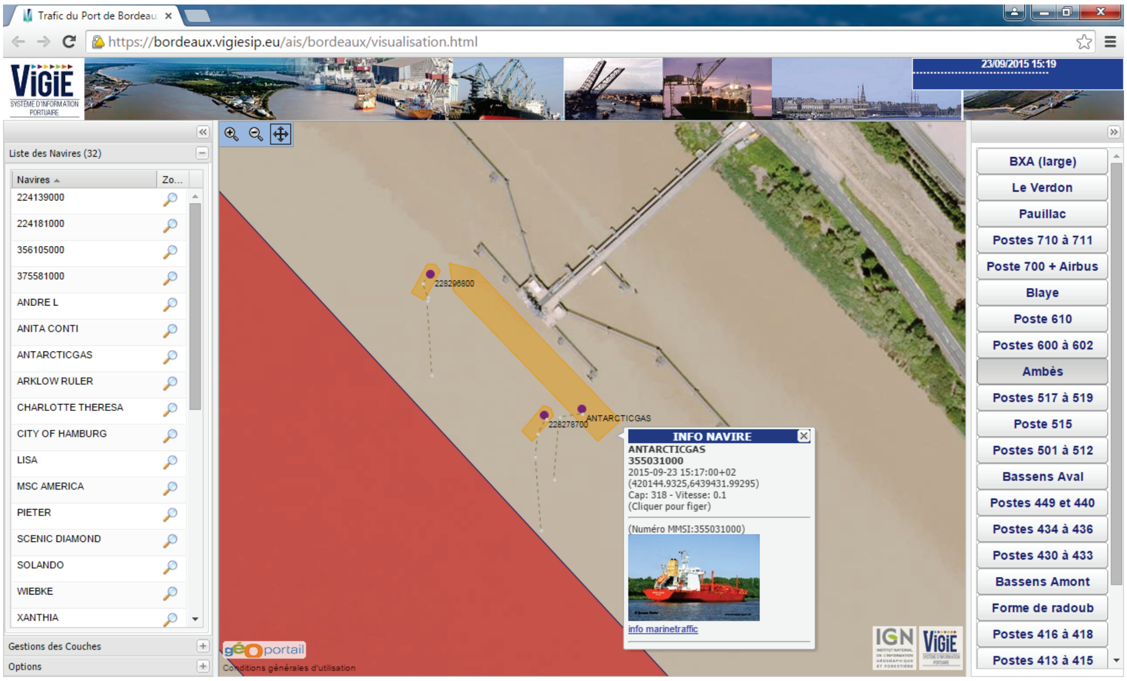


Figure 25. VIGIEsip AIS visualization

4.2.5. AISHub

AISHub[36] is a data exchange platform for sharing AIS data. It comprises a network of AIS receivers, mostly provided by enthusiasts. The data is shared as a raw NMEA 0183 data or via API. To get access to the network, one needs to connect an AIS receiver station to AISHub. Currently there are more than 28000 ships online, captured by more than 600 AIS receiving stations. Stations and coverage in EU area of interest, are presented in Figure 26.

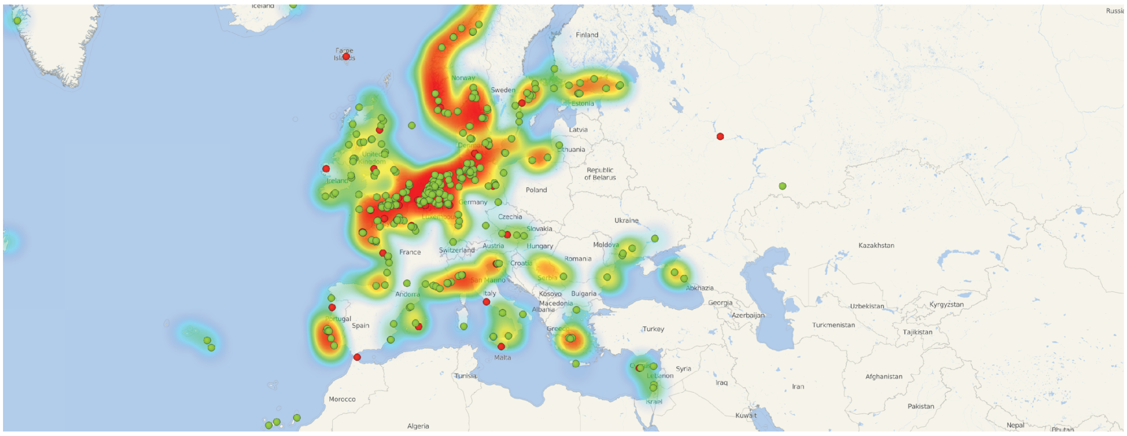


Figure 26. AISHub EU stations and coverage[37]

There is a lot of stations around Rotterdam area and English Channel, but not so many in the Mediterranean region. Port of Monfalcone is covered from Koper AIS receiving station, which also captures traffic in Port of Koper and Port of Trieste. Furthermore, there is a station in Piraeus, which captures ship traffic around the Port of Piraeus, while there is also a station available for Port of Bordeaux. On the other hand, there is no station available on AISHub that captures ship traffic around the Port of Thessaloniki.

[36] http://www.aishub.net/
[37] http://www.aishub.net/coverage

4.2.6. DEBS 2018 Challenge: Destination port and ETA prediction

DEBS 2018 Challenge was organized as part of the 12th ACM International Conference on Distributed and Event based Systems. The competition was organized as part of the EU H2020 BigDataOcean[38] and EU H2020 HOBBIT[39] projects. The goal of the competition was to predict destination port and estimated time of arrival (ETA) to the predicted port. For the competition, specific fields from the AIS data were used and the dataset was limited to the ship trips in Mediterranean Sea between ports as presented below.

- Alexandria
- Gemlik
- Palermo
- Augusta
- Genova
- Palma de Mallorca
- Barcelona
- Gibraltar
- Piombino
- Cagliari
- Gioia Tauro
- Piraeus
- Cartagena
- Haifa
- Port Said
- Castellon
- Ibiza
- Tarragona
- Ceuta
- Iskenderun
- Tobruk
- Damietta
- Livorno
- Tuzla
- Diliskelesi
- Marsaxlokk
- Valencia
- Formia
- Monaco
- Valletta
- Fos Sur Mer
- Nemrut
- Yalova

Figure 27 visualizes the data that was obtained from DEBS 2018 challenge organizers. This data consists of almost 400.000 reported AIS positions along the Mediterranean Sea. Certain shipping routes can be clearly seen. Most of the data was captured between March and May 2015.

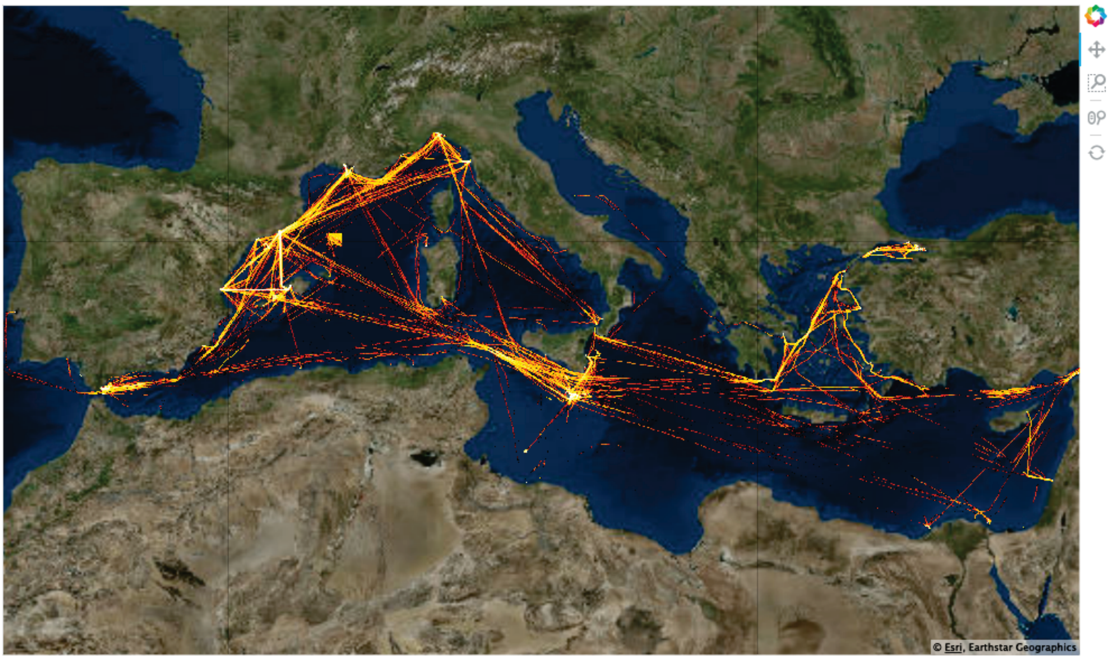


Figure 27. Visualization of DEBS 2018 data that was obtained from challenge organizers.

The competition was designed around HOBBIT platform such that data was transmitted to the competitors as stream of events (tuples), which simulated live stream of AIS data. Metrics for calculating prediction accuracy are presented in equations (1) and (2). The competitors are ranked according to this metrics and then the final

[38] http://www.bigdataocean.eu/site/
[39] https://project-hobbit.eu/

rank is computed as a weighted sum of rank and total runtime of the system, thus considering also the performance.

$$\text{accuracy} = \frac{\textit{The length of the last correctly predicted sequence}}{\textit{The total number of tuples in a trip}} \tag{1}$$

$$\text{accuracy} = \frac{\sum|\textit{predicted arrival time} - \textit{real arrival time}|}{\textit{The total number of tuples in a trip}} \tag{2}$$

Different solutions that used different Machine Learning approaches were presented. One of the approaches that was presented, used an ensemble learning based on Random Forests, GBDT (Gradient Boosting Decision Trees), XGBoost Trees and ERT (Extremely Boosting Decision Trees), to provide prediction for a destination port. For ETA prediction Feed-forward Neural Networks were used. In their evaluation, a 97% accuracy for destination port prediction was achieved and 90% accuracy (in minutes) for ETA prediction[40]. One the most interesting approaches, used sequence-to-sequence (seq2seq) approach, by representing spatiotemporal data in a grid-like structure, followed by encoding of each of the grid cells by a unique label, thus transforming stream of spatial data to word sequence[41]. One of the approaches, used Bayesian inference and heuristics[42], while an approach that uses a simple nearest neighbour search[43], was also presented.

[40] Oleh Bodunov, Florian Schmidt, André Martin, Andrey Brito, and Christof Fetzer. 2018. Real-time Destination and ETA Prediction for Maritime Traffic. In Proceedings of the 12th ACM International Conference on Distributed and Event-based Systems (DEBS '18). ACM, New York, NY, USA, 198-201. https://arxiv.org/pdf/1810.05567.pdf

[41] Duc-Duy Nguyen, Chan Le Van, and Muhammad Intizar Ali. 2018. Vessel Destination and Arrival Time Prediction with Sequence-to-Sequence Models over Spatial Grid. In Proceedings of the 12th ACM International Conference on Distributed and Event-based Systems (DEBS '18). ACM, New York, NY, USA, 217-220. https://www.researchgate.net/publication/325886915_Vessel_Destination_and_Arrival_Time_Prediction_with_Sequence-to-Sequence_Models_over_Spatial_Grid

[42] Hyungkun Jung, Kang-Woo Lee, Joong-Hyun Choi, and Eun-Sun Cho. 2018. Bayesian Estimation of Vessel Destination and Arrival Times. In Proceedings of the 12th ACM International Conference on Distributed and Event-based Systems (DEBS '18). ACM, New York, NY, USA, 195-197. https://www.researchgate.net/publication/325884194_Bayesian_Estimation_of_Vessel_Destination_and_Arrival_Times

[43] Valentin Roşca, Emanuel Onica, Paul Diac, and Ciprian Amariei. 2018. Predicting Destinations by Nearest Neighbor Search on Training Vessel Routes. In Proceedings of the 12th ACM International Conference on Distributed and Event-based Systems (DEBS '18). ACM, New York, NY, USA, 224-225. https://dl.acm.org/citation.cfm?doid=3210284.3220509

4.3. Predictions and analytics for PIXEL

As already presented in preceding sections, AIS data represents a rich source of data about maritime transport and its potential can be utilized to solve a plethora of different problems, in maritime domain, to improve safety, optimize logistics and reduce environmental impacts. The use cases with such data are unlimited. As already presented, this kind of data was also used outside maritime domain in ESS Net Big Data project, to improve the quality and internal comparability of existing statistics and to develop new statistical products[44].

In PIXEL project focus will be on following use cases regarding AIS data:

- Visualization and analysis of AIS data around the ports
- Port congestion indicators
- ETA prediction from AIS data & other sources

4.3.1.Visualization and analysis of AIS data around the ports

AIS data represents the source of data in maritime domain, which is spatially and temporally very rich and frequent. As described, besides spatial information, there is also information captured about the ship and the voyage (despite not always reliable). With spatial information, traffic flows can be visualized. An example of traffic density is presented in Figure 28. This data has been gathered from Marine Cadastre[45], which was earlier presented as one of the data sources of historical AIS data. The data represents AIS data collected around Port of Long Beach, in August 2017. Only the ships that are longer than 50m are visualized.

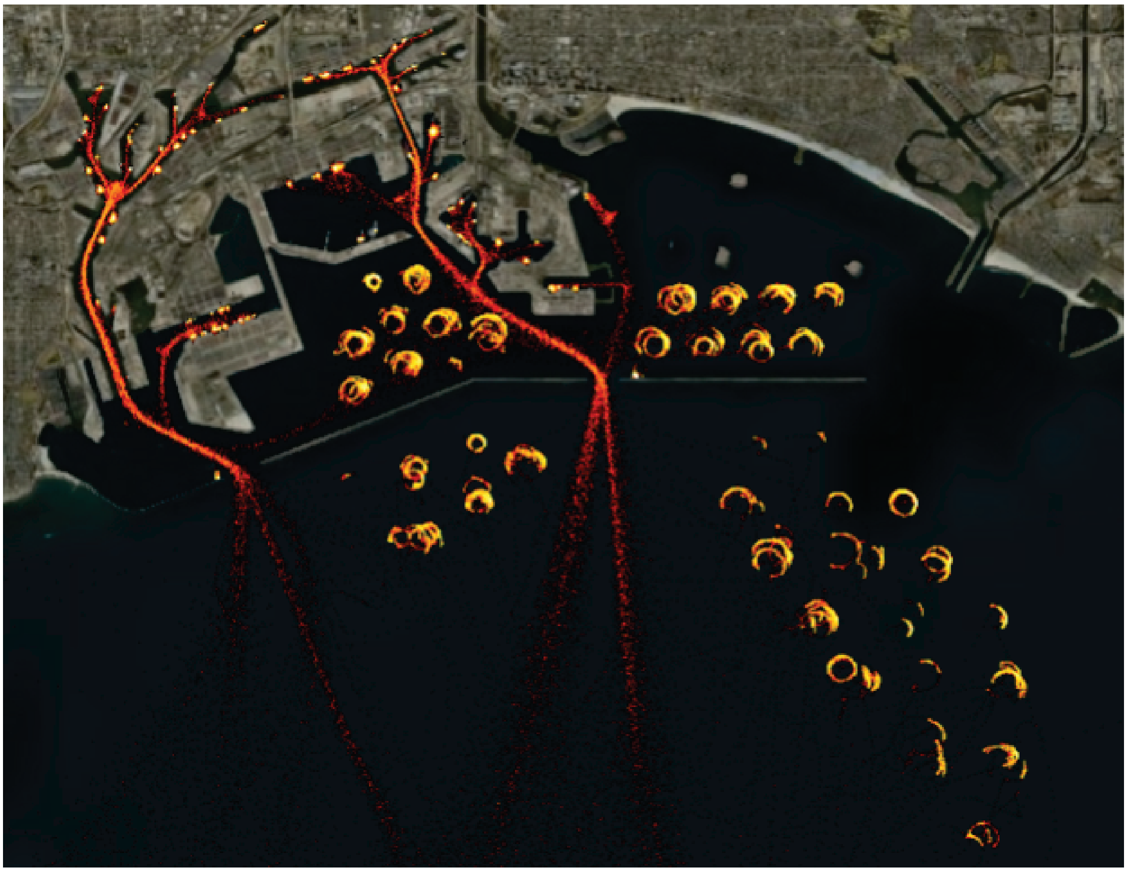


Figure 28. Visualization of ship traffic density in Port of Long Beach.

Traffic patterns can clearly be seen; especially interesting are the clearly visible positions, where ships are anchored. Similarly, visible are docking stations, where density is much higher. This kind of information can be used by other modelling tasks in WP4, to model environmental pollution more accurately (T4.4), while it

[44] https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php/WP4_Overview1

[45] https://marinecadastre.gov/ais/

will also be useful for more accurate noise pollution modelling (T4.4). Ship traffic spatial-temporal analysis can be also used to detect hotspots of high-speed ship sailing, hotspots where ship speed varies greatly, while traffic evolution over time, can also be investigated. AIS data will be collected from the ports that are part of PIXEL, from different AIS data sources that were presented and analysed.

Besides spatial information, AIS data contains also navigational status. With such information, statistics will be computed, depicting average waiting times for certain type of a ship, size or any other parameters that are captured in AIS messages. This will not only help to model environmental pollution more accurately, but on the same time, will represent a useful statistic for the port authorities. This data will be correlated with weather data, to see how weather influences the waiting time to enter the port.

Possibilities of using AIS data for port traffic analysis are limitless, though, there is not a lot of research work around, acknowledged by some of the authors that performed this kind of research[46,47]. Most of the scientific work is focused on major hubs, especially Port of Singapore. The goal is to apply such data analytics, to small and medium sized ports, which are part of PIXEL project.

4.3.2.Data analytics of the port area

As previously stated in this section, the application possibilities of AIS analysis and prediction are numerous, especially considering the lower degree of development, in comparison with other transport means. Detection of events, predictions and analytics for the port and immediately surrounding areas, is a promising field of research.

As shown in Figure 29, AIS information can be useful to compute some statistics, which would capture the congestion rate in the port area. The first research work that presented Port Congestion Indicators, used spatiotemporal mining algorithms[48]. Port Congestion Indicators (PCI), were computed from convex hull area, geohash area and average vessels proximity. Presented PCIs capture port spatial complexity, spatial density and port service time criticality. This PCIs were clustered using K-Means++, to be able to predict future port status. Such metrics could be informative for port authorities, to make informative decisions based on actual data. Different anchoring strategies can be put into place, ships can be anchored further away from the port and let in when congestion indicators improve, or actions can be made to speed up processing of the ships, which are docked.

AIS can also be used for detection of patterns, to classify into operational events and situations. In the figure below, trajectories of ships are shown, and some port events have been inferred. There is a wide range of events that can be detected exclusively using AIS data, and even more if the data is merged with port management systems. Some of the events relevant to PIXEL that can be detected are:

- **Anchor dragging:** When an anchored vessel drifts unexpectedly, dragging the anchor due to bad sea conditions.
- **Pilotage of a ship:** Marine pilot is dispatched to meet and lead the vessel at a designated compulsory pilotage area.
- **Towage:** Towing of ships and vessels by means of a small steamer, called a tug.
- **Fuel bunkering:** The act of refilling the fuel tank of a boat or a vessel.
- **Excess of speed in port area:** Cases that the vessel speed detected, has overpassed the allowed.

---

[46] Zhang, Liye & Meng, Qiang & Fang Fwa, Tien. (2017). Big AIS data based spatial-temporal analyses of ship traffic in Singapore port waters. Transportation Research Part E: Logistics and Transportation Review. https://www.sciencedirect.com/science/article/pii/S1366554516309516

[47] Meng, Q., Weng, J., & Li, S. (2014). Analysis with Automatic Identification System Data of Vessel Traffic Characteristics in the Singapore Strait. Transportation Research Record, 2426(1): 33–43. https://www.researchgate.net/publication/269140896_Analysis_with_Automatic_Identification_System_Data_of_Vessel_Traffic_Characteristics_in_the_Singapore_Strait

[48] Abualhaol, Ibrahim Y. et al. “Mining Port Congestion Indicators from Big AIS Data.” 2018 International Joint Conference on Neural Networks (IJCNN) (2018): 1-8. https://www.researchgate.net/publication/326271713_Mining_Port_Congestion_Indicators_from_Big_AIS_Data

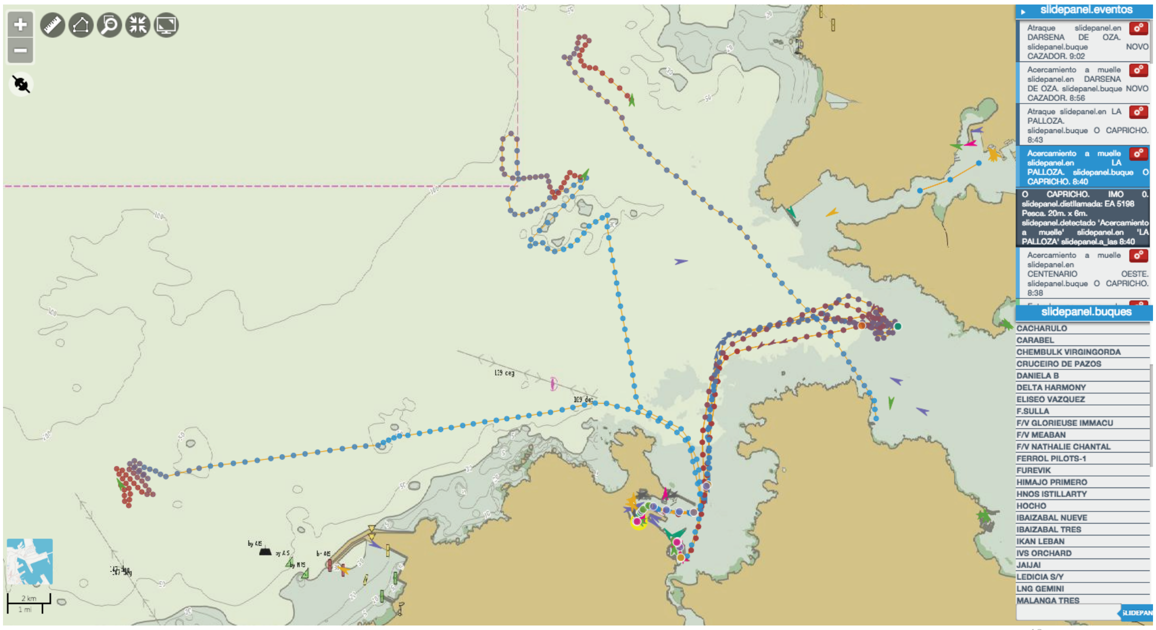


Figure 29. Detection of port events using CEP techniques in Posidonia Operations.[49]

Although, as shown in the picture, event and situations detection is something already done in commercial products, currently done with static algorithms or by using CEP (Complex Event Processing). The application of artificial intelligence to improve the accuracy and decrease the errors, is a relevant research field that will be done in PIXEL. Furthermore, these new techniques are expected to be able to detect more complex or subtle events, such as anchor dragging.

In PIXEL project, Port Congestion Metrics will be implemented and further extended, by additional metrics or modalities, such as weather and verified by the PIXEL ports, which would also give additional inputs that can be considered, to make such indicators more reliable.

As what regards event detection and classification, ML (Machine Learning) and AI (Artificial Intelligence) techniques will be applied to 1) detect the event: find a sequence of positions and data that represent a valid operation or event; 2) classify the event: decide what kind of event is being produced among a list of events; and 3) predict events from new data: using trained algorithms, predict the next step for a vessel.

4.3.3.ETA prediction

Accurate estimations about estimated time of arrival of the ship is very important, as it has influence on the whole logistics chain, not only on the operations in the port. ETA information is provided in AIS messages but is often unreliable, as it is provided manually by the crew and often not updated. EU ships are requested to send ETA information 72 hours before arrival to the port, which is often too soon and can be quite unreliable. The problem of ETA prediction, in connection with DEBS 2018 Challenge, has already been described. As acknowledged by a recent research work[50] in this domain, there is not a lot of research work around regarding ship ETA prediction, compared to other domains, where ETA prediction is useful (aviation, cars, road public transport, etc.). The scarce amount of research work is probably also due to data unavailability or its high price, especially for larger areas. Most of the research work in larger areas is performed by national agencies, while

[49] https://suppliers.porttechnology.org/prodevelop/products/9446/Port-Operation-Management-System/?pl=port-operation-management-system

[50] Alessandrini, Alfredo & Mazzarella, Fabio & Vespe, Michele. (2018). Estimated Time of Arrival using Historical Vessel Tracking Data. IEEE Transactions on Intelligent Transportation Systems. PP. 1-9. https://www.researchgate.net/publication/323233219_Estimated_Time_of_Arrival_using_Historical_Vessel_Tracking_Data

in the case of EU, most the research work in this domain is performed by JRC (Joint Research Center), which can obtain the data from national authorities or EMSA much easier.

As can be seen from DEBS 2018 visualization of the data in Mediterranean sea in Figure 27, or from a smaller area for the Port of Newark (Figure 30), where data is unfiltered and visualized for all the vessels larger than 200m, for a period of 14 days in August 2017, shipping routes can clearly be distinguished, especially the ones closer to the port. This kind of historical data can be used to predict how much time will be required for the vessel to arrive from a specific location, to the port area. Some of the methods that are used to address this kind of problems, were presented in DEBS 2018 section.

In PIXEL project focus will be on short term ETA prediction, that is at most a couple of hours distance from the port. This is due to availability of AIS data, for the development and deployment, as it cannot be realistically expected that small and medium sized ports will invest the amounts of money, which is required to obtain live AIS data for a larger area. Also, most of the research work is focused on long range ETA prediction, while according to the consortium's opinion, short term ETA prediction, deserves more attention, since it is more relevant for small and medium sized ports, with less traffic.

ETA prediction has recently become a very relevant topic in the maritime industry. Port of Rotterdam is one of the most active ports in this field and has also deployed solutions such as ShipTracker[51], which provides ETA estimation out of historical AIS data and other sources (e.g. weather information), with Machine Learning methods. Application was developed internally within Data Science department, which bigger ports can afford to have. ShipTracker tracks all the ships that will arrive to the port in the next 48 hours.

The goal of PIXEL here, is to help smaller and medium sized ports, utilize this kind of data, to make their operations more efficient and at the same time, more environmentally friendly. Indeed, this short-term ETA prediction could then be used in task 4.1 to help optimizing port activities and their impact on the environment.

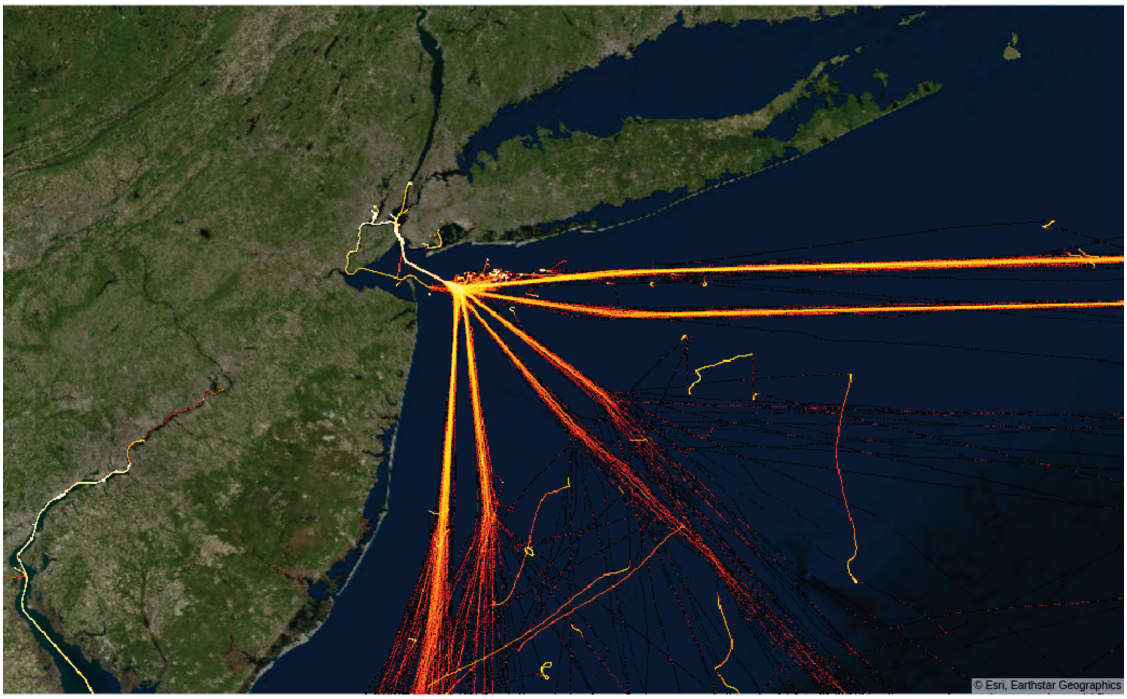


Figure 30. Visualization of ship traffic density entering Port of Newark, NY.

4.4. Initial results

The main problem addressed up to now, is the identification of potential data sources of AIS data. Three data sources that are openly available have been identified. AIS data from USA and Danish coastline, offer access to big amount of historical data, as well as live data with a slight delay, in the case of Danish Maritime Authority.

[51] https://shiptracker.portofrotterdam.com

Besides that, AISHub was identified as a potential source of live AIS data from a larger area, as a network of amateur AIS receiving stations. A 30-day trial account from AISHub was obtained and if it proves satisfactory, an AIS receiving station will need to be connected to the network, to have unlimited access to data. Unfortunately, there is no such open data available for the EU coastline by EMSA, as given by the USA agencies. In addition, data that was used in the competition for DEBS 2018 Challenge was obtained and explored. Hellenic Coast Guard was also asked, to provide data for Greek waters and currently a response is expected. Port of Bordeaux has also its own AIS receiver installed and connected to its own PMIS and will be made available.

All the data that was gathered up to now, was analysed and code was developed to visualize it. All the visualizations of AIS data presented in this deliverable, are the results of the initial work that focused on understanding AIS data and its potential use for PIXEL project.

4.5. Plan and future work

The plan for now is to focus on the data that is available, mainly from AISHub and the data from USA and Denmark. Methods that will be developed, will be general and applicable to any AIS data source. At the beginning, the focus will be placed on the task of AIS data visualization and analysis around the ports, to get a better understanding of the data before moving to other tasks. Some initial work regarding visualization was already performed; focus will now shift towards gathering insights from this data, as described in the preceding section. Besides operational insights that were described in preceding sections, such as computing average waiting times, speed maps, hotspots, etc. the robustness of AIS data captured from different sources will be estimated, especially how reliable are different fields in AIS messages that are captured and what is their value. AIS data will be available from the Port of Bordeaux (their own AIS receiver) and from Port of Piraeus and Monfalcone that are covered sufficiently enough by the AIS receivers in AISHub network.

In the next step, focus will be placed on Port Congestion Indicators as they represent a logical step forward, from a statistical analysis of AIS data from which, ideas for new metrics can be gathered and implemented. Different ports might have specific use cases, either for AIS data analysis or Port Congestion Indicators and will be discussed further.

Lastly, ETA estimation as the most challenging problem will be addressed, especially due to lack of wide scale data. As presented, focus will be on short term ETA prediction, to provide ports with more accurate prediction of berth time, to boost their efficiency and reduce environmental impacts. State-of-the-art methods will be implemented and tested on real life scenarios.

The main risk that can be identified is that of the data availability and its usefulness for a specific use case for the ports that are of interest. Particularly problematic might be ETA prediction, as AIS data is quite scarce on AISHub for the ports that we are interested in. Nevertheless, data that was collected from USA coastline, is of enough quantity and quality, to demonstrate the use case, while the developed methods will be general and applicable to any port.

5. Use of satellite imagery

5.1. Introduction

Observing Earth from space presents a new dimension of information that offers an unprecedented global view for various domains and industries. Earth observation capabilities were till recently, mostly in the domain of the governments that could afford to put satellites into space. Technological advancements have made possible to put more satellites into space, at a much lower cost and at the same time offering much higher spatial resolution and revisit times. This is becoming to be known as "democratization of the space" as more and more commercial providers are offering satellite imagery and at the same time governments or its institutions are opening its satellite constellations to the public, as open data. The most prominent example of open data satellite imagery providers is Copernicus program from ESA (European Space Agency), offering constellations of satellites (Sentinels) for different domains with a free and open data policy[52].

The biggest factor that is opening this field, for operational use cases and general use, is the exponential growth in number of satellites in space. The so called "SmallSat" revolution, has driven the prices of the satellites down by using commercial-off-the-shelf components in a much smaller form factor. Compared to traditional satellites that were in size of a bus, with a price tag usually in the hundreds of millions of euros, these small satellites can be as small as 10x10x10 cm (CubeSats) or even smaller. Their cost is a fraction of the money required by traditional satellites. The cost connected with launching the satellites, went also down by reducing the size and weight of the satellites and by emergence of ride sharing capabilities, offered by commercial providers, such as SpaceX. As of the end of 2018, there were 1900 satellites in space, 1200 of them used for EO (Earth Observation), while 3000 are to be launched between 2016 and 2022.[53]

One of the main impacts of the exponential growth in the number of satellites is that, they are usually part of a larger constellation, offering much higher revisit times compared to traditional satellites, which are usually deployed alone or in small constellations due to costs. Traditional satellites offer revisit time in the range of at least few days, or even weeks, compared to current or planned commercial constellations that are offering daily revisit times or even better.[54,55,56,57] This opens plenty of potential use cases that were before not practical or feasible, due to insufficient frequency or spatial resolution of the data.

The amount of imagery that is captured with such amount of satellites is vast and it is increasing exponentially. The data that is captured also needs to be analysed, as the focus is shifting towards obtaining useful operational insights, which can be gathered from satellite imagery, compared to obtaining raw imagery directly. These operational insights need to be found in an automatic way, using Machine Learning techniques and this AI driven data analytics is more and more recognized as the main value of the satellite imagery.

5.1.1. Satellite imagery in maritime domain

Due to its global coverage capabilities, satellite imagery presents a source of data that can be utilized in maritime domain for variety of use cases, especially with constellation of satellites that have multiple different sensors installed, for different purposes. Particularly used in maritime domain are SAR imagery (satellites with aperture radar imaging), due to its weather resistance (i.e. clouds and day & night operational capability). This kind of satellite imagery is used in different applications, some of them are listed below.

Ice monitoring: This kind of satellite imagery is used to produce navigational charts with iceberg data included, assuring safe travel in the ice-covered Arctic.

Ship monitoring: SAR imagery is predominately used for ship detection as a cost-effective solution compared to other means of monitoring.

[52] https://sentinel.esa.int/documents/247904/690755/Sentinel_Data_Legal_Notice
[53] https://www.geospatialworld.net/blogs/key-trends-in-earth-observation/
[54] https://www.planet.com/products/hi-res-monitoring/
[55] https://www.planet.com/products/planet-imagery/
[56] https://www.capellaspace.com/technology/
[57] https://www.blacksky.com/

Oil pollution monitoring: Oil spills are visible in SAR imagery as dark features and can be detected and monitored.

Some services such as oil spill detection, vessel detection and activity detection are operational and available by the European Maritime Safety Agency (EMSA) through CleanSeaNet[58] and CMS (Copernicus Maritime Surveillance) service[59]. Different commercial SAR and optical imagery providers are used for these services, as well as Copernicus Sentinel constellation, especially Sentinel-1 SAR satellites. Provided methods for added value products are mostly semi-automatic, requiring at least partial supervision of the operator. The CMS services are only available to national authorities and selected EU organisations working in the maritime domain and as such, not accessible to Port Authorities for use in their operations. These services are addressing the needs of national authorities to improve monitoring capabilities in the following areas:

- Fisheries control
- Maritime safety and security
- Law enforcement
- Marine environment (pollution monitoring)
- Support to internal organizations

5.1.2.Use for PIXEL

The goal of this task in PIXEL project is to provide some of the capabilities that satellite imagery is offering to the ports, to monitor their operations and improve their situational awareness. Our focus will be on monitoring ship traffic in and around the port. Port and bay area will be analysed in terms of number, types and sizes of the ships. Similar work was performed in ITEA APPS (Advancing Plug & Play Smart Surveillance) project[60] where cameras were installed to detect and classify ships in the port area. The goal is to advance this, by using satellite imagery instead, providing Port Authorities new dimensions of gathering operational insights. Satellite imagery offers historical data depending on satellite constellation at least 5 years back. ESA Copernicus satellite constellation data is available and is free-of-charge, including historical data from 2014 onwards and can be used for commercial purposes.

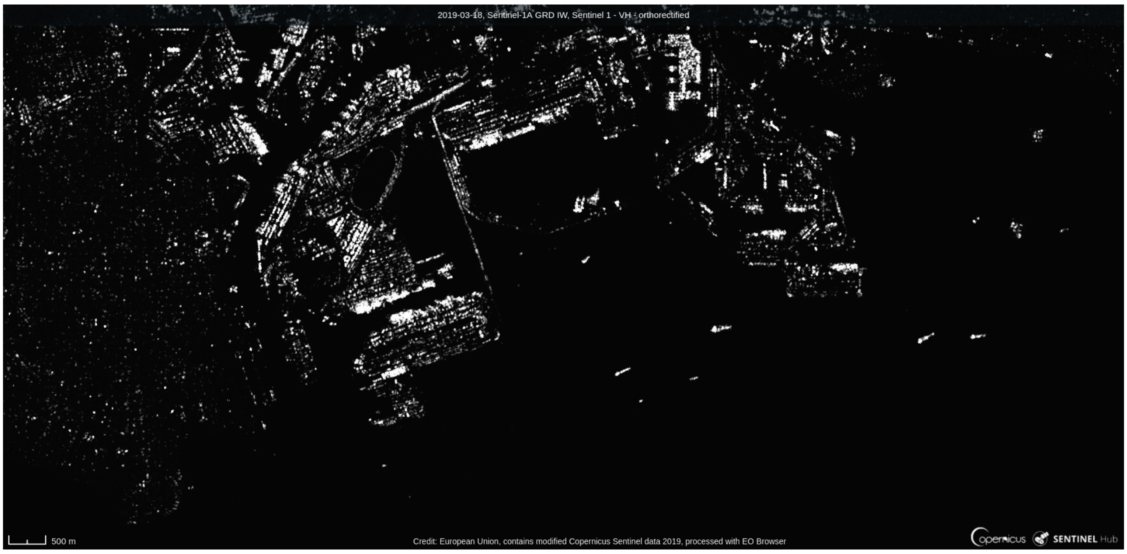


Figure 31. Sentinel-1 SAR, VH polarization, imagery of Port of Long Beach, USA.

Figure 31 presents an example of ESA Copernicus Sentinel-1 SAR imagery that is predominately used for maritime surveillance. This example shows the port area of Port of Long Beach, USA to have a comparison to some of the commercial satellite providers that offer open access to their higher resolution imagery over

[58] http://www.emsa.europa.eu/csn-menu.html
[59] http://www.emsa.europa.eu/copernicus/item/2880-copernicus-maritime-surveillance-service-overview.html
[60] https://itea3.org/project/apps.html

California and will be presented in the next section. Popularity of SAR imagery in maritime domain is due to its independence to weather conditions and night and day operational capabilities. SAR imagery has its limitations, especially near the coastline and in harbour areas, due to other objects present and their spatial density, causing SAR backscatter and consequently, it is hard to distinguish different objects. There is also speckle noise that is present in SAR imagery, due to coherent interaction of the transmitted microwave with the targets.[61] Overall, SAR imagery for ship detection is mostly used on open-seas, as it remains a challenging research problem to detect ships in harbour area[62,63,64].

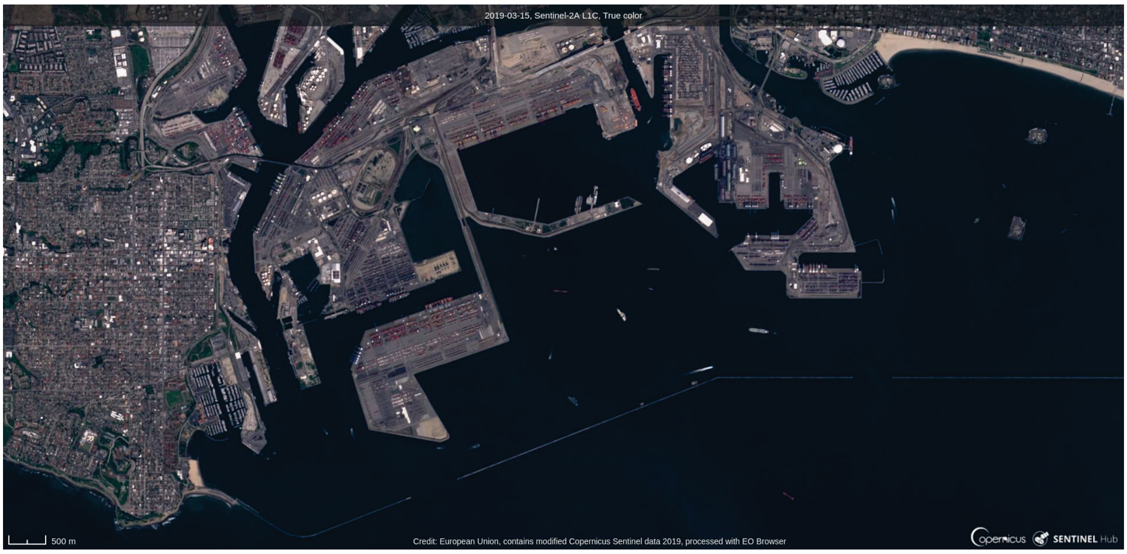


Figure 32. Sentinel-2 optical imagery of Port of Long Beach, USA.

Figure 32 presents same area with optical imagery taken from ESA Copernicus Sentinel-2 satellite. This kind of imagery is much better for analysis of port traffic. Optical imagery will thus represent, the main input for this task. Optical imagery is utilized also by Copernicus Maritime Surveillance (CMS) at EMSA. They mainly use VHR (very high resolution) imagery (i.e. < 1m) from commercial providers, with semi-automatic[65] operational tools for ship detection. ESA Sentinel-2 imagery (10m resolution) is not utilized[66]. According to the report, optical imagery is not processed on massive scale, as between 2013 and 2015 service was requested around 100 times with 800 vessels being detected.[67]

Satellite imagery can also present input for some of the analytics that uses AIS data and was presented in preceding chapter. Compared to costly AIS data, satellite imagery is provided free-of-charge (Sentinels), while

---

[61] https://crisp.nus.edu.sg/research/tutorial/sar_int.htm
[62] Li Zhi, Qu Changwen, Zhou Qiang, Liu Chen, Peng Shujuan and Li Jianwei, "Ship detection in harbor area in SAR images based on constructing an accurate sea-clutter model," 2017 2nd International Conference on Image, Vision and Computing (ICIVC), Chengdu, 2017, pp. 13-19. https://ieeexplore.ieee.org/document/7984450
[63] W. Ao, F. Xu, Y. Li and H. Wang, "Detection and Discrimination of Ship Targets in Complex Background From Spaceborne ALOS-2 SAR Images," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no. 2, pp. 536-550, Feb. 2018. https://ieeexplore.ieee.org/abstract/document/8259236
[64] W. Ao and F. Xu, "Robust Ship Detection in SAR Images from Complex Background," 2018 IEEE International Conference on Computational Electromagnetics (ICCEM), Chengdu, 2018, pp. 1-2. https://ieeexplore.ieee.org/abstract/document/8496647
[65] R. Muller, M. Berg, S. Casey, G. Ellis, C. Flingelli, R. Kiefi, A. Kornhoff, K. Lechner, T. Reize, G. S. Mattyus, E. Schwarz, E. Simon, and A. Twele, Optical Satellite Services For EMSA (Opsserve)- Near Real-Time Detection Of Vessels And Activities With Optical Satellite Imagery, ESA Living Planet Symposium, ESA Special Publication, vol. 722, December 2013, p. 309. http://adsabs.harvard.edu/abs/2013ESASP.722E.309M
[66] http://emsa.europa.eu/news-a-press-centre/external-news/item/3025-copernicus-maritime-surveillance-product-catalogue.html
[67] Enrico Stein; Robin Nitsche; Konstanze Lechner; Egbert Schwarz; Matthias Berg; Ralph Kiefl; Rupert Müller, Operational Service for the Detection of Vessels and Maritime Activities with Optical Satellite Imagery in Near Real Time – Experiences and Future Aspects Proceedings of 1st International Electronic Conference on Remote Sensing. https://sciforum.net/paper/view/conference/2824

historical data is also available. The only downside is time resolution which will be presented in detail in the next section, but bigger commercial constellations of satellites are emerging, promising even hourly revisit times. Some of the tasks such as discovering traffic patterns or analyse of long-term trends in the port traffic do not require high frequency, as is complemented with enough historical data.

In the next sections satellite imagery data sources will be presented, as well as some of the data sources to be used in developing the methods for ship detection. Novel methodology to develop state-of-the-art methods and the use for PIXEL project will also be presented, along with the results of our initial work. At the end, plans for future work are presented.

5.2. Satellite imagery data sources

As presented in the introduction, the number of satellites in space is growing exponentially as well as the number of different commercial and institutional providers. Different providers offer different satellite constellations with different sensors installed, revisit times and spatial resolutions. In this section we present ESA Copernicus Sentinel constellation, as an institutional provider of free-of-charge satellite imagery, with different sensors and Planet Labs as a commercial provider of higher resolution imagery with daily revisit times.

5.2.1.ESA Copernicus Sentinel constellation

ESA Sentinels are constellation of satellites with different sensors installed, which primary role is to ensure availability of Earth observation data for environmental and security services. The program is managed by the European Commission and implemented by European Space Agency (ESA), member states and other European agencies that rely on space data. The data from ESA is publicly available, including for commercial use. There are five Sentinels missions, each mission is a constellation of two or more satellites in the same orbital plane, to have satisfying revisit time and coverage. Bellow we briefly present all the missions.

- **Sentinel-1:** This satellite has Synthetic Aperture Radar (SAR) installed, which is particularly well suited to monitor wide maritime areas. Radar beams are transmitted and reflected pulses are captured. Radar pulses are reflected by waves in the sea, as well as objects such as ships. SAR imagery has operational capabilities day & night, independent of weather conditions such as clouds. EMSA is using Sentinel-1 imagery to detect oil spills, as well as to detect ships on open seas.[68]
- **Sentinel-2:** This satellite constellation has multi-spectral imaging device installed, that can observe the Earth in 13 different spectral bands which offers operational data for different use cases, primarily for LULC (Land Use and Land Cover) analysis. Different bands also offer possibility to monitor crop growth by monitoring different biogeophysical variables, like NDVI (Normalized Difference Vegetation Index) or LAI (Leaf Area Index). Sentinel-2 imagery in RGB (Read Green Blue) bands is underutilized in maritime domain and can be of use in harbour areas.
- **Sentinel-3:** The main objective of Sentinel-3 is to observe oceans from space with different sensors installed to measure sea-surface topography, height, wave height. Measuring ocean and land-surface temperature, colour, sea water quality and pollution etc.[69]
- **Sentinel-4, Sentinel-5:** The mission of Sentinel-4/5 constellation is to provide remote sensing data about gas concentrations and aerosols in the atmosphere to support air-quality operational services.[70,71]

The most important missions for maritime monitoring are Sentinel-1 and Sentinel-2 missions. Sentinel-1 mission is already utilized by EMSA for monitoring maritime traffic and pollution. The goal of this task for PIXEL project is to monitor harbour and port areas which is not feasible by Sentinel-1 SAR mission, as presented in the introduction. Our plan is to utilize Sentinel-2 imagery that is underutilized in maritime domain.

Sentinel-2 mission offers global coverage of all the land territory, as well as some of the maritime regions. All the coastlines are covered globally (20 km from the shore), as well as all inland water bodies, all closed seas and the whole Mediterranean Sea and is as such well suited to monitor coastal and harbour regions. Sentinel-2

[68] http://earth.esa.int/seasar2010/5_chintoa_uta.pdf
[69] https://sentinel.esa.int/web/sentinel/missions/sentinel-3/overview/mission-summary
[70] https://earth.esa.int/web/guest/missions/esa-future-missions/sentinel-4
[71] https://earth.esa.int/web/guest/missions/esa-future-missions/sentinel-5

mission revisit frequency at the equator is 5 days, with two satellites. Sentinel-1 mission revisit frequency is similar with 6 days. Revisit rate is significantly higher at higher latitudes. Revisit rates and coverage for Mediterranean Sea are presented in Figure 33, produced by JRC paper on mass processing of Sentinel-1 imagery for maritime surveillance[72]. Figure 33 clearly shows that revisit rates are much higher at this latitude as on average every 2-3 days, there is a new imagery available.

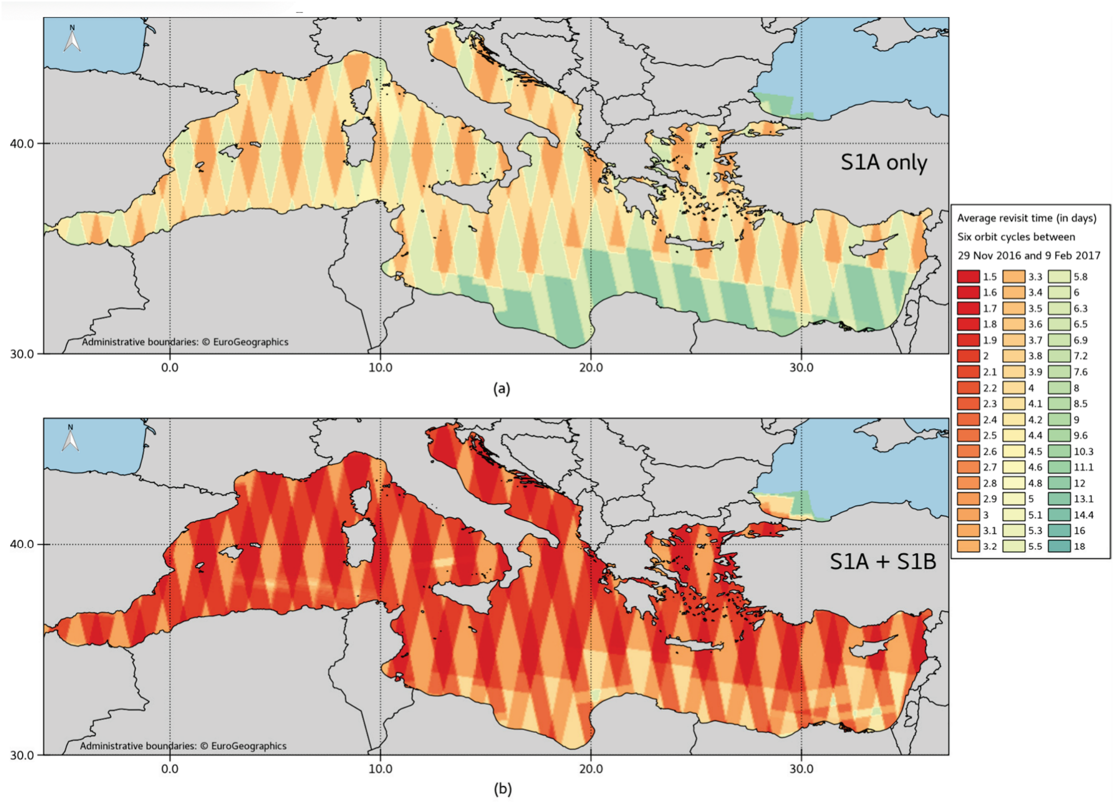


Figure 33. (a) Revisit rate of Sentinel-1 when only 1 satellite is used. (b) Revisit rate when both satellites are working in a constellation. Sentinel-2 revisit rates are similar.[73]

Besides revisit times, spatial resolution is also important. Sentinel-2 imagery offers high resolution imagery with 10m resolution, which is minimum size of the object to be recognized is 10m. Sentinel-2 resolution cannot be directly compared to Sentinel-1 due to different types of sensors but the resolution of Sentinel-1 is around 20m.

Sentinel imagery can be accessed through ESA portals[74], or through API access. More intuitive access is available through 3rd party providers such as Sinergise Sentinel Hub[75] that is also integrated to open-source Earth observation library eo-learn[76].

[72] Carlos Santamaria, Marlene Alvarez, Harm Greidanus, Vasileios Syrris, Pierre Soille, and Pietro Argentieri, Mass processing of sentinel-1 images for maritime surveillance, Remote Sensing 9 (2017), no. 7. https://www.mdpi.com/2072-4292/9/7/678
[73] Source: https://www.mdpi.com/2072-4292/9/7/678
[74] https://scihub.copernicus.eu/
[75] https://www.sentinel-hub.com/
[76] https://eo-learn.readthedocs.io/en/latest/

5.2.2.Planet Labs satellite imagery

Planet Labs is a commercial provider of satellite imagery. They emerged from "SmallSat" revolution, as the biggest operator of satellites in space. Their main goal (or "mission 1") was to image the whole Earth every day. That goal was reached on July 14th 2017[77].

Two constellations of satellites are operating:

- **PlanetScope Dove:** CubeSat 3U, form factor (10 x 10 x 30 cm, 4kg), constellation of 120+ satellites able to image entire Earth's landmass every day, at 3m and 5m resolution, at blue, green, red and NIR (Near-Infrared imaging) spectral bands.
- **SkySat:** Constellation of 15 very-high-resolution (VHR) satellites, offering sub-meter resolution imagery and video that represents the smallest satellites in space, to offer VHR capability. This represents the largest constellation of VHR satellites providing satellite imagery up to twice daily.

Compared to ESA Copernicus, Planet Labs offers better spatial resolution and revisit times. They also offer open access to California dataset[78], which will be utilized in PIXEL project, to develop and test the methods presented in the next section. Increased capabilities offered by PlanetScope constellation, complementing the methods developed on ESA Sentinel imagery, will be utilized. There are 11 public ports in California, including 3 megaports[79], which makes this freely accessible data even more interesting to develop Earth Observation (EO) analytics for the ports. Planet Labs pricing is not publicly available, but studies have been done in agriculture, making it the most cost-effective solution among commercial providers.[80] Daily revisit times and increased resolution are unique features, not available by public (free-of-charge) providers and the benefits of that will be investigated on a publicly available California dataset. Figure 34 and Figure 35 present the difference between Sentinel-2 and PlanetScope imagery over Port of Long Beach, CA.

Figure 34. Sentinel-2 satellite imagery (10m) over Port of Long Beach, CA.

[77] https://www.planet.com/pulse/mission-1/

[78] https://www.planet.com/products/open-california/

[79] http://www.dot.ca.gov/hq/tpp/offices/ogm/seaports.html

[80] https://www.researchgate.net/publication/326417596_Benchmark_of_Satellites_Image_Services_for_Precision_Agricultural_use

Figure 35. PlanetScope satellite imagery (3m) over Port of Long Beach, CA.[81]

PlanetScope imagery can complement ESA Sentinel-2 in at least two ways. Increased spatial resolution (3m vs. 10m), offering additional opportunities for new use cases, such as ship classification, which is limited with Sentinel-2 imagery, especially for smaller cargo ships. Secondly, daily revisit times offers 2-3 times better time resolution for more reliable statistics.

5.3. Predictions and analytics for PIXEL

Advancements in Earth Observation (EO) capabilities, together with advancements in AI domain, especially with the advent of Deep Learning, has opened new ways to gather operational insights from remote sensing data. The goal of this task in PIXEL project is to utilize all these advancements for the benefit of the ports but not limited to them, as the proposed methods are widely applicable to general monitoring of oceans.

The focus of this task will be the use of optical imagery, which is particularly underutilized in maritime domain and particularly well suited to monitor harbour areas. It also offers additional capabilities such as the ability to classify ships or to gather any additional contextual information, not feasible by using SAR imagery. Medium resolution imagery, of which Copernicus Sentinel-2 is particularly well suited is to be utilized, however until recently[82], there was no research work, which utilized it in any way for ship detection. There is no work around, utilizing state-of-the-art Deep Learning based methods for ship detection using Sentinel-2 imagery, which is especially challenging in port area, not only for medium resolution but also for VHR imagery.

The use of state-of-the-art methods is mostly limited, due to limited availability of annotated data, which is hard to obtain and needed to train the methods. This is especially the case for medium resolution imagery, for which there is almost no annotated datasets of enough size for ship detection. The goal is to fill this gap, with application of existing very-high-resolution (VHR) imagery to medium resolution imagery and by a novel approach of data fusion with AIS data.

The results of the application of these methods to satellite imagery around the ports are positions of the ships, types of ships, sizes of the ships and any other contextual information that can be gathered. The most important are locations of the ships, which can be utilized similarly as AIS data presented in preceding chapter for all sorts of predictive tasks. Compared to AIS data, Sentinel-2 imagery is free-of-charge, while historical data is also available for 5 years back. Compared to AIS data, information about the ships gathered from satellite imagery is obtained in a non-cooperative way and can merged with AIS data, serve also as tool to monitor proper use of

[81] Open California Satellite Imagery 2019 Planet Labs Inc. licensed under CC BY-SA 4.0

[82] Kurekin, Andrey A., et al. "Operational monitoring of illegal fishing in Ghana through exploitation of satellite Earth observation and AIS data." Remote Sensing 11.3 (2019): 293. https://www.mdpi.com/2072-4292/11/3/293

AIS transmitter. This can improve safety on sea and in port areas, especially reducing the risk of collision and possible environmental pollution.

Figure 36 presents ship detection density map from Sentinel-1 SAR imagery.[83] SAR imagery is predominately used for ship detection on open seas. From density map, frequent shipping routes can clearly be seen. SAR imagery is not suitable to detect ships near the coastline or in the ports. In this work, they have buffered the coastline by 250m, which makes it unusable to detect ships in port area. Compared to large scale analysis, traffic analysis in harbour/port areas is important for PIXEL and similar work will be performed in those areas.

In the next two subsections, the main sources of annotated data for ship detection are briefly presented, as well as a novel approach that will be used to create large scale dataset for ship detection from medium resolution imagery.

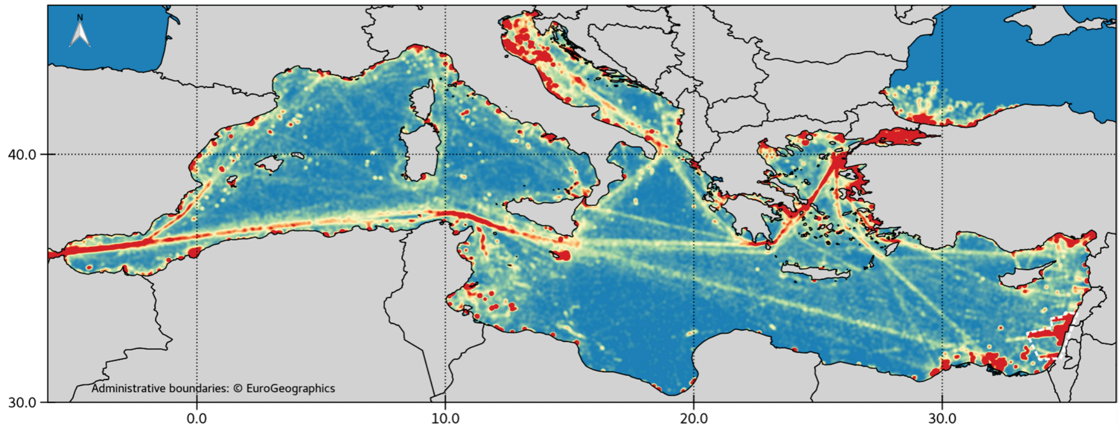


Figure 36. Density map of reliable ship detections (608.326) on Sentinel-1 SAR imagery from 3. October 2014 to 30. September 2016.[84]

5.3.1.Datasets for ship detection

In this subsection, some of the datasets that can be used to train state-of-the-art methods to detect ships are briefly presented. Datasets are mostly part of research papers or different challenges regarding EO analytics.

- **Kaggle Airbus ship detection challenge:**[85] Airbus has in 2018 organized a challenge, to detect ships on Kaggle out of their satellite imagery. This presents the biggest dataset of annotated ship locations (i.e. masks for rotated bounding boxes are provided). The dataset contains 150.000 JPEG images of 768x768 dimensions extracted from SPOT satellite which has 1.5m resolution. Most of the images do not contain ships but negative examples. There are 81723 annotated ships, but most of them are small ships, though, there is also quite a bit of cargo ships present that can be filtered out. After filtering, around 30.000 remain which is still enough, compared to other datasets. An example from the dataset is presented in Figure 37, together with provided annotations.
- **HRSC 2016:**[86] This dataset contains annotations of ships on open-sea as well as in harbour area. Images were downloaded from Google Earth with resolution ranging from 0.4m to 2m. Dataset contains 1061 images, 70 sea images with 90 samples of annotated ships and 991 sea-land images with 2886 annotated ships. Besides bounding box information rotated bounding box annotations are also provided and detailed segmentations masks for some of the data.

[83] Carlos Santamaria, Marlene Alvarez, Harm Greidanus, Vasileios Syrris, Pierre Soille, and Pietro Argentieri, Mass processing of sentinel-1 images for maritime surveillance, Remote Sensing 9 (2017), no. 7. https://www.mdpi.com/2072-4292/9/7/678

[84] Source: https://www.mdpi.com/2072-4292/9/7/678

[85] https://www.kaggle.com/c/airbus-ship-detection

[86] Liu, Z.; Yuan, L.; Weng, L. and Yang, Y. (2017). A High Resolution Optical Satellite Image Dataset for Ship Recognition and Some New Baselines. In Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods - Volume 1: ICPRAM, pages 324-331. http://www.scitepress.org/Papers/2017/61206/61206.pdf

- **DOTA:**[87] DOTA dataset is one of the first general purpose remote sensing object detection datasets, with enough diversity of classes and instances per class. The dataset consists of 2806 images with 188.282 annotated object instances in 15 different categories, including ship class.
- **xView dataset:**[88] This dataset is one the largest datasets available for remote sensing object detection, with over 1 million annotated objects across 60 classes in over 1400km^2 of imagery. Imagery is taken from WorldView-3 satellite which offers 30 cm resolution. This dataset contains different classes of annotated vessels that can be utilized for ship detection.

Figure 37. An example taken from Kaggle Airbus dataset, visualized together with the provided ground truth annotations.

5.3.2.Use of AIS data

The datasets mentioned in preceding subsection include plenty of annotations for ships in terms of location, but the dataset include imagery from different sensors, mostly from very-high-resolution (VHR) imagery. The goal is to develop the methods for medium resolution imagery, which is either free (e.g. Sentinel-2) or provided much cheaper than VHR imagery, due to large constellation of satellites. One way that existing VHR datasets can be utilized is, to artificially decrease their resolution, which will be explored and utilized.

Besides utilizing known datasets that were designed for VHR imagery, new dataset for medium resolution imagery, for the areas of interest, can be constructed with the help of AIS data in a novel way. As presented in the chapter about AIS data, AIS messages are sent every couple of seconds, containing plenty of information, not only about the location but also other characteristics about the ship, such as type of the ship, dimensions, etc. With the help of AIS, large scale dataset can be constructed for any kind of imagery, complementing existing location-based annotations with contextual information about the ship, present in AIS messages. That kind of a dataset will not only help to boost detection capabilities and accuracy but will also provide additional capabilities such as ship classification, which is currently not available on existing datasets. The use of AIS data for a construction of such a dataset is especially useful for harbour and port areas, where detection is the most challenging problem but merging AIS data with satellite data is the most accurate, due to slow, docked or moored vessels.

[87] Xia, Gui-Song et al. "DOTA: A Large-Scale Dataset for Object Detection in Aerial Images." 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (2018): 3974-3983. https://captain-whu.github.io/DOTA/

[88] Lam, Darius et al. "xView: Objects in Context in Overhead Imagery." CoRRabs/1802.07856 (2018). https://arxiv.org/pdf/1802.07856.pdf

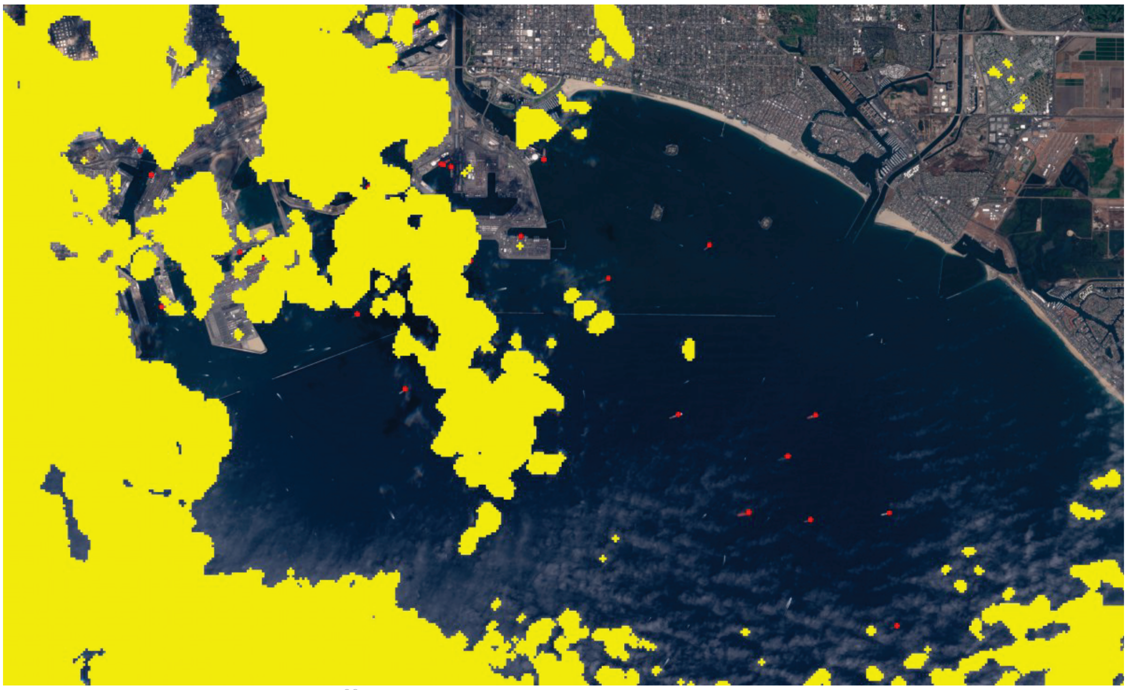


Figure 38. Sentinel-2 imagery[89] over Port of Long Beach, merged with AIS data and detected clouds.

Figure 38 and Figure 39 represents satellite imagery merged with AIS data for Sentinel-2 and PlanetScope imagery respectively. The satellite imagery is taken around Port of Long Beach, CA due to availability of Open California Planet imagery and AIS data from Marine Cadastre[90]. It can be clearly observed that GPS coordinates obtained through AIS match ship locations in satellite imagery. In this way, large scale dataset can be constructed to detect ships in port area, combined with additional information, which is present in AIS messages and can be used to classify ships, compare measurements of the length of the ships from satellite imagery to the actual ones captured in AIS messages etc. The data can be captured from different locations, taking into consideration the type of cargo terminals and consequently ships of interest. As shown in Figure 38, cloud detection[91] can also be utilized, to fully automate the procedure and not capture ships under the clouds that are still emitting AIS messages. Combined with free AIS data from AISHub, presented in AIS chapter and global availability of free Sentinel-2 imagery, this can be scaled to any port in the world, regardless of their size in a cost-efficient way.

[89] Contains modified Copernicus Sentinel data 2019, obtained with Sentinel Hub licensed under CC BY-NC 4.0
[90] https://marinecadastre.gov/ais/
[91] https://github.com/sentinel-hub/sentinel2-cloud-detector

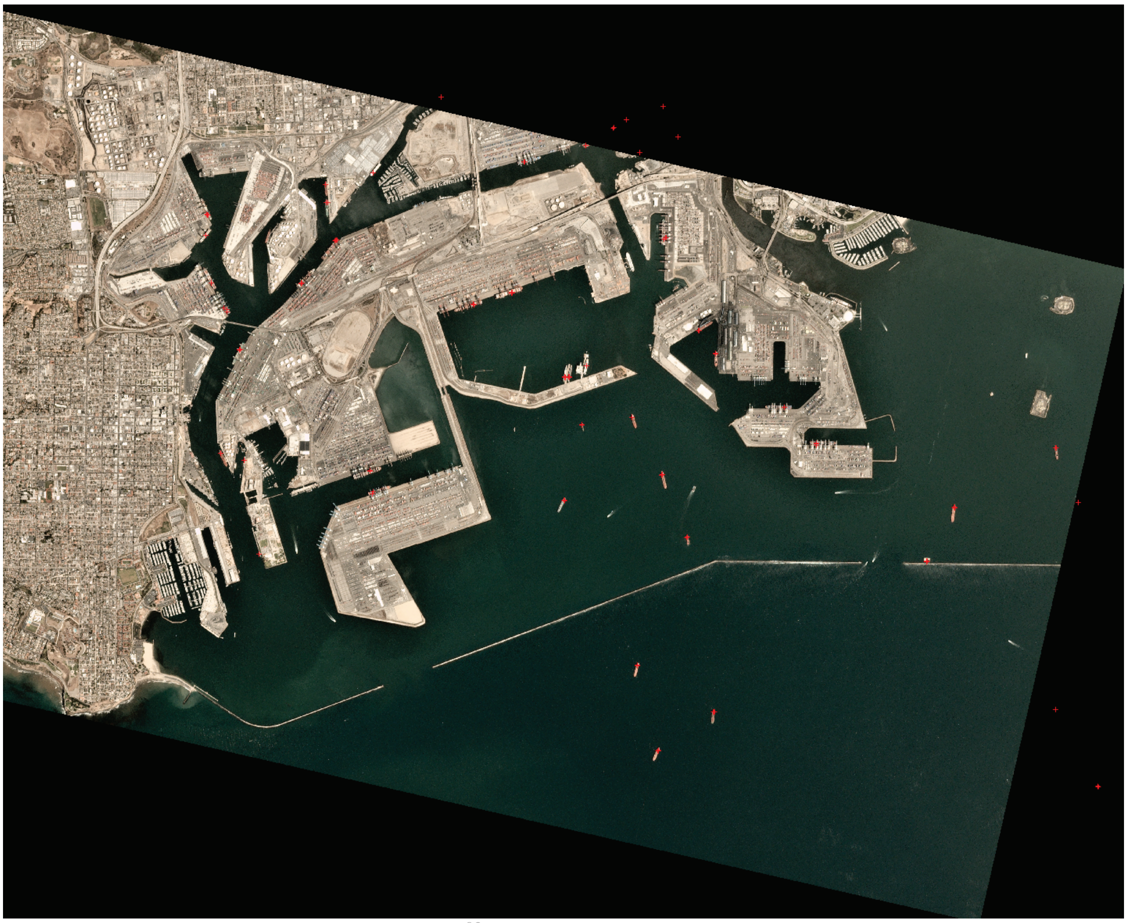


Figure 39. PlanetScope imagery[92] over Port of Long Beach, merged with AIS data.

5.4. Initial work

Most of the initial work was performed around obtaining the data from different satellite imagery providers, especially open data, that is freely available either for commercial or research purposes. ESA Copernicus and Planet OpenCalifornia open data sources were identified and described in this chapter. Ship detection is one of the most important problems identified as it is needed for all the proposed added-value products on top of satellite imagery. All the existing datasets that contain different vessel classes that were presented, were obtained and explored in terms of their usability. Novel approach towards creating large-scale dataset from AIS data was also explored and prototypes implemented, as presented in the preceding section.

To process ESA Sentinel imagery, open-source library eo-learn[93] was identified, that makes retrieving and preparing the data for processing extremely easy. As presented in section 3, density maps can be computed, to obtain historical spatial distribution of the vessel in port or larger areas. Pipeline for gathering and processing ESA Sentinel imagery was prepared with eo-learn library. Figure 40 shows an example of this processing pipeline. GIS was obtained for Trieste bay area[94], which was sliced into regions, that are manageable to process, and imagery was obtained for each of the patches for the selected time frame (e.g. 1 month of data). Different filters can be added to select only the imagery that is not covered with clouds or to mask the clouds to obtain as much usable imagery as possible. This processing pipeline can be implemented for any region in the world that is covered by satellite imagery. Developed ship detection methods will be deployed on such imagery, to process large scale imagery in both spatial and time domain. Access to Planet satellite imagery was also explored through their APIs and Python client library[95]. Similarly, as with Sentinel imagery, prototype of the pipeline for

[92] Contains Modified Open California Satellite Imagery ©2019 Planet Labs Inc. licensed under CC BY-SA 4.0
[93] https://github.com/sentinel-hub/eo-learn
[94] http://www.marineregions.org/gazetteer.php?p=details&id=3385
[95] https://github.com/planetlabs/planet-client-python

gathering the data was developed and tested out. All the imagery in preceding section about AIS data was obtained through this pipeline and the methods for merging AIS data with both providers of satellite imagery was also developed.

Different state-of-the-art methods for object detection were explored, such as Faster-RCNN (Region-Convolutional Neural Network), Mask-RCNN, and RetinaNet that are all implemented in Facebook Detectron[96]. Selected methods will be modified for the domain of ship detection, together with ship classification for the higher resolution imagery provided by Planet OpenCalifornia. Initial tests were performed with a modified Faster-RCNN method on Sentinel-2 imagery to test the feasibility of using already annotated datasets on VHR imagery on medium resolution imagery, showing promising results.

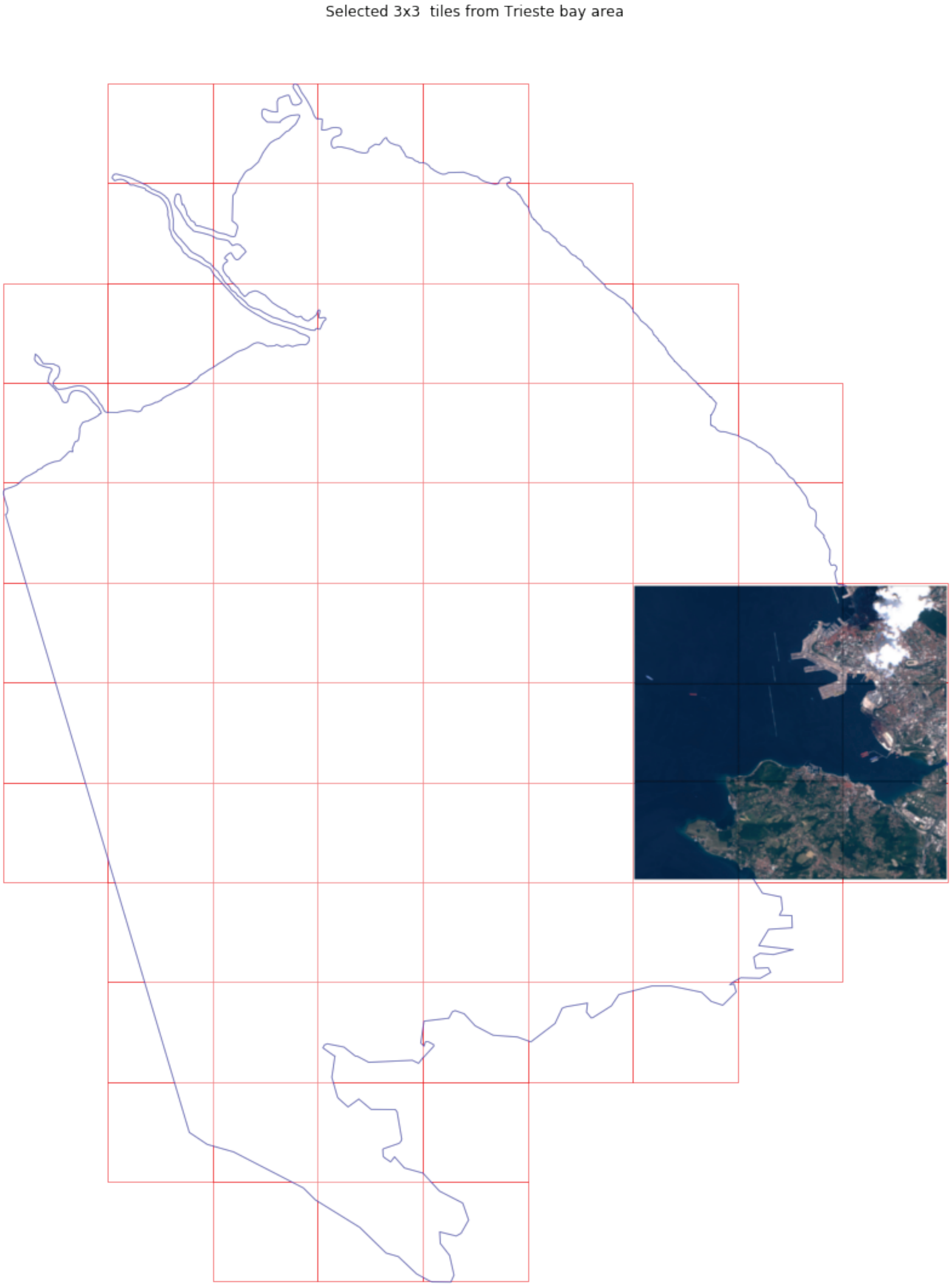


Figure 40. An example of Trieste bay area divided into patches out of provided GIS polygon together with some of the obtained patches around Port of Trieste. This pipeline was prepared using eo-learn library.[97]

[96] https://github.com/facebookresearch/Detectron

[97] Contains modified Copernicus Sentinel data 2019, obtained with Sentinel Hub licensed under CC BY-NC 4.0.

5.5. Plan and future work

In this chapter the use of satellite imagery for the ports, which presents a novel source of data that can be utilized by the ports and other entities interested in port operations in general, is presented. The use of satellite imagery is becoming more and more useful, especially because of the increasing number of companies, as well as institutional providers, offering satellite imagery with increasing spatial resolution, as well as revisit times. Two main providers of satellite imagery that are offering at least some of the data free-of-charge, which will be utilized in PIXEL project, have been identified. ESA Copernicus constellation is offering satellite imagery worldwide under completely open policy, including commercial use. Planet Labs Inc., as a commercial provider is offering open policy access to its PlanetScope constellation imagery over California.

The main problem that needs to be addressed, to create added-value products for the ports, is that of ship detection. Different datasets on very-high-resolution imagery have been presented, that will be utilized for ship detection out of medium resolution imagery provided by ESA Sentinels and Planet imagery. State-of-the-art methods for object detection will be modified and trained on these datasets to provide ship detection capabilities. Besides using already existing datasets, the novel idea to merge AIS data with satellite imagery, has been described. Combining AIS data with satellite imagery will not only provide annotated data for ship detection, but also additional data about the ship, that is captured in AIS messages. To obtain large scale dataset of this kind, regions will need to be defined around the ports or other areas with ship traffic. Imagery will be obtained for these areas for a larger period and combined with AIS data, using datasets that were presented in the chapter about AIS data.

Similar added-value products will be implemented as presented with AIS data, such as port congestion indicators, visualization and statistical analysis about the ship traffic. Compared to AIS data, this data is available with a limited time resolution, especially in the case of ESA Sentinel data, that is provided free-of-charge, but can be deployed globally for any place at any time and can also serve as a comparative analytics tool among ports. Additionally, the use of AIS transmitter can be monitored in port area, by merging visual detections with AIS data (if available in the port) and identifying ships that are not transmitting properly.

The use of satellite imagery is not limited to the use cases that were presented and are to be addressed but represents a much wider pool of opportunities that can be explored, especially with very-high-resolution (VHR) optical imagery. VHR optical imagery can be for example used, to monitor the number of containers in port area or the number of cars parked in the car terminal and with increasing time resolution, can serve as a usable operational tool, not only for the interested port but also for the comparison with other competitive ports of interest. VHR SAR imagery that is slowly coming forward with planned launches of large constellations of SAR satellites, promising to provide even hourly revisit times, is another direction that can be explored. With VHR capabilities and resistance to weather conditions, it can prove to be very useful to monitor port operations, despite limitations presented in this chapter. The goal here is to increase ports awareness to these emerging technologies and its capabilities.

6. Analysis and prediction of road traffic conditions with connection to port operations

6.1. Introduction

Road traffic prediction consists of knowing, prior to occurrence, the level of traffic on a certain spatial area of a road at a moment of time. Drawing from current and past traffic information, its aim is to estimate the future, to act in advance, to prevent harmful situations. Traffic prediction is an important field of research which is gaining increasing attention among the academic and industrial community.

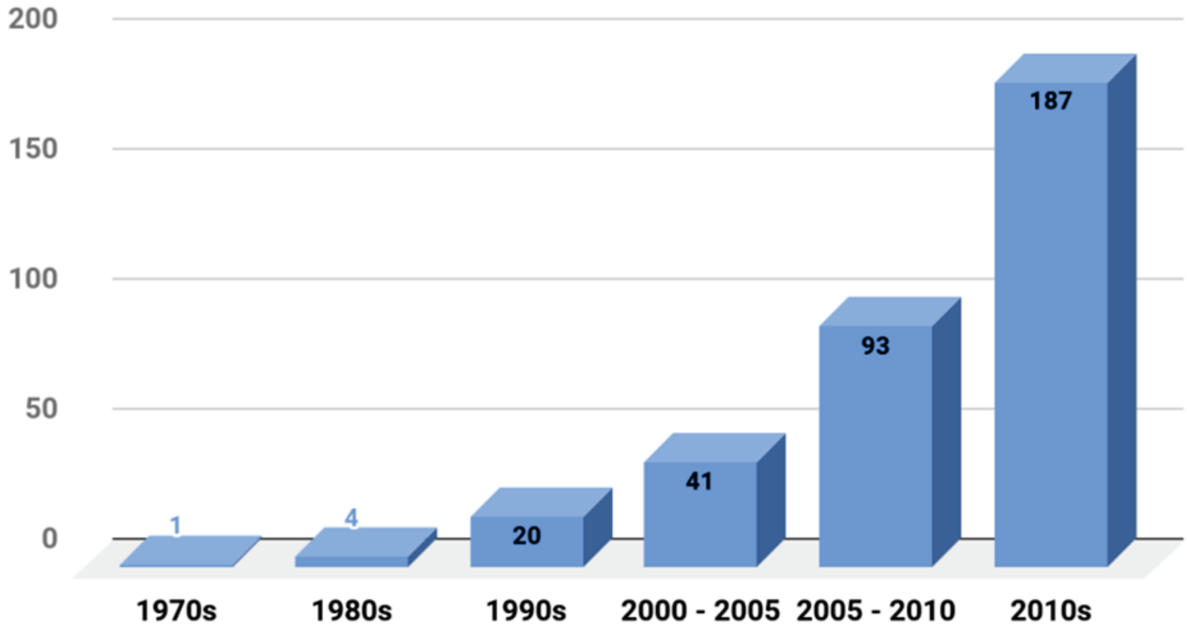


Figure 41. Basic article counts on traffic prediction during the last decades.[98]

Generically, what a road traffic prediction algorithm aims to achieve is to forecast density, queues, volume or speed of traffic. For some applications, travel times from one point to another is a major outcome of these algorithms. Techniques used to forecast have evolved since the "beginning" of this field of study, which is dated 40 years ago. The most used techniques in modern approaches are Neural and Bayesian Networks, Fuzzy and Evolutionary techniques, applied to specific application environments with different accuracy outcomes.

Nevertheless, the major problem encountered in real applications for accurate predictions is two-fold: (i) lack of real-time data (data collection at single points and availability of extensive datasets in real scenarios is harsh), needed to establish good models and execution environments and (ii) unpredictability of phenomena that influence status of traffic at a certain moment of time (e.g. crashes or other incidents). These are added to other hindrances that make the field open for research, especially now (and in the future) with more powerful and sophisticated computation technology, are being wide-spread to the community. An effort was made in the literature, to synthesize the 10 most important challenges on road traffic prediction, relevant literature and possible future research directions, presented in the table below.

[98] http://www.lsi.upc.es/~gavalda/papers/ewgt2017-slides.pdf

Table 4. Directions for further research in relation to the 10 challenges.[99]

Challenges	Further research directions
Developing responsive algorithms and prediction schemes	Weather and incident responsive algorithms, enhancing the efficiency of online computations using artificial intelligence, standardizing the requirements regarding spatial and temporal data coverage.
Freeway, arterial and network traffic predictions	Focus on network level predictions, Synergy with traffic flow theory and models.
Short-term predictions: from volume to travel time	Producing existing or novel measures of traffic performance, using data from multiple sources or using novel technologies for collecting and fusing data.
Data resolution, aggregation and quality	Determining the optimal degree of aggregation, in relation to the short-term forecasting application, Quality of probe data.
Using new technologies for collecting and fusing data	Testing the efficiency of new technologies for collecting traffic data, Reliability under all types of traffic flow (constrained, unconstrained), market penetration, standardization, cost, privacy issues, effectiveness of fusing strategies.
Temporal characteristics and spatial dependencies	Focus on network level spatio-temporal approaches, fusing modelling and data-driven algorithms.
Model selection and testing	Establishing synergies with statistics for estimating model specification and fit.
Compare models or combine forecasts	Introducing combinations of forecasts for multiple steps ahead predictions, testing the reliability of combinations of forecasts over single model predictions.
Explanatory power, associations and causality	Synergy with statistics and computationally intelligent algorithms to enhance the transparency of data driven approaches.
Realizing the full potential of artificial intelligence	Introducing intelligence to data collection and storage, traffic analysis, optimization modelling and decision making.

In the state of the art, few consolidated tools have been found on relation of road traffic prediction for ports. Examples of traffic prediction in ports in the literature, are focused more on maritime traffic rather than road traffic. However, some general tools for traffic analysis are worth mentioning:

- AIMSUN:[100] Aimsun Live simulates large areas in real time, allowing traffic operators to envisage traffic conditions before they unfold and to evaluate the impact of congestion management strategies.
- PTV Vissim:[101] Through a visualization tool, this product creates a model of traffic for the location indicated and then links several data sources to dynamically apply traffic patterns and inform the user. It is accompanied by other PTV products within an urban-oriented software suite.
- IBM traffic prediction tool:[102] IBM TPT (Traffic Prediction Tool) is a statistical model for the near-term prediction of traffic conditions. TPT was developed at IBM and has been tested in Singapore, where the Land Transport Authority is working with IBM and others to develop technology that will provide one-hour traffic predictions.

Regarding open tools, some initiatives that have been used by research projects and explained in articles that could be used by ports, were found. The one selected to comment in this chapter is VEINS (Vehicles in Network

[99] Vlahogianni, Eleni I., Matthew G. Karlaftis, and John C. Golias. "Short-term traffic forecasting: Where we are and where we're going." Transportation Research Part C: Emerging Technologies 43 (2014): 3-19. https://www.deepdyve.com/lp/elsevier/short-term-traffic-forecasting-where-we-are-and-where-we-re-going-0J8iZgyRBC
[100] https://www.aimsun.com/aimsun-live/
[101] http://vision-traffic.ptvgroup.com/en-us/products/ptv-vissim/
[102] https://researcher.watson.ibm.com/

Simulation)[103]. The Veins framework includes a comprehensive suite of models to make vehicular network simulations as realistic as possible, without sacrificing speed.

Underlying this tool there are two modelling/simulation frameworks that are used to build VEINS utility. The GUI (Graphical User Interface) and IDE (Integrated Development Environment) of OMNeT++ and SUMO (Simulation of Urban MObility) can be used for quickly setting up and interactively running simulations. More concretely:

- Road traffic simulation is performed by SUMO, which is well-established in the domain of traffic engineering.
- Network simulation is carried out by OMNeT++, which is a software that create networks and interfaces based on maps, objects and other spatial references.

In the image below there are two screenshots that provide vision about the aspect of the tool:

Figure 42. Simulation framework with Veins-Omnet

Being able to predict (with a reasonable degree of certainty) the amount of traffic, in a certain moment, can provide a series of benefits to a port. Generically, what could be measured/identified when successful algorithms are applied, are the following traits:

- Trends per year/season and other historical patterns
- A sudden change on outbound traffic conditions
- Gradual change of outbound traffic conditions

And what that information could mean for ports is, at least:

- To find correlations on external traffic conditions and efficiency of internal operations.
- To predict waiting times for trucks to enter/exit the port through gate access.
- To forecast efficiency of the logistic chain taking into consideration the road interface with the port as an essential node. This is crucial because it involves multiple players (e.g. logistics companies) and the better something can be planned, the more (cost)-effective is to all of them.
- To make short, medium and long-term planning on fleet needs.
- To make accuracy reports analysing traffic prediction data.
- Utilise other transport means for output flows (e.g. train):
    - Open/close entry/exit gates for trucks
    - Increase/decrease the fleet of trucks to be used in a moment of time by the port

[103] http://omnetsimulator.com/veins-omnet/

6.1.1.ASPM/SDAG integrated road traffic predictions

Port of Monfalcone, together with inland terminals, such as SDAG and the whole FVG (Friuli Venezia Giulia) Region, presents an “intermodal transport use case”. Geographically located in the crossroads of important transport corridors, presents an interesting use case to apply predictive analytics methods, to gather additional operational insights for stakeholders, to improve their operational efficiency or to provide additional information and make data driven decisions. This is especially interesting, because of the availability of the regional logistic information system, that also integrates traffic information from the ports, as well as regional road network. Port of Monfalcone can be reached by train or by road. In this task, analysing and predicting road traffic, in correlation with operations in the port and inland terminals, such as SDAG, will be addressed. As part of the SEC (Safe and Efficient Cargo), regional logistic information system SILI was developed in FVG region, that will present the main input for this task, along with data from the port and SDAG, which is currently not connected to SILI system. SILI information system was deployed in 2008, to improve efficiency and safety of intermodal transport, allowing to monitor access gates to the ports, inland terminals, as well as different parts of the regional road networks that are used by trucks entering ports of interest.

Analysing and predicting traffic conditions in the port and areas around it, will represent the main task to be performed for the ASPM/SDAG use case. This task will be addressed with the help of data provided by SILI system, data provided by camera installed in SDAG premises, vessel call data from the port and any other data obtained from highway operators, if available. This work will also be used by task T4.3.

6.1.2.PPA road traffic predictions

The port of Piraeus, one of the busiest ports in the Mediterranean, is affected by all the operations related with their activities. Among these activities, there are:

- Container traffic operations
- Coastal (passenger) traffic operations

The focus of the Piraeus use case will be passenger traffic operations.

Due to the quantity of passengers that arrive in cruises, the city is not able to assimilate the generated traffic and situations of traffic congestion, which also lead to occur noise pollution and other environmental impacts.

To anticipate this congestion traffic situations, ports authorities need to know where these situations occur and to be able to simulate these situations to optimize the traffic.

PPA wishes to monitor the traffic within the predefined following port coverage area, shown in the image below. The area includes the main port coastal passenger terminal, gates and the two cruise terminal gates. Also, west road arteries connect container and car terminals, through which most of the truck container traffic is passing.

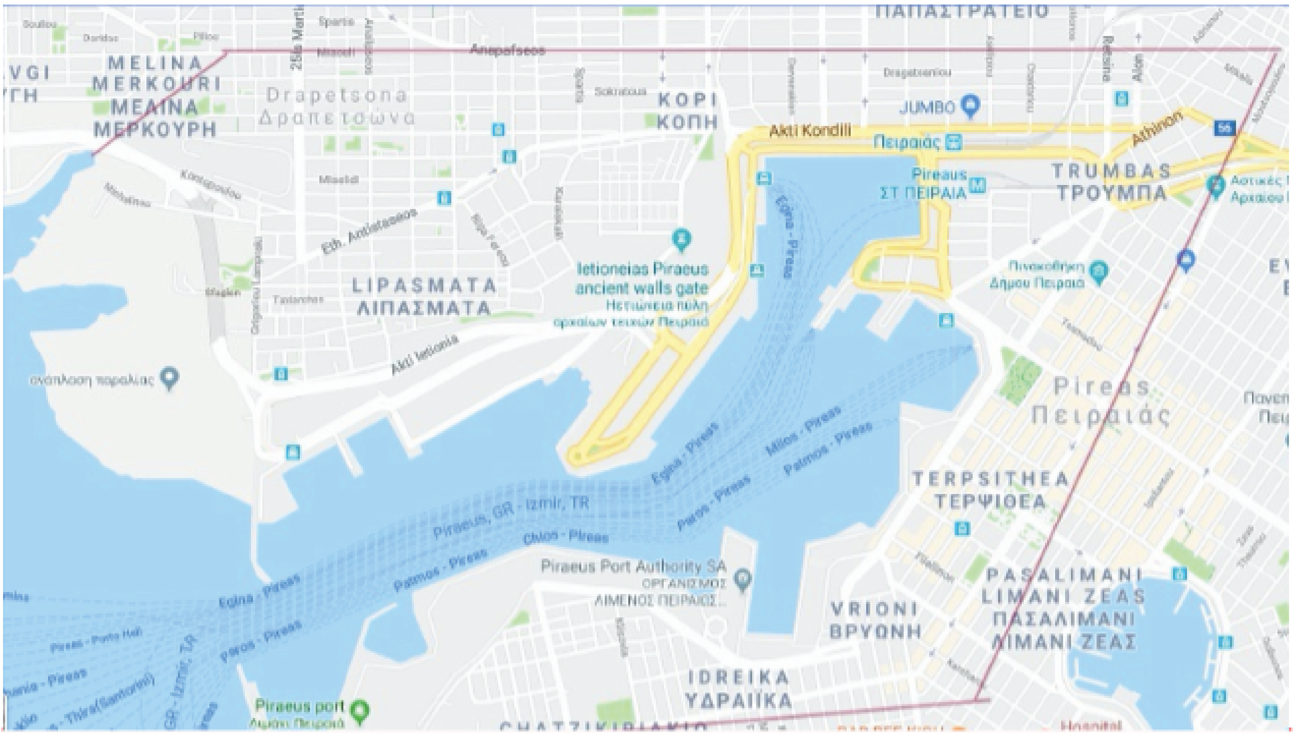


Figure 43. Port of Piraeus traffic coverage area (enclosed inside the red lines)

In PIXEL, the issue related to road traffic predictions could be addressed to one of the use cases of PPA.

Intermodal transport use case: One of the advantages that PPA wants to extract from PIXEL, is to improve the access to the seaport and better handle traffic congestion (due to the tourism) in the port surrounding area.

PIXEL aims at predicting congestions with the help of predictive algorithms (PIXEL Operational Tools). To predict correctly the situation, it is necessary to consider as much data available as possible. The objective will be to provide the port with the necessary tool to predict this situation, when the necessary conditions are met (traffic congestion due to the arrival of cruise ships at the port).

Although not belonging to the same use case, a reduction in traffic congestion situations, cause a reduction in certain environmental indicators that influence the PEI (Port Environmental Index) and therefore related to the other use case that occurs in Piraeus: **Port - City integration**.

6.1.3.THPA road traffic predictions

The Port of Thessaloniki through PIXEL, and combining all available resources, will collect, monitor and publish real time traffic and environmental data. Data analysis will enable the port to identify periods of time with emissions above accepted levels and optimise inbound and outbound truck flows and mechanic equipment movements.

The port of Thessaloniki has already been described in deliverable D3.4, but for the traffic scenario it is important to highlight and summarize, how current operators are managing the traffic in the neighbourhood of the city. This is mainly done via a web application from TrafficThess[104], as depicted below for the port surroundings.

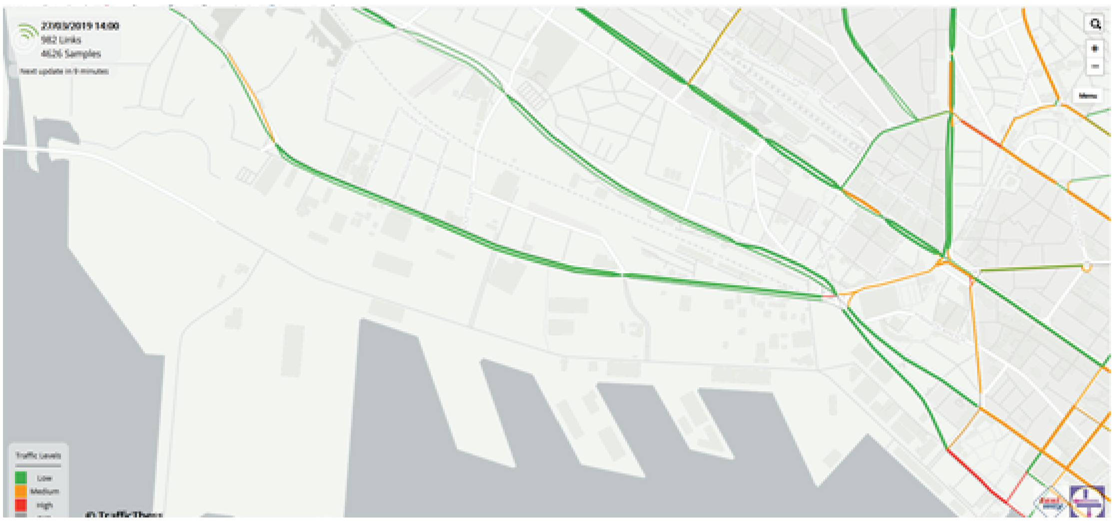


Figure 44. Traffic monitoring in the city of Thessaloniki

It is important to locate the gates for accessing/leaving the port, as well as the relevant roads to track, as depicted in the figure below.

Figure 45. Relevant streets and gates for ThPA

ThPA SA will exploit the information gathered from the devices already installed and new, to render traffic and port operations in an eco-friendly way. The PIXEL project will provide ThPA SA, an advanced decision support tool.

Finally, THPA defined four impacts to be achieved by the deployment of PIXEL. One of them depends on creating a successful prediction that could help Terminal Operator to address better decision making. More concretely, it was stated as: "*Monitoring and optimization of inbound and outbound truck traffic*". This may

[104] https://www.trafficthess.imet.gr/

imply effects on climate change and environment, as well as in logistics efficiency of the port (via alleviating bottlenecks and congestion at the entry/exit of the port).

Therefore, the essence of the THPA use-case in relation with prediction, can be summarised as follows:

Road traffic prediction algorithm will allow ThPA to identify traffic bottlenecks, optimise traffic flows and reduce the environmental impact of the port towards the city of Thessaloniki.

Besides, in deliverable D3.4 (aiming at condensing T3.3 work), the Consortium defined several user-stories per port, in which, each port specified expected goals for operations/processes. For THPA, six stories were described. One of those was to predict and estimate the impact of traffic entering/exiting the port. To this purpose, predictive algorithms are needed.

Table 5. User story of THPA related with road traffic prediction

As a/an	I want to …	So that …
Terminal Operator	To estimate the impact of the current inbound / outbound flow of trucks entering /exiting the port, considering the actual traffic in the nearby (city)	We can regulate the number of working entry/exit gates to optimize the inbound/outbound traffic without impacting too much in the city traffic (alleviate bottleneck).

THPA wishes to predict/forecast the road traffic implying potential bottlenecks on their entry/exit gates. This information will serve as a tool for decision making to the Terminal Operator. To address this user-story, the following aspects are needed: (i) historical data to create a solid baseline for forecasting, (ii) current data of traffic level in the city, so that upcoming jam situations are detected, (iii) sensors/information of gates status to map it with surrounding situation.

6.2. Data sources

6.2.1.ASPM/SDAG

Data from SILI system represents the main data source. Besides supporting electronic exchange of the documents needed to enter the ports, SILI system also captures traffic and gate access information from ports, inland terminals as well as regional roads.

Locations of stations in the regional road network are presented below:

- Valico Rabuiese
- Valico di Fernetti
- Prosecco
- Raccordo Villesse - Gorizia
- Interporto di Cervignano
- Stazioni doganali Autoporto di Gorizia

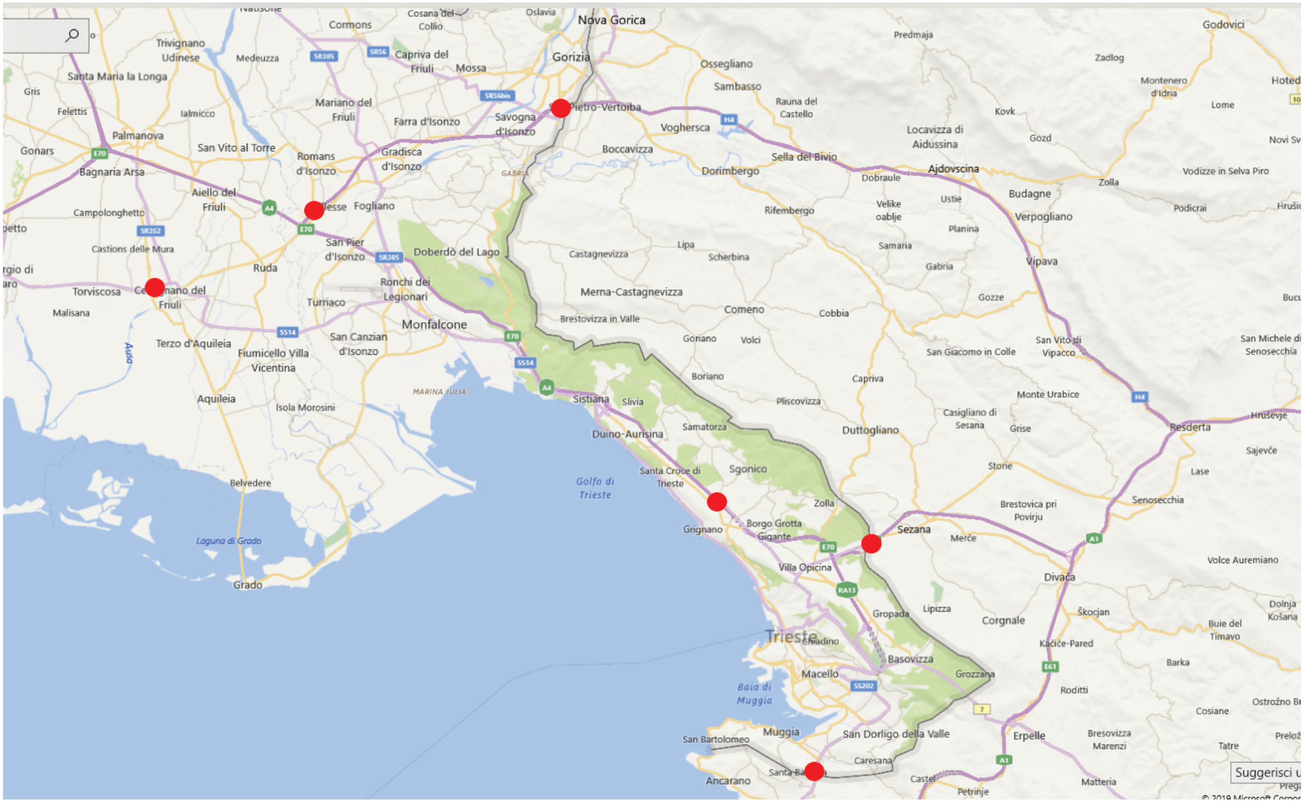


Figure 46. Locations of traffic cameras connected to SILI system in regional road network.

Besides regional road network, port access gates in the Port of Monfalcone are connected to the SILI systems, as well as some of the inland terminals. SDAG premises are not connected to the SILI system, although some sensors and cameras are installed there, constituting another possible source of data. An effort will be made to be obtained from SDAG. Additional sensors will be installed in the parking area in front of the Port of Monfalcone, providing additional data.

An example of the data that is captured by cameras connected to the SILI system, is provided below. The direction of the traffic is captured, along with the exact time, license plate, nationality and type of the vehicle. This kind of data is available for all 6 locations along regional road network, as well as for access gates to the port. Apart from live data, historical data is also available and will be provided, to analyse and predict traffic volumes and congestion rates.

Varco (gate)	Corsia (lane)	Data e ora	Targa (license plate)	Nazione	Classe (type of vehicle)
Interporto Cervignano OUT	1	Jun 17, 2018 9:01:55 AM	[illegible]	I	Autovettura
Interporto Cervignano OUT	1	Jun 17, 2018 8:51:08 AM	[illegible]	I	Autovettura
Interporto Cervignano IN	1	Jun 17, 2018 8:37:46 AM	[illegible]	I	Autovettura
Interporto Cervignano OUT	1	Jun 17, 2018 8:35:17 AM	[illegible]	I	Autovettura
Interporto Cervignano IN	1	Jun 17, 2018 8:33:05 AM	[illegible]	I	Autovettura
Interporto Cervignano IN	1	Jun 17, 2018 8:32:00 AM	[illegible]	I	Autovettura
Interporto Cervignano OUT	1	Jun 17, 2018 7:26:33 AM	[illegible]	I	Autovettura
Interporto Cervignano IN	1	Jun 17, 2018 7:26:01 AM	[illegible]	I	Autovettura
Interporto Cervignano IN	1	Jun 17, 2018 7:24:40 AM	[illegible]	I	Autovettura

Figure 47. Example of traffic data provided by INSIEL gathered in SILI system

6.2.2.PPA

For the measuring of road traffic, PPA can use three different alternatives:

- **Live data report from the police (traffic department):**[105] The data provides the traffic situation (middle column of the table below and the rated traffic as: normal, congested and very congested) of the main Athens - Piraeus road arteries (listed in the left column) with special explanatory notes per each main road (right column). These are only live traffic data, while no historical info is provided.

ΔΙΕΥΘΥΝΣΗ ΤΡΟΧΑΙΑΣ ΑΤΤΙΚΗΣ - Θ.Ε.Π.Ε.Κ.
Δελτίο Κυκλοφορίας - ΑΘΗΝΑ

Τελευταία Ενημέρωση: 08/04/2019 - 15:07

ΟΔΙΚΟΙ ΑΞΟΝΕΣ	ΚΥΚΛΟΦΟΡΙΑ			ΕΠΙΣΗΜΑΝΣΕΙΣ
	Ομαλή	Αυξημένη	Πολύ Αυξημένη	
Αθηνών-Κορίνθου/Λ. Αθηνών (έξοδος)	✓			
Αθηνών-Κορίνθου/Λ. Αθηνών (είσοδος)	✓			
Αθηνών-Λαμίας(Αθηνών έως κόμβος Καλυφτάκη -έξοδος)		✓		ΥΨΟΣ ΝΕΑΣ ΦΙΛΑΔΕΛΦΕΙΑΣ
Αθηνών-Λαμίας(κόμβος Καλυφτάκη έως Αθηνών-είσοδος)		✓		ΥΨΟΣ ΝΕΑΣ ΦΙΛΑΔΕΛΦΕΙΑΣ
Λ. Κηφισού (από παραλιακη εως Αθηνών)	✓			
Λ. Κηφισού (από Αθηνών έως παραλιακή-κάθοδος)	✓			
Θηβών	✓			
Ιερά Οδός	✓			
Πέτρου Ράλλη	✓			
Πατησίων	✓			
Αχαρνών	✓			
Αλεξάνδρας	✓			
Μεσογείων	✓			
Κατεχάκη (ρεύμα από Ηλιούπολη)	✓			
Κατεχάκη (ρεύμα προς Ηλιούπολη)	✓			
Λ. Κηφισίας (άνοδος)		✓		ΥΨΟΣ ΨΥΧΙΚΟΥ
Λ. Κηφισίας (κάθοδος)	✓			
Βασ. Σοφίας (άνοδος)		✓		
Βασ. Σοφίας (κάθοδος)	✓			
Βασ. Κωνσταντίνου	✓			
Καλλιρρόης	✓			
Λ. Συγγρού (άνοδος)	✓			
Λ. Συγγρού (κάθοδος)	✓			
Ακαδημίας	✓			
Πανεπιστημίου	✓			
Σταδίου	✓			
Βασ. Αμαλίας		✓		
Φιλελλήνων	✓			
Μιχαλακοπούλου	✓			

Figure 48. An example of the data that police department offers

- **TELENAVIS road traffic data:** Data may also be obtained from a third-party subscription database, in the format of .csv and for a port specific area, as graphically represented in the images below, for one of the providers. Demo data sets are being prepared by the provider, to be delivered to PPA for evaluation purposes, within May 2019. No historical data can be provided from this supplier, but future

---

[105] http://www.astynomia.gr/traffic-athens.php

data can be stored when subscription is made available. The data to be generated and the cost to obtain them, is a parameter to be examined and the final decision on the usefulness of the data, will be determined within May 2019. This will represent the main source of data for this task.

Figure 49. A Live map from ***http://navigation.gr***

- **Traffic reports from the Region of Attica observatory on traffic congestions:**[106] These data reports are published every 6 months. These reports are available in the Greek language and describe the findings and actions concerning the traffic on the main streets of Attica based on traffic jams, march, blockade of roadworks etc.), as well as other incidents, such as vehicle stoppages due to damage, accidents, severe weather events, etc.

6.2.3.THPA

Several data sources have been identified that are either useful, or have the potential of being useful, for the implementation of the predictive algorithm and will be listed in the following subsections.

6.2.3.1. Open data sources

The main data source is available as web application, which provides near real-time information every 15 minutes[107]. Such information is already presented on a map (see figure below), showing those lanes that are low (green), medium (yellow) or highly (red) congested. The coverage area is big enough to cover both the Thessaloniki port and city.

[106] http://www.patt.gov.gr/site/index.php?option=com_content&view=article&id=29298:35i-ekthesi-leitourgias-tou-kentrou-diaxeirisis-kykloforias&catid=301&Itemid=320

[107] https://www.trafficthess.imet.gr/

Figure 50. Near real time data from Trafficthess.

The web application also provides speed and unique vehicles of the previous hour, when clicking on a specific lane, as depicted in the Figure below. Historical data is also available.[108]

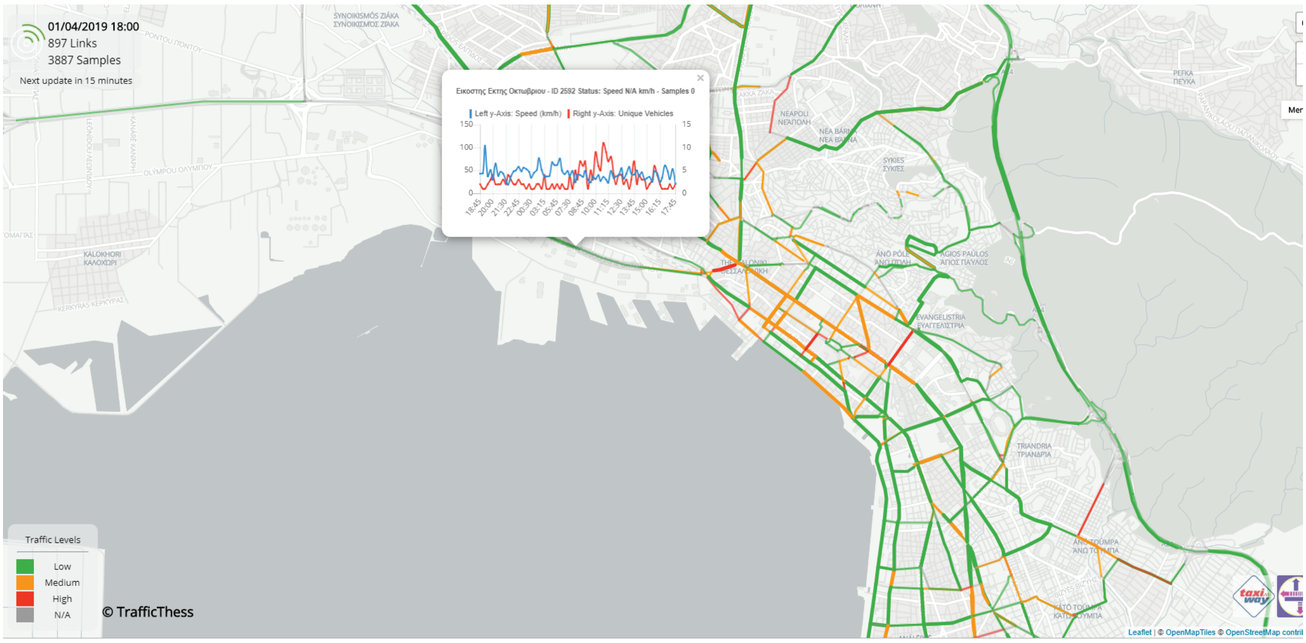


Figure 51. Historical (last hour) data from Trafficthess.

It is important to highlight the fact that, according to the previous picture where the location of the two gates for the port of Thessaloniki are depicted, that for one of the gates, there is no direct traffic data (west gate). For that gate, some relationship with the nearest main road will be searched.

Another relevant traffic data source is the open data portal from the CERTH-HIT (Hellenic Institute of Transport)[109].

Two important datasets are identified:

[108] https://www.trafficthess.imet.gr/exporter.aspx
[109] http://opendata.imet.gr/dataset?tags=Thessaloniki

- **Network-speed**: Current speeds on OpenStreetMap network links, available as json, xml and csv. The Figure below, shows an example of the data format that can be retrieved.

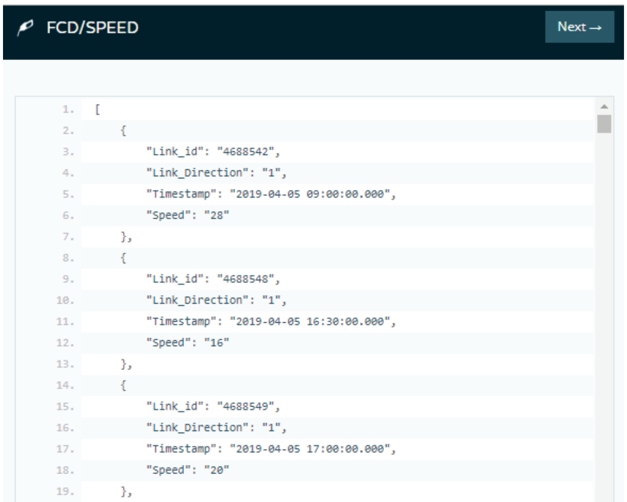


Figure 52. OpenStreetMaps current speed json example.

- **Network-congestion:** The current traffic conditions on OpenStreetMap network links, available as json, xml and csv. The Figure below shows an example of the data format that can be retrieved. In fact, it seems at first sight that the web application at Trafficthess may obtain the data about congestions from this source.

Figure 53. OpenStreetMaps traffic congestion json example.

There is also real time and historical data (from 2016), from a floating fleet of cars in Thessaloniki. This data is especially interesting, due to amount of it and its accuracy, as it is extracted from GPS devices installed in the cars. The Figure below, shows an example of the data format that can be retrieved.


```
[
    {
        "recorded_timestamp": "2019-04-05 17:25:02.740",
        "lon": "22.9406216666667",
        "lat": "40.6376966666667",
        "altitude": "2.2",
        "speed": "0",
        "orientation": "131.800003051758"
    },
    {
        "recorded_timestamp": "2019-04-05 17:25:02.763",
        "lon": "22.9534616666667",
        "lat": "40.6070116666667",
        "altitude": "4.1",
        "speed": "0",
        "orientation": "343.100006103516"
    },
```


Figure 54. JSON data example from the floating fleet of cars in Thessaloniki

The main objective is to gather relevant insights if a correlation between this data with the traffic in the port can be found.

For example, Figure 55 visualizes the number of vehicles captured throughout a whole day. Note that, as not all vehicles can be really printed (1 million points from this day), the point density is converted into a different colour to visualize low, medium and heavy loaded roads.

Figure 55. Data traffic visualization by processing historical data

What makes also sense is to get real time and historical information about the weather, as probably it is a factor which has an impact on port operations (ships berthing or not, trucks arriving or not, etc.). This does not seem to be a major problem, as there are various open APIs offering such service, such as OpenWeatherMap[110].

6.2.3.2.Internal ports' data

ThPA is providing a dataset (for the moment API is not available) with information about the vehicles

[110] https://openweathermap.org/api

entering and leaving the gates. Therefore, it is known:

- When someone enters port, and from which Gate.
- When someone leaves the port, and from which Gate.
- Dwell time spent within the port.

The figure below shows an extract of such data set.

Registration Plates	Vehicle Type	RFID	company	Maker	Model	Colour	ISPS	Entry	Entry Gate (ACU)	Exit	Exit Gate (ACU)	reason	Comments	Dwell time
								02/02/2019 23:37	Gate 16 - Entry #2	**02/02/2019 23:42**	Gate 10A - Exit			4' 22"
								02/02/2019 23:08	Gate 10A - Entry	**02/02/2019 23:33**	Gate 16 - Exit			25' 12"
	Passenger vehicle			FORD	FOCUS	ΜΑΥΡΟ		**02/02/2019 21:30**	Gate 10A - Entry	**02/02/2019 21:33**	Gate 16 - Exit			3' 37"
	Passenger vehicle			TOYOTA	YARIS	ΓΚΡΙ		**02/02/2019 21:29**	Gate 10A - Entry	03/02/2019 7:29	Gate 10A - Exit			9h 59' 57"
	Passenger vehicle			SEAT	MINI	ΛΕΥΚΟ		**02/02/2019 21:15**	Gate 10A - Entry	04/02/2019 11:21	Gate 10A - Exit			1d 14h 6' 0"
	Passenger vehicle			CITROEN	C1	ΜΑΥΡΟ		**02/02/2019 21:14**	Gate 10A - Entry	03/02/2019 7:35	Gate 10A - Exit			10h 21' 25"
	Passenger vehicle			FORD	FIESTA	ΕΡΥΘΡΟ		**02/02/2019 20:56**	Gate 10A - Entry	03/02/2019 5:43	Gate 16 - Exit			8h 46' 45"
	Passenger vehicle			PEUGEOT	307	ΜΑΥΡΟ		**02/02/2019 20:44**	Gate 10A - Entry	03/02/2019 0:50	Gate 10A - Exit			4h 5' 56"
	Passenger vehicle			HONDA	E 19	ΜΠΛΕ		**02/02/2019 20:26**	Gate 10A - Entry	03/02/2019 5:42	Gate 16 - Exit			9h 16' 50"

Figure 56. Vehicles incoming/outgoing in the port of Thessaloniki

From this information, we can count vehicles entering and leaving, and correlate it with parking availability within the port, but a prediction of average waiting times at the gates, is still difficult to achieve.

To facilitate this, an estimation of the average throughput of each gate is required (gates 10 and 16), so that, by counting vehicles, some heavy or low traffic may be inferred. The case of Gate 16 is special, as it has three entrance lanes, thus it is important to know when each entrance is open.

Another source of information is, whether a vessel is arriving or departing at the port of Thessaloniki, to correlate this data with trucks entering and leaving the port. ThPA provides datasets (for the moment no available API), an extract of which, can be seen in the figure below.

	A	B	C	D	E	F	G	H
1	VO_CODE	VESSEL_NAME	ORIGIN_PORT_ID	DESTIN_PORT_ID	ARRIVAL_DATE	DEPARTURE_DATE	ARRIVAL_DATE_SCHED	DEPARTURE_DATE_SCHED
2	MXPRIDE1908	MAX PRIDE	GRSKG	CYLMS	2019-02-24 17:24:01.697	2019-02-24 17:24:01.697	2019-02-24 00:00:00.000	2019-02-25 00:00:00.000
3	TGS006W19	AURETTE A	GRPIR	GRPIR	2019-02-23 14:55:41.877	2019-02-23 14:55:41.877	2019-02-22 07:00:00.000	2019-02-25 06:00:00.000
4	11577DE908	MERITO	GRVOL	GRPIR	2019-02-23 13:41:38.377	2019-02-24 15:47:42.697	2019-02-22 00:00:00.000	2019-02-23 00:00:00.000
5	12318AD907	MED TEKIRDAG	TRTEK	EGDAM	2019-02-23 02:47:44.383	2019-02-24 15:44:49.830	2019-02-22 00:00:00.000	2019-02-24 00:00:00.000
6	OTF4RS1MA	LION	NULL	NULL	2019-02-22 08:14:38.383	2019-02-22 08:14:38.383	2019-02-18 10:00:00.000	NULL
7	TGS007E19	GABRIEL A	MTMAR	GRPIR	2019-02-21 23:40:07.973	2019-02-23 04:31:13.813	2019-02-21 02:00:00.000	2019-02-23 06:00:00.000
8	OTF4RS1MA	LION	GRPIR	TRIST	2019-02-21 14:54:11.373	2019-02-23 00:29:25.840	2019-02-21 07:06:00.000	2019-02-23 13:00:00.000
9	04607NC908	MSC LISA	GRPIR	TRIZM	2019-02-20 22:03:02.223	2019-02-21 21:56:00.983	2019-02-20 00:00:00.000	2019-02-22 00:00:00.000
10	VOASPR517W	UNI-ASPIRE	GRSKG	GRPIR	2019-02-19 14:52:36.290	2019-02-20 14:44:28.690	2019-02-19 16:30:07.000	2019-02-21 16:30:07.000
11	CD016N19	CAFER DEDE	TRGEM	TRIZM	2019-02-18 18:48:45.180	2019-02-19 23:20:46.160	2019-02-18 13:16:50.000	2019-02-19 13:16:50.000
12	12590AD906	CONTSHIP TOP	TRTEK	EGDAM	2019-02-18 07:54:49.883	2019-02-18 16:48:43.770	2019-02-18 00:00:00.000	2019-02-20 00:00:00.000
13	MXPRIDE1907	MAX PRIDE	GRSKG	CYLMS	2019-02-17 07:50:06.377	2019-02-18 15:42:10.947	2019-02-16 00:00:00.000	2019-02-17 00:00:00.000
14	TGS006E19	AURETTE A	MTMAR	GRPIR	2019-02-17 02:12:36.843	2019-02-18 23:54:25.410	2019-02-15 00:00:00.000	2019-02-17 00:00:00.000
15	11577DE907	MERITO	GRVOL	GRPIR	2019-02-16 20:06:05.567	2019-02-17 19:37:47.700	2019-02-15 00:00:00.000	2019-02-16 00:00:00.000
16	04217NC907	MSC FABIENNE	GRPIR	TRIZM	2019-02-15 19:13:42.037	2019-02-17 00:09:03.057	2019-02-14 00:00:00.000	2019-02-15 00:00:00.000
17	TGS005W19	GABRIEL A	GRPIR	MTMAR	2019-02-15 18:12:35.033	2019-02-16 18:10:32.220	2019-02-13 00:00:00.000	2019-02-14 00:00:00.000
18	VOWNBL024W	WARNOW BELUGA	GRSKG	GRPIR	2019-02-13 20:35:23.510	2019-02-15 17:06:37.203	2019-02-12 23:03:18.000	2019-02-14 09:03:18.000

Figure 57. Vessel arrival and departure times and the port of Thessaloniki

As can be seen, this piece of data shows real and scheduled times; as they are not the same, it is unclear whether scheduled times for future predictions can be "trusted", since there seems to be a timeframe of uncertainty. Predictive algorithms for maritime traffic may reduce such uncertainty and provide the link between both maritime and road traffic predictive algorithms.

So far, it has been difficult to get additional data sources that may be relevant for the use case of the port of Thessaloniki (prediction of road traffic near the port, especially at the gates):

- There is no video camera at the entrance of the gates with AI capabilities to count vehicles and detect congestion.
- There is no available information about cameras located at relevant lanes (highways), with access to the port to consider them.
- Google Maps does not provide an open API to access traffic information on the lanes.

6.3. Predictions and analytics for PIXEL

All the presented use cases have a common task of traffic volume prediction and correlation with operations in the port, either ship traffic or truck traffic at entrance to the port. The main task will be short-term traffic volume prediction, which is a well-established research topic with common approaches, which have been established in the literature in recent years.

Approaches can be roughly divided into two categories. The most common approaches are parametric, with the most prominent example of ARIMA Autoregressive Integrated Moving Average) based methods, which serve today as a baseline in the domain of traffic forecasting. The main downside of such methods is that they are prone to errors, when there is a lot of variations in the traffic patterns, which is especially common in short-term forecasting. This downside is addressed with non-parametric approaches such as SVM[111], k-nearest neighbours[112] and lately Deep Learning based methods[113,114].

The data that will be provided, will be very similar to the data that is used in literature[115] regarding short-term traffic prediction, with multiple points capturing traffic flow. To begin with, an estimation of traffic volume for each point separately will be addressed, using classical parametric approaches (e.g. ARIMA), included in an open-source library, provided by Facebook[116]. This information, from all the points that can influence traffic in a certain location will be utilized, as well as some of the more sophisticated state-of-the-art methods, especially Deep Learning based methods, such as those utilizing RNNs[117].

The amount and kind of data that will be collected for different use cases, will be diverse and as such appropriate methods will be selected for each of the use cases separately, considering also their specific nature. In the next subsections, the specifics for each use case are presented.

6.3.1.ASPM/SDAG

Two main tasks, which are connected to ASPM/SDAG use case, will be addressed. The first task includes the analysis of traffic data from access gates, to the Port of Monfalcone. Similarly, as presented in the vessel call data chapter, exploratory data analysis will be performed to visualize, and present insights gathered from raw data, which will be obtained from SILI system and ASPM video monitoring system, when available. Seasonal long-term trends, as well as short-term trends on a weekly or daily basis, will be investigated and correlated with vessel call data. Similar methods for visualization and data exploration will be used as in vessel call data chapter.

In the second task, traffic data from the regional road network that is available through SILI system will be utilized and possibly external providers, if publicly available. The goal of this task will be, to predict traffic volumes on the regional road network, to support the decision-making process, regarding rerouting the trucks to the inland parking premises, such as SDAG. Traffic data will also be correlated with traffic at the gates. Given the availability of license plates, time needed to entry the port can be computed from a point in the regional road network, where cameras are installed, and congestion rates can be computed, either on the road or in the port

[111] Zhang, Yang & Liu, Yuncai. (2009). Traffic forecasting using least squares support vector machines. Transportmetrica. 5. 193-213. https://www.tandfonline.com/doi/abs/10.1080/18128600902823216

[112] Lun Zhang, Qiuchen Liu, Wenchen Yang, Nai Wei, Decun Dong. (2013). An Improved K-nearest Neighbor Model for Short-term Traffic Flow Prediction. Procedia - Social and Behavioral Sciences. Volume 96, pages 653-662. https://www.sciencedirect.com/science/article/pii/S1877042813022027

[113] G. Polson, Nicholas & Sokolov, Vadim. (2017). Deep learning for short-term traffic flow prediction. Transportation Research Part C: Emerging Technologies. 79. 1-17. https://arxiv.org/pdf/1604.04527.pdf

[114] Yisheng, Lv & Duan, Yanjie & Kang, Wenwen & Li, Zhengxi & Wang, Fei-Yue. (2014). Traffic Flow Prediction With Big Data: A Deep Learning Approach. IEEE Transactions on Intelligent Transportation Systems. 16. 865-873. https://www.researchgate.net/publication/273564489_Traffic_Flow_Prediction_With_Big_Data_A_Deep_Learning_Approach

[115] http://new.portal.its.pdx.edu:8080/downloads/

[116] https://facebook.github.io/prophet/

[117] Chen, Weihai & Zhao, Zheng & Liu, Jingmeng & C. Y. Chen, Peter & Wu, Xingming. (2017). LSTM Network: A Deep Learning Approach For Short-Term Traffic Forecast. IET Intelligent Transport Systems. 11. https://ieeexplore.ieee.org/document/7874313

area. Combined with traffic volume information and predictions on the regional road network, bottlenecks can be identified, either in the port or in parts of the regional road network respectively.

The focus of both tasks is connected and can be considered as a short-term traffic volume prediction, which is a well-established research topic with different common approaches. Short-term traffic volume prediction usually predicts traffic volume in the next few minutes, up to the next few hours. To support close to real-time decision making to reroute traffic, short-term prediction is needed in PIXEL project, compared to daily or even longer predictions, which are not relevant for this task.

External data, provided through traffic portals regarding special events (e.g. accidents) will be included, as well as weather data. The prediction of short-term traffic volume based on historical data will be achieved through the provided methods, while the exact methods to be used and the size of the short-term window, will be defined when actual historical data are made available and after the first methods are evaluated against different short-term window sizes.

6.3.2.PPA

Simulation modelling of traffic in PPA cruise and passenger terminals will be used for performance evaluation and optimization of the operational policy. During the summer, the traffic intensity of the cruise ships arrivals is higher compared to other periods of the year. The increased traffic intensity causes congestion at anchorages, which can lead to significant dissatisfaction of the cruise ship operators and passengers. Other than economic and social benefits, port traffic generates negative externalities, which reduce overall benefits from port activities. Piraeus is no exception to the rule and traffic density at the port implies that these are significant. Cruise is an important contributor, generating especially air pollution and extra-urban congestion.

PIXEL must be able to forecast the probability of a traffic peak in a specific time frame (e.g.: daily) on the roads connecting the port of Piraeus with the city of Athens. The PIXEL platform should be able to:

- Estimate the impact of passenger ships on city traffic, in the next few hours.
- Identify traffic congestions which may take place inside or outside the port due to traffic jam, accidents or road works.
- Estimate the impact of tourist buses that is a strategically critical activity of cruise companies; that is, the sale of excursions.

PIXEL must be able to integrate information gathered from different sources to identify already planned or unforeseen traffic peaks, accidents or temporary driving bans, to better predict the volume of traffic and influence of port operations on the road traffic. The seaport city will be benefited from the decreased road traffic, saturating the streets and decreasing the quality of life for the citizens, leaving on the same hand, valuable area around the city centre for purposes other than traffic.

6.3.3.THPA

Drawing from the previous explanations, the team assigned to task "*T4.5 – Predictive algorithms*" proceed, to delimit the scope of prediction for THPA, in relation to road traffic. Besides, after several conversations with the port, the technical team has tried to merge both worlds (technical feasibility and port expectations) to align everything with PIXEL objectives and innovation aims. Traffic prediction is under continuous research and no universal solution has been validated for this purpose yet. Traffic jams depend on a huge number of variables (some of them unidentified) and its impact is utterly varying. Therefore, prior to define the methodology for development, the objectives and deadlines have also been analysed. Depending on the level of accuracy, data availability and impact of associated actions (what will happen after a prediction is made) the approach and techniques to be used will be different.

The objectives and scope of road traffic prediction in THPA, are the following:

- To visualize current traffic status at the surroundings of port gates. Initially THPA is only interested into monitoring traffic close to gate 16 and gate 10. Being integrated into PIXEL framework, this will help the Terminal Operator to observe at a glance the density of movement in the vicinity, while monitoring other operations in the same "visualization environment". This action will not require development of predictive algorithm, as it consists of visualization of data that could be extracted from already identified sources.

- Besides visualizing the status of traffic at the surroundings of the gate, the system should provide an indicator (it can be semaphore-like) about the congestion at the external side of THPA's entry/exit gates. This action will not require a predictive algorithm; setting scales plus being able to measure with a number the "quantity of traffic" will suffice to satisfy this point.
- To be able to have a prediction of amount of traffic at the entry/exit gates of the port, at a certain moment. The intention is to have a short-term traffic prediction, based on the available information from the port and other data providers. This action requires a predictive algorithm, which may need to take into consideration the correlation of different items (vessel calls, traffic in the city, seasonality, etc.), to provide a forecast.
- To provide an estimated waiting time for a truck entering and exiting the port. The potential of predicting waiting times, may provide ThPA relevant insights to be able to apply useful strategies or policies, from alerting truck drivers and logistics companies, up to scheduling available entering slots for them, or even increasing the number of entrances.

6.4. Plan and future work

As presented, the focus of described use cases will be the analysis and predictions regarding traffic around the port area or in the regional road network. The results of this task will provide additional operational insights and forecasts, which will help operators at rerouting trucks to the inland terminals, gather information about road conditions in general or to simply provide operational insights about traffic out of historical data, which can assist in making ports more sustainable. Road traffic data will be correlated with the vessel call data in the port, to explore possible correlations between road and port traffic, as well as with any other external data that might prove useful, such as weather information or traffic events.

Current work was mainly focused on gathering available data from the ports and regional operators. For the time being, only samples of data were provided to get a feeling about the structure of it.

A risk has been detected with data for the PPA area, as it is not yet clear whether data from TELENAVIS is going to be available. In case TELENAVIS data cannot be used, PPA will, in collaboration with technical partners, assess alternatives such as Google or HERE. In case no alternative is found, predictions will be made with entry and exit gates information.

Similar research problems, with similar data structure and availability were identified in the literature and reviewed, which will be explored and translated to the data and tasks presented.

7. Prediction of renewable energy production

7.1. Introduction

To decarbonize the energy supplies, it is envisaged that by 2035 renewable energies will represent at least 50% of the global energy mix. For port ecosystems, there are many opportunities today to contribute to this goal. The use of available surfaces on hangars and buildings, is one of the tools available to initiate this energy and ecological transition.

Photovoltaic power generation has been steadily increasing in recent years, thanks to lower equipment costs, more favourable public policies and a capacity of these production units to reduce the environmental impact of energy production. To be a part of this energy and ecological transition, many ports are already taking advantage of available space, especially on warehouses to install photovoltaic power plants[118]. By taking advantage of quite large surface area for the installation, the ports contribute to the production of green energy, which emits little greenhouse gases, while benefiting from a new source of income (energy is self-consumed reducing the bill, or it is sold directly to energy distributors). In addition, the installation of photovoltaic power plants within the ports, makes it possible for ports to be even more integrated with the city, by supplying the neighbourhood districts with their energy surplus.

To implement this new way of producing energy, port authorities need to better understand the potential of solar production on their site. Furthermore, the estimation of their annual production, will assist in optimizing the location and use of photovoltaic panels and will also provide tools for a finer prediction of their energy production (hour, day, and week)[119]. All these points need to be addressed to optimize port activities which consume a lot of energy and to establish an energy management plan. Although photovoltaic power generation is booming, there are still many challenges to make it reliable and efficient. One of the main difficulties lies in the fact that there is some uncertainty and variability of the solar resource, mainly due to the variability of weather conditions and especially cloud cover.

To help GPMB, but also all small and medium ports, to better understand their photovoltaic energy production capacity, in PIXEL a methodology will be provided, to predict this kind of production considering the available data, the forecast's objective, the horizon time and the spatial resolution of the prediction. The objective is to provide the capacity for a port to check, whether using renewable energies, can be one of the actors on the EU strategy of reducing emissions and environmental impact of energy use.

In the deliverable D4.1 (PIXEL Models v1.0), first insights of a photovoltaic energy production model have already been addressed. In the following section, there is a more in-depth investigation on what could be done to forecast photovoltaic energy production. This forecasting methodology will be used in task T4.2.

7.1.1. Forecasting scope and objectives

The first step before setting up a methodology for predicting photovoltaic energy is, to define as precisely as possible the objectives of the study. In view of the exchanges with GPMB and previous deliverables of the project, it is suggested to consider the following points in the scope of the prediction:

- To be able to forecast typical solar energy production for the following horizon time: week, month and year. Typical production is based on weather conditions for a week, month, year that are representative of typical weather conditions, so historical data will be used. This corresponds to the need of GPMB to better understand the potential of photovoltaic energy production. This goal is also consistent with the accuracy of the open data currently available.

[118] https://www.meretmarine.com/fr/content/nantes-15000-m2-de-panneaux-solaires-sur-les-entrepots-de-sogebras https://www.greenport.com/news101/energy-and-technology/solar-power-as-an-option-for-decarbonising-terminals

[119] J. M. Morales, A. J. Conejo, H. Madsen, P. Pinson, and M. Zugno, "Integrating renewables in electricity markets - Operational problems," Springer, vol. 205, p. 429, 2014. https://www.springer.com/gp/book/9781461494102

- Since there is no photovoltaic installation inside GPMB, the prediction models that will be put in place, cannot be compared with real production within the port, so to validate this predictive approach, open data will be used. This is expected to have an impact on the accuracy of the models, but the methodology will be easily adaptable.
- The purpose of the prediction is not to provide a fine real-time forecasting for production control. It is rather to provide an estimation of the photovoltaic production profile and what there is to be done, to balance it with the consumption profile. That is why a high precision model is not a requirement.

These prediction objectives are expected to evolve over time, based on the actual data available within the port.

7.1.2.Description of the solar energy prediction problem

The production of solar energy involves many complex physical processes, which are coupled with both uncertainties about our understanding of these phenomena, but also with the intrinsic variability of weather conditions. Indeed, the production of solar energy depends on many factors. Among those identified as having a major impact on production are:

- Real irradiance on photovoltaic panels.
- The climatological context: the efficiency of a panel is effectively directly correlated to the ambient temperature and the wind speed.
- The technical characteristics of photovoltaic panel: the performance of a panel depends on the technology used, the aging laws of the panel, its orientation, its inclination, etc.
- The types of inverters and energy losses related to its transport.

As already described in deliverable D4.1, there are two mains approaches to predict the photovoltaic energy production:

- Use of the irradiance, as an input of the prediction of photovoltaic production. In this approach, irradiance could be directly measured on-site or could be predicted through weather data, satellite imagery, sky image or trough numerical simulation. Then, the irradiance in the plan of PV panels, can be coupled with physical models of the panels, to predict PV energy production.
- Use of historical dataset and contextual data (weather data, sun elevation, etc.). In this approach, only past conditions of production and contextual data to predict the future were considered.

Both approaches have their drawbacks and advantages, while their use depends on the available data (time and spatial resolution, correctness, completeness, etc.), the horizon time of forecasting and the purpose of forecasting. Figure 58 shows the different possibilities for solar energy prediction.

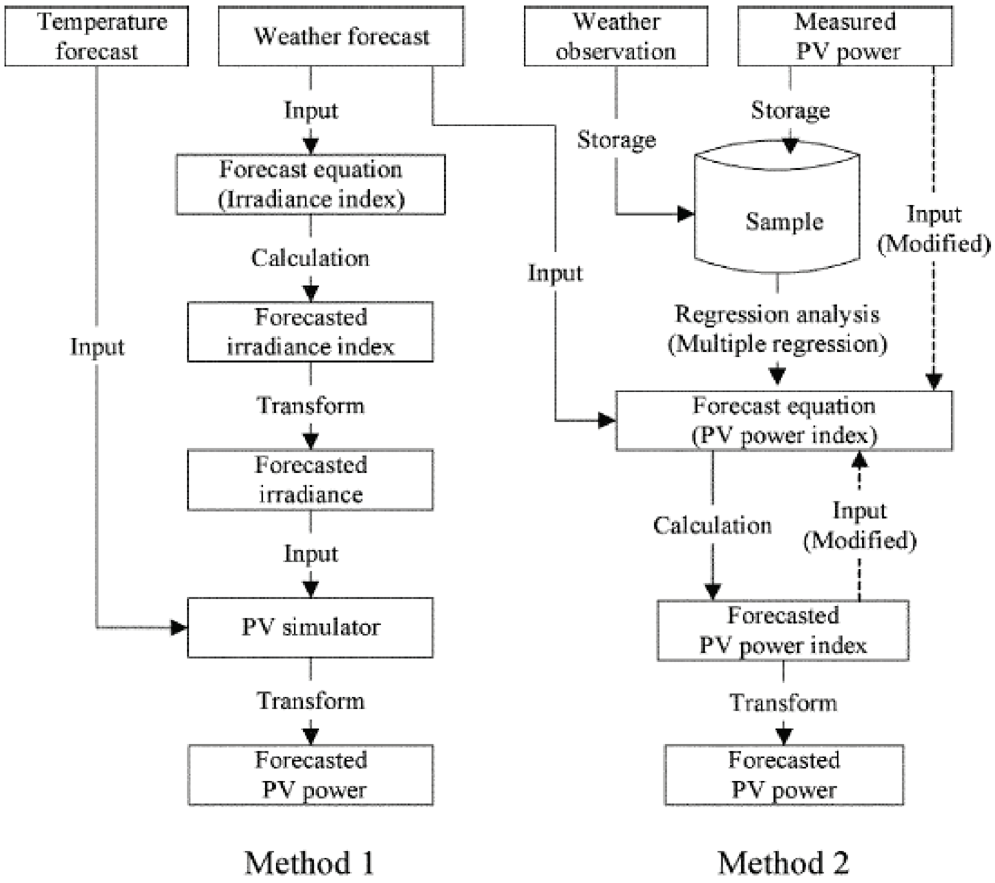


Figure 58. Summary of the different possibility to predict photovoltaic energy production[120]

The followings problems will probably occur when forecasting solar PV production, in the scope of PIXEL:

- There is a lack (or no information at all) on the exact location, orientation of solar PV.
- There is a lack of live data.

These two points can be explained by the fact that small and medium ports are not equipped with PV yet and with an IoT platform, able to store and share data. In the framework of PIXEL, models, predictive algorithms and the PIXEL IoT platform will help to provide an answer.

7.2. Data sources

As previously presented data sources to predict solar energy could be divided in two categories; forecasted data or observed/measured data:

- Temperature forecast
- Weather forecast
- Weather observation
- Measured PV power

A temporal and spatial coherence of the data is the difficulty to be encountered, in the data collection process. Indeed, if, for example, the weather forecast data are obtained via a specialized service and the production

[120] Forecasting Electric Power Generation in a Photovoltaic Power System for an Energy Network, Electrical Engineering in Japan, Vol.167, No. 4, 2009, Misturu Kudo, Akira Takeuchi, Yousuke Nozaki, Hisahito Endo, Jiro Sumita. https://www.jstage.jst.go.jp/article/ieejpes/127/7/127_7_847/_article

measurements are made on the site, it must be ensured that the weather forecasts can be transposed to the studied site. The following types of data sources could be used to forecast solar energy production:

- **Live data:** When there are no big changes in weather conditions and for a very short-term horizon, this kind of data could be used to estimate the solar energy production.
- **Metadata:** Includes generation capacity, technical specifications, orientation data and location data (including shadows).
- **Historical data:** they can be used to fine-tune prediction models. They will be used for example, to forecast energy production for similar weather conditions.

Today, none of the PIXEL ports are equipped with photovoltaic panels and therefore, no real production measurements or historical data from the ports are available. Therefore, the methodology that will be put in place for the prediction of the production of photovoltaic panels, will remain rather generic and mainly based on open data. Although this will not allow a high accuracy on the production, the methodology used will be more easily transposable from one port to another.

7.2.1.Temperature and weather forecast

In the framework of PIXEL, it is not envisaged to set up methods for forecasting the temperature and the different weather conditions. These methodologies involve complex numerical methods and are the subject of many research and entirely dedicated services. However, nowadays, there are many services for obtaining a prediction of meteorological data. For example, OpenWeatherMap provides weather forecast with temperature, wind speed and cloud cover. There is a free API[121] to obtain 5-days / 3-hours forecasts (13 days are available with subscription), but this kind of service (OpenWeatherMap or a similar one) to obtain forecast of weather conditions, will not be used.

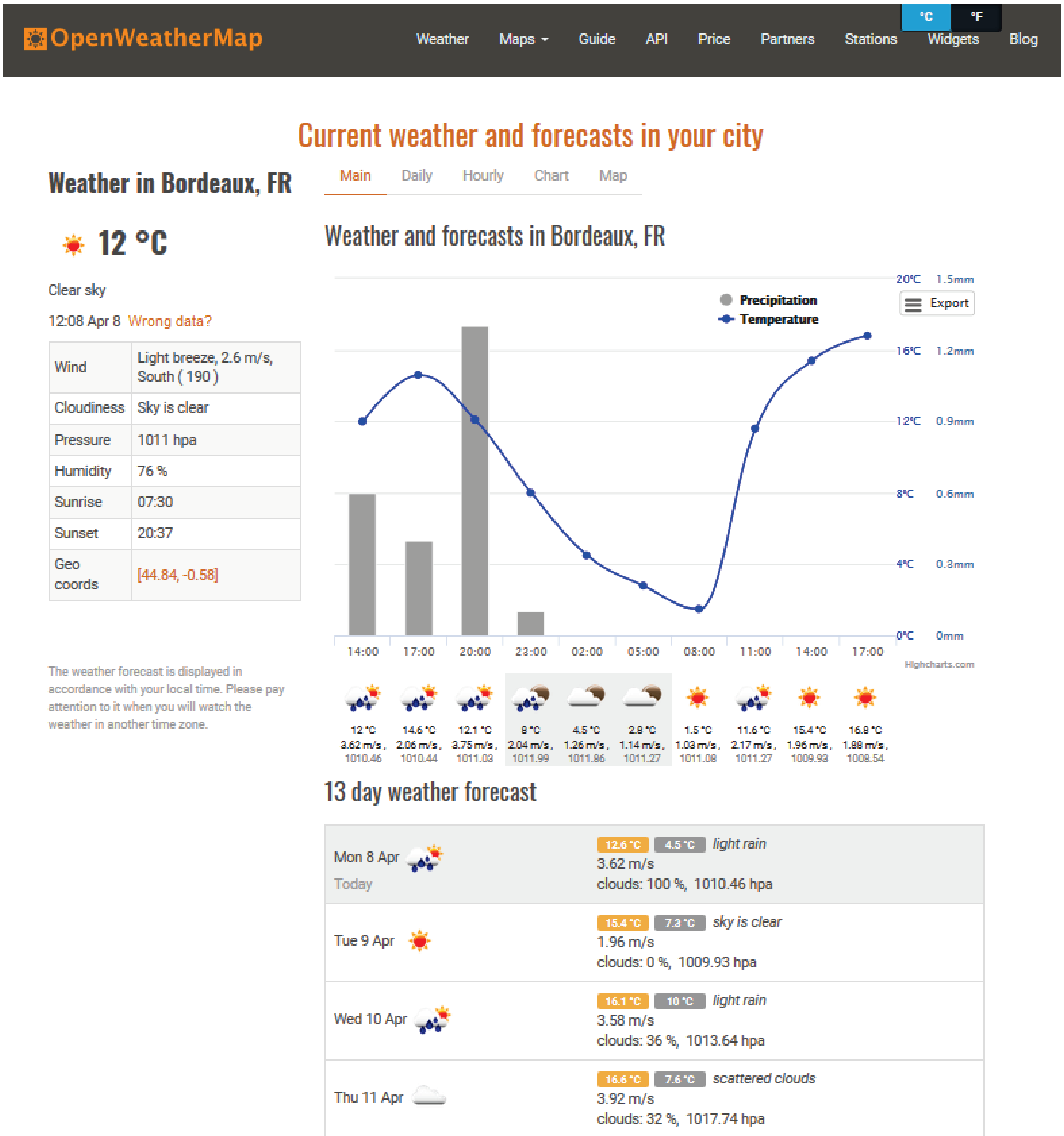


Figure 59. Weather conditions forecast web-service (https://openweathermap.org/api/)

[121] https://openweathermap.org/api

7.2.2.Temperature and weather real-time and historical data set

As presented in deliverable D6.1 (PIXEL Information System Architecture and design v1), weather measurements are available through Météo France (official French weather forecasting service). These data are open data[122]. They have a frequency of 3h and are observational data from international SYNOP (surface synoptic observations) messages, circulating on the GTS (Global Telecommunication System) of the WMO (World Meteorological Organization). The atmospheric parameters measured are temperature, humidity, wind direction and force, atmospheric pressure, precipitation height. The atmospheric parameters observed are sensitive time, cloud description, visibility from the Earth's surface. Depending on instrumentation and local specificities, other parameters may be available (snow depth, soil condition, etc.).

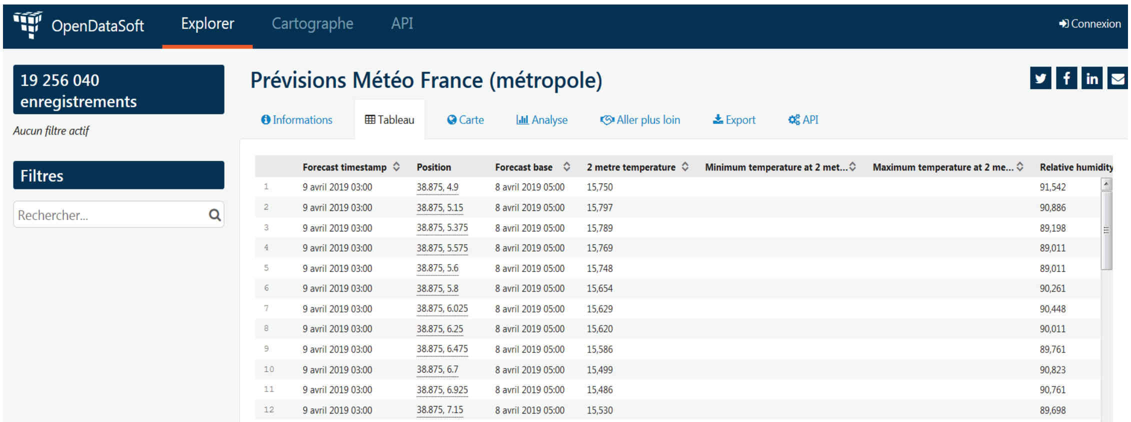


Figure 60. Weather data access through opendatasoft.com[123]

The direct Météo France's open data part (free services) does not seem suitable, since data need to be manually downloaded. On the contrary, subscription services with API are proposed. However, an indirect API access, using opendatasoft.com,[124] may be interesting.

Moreover, according to D3.4 GPMB plans to purchase a weather station. If these data are made available, will be of great interest, since they will directly be reflected in real data on port's premises. It is moreover suggested strongly that this weather station is equipped with a pyranometer to measure real irradiance.

[122] Weather data are accessible following this link: https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=90&id_rubrique=32 The Bordeaux station code is 07510 (filter rows on this ID).

[123] https://public.opendatasoft.com/explore/dataset/arome-0025-sp1_sp2/table/?location=22,44.85002,-0.575&basemap=jawg.streetshttps:%2F%2Fhelp.opendatasoft.com%2Fapis%2Fods-search-v1%2F#search-api-v1

[124] API services for weather data: https://public.opendatasoft.com/explore/dataset/arome-0025-sp1_sp2/table/?location=6,46.00538,2&basemap=jawg.streets https://help.opendatasoft.com/apis/ods-search-v1/#search-api-v1

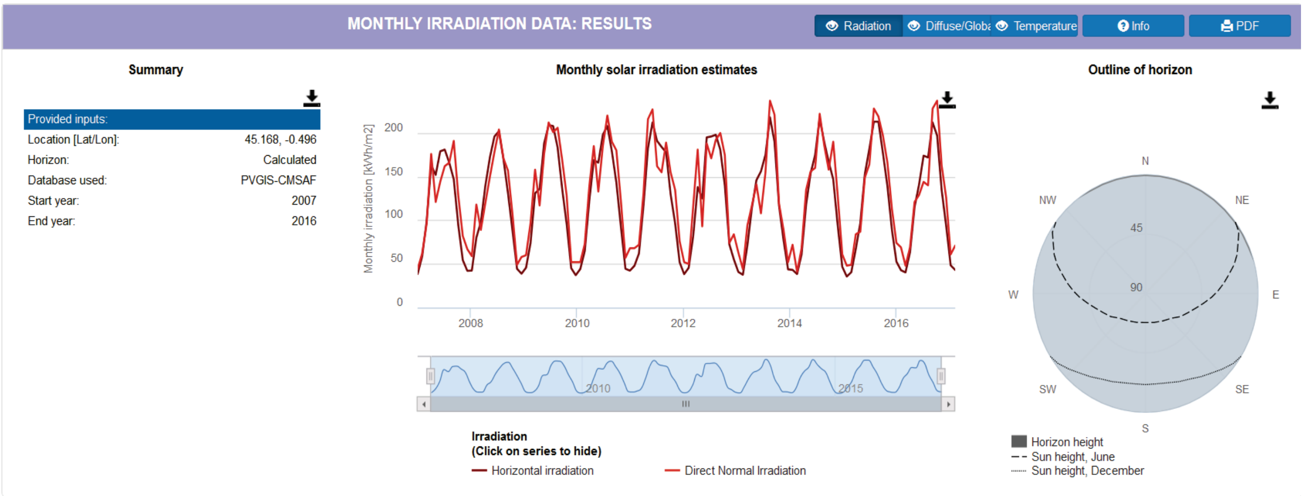


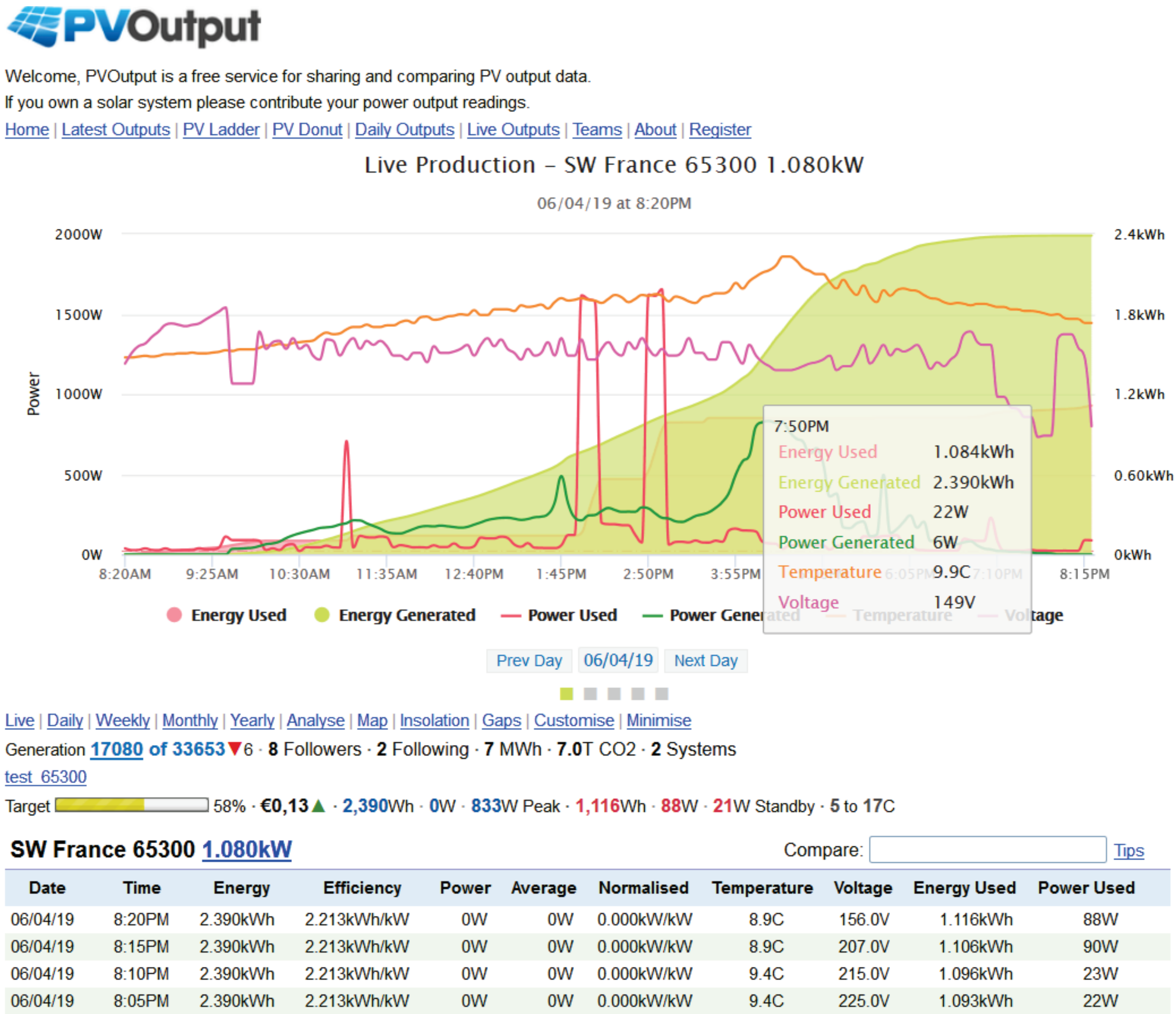
Figure 61. Historical irradiance in Bordeaux from 2007 to 2017[125]

Past and average values of real irradiance can also be obtained through open data portal, such as, CMSAF (Climate Monitoring Satellite Application Facility) or NSRD (National Solar Radiation Database), which can be used for future irradiance estimation, based on average/typical values. Such feature is frequently embedded in tools as PVGIS (PhotoVoltaic Geographical Information System). The data will be utilized, to obtain historical data of irradiance in GPMB.

125 PVGIS http://re.jrc.ec.europa.eu/pvg_tools/en/tools.html

7.2.3.Real-time data and historical data set of PV production

As mentioned before, since the port of Bordeaux does not have photovoltaic panels installed yet, currently there are no real data for solar PV production. However, many real production data are available as "open data"[126]. With this web-service, access can be granted, both to live and historical data of the generated power, the efficiency (that is directly linked with weather conditions: high efficiency means a sunny day) and temperature, with resolution of 5 min in time and from different photovoltaic power plants that are available.



Date	Time	Energy	Efficiency	Power	Average	Normalised	Temperature	Voltage	Energy Used	Power Used
06/04/19	8:20PM	2.390kWh	2.213kWh/kW	0W	0W	0.000kW/kW	8.9C	156.0V	1.116kWh	88W
06/04/19	8:15PM	2.390kWh	2.213kWh/kW	0W	0W	0.000kW/kW	8.9C	207.0V	1.106kWh	90W
06/04/19	8:10PM	2.390kWh	2.213kWh/kW	0W	0W	0.000kW/kW	9.4C	215.0V	1.096kWh	23W
06/04/19	8:05PM	2.390kWh	2.213kWh/kW	0W	0W	0.000kW/kW	9.4C	225.0V	1.093kWh	22W

Figure 62. Open data of a solar system[127].

This data will be used to set up a full methodology for forecasting solar PV production, knowing date and type of weather conditions. The models will be based on past data of a solar PV production and the associated weather conditions. For all the reasons mentioned above, models are not expected to be able to predict solar production in GPMB, but the methodology will be easily transferable.

[126] www.PVOutput.org
[127] https://pvoutput.org./intraday.jsp?id=15556&sid=13412&dt=20190407&gs=0&m=0

7.3. Predictions and analytics for PIXEL

In the following, a short state-of-the-art of solar PV production forecasting is presented. Currently, for the first approach of forecasting, based on the use of irradiance, the main challenge is to obtain the best reliable prediction of the actual irradiance on the panels. This prediction must then take into consideration the influence of cloud cover, atmospheric conditions (humidity rate in particular) and other weather data (like the presence of dust for example).

There is in fact, a very strong variability of the solar resource, making a precise prediction, for all time horizons, almost impossible. To know the irradiance in the plane of the panels, several approaches are possible[128]:

- **Time series prediction with statistical learning methods**: Direct measurement of actual irradiance, using a pyranometer, can be coupled with statistical time series methods, to predict future evolution based on past conditions. In this case the historical irradiance and weather data of the concerned site are used to drive a predictive model and then the real-time meteorological data are used as input for the prediction. In these methods are included: neural networks, regression models, support vector machines, Markov chains, etc. These models show the best results for a prediction horizon of less than one hour but can still provide good results for a time horizon of 2 to 3 hours. This kind of methods could be used in PIXEL, in case irradiance and other weather data inside the port can be measured. However, for now, real data of irradiance is not available and will not be considered in a first approach. However, as GPMB is planning to purchase a weather station, if data are made available soon, a more in-depth exploration of such methods can be conducted.
- **Sky image:** It is possible to use cameras producing high quality images of the sky, to detect clouds and thus, estimate their heights and movements. By this method, it is also possible to categorize the clouds according to their thickness and to differentiate them from dust for example. Consecutive images also make it possible to know the speed of convection of the clouds and thus, to provide a very short-term forecast. However, these prediction methods show a rapid decrease in their performance beyond 30 min. This kind of solutions seem to be out the PIXEL scope, since the purchase of high-quality equipment is required, which are not precise enough for day, as time horizon.
- **Image satellite:** Firstly, a physical model can be used to calculate the irradiance conditions under clear sky conditions for a specific site, by considering data such as elevation, humidity, ozone level, presence of aerosols. This irradiance is then modulated, by the estimated irradiance derived from the satellite images. Several satellite images can be combined to calculate a velocity vector for clouds and predict their future evolution. This method has shown good results for time horizons ranging from minutes to hours. These methods are less relevant in conditions of formation and rapid dissipation of clouds (which is typically the case in the marine environment), however, can be used to obtain historical data of irradiance. In PIXEL this kind of data, provided by PVGIS, will be used, to have a historical data set of solar irradiances.
- **NWP (Numerical weather prediction):** NWP have been used for many years to predict weather conditions. These models showed a good prediction for a time horizon ranging from 6 hours to 2 days. These methods are decrepit in the literature and involve the resolution of Navier-Stokes equations coupled with those of radiative transfer, cloud microphysics, etc. Historically, these methods have been used and parameterized to better predict the temperature, humidity, the probability of rain and the wind speed. Precise prediction of real irradiance is a topic that is being addressed currently. These models are now coupled with statistical learning methods, to correct modelling errors. In PIXEL, this kind of method will not be explored, because it requires lot of computational power and there is a need for fine-tuning the models with precise data, which small and medium ports do not have.

[128] A. Tuohy, J. Zack, S. E. Haupt, J. Sharp, M. Ahlstrom, S. Dise, E. Grimit, C. Mohrlen, M. Lange, M. G. Casado, J. Black, M. Marquis, and C. Collier, "Solar Forecasting: Methods, Challenges, and Performance," IEEE Power Energy Mag., vol. 13, no. 6, pp. 50–59, 2015. https://ieeexplore.ieee.org/abstract/document/7299804

There is thus a variety of methods of predicting irradiance, whose use depends on: the prediction horizon, the available data for the prediction and the end use of this prediction. This method is synthesized in Figure 63 on an axis, translating the horizon of prediction.

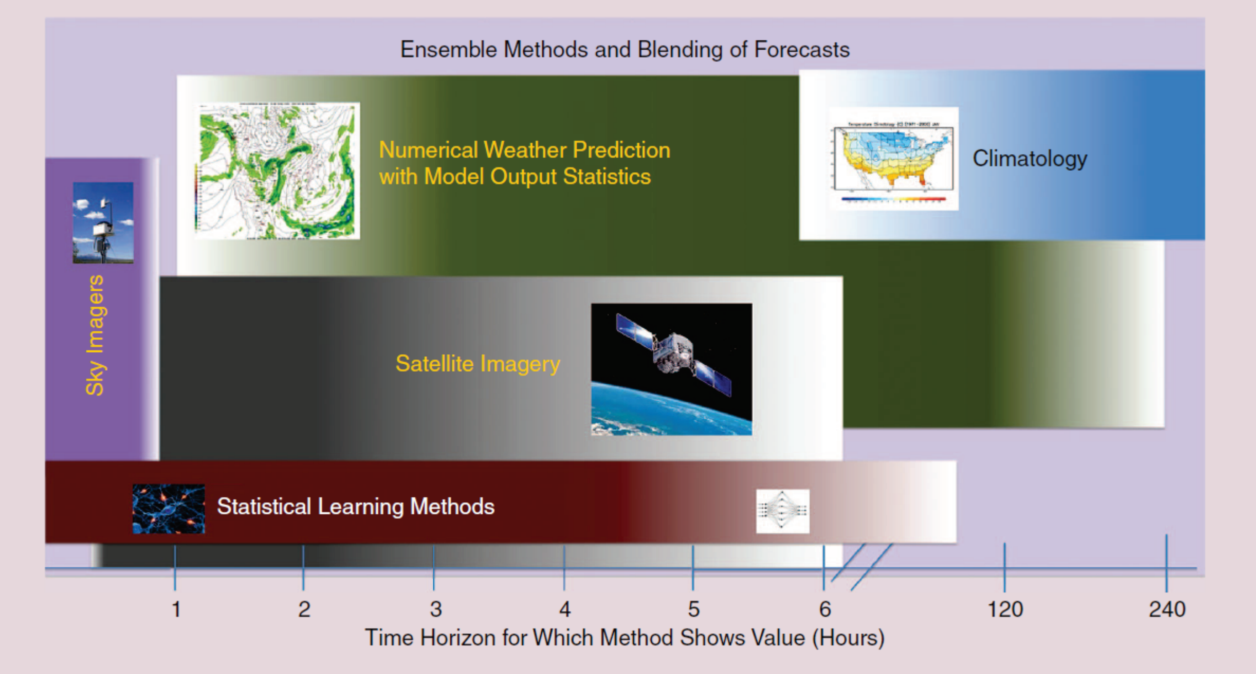


Figure 63. Overview of solar forecasting methods based on time horizon for which they show values [129]

For the second approach based on the use of historical data of PV production, statistical learning methods are used. There is a wide range of statistical methods, from simple persistent model to complex hybrid neural network. Each of those models have strong points and weaknesses (see Das 2018[130] for an extensive review with pertinent consideration to available data and forecast horizon).

[129] A. Tuohy, J. Zack, S. E. Haupt, J. Sharp, M. Ahlstrom, S. Dise, E. Grimit, C. Mohrlen, M. Lange, M. G. Casado, J. Black, M. Marquis, and C. Collier, "Solar Forecasting: Methods, Challenges, and Performance," IEEE Power Energy Mag., vol. 13, no. 6, pp. 50–59, 2015. https://ieeexplore.ieee.org/abstract/document/7299804

[130] Das, Utpal Kumar, Tey, Kok Soon, Seyedmahmoudian, Mehdi, Et Al. Forecasting Of Photovoltaic Power Generation And Model Optimization: A Review. Renewable and Sustainable Energy Reviews, 2018, Vol. 81, P. 912-928. https://ieeexplore.ieee.org/abstract/document/8448633

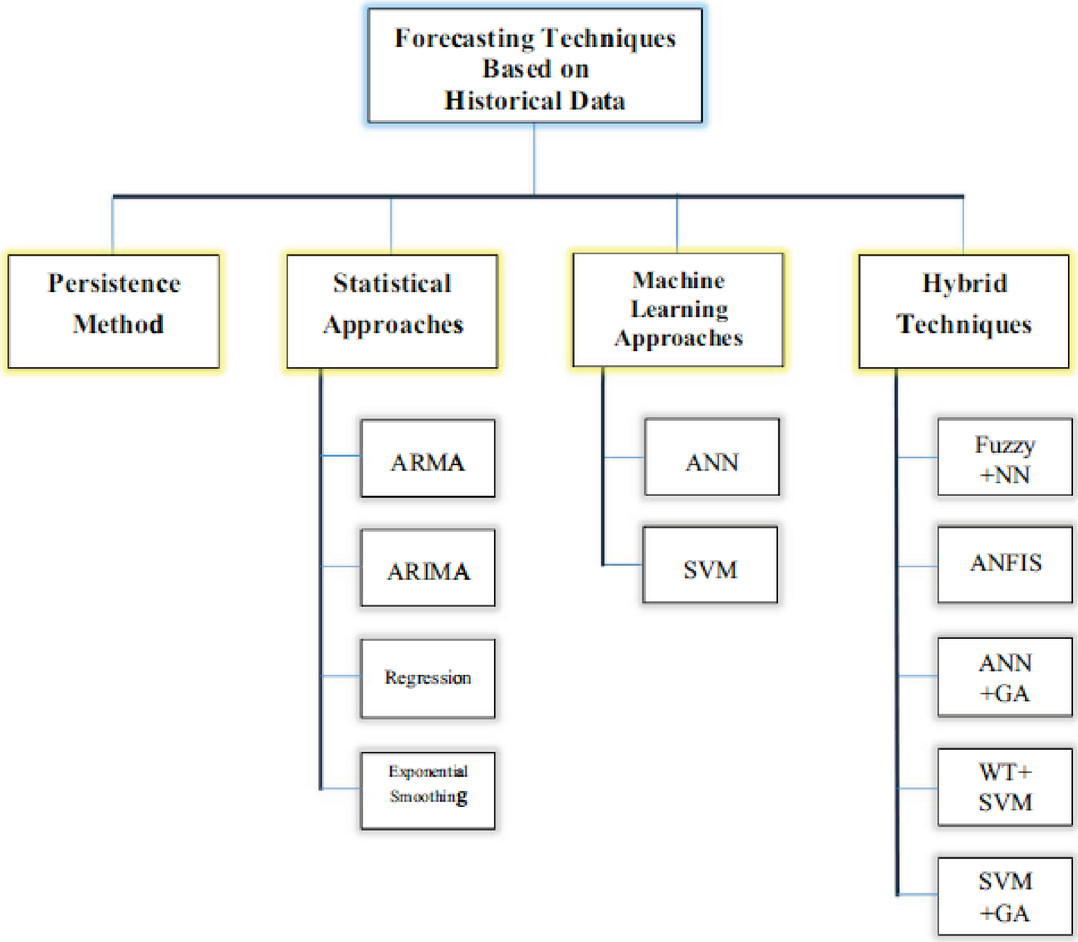


Figure 64. Classification of PV power forecasting based on historical data (Das 2018)

7.4. Initial work

To build a statistical model of the solar energy production for a specific photovoltaic installation, historical datasets are required. They correspond to three elements observed at the same time: the extra-terrestrial radiation, the weather conditions and the corresponding solar energy production.

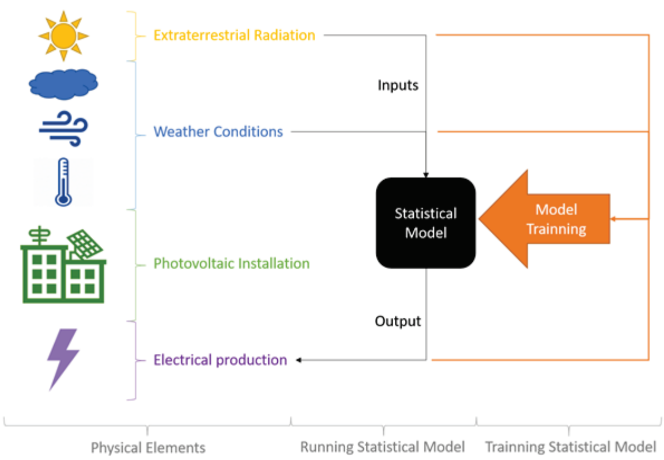


Figure 65. A statistical model can predict a photovoltaic installation production for a given date and weather forecast.

7.4.1. Photovoltaic production (historical data)

This data-set corresponds to the recorded energy production of the PV system. It covers a period of 2.281 days (corresponding to 81 months or 6.8 years), from the 14th of December 2012 to the 5th of March 2019. For each day, the energy generated (kWh), is available with a 5/10 min time resolution. Data are available for every day, with only few missing values for some hours.

Daily Energy Generated (Wh)	
# Days Recorded	2281
Mean	2810
Std	1669
Min	0
25%	1371
50%	2825
75%	4278
Max	6170

Table 6. Daily photovoltaic production data-set summary

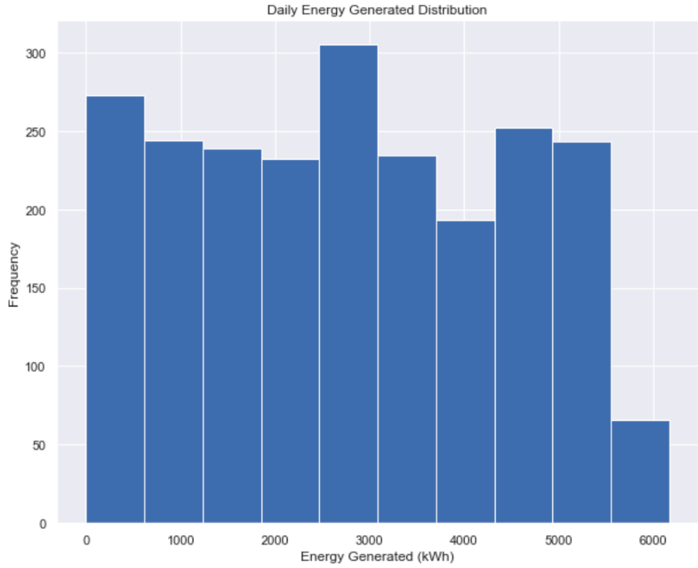


Figure 66. Photovoltaic daily production distribution

7.4.2. Solar irradiance (historical data)

For a known time and position on Earth, the solar irradiance can be calculated to produce a historical dataset. Note that those values are theoretical and do not correspond to the actual radiation flux received at the photovoltaic panel, since atmosphere reduces it. The solar irradiance data are available for every energy generated data point.

Daily Insolation (Wh)	
# Days Recorded	2281
Mean	5393
Std	1246
Min	3182
25%	4134
50%	5848
75%	6531
Max	6819

Table 7. The extra-terrestrial radiation data-set summary

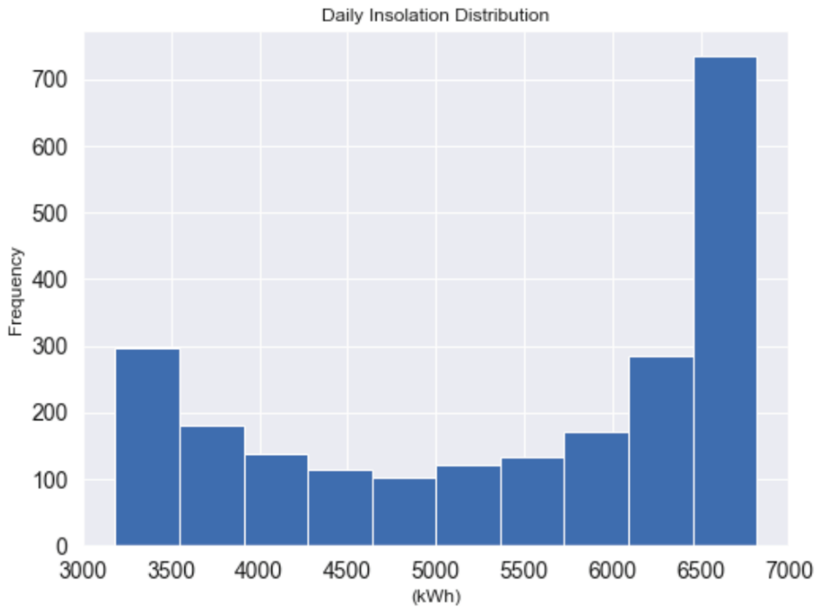


Figure 67. The extra-terrestrial radiation distribution

7.4.3. Weather conditions (historical data)

Weather condition has two major impacts. On the one hand, cloud coverage has a strong impact on the effective irradiance reaching the photovoltaic panel, and thus the energy generated. On the other hand, indirect effects can occur, as the rainfall removing dust from panels for example ("shower" effect). Furthermore, the temperature (and thus wind) impacts the panels' energy conversion rate. Both cloud coverage and temperature are available on this historical dataset.

Weather Condition		
Cloud Cover	# Days	Temp. (°C)
Fine	963	25.8
Partly Cloudy	402	23.2
Mostly Cloudy	315	20.9
Cloudy	311	17.4
Showers	241	11.8
Not Sure	49	22.8

Table 8. The weather conditions data-set summary

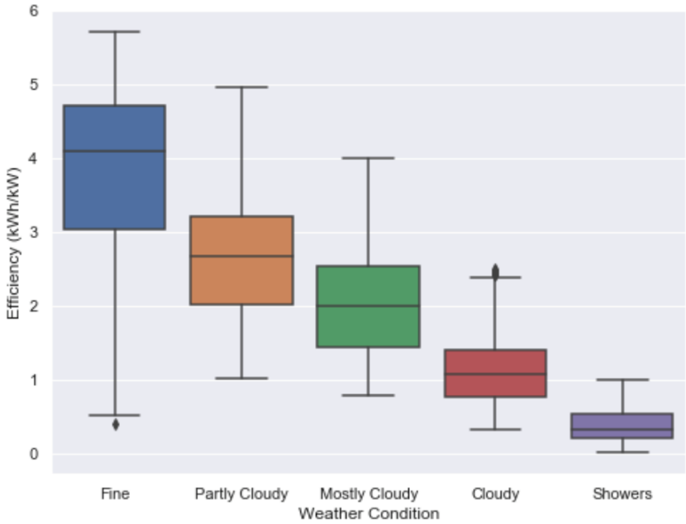


Figure 68. Relation between weather conditions and PV system efficiency

7.5. Plan and future work

In view of the available data, the following approaches will be used to forecast solar energy production:

- Based on historical irradiance data and associated weather conditions, obtained either by measurement or by satellite-based tools (PVGIS), a full methodology will be proposed, to predict one-point irradiance for a time horizon from a day to year. This prediction will in fact reflect a typical day, week, month, year based on past data. The main work here will be to interact with web services like PVGIS, to obtain historical data and extract a typical irradiance.
- Then, well-known physical models of solar panels (based on technical specifications such as orientation or panel technology) will be used, to predict a typical photovoltaic production. This first work will be useful for small and medium ports, to easily estimate their potential of PV production and a have first idea of their production profile.
- Based on historical production data, mainly obtained as open data, predictive algorithms for photovoltaic production based on past data of real production and associated weather conditions, will be implemented. Even if the models will not be directly designed for a PV system in site, the full methodology will be reproducible. The objective here is to provide ports with a tool to estimate their production, based on their real installation.

8. Conclusions

In this deliverable, the outline of the work that has been performed in T4.5, as well as future work regarding the use of predictive algorithms in PIXEL project, was presented. Overall, it was concluded that there is plenty of data captured in the ports for different operational processes; however, this data is underutilized or even not utilized at all, to provide any operational insights. What is even a more important finding is that, there is a lack of awareness in smaller sized ports about the importance of properly capturing, storing and creating added value products from operational data. This awareness could help ports at optimizing their operational efficiency in different domains that create impacts on their operational performance but also on the environment.

The use of emerging technologies such as AI, is one of the areas where the gap between large and small ports is especially noticeable. Larger ports are increasingly investing in AI based solutions. In addition to acquiring solutions that can already be found on the market and quickly applied, they also invest in research, to find new application areas and means to address them. The potential is huge and the problems that are addressed by larger ports are ambitious and challenging. Some of ports have introduced AI in automated loading cranes, to decide which containers to process first, along with the data about the shipments to manage entire delivery cycles (optimally with the use of autonomous trucks) and predictive maintenance of port equipment, to make the delivery chain more reliable. Some of the ports are addressing water and weather conditions, to better utilize their berths and to predict visibility, among other parameters, to determine the smoothness of ship entry, which results in lower fuel consumption and environmental impacts. AI has reached the world of transport, shipping industry and ports and is here to stay.

The solutions that were proposed in this deliverable, are more cost-effective when compared to solutions in larger ports, considering the limited resources of smaller ports and feasibility of implementation. The main goal is to utilize internal data that is captured, stored and available in ports or by other stakeholders. Data will be combined with external sources that can be obtained affordably or free-of-charge. Additionally, the use of advanced emerging technologies for analytics such as AIS data and satellite imagery, already used by some of the larger ports was introduced, however, in the scope that is applicable and affordable for smaller ports.

According to this, predictive tasks were identified and presented to support real-life scenarios and PIXEL objectives, while at the same time, closing the gap between small and large ports and helping them to make a step towards the Port of the Future.

Appendix 1: Data sources description

Dataset name	**GPMB vessel call data**
Data Source	Project partner CATIE provided the data owned by GPMB. Data has been shared through consortium document collaboration platform.
Description	7 years (2010-2017) of vessel call data for GPMB. Name of the ship, type of cargo, entry and exit times and amount of cargo is provided.
Usage in PIXEL	Analysis and prediction of vessel calls: (1) General statistical analysis and visualisations of the provided data; (2) Predicting call time; (3) Predicting vessel call.
Algorithms	Different supervised machine learning methods for regression, time series analysis and classification.
Sharing of results	Results initially available to project partners. External use to be decided. Data available to all project partners.
License and terms of use	Not specified by the data provider that made the data available to the consortium
DPO assessment (XLAB)	Considering the GDPR (General Data Protection Regulation) definition of personal data and considering that the dataset relates to cargo vessels (which are large ships, owned by legal entities and not individuals), this dataset only includes non-personal data: vessel name, cargo type, entry/exit times, cargo amount. These parameters cannot be used to, by reasonable means, directly or indirectly identify any data subject. The same holds for the 'IMO number' parameter. **The dataset can be processed as indicated above without any limitation, considering that the processing is in line with the data sharing agreement or licence of the dataset owner/provider.**

Dataset name	**PPA vessel call data**
Data Source	"PPA provided sample data for the last 3 months of 2018.
Description	Similar fields provided as with GPMB data except: IMO number, berthing area, flag.
Usage in PIXEL	Analysis of vessel calls: Impact of vessel calls on road traffic (congestion at port).
Algorithms	Different supervised machine learning methods for regression, time series analysis and classification.
Sharing of results	Results initially available to project partners. External use to be decided. Data available to all project partners.
License and terms of use	Not specified by the data provider that made the data available to the consortium
DPO assessment (XLAB)	Considering the GDPR definition of personal data and considering that the dataset relates to cargo vessels (which are large ships, owned by legal entities and not individuals), this dataset only includes non-personal data: vessel name, cargo type, entry/exit times, cargo amount, IMO number, berthing area, flag. These parameters cannot be used to, by reasonable means, directly or indirectly identify any data subject. **The dataset can be processed as indicated above without any limitation, considering that the processing is in line with the data sharing agreement or licence of the dataset owner/provider.**

Dataset name	ThPA vessel call data
Data Source	ThPA (data source). They sent sample data for 2-3 months and are prepared to share more historical data.
Description	Provided fields are: voyage code, vessel name (no IMO provided), origin and destination port, arrival and departure time, scheduled arrival and departure date.
Usage in PIXEL	Analysis of vessel calls: Impact of vessel calls on road traffic (congestion at port).
Algorithms	Algorithms used will relate to road traffic, but will consider vessel calls as input parameter
Sharing of results	Results initially available to project partners. External use to be decided. Data available to all project partners.
License and terms of use	Not specified by the data provider that made the data available to the consortium
DPO assessment (XLAB)	Considering the GDPR definition of personal data and considering that the dataset relates to cargo vessels (which are large ships, owned by legal entities and not individuals), this dataset only includes non-personal data: vessel name, origin port, destination port, arrival date, departure date. These parameters cannot be used to, by reasonable means, directly or indirectly identify any data subject. **The dataset can be processed as indicated above without any limitation, considering that the processing is in line with the data sharing agreement or licence of the dataset owner/provider.**

Dataset name	ASPM vessel call data
Data Source	Project partner INSIEL supported ASPM to provide the data. Data has been shared through consortium document collaboration platform.
Description	Data about the vessel calls in ASPM for the year 2018. Compared to other vessel call data, the following data is additionally provided: mooring time, flag of the ship, draught, length of the ship, (ETD estimated time of departure), last and next port of call, shipping agent.
Usage in PIXEL	Analysis and prediction of vessel calls: (1) General statistical analysis and visualisations of the provided data; (2) Predicting call time; (3) Predicting vessel call.
Algorithms	Different supervised machine learning methods for regression, time series analysis and classification.
Sharing of results	Results initially available to project partners. External use to be decided. Data available to all project partners.
License and terms of use	Not specified by the data provider that made the data available to the consortium
DPO assessment (XLAB)	Regarding the vessel call data, we should consider the fact that, by law, only large vessels (ships of more than 300 gross tonnage, passenger ships, and fishing vessels with a length of above 15 metres) are required to carry an AIS transponder that makes calls to the ASPM dataset. These vessels are owned by legal entities and not natural persons. **Hence, data from these large vessels do not contain any personal information.** Other, smaller vessels (including those, which could be owned by natural persons) are not required to be fitted with AIS. If they are fitted with AIS, they provide data to the maritime authorities on voluntary basis considering and accepting their terms of use and privacy policies (which explain that the collected data are available as open data). Nevertheless, since no direct personal data relating to natural persons are included in the AIS calls (only vessel ID), re-identification of natural persons would require significant (financial, technological, time-wise) resources. We can assume that, currently, such means would not reasonably likely to be used. According to the Recital 26 of the GDPR, we can conclude that, **even in the case of smaller vessels, this dataset is anonymized and the GDPR does not apply.**

	Considering also the fact that the shipping agents are legal entities and not natural persons, we can conclude that this **dataset does not contain any personal data. The dataset can thus be processed as indicated above without any limitation as long as the processing is aligned with the data sharing agreement of the licence of the data provider.**

Dataset name	**DEBS 2018 Challenge dataset**
Data Source	BigDataOcean EU project was contacted as an organizer of this challenge and the data was provided by one of its project partners. http://www.cs.otago.ac.nz/debs2018/calls/gc.html
Description	Data that was used for the DEBS 2018 Challenge (12th ACM International Conference on Distributed and Event-Based Systems, 25 – 29th June 2018, Hamilton, NZ). Data was prepared out of AIS (Automatic identification system) data from the Mediterranean Sea. Only relevant fields are provided such as GPS (Global Positioning System) locations of the ship along the route, departure and arrival ports. Data regarding ship is anonymized. More information about DEBS 2018: https://dl.acm.org/citation.cfm?id=3220510
Usage in PIXEL	DEBS 2018 data will be mainly used to show-case ETA prediction for the ports. Same task was required from challenge participants.
Algorithms	Different supervised machine learning methods for regression, time series analysis and classification.
Sharing of results	Results initially available to project partners. External use to be decided. Data only available internally.
License and terms of use	https://hobbitdata.informatik.uni-leipzig.de/DEBS_GC_2018/BDO-Data%20Access%20and%20License%20agreement%20for%20academic%20purposes%202018.pdf
DPO assessment (XLAB)	According to the data description above and the one provided on the DEBS2018 website (http://www.cs.otago.ac.nz/debs2018/calls/gc.html), this dataset only includes non-personal data: ports locations, anonymized ship ID, ship type, ship speed, ship position (longitude and latitude), ship direction (course and heading), timestamp, departure port name, draught. **The dataset can be processed as indicated above without any limitation, considering that the processing is in line with the data sharing agreement or licence of the dataset owner/provider.**

Dataset name	**DMA (Danish Maritime Authority) AIS data**
Data Source	Danish Maritime Authority https://www.dma.dk/SikkerhedTilSoes/Sejladsinformation/AIS/Sider/default.aspx# ftp://ftp.ais.dk/ais_data/
Description	Historical AIS data for Danish coastal waters. AIS data is already decoded and provided in csv files with all the fields present. Compared to DEBS 2018 data, this includes IMO and MMSI numbers, as well as ship name and call sign.
Usage in PIXEL	This data will be used to support tasks regarding the use of AIS data, mainly in port area. Usability for short-term ETA prediction will also be investigated.

Algorithms	Different supervised machine learning methods for regression, time series analysis and classification.
Sharing of results	Results initially available to project partners. External use to be decided.
License and terms of use	https://www.dma.dk/SikkerhedTilSoes/Sejladsinformation/AIS/Sider/AIS%20datamanagementpolitik.aspx http://ec.europa.eu/newsroom/document.cfm?doc_id=1262 (PSI act)
DPO assessment (XLAB)	By law, only large vessels (ships of more than 300 gross tonnage, passenger ships, and fishing vessels with a length of above 15 metres) are required to carry an AIS transponder. These vessels are owned by legal entities and not natural persons. **Hence, data from these large vessels do not contain any personal information.** Other, smaller vessels (including those, which could be owned by natural persons) are not required to be fitted with AIS. If they are fitted with AIS, they provide data to the maritime authorities on voluntary basis considering and accepting their terms of use and privacy policies (which explain that the collected data are available as open data: https://www.dma.dk/SikkerhedTilSoes/Sejladsinformation/AIS/Sider/AIS%20datamanagementpolitik.aspx). Nevertheless, since no direct personal data relating to natural persons are included in the AIS calls (only vessel ID), re-identification of natural persons would require significant (financial, technological, time-wise) resources. We can assume that, currently, such means would not reasonably likely to be used. According to the Recital 26 of the GDPR, we can conclude that, even in the case of smaller vessels, this dataset is anonymized and the GDPR does not apply. **The dataset can thus be processed as indicated above without any limitation as long as the processing is aligned with the licence of the data provider.**

Dataset name	**U.S. Coast Guard AIS data**
Data Source	https://marinecadastre.gov/ais/
Description	Historical AIS data for USA coastal waters. AIS data is already decoded and provided in csv files with all the fields present. Compared to DEBS 2018 data, this includes IMO and MMSI numbers, as well as ship name and call sign.
Usage in PIXEL	This data will be used to support tasks regarding the use of AIS data, mainly in port area. Usability for short-term ETA prediction will also be investigated.
Algorithms	Different supervised machine learning methods for regression, time series analysis and classification.
Sharing of results	Results initially available to project partners. External use to be decided.
License and terms of use	https://www.navcen.uscg.gov/?pageName=NAISdisclaimer
DPO assessment (XLAB)	By law, only large vessels (ships of more than 300 gross tonnage, passenger ships, and fishing vessels with a length of above 15 metres) are required to carry an AIS transponder. These vessels are owned by legal entities and not natural persons. **Hence, data from these large vessels do not contain any personal information.** Other, smaller vessels (including those, which could be owned by natural persons) are not required to be fitted with AIS. If they are fitted with AIS, they provide data to the maritime authorities on voluntary basis considering and accepting their terms of use and privacy policies (which explain that the collected data can be shared with other organisations). Nevertheless, since no direct personal data relating to natural persons are included in the AIS calls (only vessel ID), re-identification of natural

	persons would require significant (financial, technological, time-wise) resources. We can assume that, currently, such means would not reasonably likely to be used. According to the Recital 26 of the GDPR, we can conclude that, **even in the case of smaller vessels, this dataset is anonymized and the GDPR does not apply.** **The dataset can thus be processed as indicated above without any limitation as long as the processing is aligned with the licence of the data provider.**

Dataset name	**AISHub data**
Data Source	http://www.aishub.net/
Description	AIS data sharing platform. To receive data, one needs to connect own raw AIS data feed. Mostly amateur AIS receiving stations are present, all over the world. The data is provided through an API (decoded, processed AIS data) and through a dedicated TCP (Transmission Control Protocol) port (raw NMEA AIS data). Information that is captured is the same as provided with historical datasets already presented. This data will be used to support tasks regarding the use of AIS data, mainly in the port area and for short-term ETA prediction.
Usage in PIXEL	This data will be used to support tasks regarding the use of AIS data, mainly in port area and for short-term ETA prediction.
Algorithms	Different supervised machine learning methods for regression, time series analysis and classification.
Sharing of results	Results initially available to project partners. External use to be decided.
License and terms of use	http://www.aishub.net/
DPO assessment (XLAB)	By law, only large vessels (ships of more than 300 gross tonnage, passenger ships, and fishing vessels with a length of above 15 metres) are required to carry an AIS transponder. These vessels are owned by legal entities and not natural persons. **Hence, data from these large vessels do not contain any personal information.** Other, smaller vessels (including those, which could be owned by natural persons) are not required to be fitted with AIS. If they are fitted with AIS, they provide data to the maritime authorities on voluntary basis considering and accepting their terms of use and privacy policies (which explain that the collected data can be shared with other entities). Nevertheless, since no direct personal data relating to natural persons are included in the AIS calls (only vessel ID), re-identification of natural persons would require significant (financial, technological, time-wise) resources. We can assume that, currently, such means would not reasonably likely to be used. According to the Recital 26 of the GDPR, we can conclude that, **even in the case of smaller vessels, this dataset is anonymized and the GDPR does not apply.** **The dataset can thus be processed as indicated above without any limitation as long as the processing is aligned with the licence of the data provider.**

Dataset name	**Thessaloniki car fleet data**
Data Source	Historical data: http://opendata.imet.gr/dataset/fcd-gps-historical Live feed: http://opendata.imet.gr/dataset/fcd-gps
Description	Car fleet equipped with GPS providing location and speed information. Each observation cannot be linked to a specific car (no car id provided).
Usage in PIXEL	Correlation of port operations (e.g. through vessel call data) and regional road network for analysis and prediction purposes.

Algorithms	Different supervised machine learning methods for regression, time series analysis and classification.
Sharing of results	Results initially available to project partners. External use to be decided.
License and terms of use	Citation is required: http://opendata.imet.gr/about. At the bottom of the website, link to open data license is provided: https://opendefinition.org/od/2.1/en/
DPO assessment (XLAB)	This dataset specifies the following parameters: Timestamp, GPS coordinates (longitude, latitude, and altitude), speed, and orientation. This dataset may contain personal information. Namely, even though there are no car IDs provided, GPS coordinates may reveal how a specific car travels from point A to point B, where each of these points A or B may belong to a residential address. In this sense, it may happen, that one could track how an identifiable natural person moves. Nevertheless, such re-identification of individuals would require significant (financial and technological) resources and time, and we can assume that, currently, such means would not reasonably likely be used. **According to the Recital 26 of the GDPR, we can conclude that the dataset is anonymized and the GDPR does not apply. The dataset can be freely processed without any limitation, given that the processing is in line with the license of the dataset owner/provider.**

Dataset name	**Traffic data form the Thessaloniki stationary sensor network**
Data Source	http://opendata.imet.gr/dataset?tags=Thessaloniki
Description	Besides car fleet data, stationary sensor network is also available (traffic counters, speed sensors) and other product derivatives.
Usage in PIXEL	Correlation of port operations (e.g. through vessel call data) and regional road network for analysis and prediction purposes.
Algorithms	Different supervised machine learning methods for regression, time series analysis and classification.
Sharing of results	Results initially available to project partners. External use to be decided.
License and terms of use	Citation is required: http://opendata.imet.gr/about. At the bottom of the website, link to open data license is provided: https://opendefinition.org/od/2.1/en/
DPO assessment (XLAB)	DPO checked the following datasets: • **network-speed**, current speeds on OpenStreetMap network links (http://opendata.imet.gr/dataset/network-speed): This dataset specifies the following parameters: link ID, link direction, timestamp, speed. This dataset does not contain any personal data. It can be processed as indicated above without any limitation, considering that the processing is in line with the licence of the dataset owner/provider. • **network-congestion**, the current traffic conditions on OpenStreetMap network links (http://opendata.imet.gr/dataset/network-congestion): This dataset specifies the following parameters: link ID, link direction, timestamp, congestion. • **itravel-devices**, iTravel devices characteristics (http://opendata.imet.gr/dataset/itravel-devices): This dataset specifies the following parameters: device ID, device name, GPS coordinates (longitude, latitude). • **itravel-detections**, current total number of detections per iTravel device (http://opendata.imet.gr/dataset/itravel-detections): This dataset specifies the following parameters: device ID, number of records, timestamp. • **itravel-paths**, predefined paths between the iTravel devices (http://opendata.imet.gr/dataset/itravel-paths): This dataset specifies the

	following parameters: path ID, path name, path origin device ID, path destination device ID, polyline (coordinates for the track between origin and destination devices). • **itravel-traveltimes**, current travel times for selected paths in Greece produced using iTravel directions (http://opendata.imet.gr/dataset/itravel-traveltimes): This dataset specifies the following parameters: path ID, timestamp, duration. Considering the GDPR definition of personal data, these datasets only contain non-personal data. **The datasets can be processed as indicated above without any limitation, considering that the processing is in line with the licence of the dataset owner/provider.**

Dataset name	Thessaloniki port traffic data
Data Source	ThPA (data source). Sample data for 1 day was sent and are prepared to share more historical data. However, real time data collection is more of interest.
Description	List of vehicles entering and leaving the port, including gate number and timestamps (so as dwell time). Registration plates, RFID (Radio-frequency identification) and company have been excluded.
Usage in PIXEL	Correlation of timing operations to detect congestion at the gates
Algorithms	Both real time estimation and predictive algorithms
Sharing of results	Results initially available to project partners. External use to be decided.
License and terms of use	Not specified by the data provider that made the data available to the consortium
DPO assessment (XLAB)	This dataset specifies the following parameters: vehicle type, marker, model, colour, entry time, entry gate, exit time, exit gate, dwell time **According to the Recital 26 of the GDPR, we can conclude that the dataset is anonymized and the GDPR does not apply. The dataset can be freely processed without any limitation, given that the processing is in line with the license of the dataset owner/provider.**

Dataset name	**SILI system traffic data**
Data Source	Data provided by INSIEL - project partner, behind authorization of ASPM, as data owner.
Description	Sample data was provided by INSIEL for one of the gates that is monitored by a traffic camera connected to the SILI system. The data consists of the following fields: gate name and direction of the traffic, lane identifier, exact date and time, license plate and nationality extracted from the license plate, vehicle type.
Usage in PIXEL	This will be used for short-term traffic volume prediction and correlation with operations in the port. License plate information is available in the data and will be used to identify same vehicles among different locations, for travel time computation and congestion rate estimation.
Algorithms	Different supervised machine learning methods for regression, time series analysis and classification.
Sharing of results	Results initially available to project partners. External use to be decided.
License and terms of use	Not specified by the data provider that made the data available to the consortium.
DPO assessment (XLAB)	This dataset contains personal information. Namely, vehicle registration plate number can be used to, directly or indirectly, identify a natural person (e.g., a vehicle owner) and is, as such, according to Article 4.1 of the GDPR, considered as personal data. The indicated processing is (according to Article 6.1.f of the GDPR) lawful because it is in the project's (and, specifically, XLAB's) legitimate economic interest to conduct the research outlined in the Grant Agreement, signed by the European Commission and the project partners (including XLAB).

	Nevertheless, such processing does not pose any high risks to the freedoms and rights of natural persons (the processing does not produce any legal or similarly significant effects on the data subjects, there are no special categories of personal data included, there is no large-scale monitoring included, significant resources would be required for the re-identification of natural persons). In contrast, the research results stemming from such processing activities will bring significant benefits to the society and are thus considered to be in the public interest and in the interest of data subjects. Appropriate technical and organisational safeguards have been put in place to ensure adequate protection of personal data (including strict access control and local processing on private, secure infrastructure – XLAB has the ISO/IEC 27001 certificate). Finally, the collected data are processed exclusively for scientific purposes in accordance with Article 89.1 of the GDPR. It should be noted that for the fulfilment of the processing purpose (i.e. to conduct the research outlined in the Grant Agreement, which includes the development of prediction models for which vehicle registration plate numbers are necessary, as explained above), the dataset should not be anonymised. **Considering these arguments, the dataset can be processed as indicated above as long as the processing is aligned with the Data Protection Agreement between the data provider (INSIEL) and the data processor (XLAB), which governs the processing in accordance with the GDPR.**

Dataset name	**ESA Sentinel satellite imagery (Sentinel Hub)**
Data Source	ESA Sentinel imagery provided through external provider Sentinel Hub
Description	Satellite imagery provided by ESA and further distributed through external provider Sentinel Hub, which offers API for simplified access.
Usage in PIXEL	ESA Sentinel imagery will be used to develop methods for ship detection and classification from satellite imagery. Sentinel-1 (SAR) and mostly Sentinel-2 (optical) imagery will be used. Satellite imagery will also be used for data fusion with AIS data.
Algorithms	To train different CNN based object detection and classification methods.
Sharing of results	Results initially available to project partners. External use to be decided.
License and terms of use	ESA Sentinel data: https://sentinel.esa.int/documents/247904/690755/Sentinel_Data_Legal_Notice Sentinel Hub: https://sentinel-hub.com/tos
DPO assessment (XLAB)	This dataset does not include any personal information. The dataset can thus be processed as indicated above without any limitation as long as the processing is aligned with the licence of the data provider.

Dataset name	**Planet Labs satellite imagery**
Data Source	We will use openly available satellite imagery over California - OpenCalifornia. https://www.planet.com/products/open-california/
Description	Satellite imagery provided by PlanetScope (Dove) constellation of satellites with 3m resolution and daily revisit time (14-day delay in case of OpenCalifornia). Only optical imagery is available.
Usage in PIXEL	Planet Labs imagery will be used to develop methods for ship detection and classification from satellite imagery. Increased resolution (3m) and daily revisit time will offer additional capabilities especially for ship classification. Satellite imagery will also be used for data fusion with AIS data.
Algorithms	To train different CNN based object detection and classification methods.
Sharing of results	Results initially available to project partners. External use to be decided.
License and terms of use	https://creativecommons.org/licenses/by-sa/4.0/ https://www.planet.com/assets/pdfs/planet-open-ca-license-faqs.pdf

DPO assessment (XLAB)	This dataset does not include any personal information. The dataset can thus be processed as indicated above without any limitation as long as the processing is aligned with the licence of the data provider.

Dataset name	**Kaggle Airbus ship detection**
Data Source	https://www.kaggle.com/c/airbus-ship-detection
Description	Satellite imagery with annotated ships for ship detection from optical imagery.
Usage in PIXEL	To develop the methods for ship detection which results will be further used for traffic analysis and prediction in and around the port.
Algorithms	To train different CNN based object detection methods.
Sharing of results	Results initially available to project partners. External use to be decided.
License and terms of use	Section B, rule 7: https://www.kaggle.com/c/airbus-ship-detection/rules
DPO assessment (XLAB)	Considering the GDPR definition of personal data and considering that the dataset relates to cargo vessels (which are large ships, owned by legal entities and not individuals), this dataset only includes non-personal data: images and metadata (IDs, labels, coordinates, etc.). These parameters cannot be used to, by reasonable means, directly or indirectly identify any data subject. **The dataset can be processed as indicated above without any limitation, considering that the processing is in line with the data sharing agreement or licence of the dataset owner/provider.**

Dataset name	**HRSC 2016**
Data Source	Dataset: https://www.kaggle.com/guofeng/hrsc2016 Publication: http://www.scitepress.org/Papers/2017/61206/61206.pdf
Description	High Resolution Ship Collections 2016 satellite imagery with annotated ships for ship detection and classification from optical imagery.
Usage in PIXEL	To develop the methods for ship detection and classification which results will be further used for traffic analysis and prediction in and around the port.
Algorithms	To train different CNN based object detection and classification methods.
Sharing of results	Results initially available to project partners. External use to be decided.
License and terms of use	On Kaggle released under ODbL: https://opendatacommons.org/licenses/odbl/1.0/ (link in Kaggle header). No additional license in actual data.
DPO assessment (XLAB)	Considering the GDPR definition of personal data and considering that the dataset relates to cargo vessels (which are large ships, owned by legal entities and not individuals), this dataset only includes non-personal data: images and metadata (IDs, labels, coordinates, etc.). These parameters cannot be used to, by reasonable means, directly or indirectly identify any data subject. **The dataset can be processed as indicated above without any limitation, considering that the processing is in line with the data sharing agreement or licence of the dataset owner/provider.**

Dataset name	**xView dataset**
Data Source	Dataset: http://xviewdataset.org Publication: https://arxiv.org/pdf/1802.07856.pdf
Description	Satellite imagery with annotated ships (and other classes) for object detection from optical imagery.
Usage in PIXEL	To develop the methods for ship detection and classification which results will be further used for traffic analysis and prediction in and around the port.
Algorithms	To train different CNN based object detection and classification methods.
Sharing of results	Results initially available to project partners. External use to be decided.

License and terms of use	Section 5.1 (CC BY-NC-SA 4.0), research use only: http://xviewdataset.org/terms.html; CC BY-NC-SA 4.0: https://creativecommons.org/licenses/by-nc-sa/4.0/ Asked the organizers - DIUx if the data can be used outside the scope of the competition for research purposes if properly cited in publications, project reports etc. and got positive response. Confirmation to 5.1 in terms of usage.
DPO assessment (XLAB)	Considering the GDPR definition of personal data and considering that the dataset relates to cargo vessels (which are large ships, owned by legal entities and not individuals) or other objects (buildings, roads; taken from Google Earth), this dataset only includes non-personal data: images and metadata (IDs, labels, coordinates, etc.). These parameters cannot be used to, by reasonable means, directly or indirectly identify any data subject. **The dataset can be processed as indicated above without any limitation, considering that the processing is in line with the data sharing agreement or licence of the dataset owner/provider.**

Dataset name	**DOTA**
Data Source	Dataset and project page: https://captain-whu.github.io/DOTA/index.html Publication: https://arxiv.org/pdf/1711.10398.pdf
Description	DOTA (Large-scale Dataset for Object DeTection in Aerial Images) satellite imagery with annotated ships (and other classes) for object detection from optical imagery.
Usage in PIXEL	To develop the methods for ship detection and classification which results will be further used for traffic analysis and prediction in and around the port.
Algorithms	To train different CNN based object detection and classification methods.
Sharing of results	Results initially available to project partners. External use to be decided.
License and terms of use	Research use only: https://captain-whu.github.io/DOTA/dataset.html
DPO assessment (XLAB)	Considering the GDPR definition of personal data and considering that the dataset relates to cargo vessels (which are large ships, owned by legal entities and not individuals) or other objects (buildings, roads; taken from Google Earth), this dataset only includes non-personal data: images and metadata (IDs, labels, coordinates, etc.). These parameters cannot be used to, by reasonable means, directly or indirectly identify any data subject. **The dataset can be processed as indicated above without any limitation, considering that the processing is in line with the data sharing agreement or licence of the dataset owner/provider.**

Dataset name	**Kaggle Planet Labs**
Data Source	https://www.kaggle.com/rhammell/ships-in-satellite-imagery
Description	Satellite imagery from Planet Labs with ship/not ship imagery and annotations.
Usage in PIXEL	To develop the methods for ship detection and classification which results will be further used for traffic analysis and prediction in and around the port.
Algorithms	To train different CNN based object detection and classification methods.
Sharing of results	Results initially available to project partners. External use to be decided.
License and terms of use	According to Kaggle link in the header of the dataset page, dataset is released under CC BY-SA 4.0: https://creativecommons.org/licenses/by-sa/4.0/
DPO assessment (XLAB)	Considering the GDPR definition of personal data and considering that the dataset relates to cargo vessels (which are large ships, owned by legal entities and not individuals), this dataset only includes non-personal data: images and metadata (IDs, labels, coordinates, etc.). These parameters cannot be used to, by reasonable means, directly or indirectly identify any data subject.

	The dataset can be processed as indicated above without any limitation, considering that the processing is in line with the data sharing agreement or licence of the dataset owner/provider.

Dataset name	**PVGIS data**
Data Source	Dataset and project page: http://re.jrc.ec.europa.eu/pvg_download/data_download.html Publications: http://re.jrc.ec.europa.eu/pvg_static/Publications_in_proc.html Methods: http://re.jrc.ec.europa.eu/pvg_static/methods.html
Description	Web applications to browse and query GIS databases of solar radiation and other climatic parameters. With this data it is possible to estimate PV electricity generation at any location in Europe, Africa, most of Asia, North America and most of South America
Usage in PIXEL	Based on historical irradiance data and associated weather conditions, obtained either by measurement or by satellite-based tools (PVGIS), we will propose a full methodology to predict one-point irradiance for a time horizon from a day to year. This prediction will in fact reflects typical day, week, month, year based on past data.
Algorithms	To interact with web-services like PVGIS to obtain historical data and extract a typical irradiance.
Sharing of results	Results initially available to project partners. External use to be decided.
License and terms of use	The solar radiation data which are make available are long-term averages for each month and for the year, based on data with hourly time resolution from satellite. In all cases, the original data are freely available from the organizations that have produced the data sets. The use of these data is authorised, if the source is acknowledged.
DPO assessment (CATIE)	According to the data description above and the one provided on the PVGIS website (http://re.jrc.ec.europa.eu/pvg_download/data_download.html), this dataset only includes non-personal data: solar radiation, geographical data, temperature, PV technical specification. **The dataset can be processed as indicated above without any limitation, considering that the processing is in line with the data sharing agreement or licence of the dataset owner/provider.**

Dataset name	**PVoutput data set**
Data Source	https://pvoutput.org./list.jsp?id=15556&sid=13412
Description	Output is a free online service for sharing and comparing photovoltaic solar panel output data. It provides both manual and automatic data uploading facilities. Output data can be graphed, analysed and compared with other PV output contributors over various time periods. While PV output is primarily focused on monitoring energy generation, it also provides equally capable facilities to upload and monitor energy consumption data from various energy monitoring devices.
Usage in PIXEL	Based on historical production data, we will implement predictive algorithms for photovoltaic production based on past data of real production and associated weather conditions.
Algorithms	To train different machine learning models to predict solar energy production.
Sharing of results	Results initially available to project partners. External use to be decided.
License and terms of use	No license has been founded. CATIE has acquired the donation status to be able to query historical data through the PV output API. Raw data have been downloaded, stored and analysed by CATIE.
DPO assessment (CATIE)	According to the data description above and the one provided on the PVoutput website (https://pvoutput.org), this dataset only includes non-personal data: solar energy production, energy consumption, geographical data, weather conditions, temperature, PV technical specification.

	The dataset can be processed as indicated above without any limitation, considering that the processing is in line with the data sharing agreement or licence of the dataset owner/provider.

Dataset name	**OpenWeatherMap data**
Data Source	https://donneespubliques.meteofrance.fr/?fond=produit&id_produit=131&id_rubrique=51
Description	Current weather, daily forecast for 16 days, and 3-hourly forecast 5 days. Stats, graphics, and this day in history charts are available for reference. Interactive maps show precipitation, clouds, pressure, wind.
Usage in PIXEL	To use as inputs to predict solar energy production based on weather conditions.
Algorithms	
Sharing of results	Available to project partners.
License and terms of use	The OpenWeatherMap API is free to use; however, there are also paid plans available for developers that require higher levels of weather data and support.
DPO assessment (CATIE)	According to the data description above and the one provided on the OpenWeatherMap, this data only includes non-personal data: geographical data, weather conditions, temperature. **The dataset can be processed as indicated above without any limitation, considering that the processing is in line with the data sharing agreement or licence of the dataset owner/provider.**

Dataset name	**OpenDataSoft data**
Data Source	https://public.opendatasoft.com/explore/dataset/arome-0025-sp1_sp2/table/?location=6,46.00538,2&basemap=jawg.streetshttps:%2F
Description	Fields of analysis and forecasts in grid points, resulting from the atmospheric model Arome on the metropolis. Parameters, levels, deadlines and configurable domains according to various resolutions and deadlines up to 36h or 42h.
Usage in PIXEL	To use as inputs to predict solar energy production based on weather conditions. To use as historical data set for weather conditions.
Algorithms	
Sharing of results	Available to project partners.
License and terms of use	No fee under EtalabLogo's Open License for the open license. The source to indicate is "Météo-France". Some suggestions: "Source: Météo-France" or "Information created from Météo-France data".
DPO assessment (CATIE)	Considering the GDPR definition of personal data and considering that the dataset relates to cargo vessels (which are large ships, owned by legal entities and not individuals), this dataset only includes non-personal data: images and metadata (IDs, labels, coordinates, etc.). These parameters cannot be used to, by reasonable means, directly or indirectly identify any data subject. **The dataset can be processed as indicated above without any limitation, considering that the processing is in line with the data sharing agreement or licence of the dataset owner/provider.**

Dataset name	**Piraeus live data report from police department (PPA use case)**
Data Source	Police department (data source). The data of road traffic from police department are accessible from this link: http://www.astynomia.gr/traffic-athens.php
Description	Data provides the traffic situation (middle column and the rated traffic as normal - congested - very congested) of the main Athens - Piraeus road arteries (listed in the left column) with special explanatory notes per each main road (right column). These are live traffic data only and no historical info is provided

Usage in PIXEL	Correlation of port operations and regional road network for analysis and prediction purposes
Algorithms	Different supervised machine learning methods for regression, time series analysis and classification.
Sharing of results	Results initially available to project partners. External use to be decided.
License and terms of use	Open free data to anyone. No fees
DPO assessment (Prodevelop)	**No GDPR restrictions apply**

Dataset name	**Traffic reports from the Region of Attica (PPA use case)**
Data Source	Region of Attica (data source). Data is published every 6 months. Data is available at the following link: http://www.patt.gov.gr/site/index.php?option=com_content&view=article&id=29298:35i-ekthesi-leitourgias-tou-kentrou-diaxeirisis-kykloforias&catid=301&Itemid=320
Description	These data are in the Greek language and describe the findings and actions concerning traffic, on the main streets of Attica based on traffic jams, march, blockade of roadworks etc.), as well as other incidents, such as vehicle stoppages due to damage, accidents, severe weather events, etc.
Usage in PIXEL	Correlation of port operations and regional road network for analysis and prediction purposes
Algorithms	Different supervised machine learning methods for regression, time series analysis and classification.
Sharing of results	Results initially available to project partners. External use to be decided.
License and terms of use	Open free data to anyone. No fees.
DPO assessment (Prodevelop)	**No GDPR restrictions apply**

Dataset name	**Data from third party subscription database (PPA use case)**
Data Source	Third party subscription (data source). The data will be provided from a third-party subscription database in the format of .csv.
Description	The data will be in csv format and cover the port specific area (Piraeus). No historical data can be provided from this supplier
Usage in PIXEL	Correlation of port operations and regional road network for analysis and prediction purposes
Algorithms	Different supervised machine learning methods for regression, time series analysis and classification.
Sharing of results	Results initially available to project partners. External use to be decided.
License and terms of use	Customized data as per negotiated contract with the data provider for a predefined port area at a fee to be determined based on the testing demo results.
DPO assessment (Prodevelop)	**No GDPR restrictions apply**