Ecosystem for COllaborative Manufacturing PrOceSses – Intra- and Interfactory Integration and AutomaTION
(Grant Agreement No 723145)

D3.7 Computational Modelling, Simulation and Prediction of Logistics II

Date: 2018-12-21

Version 1.0

Published by the COMPOSITION Consortium

Dissemination Level: Public

Co-funded by the European Union's Horizon 2020 Framework Programme for Research and Innovation under Grant Agreement No 723145

Document control page

Document file: D3.7 Computational Modelling, Simulation and Prediction of Logistics II
Document version: 1.0
Document owner: CERTH

Work package: WP3 – Manufacturing Modelling and Simulation
Task: T3.3 – Simulation and Forecasting in Production and Logistics
Deliverable type: R

Document status: ☒ Approved by the document owner for internal review
☒ Approved for submission to the EC

Document history:

Version	Author(s)	Date	Summary of changes made
0.1	Thanasis Vafeiadis, Alexandros Nizamis (CERTH)	2018-11-13	ToC and Initial text
0.2	Thanasis Vafeiadis, Alexandros Nizamis (CERTH)	2018-11-19	Contribution on Sections 4 and 5
0.3	Thanasis Vafeiadis (CERTH)	2018-11-23	Contribution on Sections 1, 3, 6 and 7
0.4	Thanasis Vafeiadis(CERTH)	2018-11-24	Document finalization
1.0	Thanasis Vafeiadis (CERTH)	2018-12-20	Document finalization

Internal review history:

Reviewed by	Date	Summary of comments
Matteo Pardi (NXW)	2018-12-17	Minor comments and suggestions
Dimitris Gortzis (ELDIA)	2018-12-18	No comments or corrections to make

Index:

1 Executive Summary 4
2 Abbreviations and Acronyms 5
3 Introduction 6
3.1 Purpose, context and scope of this deliverable 6
3.2 Content and structure of this deliverable 6
4 Simulation and Forecasting tool in Overall COMPOSITION Architecture 7
5 Functions and methodologies 8
5.1 UC – ELDIA 1 Fill Level Notification – Contractual recyclable Material Management 8
5.2 UC – KLE 4 Scrap Metal Collection Process 8
5.3 Fill Level Sensor Deployment 8
5.4 UC – ELDIA 1 Fill Level Notification – Contractual recyclable Material Management 9
5.4.1 Methodologies 9
5.4.2 Applications 10
5.5 Time Series Prediction Engine 12
5.5.1 Data collection 12
5.5.2 Moving average 13
5.5.3 Auto-regression models 14
5.5.4 Markov chain 15
5.5.5 Fill Level Sensor Data Analysis 16
6 Conclusions 18
7 List of Figures and Tables 19
7.1 Figures 19
7.2 Tables 19
8 References 20

1 Executive Summary

The present document is a deliverable of the "***Ecosystem for COllaborative Manufacturing PrOceSses – Intra- and Interfactory Integration and AutomaTION***" - (COMPOSITION) project, funded by the European Commission's Directorate - General for Research and Innovation (D-G RTD) under Horizon 2020 Research and Innovation programme (H2020). The deliverable presents the final version of Computational Modelling, Simulation and Prediction of Logistics developed until M28 of the project.

The COMPOSITION project needs simulation and prediction in both intra- and inter-factory scenarios. This report is focused on inter-factory scenarios which are related to logistics. In this first stage of the project and *Task 3.3 - Simulation and Forecasting in Production and Logistics* the research has been conducted to use cases which were selected as the cases with the highest priority from both pilot and technical partners of the project. This report alongside with *D3.5 - Computational Modelling, Simulation and Prediction of Production II* which is published in parallel with D3.7, present the total work has done in Task 3.3.

2 Abbreviations and Acronyms

Table 1: Abbreviations and acronyms are used in this deliverable

Acronym	Meaning
BMS	Building Management System
DSS	Decision Support System
ERP	Enterprise Resource Planning
GA	Genetic Algorithm
MQTT	Message Queuing Telemetry Transport
IR	Information Retrieval
SSP	Slope Statistic Profile

3 Introduction

3.1 Purpose, context and scope of this deliverable

This document presents the computational modelling, simulation and prediction functions on production developed until M28 of the COMPOSITION project. This document is part of the Task 3.3 – Simulation and Forecasting in Production and Logistics and aims to design and implement trend analysis, statistics and probability theory for the analysis of significant and key process variables. This deliverable defines the final approaches for the core set of algorithms, techniques and methodologies dedicated on fill level notification and contractual recyclable material management. With the implementation of such techniques, we aim to provide detection of possible optimum pair of routes and weights for waste materials, an estimation for the bins fullness, prediction of median carried cargo tonnage time series, and calculation of probabilities of future states of the tested time series. The user interfaces related to SFT will be described in details at D5.2 and D3.9, which are related to visual analytics tools and DSS respectively.

3.2 Content and structure of this deliverable

The content of this deliverable is organized as follows:

Section 4 presents the position of Simulation and Forecasting tool in overall project's architecture. In Section 5 a brief description of the industrial data used for each use case that belongs to the area of computational modelling, simulation and prediction of production, according to D2.1 – Industrial Use cases for an Integrated Information Management System is provided. In Section 6 provides a description of the functions and methodologies developed (new) and utilized or modified (existing ones from scientific literature) until M28 of the project, along with their application on the use cases. In this section, several data processing techniques and transformations, functions and methodologies are described and applied on project use cases. Finally, in Section 7 we draw our conclusions.

4 Simulation and Forecasting tool in Overall COMPOSITION Architecture

This section describes the position of the Simulation and Forecasting Tool in the COMPOSITION project. We provide in this report an overview of the total COMPOSITION architecture and highlighting the Simulation and Forecasting tool's position without giving more descriptions about the components that are connected to it. This is done in more details in *D3.5 Computational Modelling, Simulation and Prediction of Production II* – a report published in parallel with D3.7 and contains the Simulation and Forecasting tool related to intra-factory use cases of the project.

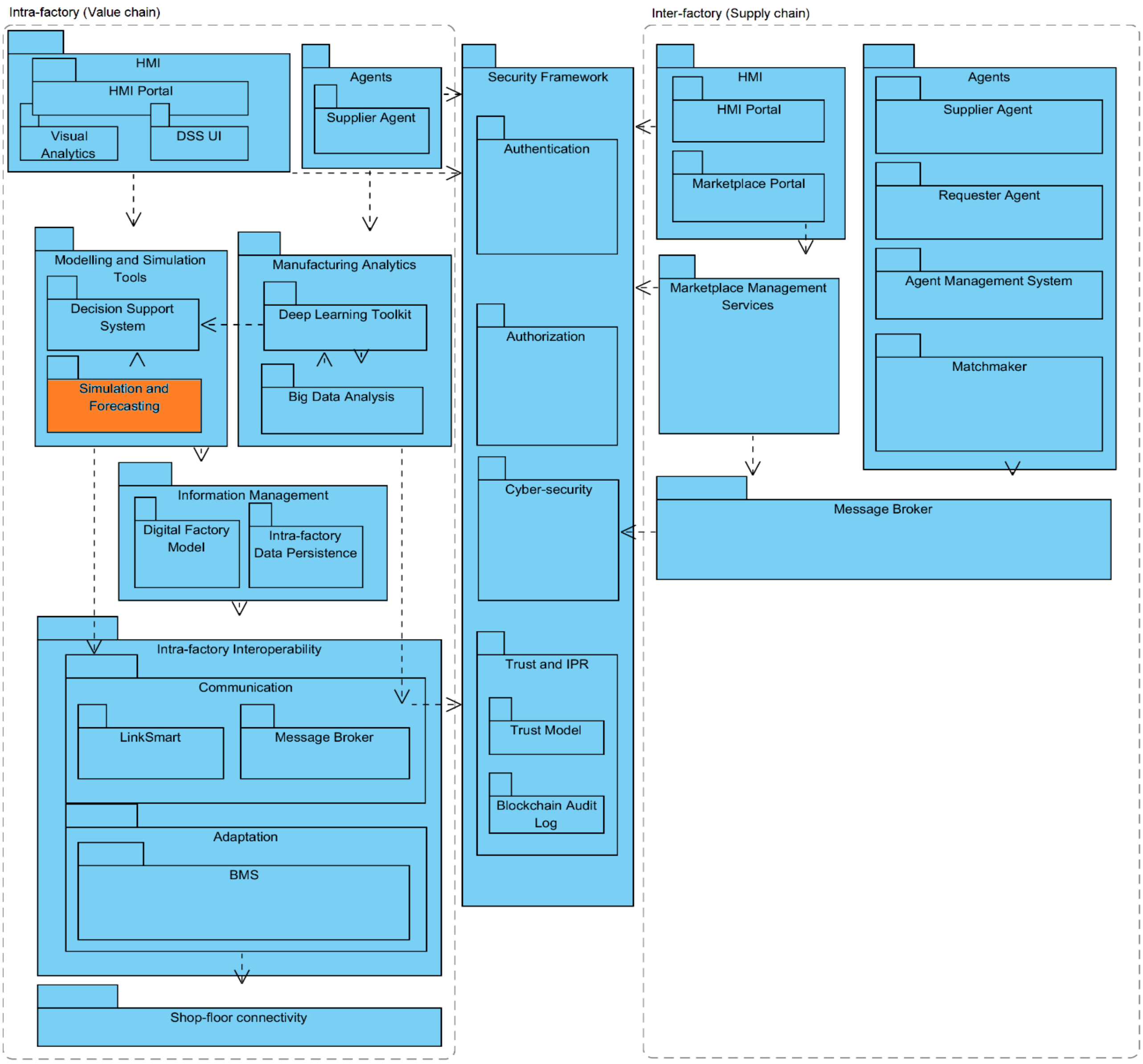


Figure 1: COMPOSITION architecture functional view

As depicted in the Figure 1, the Simulation and Forecasting tool belongs conceptually to intra-factory components. Besides that, as the Analytics components are connected with the decision support over the Marketplace and the inter-factory use cases of the project as well. In order to create forecasts and offer optimization of some processes related to supply-chain pilots the SFT uses data coming from Information Management components and export its output to Visual Analytics and DSS UI for the Marketplace. The connection of SFT and these components are presented in more details in D3.5 as mentioned before. Furthermore, information about the deployment of SFT are available on D3.5 as well, and they are not repeated in this document.

5 Functions and methodologies

This section provides a description of all the functions and methodologies utilized for logistics until M28 of the project, such as genetic algorithms, moving average and auto-regression models, Markov chains and Slope Statistic Profile, along with their application on the use cases.

5.1 UC – ELDIA 1 Fill Level Notification – Contractual recyclable Material Management

The primary goal of this use case is to be able to receive notifications of the fill level of various containers installed at ELDIA customers' facilities, thus facilitating the logistics service and improving the reaction time.

ELDIA ERP maintains a database including information about the date of pick up, type, weight and prices of various recyclable materials. These data will be used for simulation and forecasting. This information become available to the project's technical and research partners for further analysis in order to enable possible estimations of the fill level of various containers.

5.2 UC – KLE 4 Scrap Metal Collection Process

The goal of this use case is to optimize scrap metal collection and bidding process by getting better scrap metal prices, minimizing costs and receiving a fast and efficient service. A dataset is generated by the company's ERP system. The ERP maintains a database including information about the produced scrap metal and the price of it. This set of data is extracted from ERP in excel as a report file. In 2016 around 1.000 tons of scrap metal were produced with an average price of € 120 - 150/ ton.

5.3 Fill Level Sensor Deployment

For the COMPOSITION use cases related to bins fill level notification (UC-KLE 4 and UC-ELDIA 1), a fill level monitoring sensor has been developed. This sensor is able to provide data to the Simulation and Prediction tool in order to enable the estimation of the date in which the bin will be full.

Figure 2: Fill level sensor's position

For the fill sensor data the HRXL-MaxSonar-WRT MB7380 IP67 Ultrasonic Sensor is used to capture raw measurements of distance of the waste heap from the deployment point. The sensor has its own filtering mechanisms to provide reliable measurements and delivers the result via PWM output. The sensor and the communication are controlled by the STM32L053 low power microcontroller. The microcontroller performs approximately 12 measurements with a duty cycle of less than 1%, discards the invalid ones and filters the valid ones with a median filter choosing the final measurement to send. In case there are no valid measurements taken, an error value is sent. The wireless communication is carried out via the sx1272mbas LoRa Module for STM. The system is powered by 4 AA batteries.

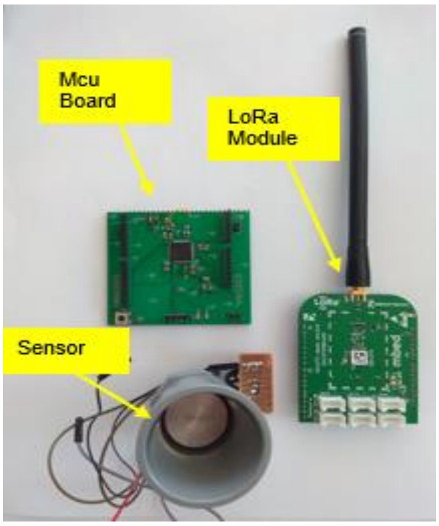


Figure 3: Fill level sensor system overview

Data are transferred via the LoRaWAN low power protocol to the LoRa Gateway. The gateway used is the LoRank 8. It has to be noted that LoRaWAN protocol allows transmission of only very small packets of data, so this means that only raw measurements of distance are transferred along with the measurement of battery level and id of bin. The gateway that is connected to the internet via Ethernet or Wi-Fi publishes those data on an MQTT topic on the cloud and a listener, connected to the same broker as the gateway, reprocesses the data to convert the raw distance measurement into a fill percentage and derive the JSON object format to be sent on the destination platform. The Building Management System (BMS) component of the project is the destination platform.

5.4 UC – ELDIA 1 Fill Level Notification – Contractual recyclable Material Management

5.4.1 Methodologies

This section provides a set of algorithms and methodologies for optimization and time series prediction. These methodologies are: genetic algorithms for the management and the transportation or recyclable materials and slope statistic profile, moving average, auto-regression models and Markov chain approaches and methodologies for the prediction of key performance indicator values.

5.4.1.1 Genetic Algorithms

People use search engines to find information they desire with the aim that their information needs will be met. Information retrieval (IR) is a field that is concerned primarily with the searching and retrieving of information in the documents and also searching the search engine, online databases, and Internet. Genetic algorithms (GAs) are robust, efficient, and optimized methods in a wide area of search problems motivated by Darwin’s principles of natural selection and survival of the fittest.

IR system searches for the matches in the document databases and, thus, retrieves search results of the matching process. However, based on the relevance, the user will then evaluate and display the search results. The relevance of the document is very important to the user. If the user feels that it is a relevant document, he finishes the search while else user continues to search in the document database by reformulating the query until the relevant documents that will satisfy users’ information needs are retrieved.

GA is a probabilistic algorithm simulating the process of natural selection of living organisms and finally coming up with an approximate solution to a problem (Holland, 1975), (DeJong, 1975) & (Goldberg, 1989). In GA implementation, the search space is composed of candidate solutions (called individuals or creatures) to an optimization problem to evolve better solutions; each represented by a string is termed chromosome. Each chromosome has an objective function value, called fitness. A set of chromosomes together with their associated fitness is called a population. This population, at a given iteration of the genetic algorithm, is called a generation. In each generation, the fitness of every individual in the population is evaluated from the current population based on their fitness value and modified to form a new population. The new population is then used in the next iteration of the algorithm.

GA terminates when either a maximum number of generations has been produced or a satisfactory fitness level has been reached for the population. If the algorithm has terminated due to a maximum number of generations, a satisfactory solution may or may not have been reached. The working of the genetic algorithm depends upon the constraint of how well we choose our initial random keywords.

5.4.2 Applications

In the current contractual recyclable material management dataset (see Table 2) there were 4 recyclable materials (organic waste = 1, wood = 2, plastic = 3 and paper = 4) recorded for the years of 2015 and 2016. In order to apply the GA we have chosen waste and wood materials, which are recorded for all customers for both years. The fitness function is the ratio of routes and weights that need to be minimized (the company needs to carry as much weights as possible with as few as possible routes). Thus, the target value of this fitness function is the minimum ratio of all customers for a specified material. In Table 4, one can see the average values of fitness function for all waste and wood materials, for all customers and for the years 2015 and 2016. The Objective bounds of routes and weights per material are given in Table 5. The ratios of routes per weights for all customers, materials and years are also given in Figure 4, based on Table 3. As input in the GA are the bounds of routes and weights along with the minimum ratio of all customers per material and per year (see the highlighted values in Table 4).

Table 2: Contractual recyclable material management dataset

A.A.	Customer	Description	January	February	March	April	May	June	July	August	September	October	November	December	Sum
1	Customer A 2015	Route_Material_1	2	4	6	6	6	8	8	5	5	7	6	6	69
		Weight_Material_1	10.070	17.960	35.420	27.070	23.840	46.860	44.040	25.190	28.240	35.300	23.830	27.520	345.340
		Route_Material_2	2	3	3	4	5	6	3	4	4	3	5	4	46
		Weight_Material_2	0.750	1.230	1.380	1.920	2.020	2.610	1.080	2.030	1.880	1.390	2.130	2.840	21.260
	Customer A 2016	Route_Material_1	3	6	7	8	8	6	6	5	8	7	6	3	73
		Weight_Material_1	14.200	29.420	36.400	40.700	39.940	32.360	33.050	23.570	41.510	34.500	24.230	13.700	363.580
		Route_Material_2	4	4	5	5	6	4	3	4	4	4	4	2	49
		Weight_Material_2	3.440	1.960	3.990	1.870	4.630	1.480	1.270	2.360	1.950	1.780	1.630	0.700	27.060
2	Customer B 2015	Route_Material_1	1	1	2	3	2	3	1	2	3	3	1	2	24
		Weight_Material_1	1.490	2.780	5.110	6.220	3.410	6.220	2.790	2.960	5.710	6.420	0.900	6.540	50.550
	Customer B 2016	Route_Material_1	1	1	1	3	1	2	2	1	2	1	1	2	18
		Weight_Material_1	3.030	3.810	3.050	6.350	2.220	5.190	4.180	2.330	5.560	2.670	2.940	5.460	46.790
3	Customer C 2015	Route_Material_1	6	11	10	10	10	11	10	6	8	13	10	12	117
		Weight_Material_1	20.540	40.210	27.970	29.750	30.790	32.310	29.440	24.110	25.000	64.700	27.390	29.950	382.160
		Route_Material_2	3	3	3	3	3	4	4	2	3	4	4	5	41
		Weight_Material_2	1.030	1.190	1.290	1.750	1.230	1.870	1.610	0.710	0.950	1.670	2.400	1.930	17.630
		Route_Material_3	2	2	2	1	2	2	3	-	1	1	1	1	18
		Weight_Material_3	0.570	0.200	0.270	0.190	0.320	0.190	0.320	-	0.190	0.190	0.190	0.040	2.670
		Route_Material_4	4	5	4	4	5	5	6	2	6	5	5	6	57
		Weight_Material_4	2.570	2.200	2.690	1.940	3.040	2.090	2.770	1.310	3.400	3.660	2.410	2.660	30.740
	Customer C 2016	Route_Material_1	8	12	12	14	10	10	12	8	14	12	14	16	142
		Weight_Material_1	25.920	37.010	32.250	37.960	26.190	35.970	36.830	29.740	43.640	33.330	40.010	53.420	432.270
		Route_Material_2	3	2	7	5	5	4	5	4	6	7	7	5	60
		Weight_Material_2	1.120	0.940	2.770	2.000	3.790	4.290	1.980	5.820	2.530	2.950	4.950	1.930	35.070
		Route_Material_3	3	-	-	1	1	-	-	-	1	-	-	2	8
		Weight_Material_3	0.380	-	-	0.250	0.210	-	-	-	0.180	-	-	0.620	1.640
		Route_Material_4	4	6	7	7	5	6	6	4	6	7	6	6	70
		Weight_Material_4	2.040	4.020	3.850	3.100	2.460	4.000	2.930	1.780	2.970	3.750	3.560	3.660	38.120
4	Customer D 2015	Route_Material_1	10	10	11	11	13	12	14	8	12	17	13	10	141
		Weight_Material_1	31.690	32.180	39.430	41.140	44.510	36.170	40.550	25.620	36.340	38.820	49.940	28.110	141.000
	Customer D 2015	Route_Material_1	9	10	10	11	11	13	15	7	14	11	12	10	133
		Weight_Material_1	26.330	37.590	40.800	34.720	34.450	37.990	45.430	20.380	43.080	32.140	36.890	37.550	427.350

Table 3: Route/weights ratio

Customer	Year	Description	Jan	Feb	Mar	Apr	May	Jun	Juj	Aug	Sept	Oct	Nov	Dec
A	2015	Ratio - Material_1	0.20	0.22	0.17	0.22	0.25	0.17	0.18	0.20	0.18	0.20	0.25	0.22
A	2015	Ratio - Material_2	2.67	2.44	2.17	2.08	2.48	2.30	2.78	1.97	2.13	2.16	2.35	1.41
A	2016	Ratio - Material_1	0.21	0.20	0.19	0.20	0.20	0.19	0.18	0.21	0.19	0.20	0.25	0.22
A	2016	Ratio - Material_2	1.16	2.04	1.25	2.67	1.30	2.70	2.36	1.69	2.05	2.25	2.45	2.86
B	2015	Ratio - Material_1	0.67	0.36	0.39	0.48	0.59	0.48	0.36	0.68	0.53	0.47	1.11	0.31
B	2016	Ratio - Material_1	0.33	0.26	0.33	0.47	0.45	0.39	0.48	0.43	0.36	0.37	0.34	0.35
C	2015	Ratio - Material_1	0.29	0.27	0.36	0.34	0.32	0.34	0.34	0.25	0.32	0.20	0.37	0.33
C	2015	Ratio - Material_2	2.91	2.52	2.33	1.71	2.44	2.14	2.48	2.82	3.16	2.40	1.67	2.59
C	2015	Ratio - Material_3	3.51	10.0	7.41	5.26	6.25	10.5	9.38	-	5.26	5.26	5.26	25
C	2015	Ratio - Material_4	1.56	2.27	1.49	2.06	1.64	2.39	2.17	1.53	1.76	1.37	2.07	2.25
C	2016	Ratio - Material_1	0.31	0.32	0.37	0.37	0.38	0.28	0.33	0.27	0.32	0.36	0.35	0.29
C	2016	Ratio - Material_2	2.68	2.13	2.53	2.50	1.32	0.93	2.53	0.69	2.37	2.37	1.41	2.59
C	2016	Ratio - Material_3	7.89	-	-	4.00	4.76	-	-	-	5.56	-	-	3.22
C	2016	Ratio - Material_4	1.96	1.49	1.82	2.26	2.03	1.50	2.05	2.25	2.02	1.87	1.69	1.64
D	2015	Ratio - Material_1	0.32	0.31	0.28	0.27	0.29	0.33	0.35	0.31	0.33	0.44	0.26	0.35
D	2016	Ratio - Material_1	0.34	0.27	0.25	0.32	0.32	0.34	0.33	0.34	0.32	0.34	0.33	0.27

Table 4: Fitness function average value for material / customer for years 2015 and 2016.

Years	2015		2016	
Materials	**Waste**	**Wood**	**Waste**	**Wood**
Customer A	**0.17**	**1.41**	**0.18**	1.16
Customer B	0.30	1.66	0.26	**0.68**
Customer C	0.20	-	0.27	-
Customer D	0.26	-	0.25	-

Table 5: Objective bounds of routes and weights per material

Years	2015				2016			
Materials	**Waste**		**Wood**		**Waste**		**Wood**	
	Routes	**Weights**	**Routes**	**Weights**	**Routes**	**Weights**	**Routes**	**Weights**
Min	1	1	2	0.71	7	20.38	2	0.94
Max	17	65	5	2.4	15	45.43	7	5.82

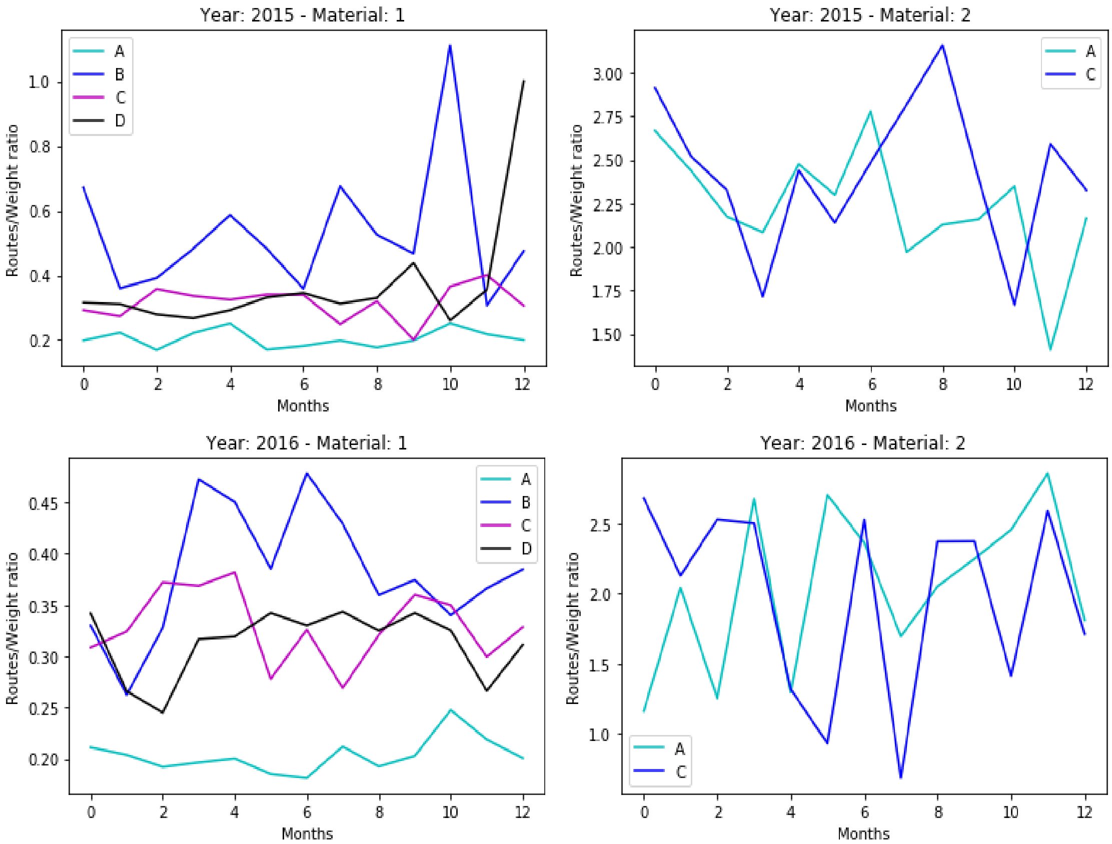


Figure 4: Ratios of routes per weights for all customers, materials and years

Thus, the results from the application of GA on the various materials (organic waste and wood) and for years 2015 and 2016 are given in Table 6. In order to train the GA algorithm, the minimum ratio (routes/weights) of all customers per material is used as the target of the fitness function. Moreover, the GA algorithm iterates among all possible values of routes and weights (based on minimum and maximum values given by the ELDIA). For the year 2015 the optimum pair of routes and weights for organic waste material is 9 routes and 53 tonnages and wood material the optimum pair of routes and tonnages is 4 and 3, respectively. For the year 2016 the optimum pair of routes and weights for waste material is 7 routes and 39 tonnages and wood material the optimum pair of routes and tonnages is 2 and 3, respectively. The optimum pair denotes what would be the number of routes and the number of carried weights so as the company to have no losses.

Table 6: GA selections

Years	**2015**				**2016**			
Materials	**Waste**		**Wood**		**Waste**		**Wood**	
	Routes	**Weights**	**Routes**	**Weights**	**Routes**	**Weights**	**Routes**	**Weights**
GA	9	53	4	3	7	39	2	3

5.5 Time Series Prediction Engine

5.5.1 Data collection

The tested time series for all the methodologies of this time series prediction engine is the calculated median of carried cargo (in tonnage) from all customers. The carried cargo statistics are given in Table 7 below:

Table 7: Customer's route and tonnage statistics

	January		February		March		April		May		June		July		August		September		October		November		December	
Customers	Routes	Tonnage	Routes	Tonnage	Routes	Tonnage	Routes	Tonnage	Routes	Tonnage	Routes	Tonnage	Routes	Tonnage	Routes	Tonnage	Routes	Tonnage	Routes	Tonnage	Routes	Tonnage	Routes	Tonnage
Customer 1	4	3.710	3	2.560	3	2.370	3	2.460	3	2.420	3	2.680	4	3.980	3	2.570	4	3.860	2	1.290	3	2.640	3	2.680
Customer 2	4	3.920	4	3.880	5	5.050	4	4.190	4	4.000	3	2.580	4	4.030	2	1.270	4	4.050	3	2.630	5	5.180	4	4.030
Customer 3	18	24.550	14	21.770	16	21.720	17	23.290	20	27.460	18	23.400	21	31.980	19	26.900	22	36.840	27	44.310	20	37.870	15	26.650
Customer 4	3	2.680	4	4.610	4	3.440	3	2.810	4	3.550	3	2.330	4	2.910	1	0.980	4	3.800	4	3.700	2	3.700	2	2.840
Customer 5	6	4.620	7	5.600	9	6.900	7	5.740	7	5.680	9	7.300	8	6.910	8	5.840	10	6.630	8	5.910	8	5.430	8	5.930
Customer 6	3	4.780	1	2.000	2	2.720	2	2.970	3	4.250	3	2.620	3	4.590	3	4.520	3	4.790	2	3.960	3	3.740	1	1.160
Customer 7	4	9.230	5	11.060	5	10.178	5	9.050	6	14.390	4	8.960	4	7.730	3	5.990	5	9.560	3	5.740	6	14.160	3	5.150
Statistics																								
Min	3	2.68	1	2	2	2.37	2	2.46	3	2.42	3	2.33	3	2.91	1	0.98	3	3.8	2	1.29	2	2.64	1	1.16
Max	18	24.55	14	21.77	16	21.72	17	23.29	20	27.46	18	23.4	21	31.98	19	26.9	22	36.84	27	44.31	20	37.87	15	26.65
Median	4	4.62	4	4.61	5	5.05	4	4.19	4	4.25	3	2.68	4	4.59	3	4.52	4	4.79	3	3.96	5	5.18	3	4.03
Average	6.00	7.64	5.43	7.35	6.29	7.48	5.86	7.22	6.71	8.82	6.14	7.12	6.86	8.88	5.57	6.87	7.43	9.93	7.00	9.65	6.71	10.39	5.14	6.92

The selected time series is the median tonnage of all customers given in green in Table 7.

$$x = [4.62, 4.61, 5.05, 4.19, 4.25, 2.68, 4.59, 4.52, 4.79, 3.96, 5.18, 4.03]$$

The size of the tested time series is 12 elements, but all the algorithms and methodologies provided in this section are designed for time series of a free-size.

5.5.2 Moving average

This algorithm helps us to forecast new observations based on a time series. This algorithm uses smoothing methods. The moving average, denoted hereafter as MA, algorithm is used only on time series that does not have a trend. This method is by far the easiest and it consists of making the arithmetic mean of the last n observations contained by the time series to forecast the next observation. We use the following formula:

$$MA_{t+1} = \frac{\sum_{i=t-n}^{t} x_i}{n}$$

where t is the period of time (or the size of the time series) and x_i is the actual time series.
In some cases, we need to find the optimal number n of observation to be used in the forecast. We can find it by checking the square error mean of multiple n observations. Basically, the range for forecasting elements is to start at 2 observations up to half of the data set size + 1.
The moving average model has been applied on the time series of median tonnage per month for $n = 2,3,4,5$, which means that the predictions based on MA model will take place from 2 to 5 months ahead. Figure 5 shows the actual time series of median tonnage (blue) and the predictions based on MA model (red).

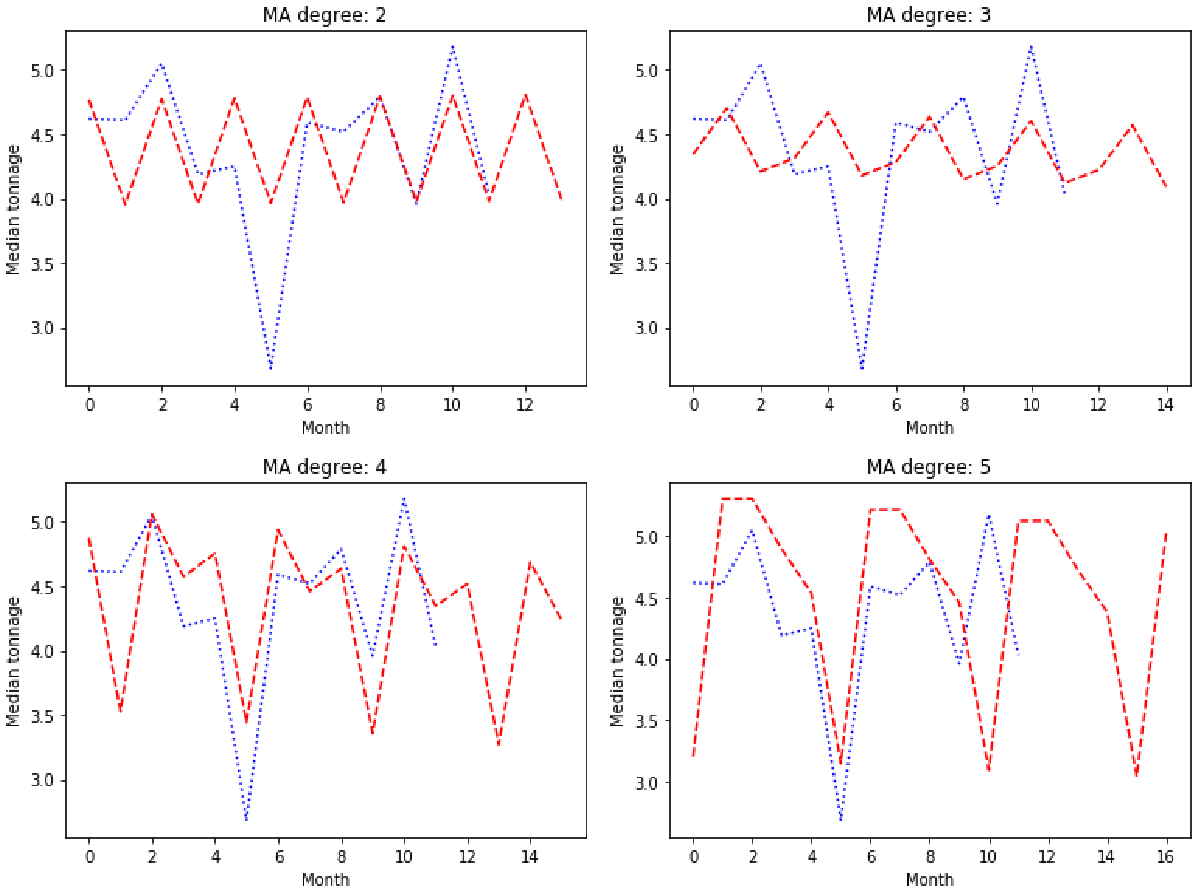


Figure 5: Predictions based on MA model from 2 to 5 months ahead.

One can see that not all MA model fit with the form of time series data and this is because the size of time series is very small in order the MA model to provide quality results. The model with $n = 4$ is the one that fits mostly with the time series data and has the form of the initial time series.

5.5.3 Auto-regression models

A **time series** is a sequence of measurements of the same variable(s) made over time. Usually the measurements are made at evenly spaced times. Let us first consider the problem in which we have a *y*-variable measured as a time series. As an example, we might have *y* a measure of global temperature, with measurements observed each year. An **autoregressive model** is when a value from a time series is regressed on previous values from that same time series, for example, y_t on $y_t - 1$:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t \quad (1)$$

In this regression model, the response variable in the previous time period has become the predictor and the errors have our usual assumptions about errors in a simple linear regression model. The **order** of an auto-regression is the number of immediately preceding values in the series that are used to predict the value at the present time. Thus, the preceding model given in eq. (1) is a first-order auto-regression, written as AR(1). If we want to predict y_t using measurements (y_{t-1}, y_{t-2}), then the autoregressive model for doing so would be:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \varepsilon_t \quad (2)$$

The model in eq. (2) is a second-order auto-regression, written as AR(2).

The application of an auto-regression model with lag 1, meaning AR(1), on time series x provides a prediction of median carried tonnage for the next month around 3.46

5.5.4 Markov chain

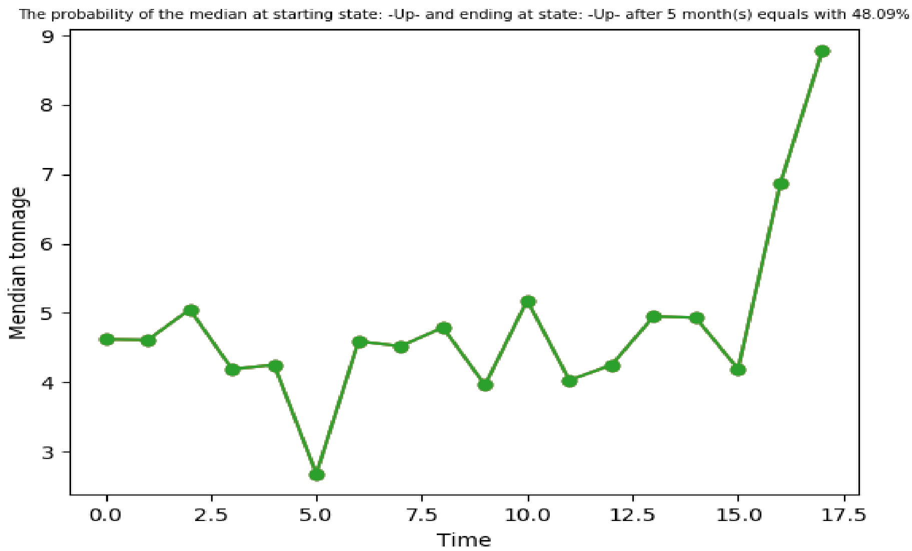
Markov chains are a fairly and simple common way to statistically model random processes. Markov Chains are conceptually quite intuitive, and are very accessible in that they can be implemented without the use of any advanced statistical or mathematical concepts. Formally, a Markov chain model is a probabilistic automaton. The probability distribution of state transitions is typically represented as the Markov chain's *transition matrix* P. If the Markov chain has N possible states, the matrix will be an $N\ x\ N$ matrix, such that entry (i, j) is the probability of transitioning from state i to state j. Additionally, the transition matrix must be a **stochastic matrix**, a matrix whose entries in each row must add up to exactly 1. This makes complete sense, since each row represents its own probability distribution.

Markov chain models can be used to determine the probability of moving either from state i to state j **for a single step** or from state i to state j **over** M **steps**. As it turns out, this is actually very simple to find out. Given a transition matrix P, this can be determined by calculating the value of entry (i, j) of the matrix obtained by raising P to the power of M.

For the time series of median tonnage per month we define two possible states: "Up" and "Down", which means that a value of the time series at its current time can be followed by a larger ("Up") or a smaller ("Down") value. We could also define the a third state when the current and the future value are the same ("Steady"), but the form of the time series does not provide any clues for the existence of this state. Thus, based on the assumptions described above, the transition matrix P for the median tonnage per month time series is calculated and given below:

$$P = \begin{bmatrix} 0.222 & 0.778 \\ 0.857 & 0.143 \end{bmatrix}$$

From the transition matrix P one can see that the probability of being at state "Up" and the next state to be the same is 0.222 and the probability of being at state "Up" and the next state to be "Down" is 0.778. Moreover, the probability of being at state "Down" and the next state to be "Up" is 0.857 and the probability of being at state "Down" and the next state to be "Down" is 0.143. Based on transition matrix P it is easy to calculate the probabilities of futures states over M steps. For example, the probability of being at state "Up" and after 5 months to be at the same state is 0.48 and the probability of being at state "Up" and after 5 months to be at state "Down" is 0.53 (see Figure 6).



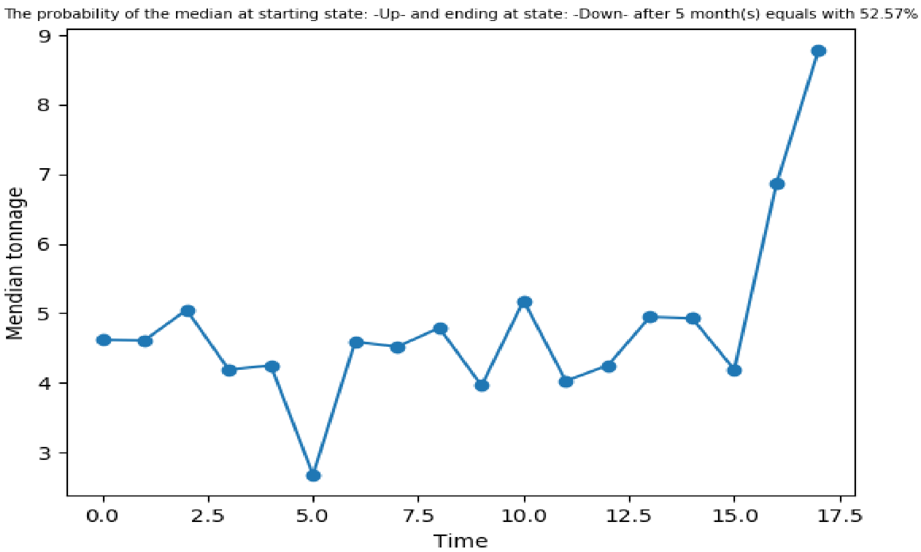


Figure 6: Time series plots with the calculated transition probabilities for M=5 and states "Up" to "Up" (left plot) and M=5 and states "Up" to "Down" (right plot).

5.5.5 Fill Level Sensor Data Analysis

The Slope Statistic Profile (SSP) method aims to estimate the change point *T* from the profile of a linear trend test statistic, computed on consecutive overlapping time windows along the time series. The test statistic for linear trend estimation is selected based on its high power compared to other test statistics for both correlated and white noise residuals. In the SSP approach, a first candidate change point T is the time point at which the profile crosses the threshold line of rejection of the null hypothesis of no trend at $\pm t_{w-2,1-a/2}$, where *a* is the significance level, *w* is the size of the sliding window and *t* follows the Student distribution with *w-2* degrees of freedom. The search of change point T is confined in a time interval corresponding to the profile segment bounded by $t_{w-2,1-a_1/2}$ and $t_{w-2,1-a_2/2}$ for positive trends and by $-t_{t_{w-2,1-a_2/2}}$ and $-t_{w-2,1-a_1/2}$ for negative trends, where significance levels a_1 and a_2 for two side test are 0.20 (or 20%) and 0.05 (or 5%), respectively. The t-statistic for the parametric linear trend test that is used in SSP approach is $t = \frac{\hat{\beta}}{s(\hat{\beta})} \sim t_{w-2}$, where $\hat{\beta}$ is least square estimator for the trend parameter and $s(\hat{\beta})$ is the estimated standard error of $\hat{\beta}$. The null hypothesis of no trend is rejected at a significance level a if $|t| \geq t_{w-1,1-a/2}$. The selection of two significant levels is based on the assumption that there are not sudden changes in natural variations, which means that some time is needed in order a time series, which has a not obvious structural break to pass from no linear trend to linear trend condition. Thus, the existence of two significant levels describes the transition between these conditions. Thus, $t_{w-2,1-a_1/2}$, $t_{w-2,1-a_2/2}$ will be denoted as *Upper₁* and *Upper₂* thresholds, respectively and $-t_{w-2,1-a_1/2}$, $-t_{w-2,1-a_2/2}$ will be denoted as *Lower₁* and *Lower₂* thresholds, respectively. The SSP method is initially created in order to detect significant changes in linear trend of known time series. The version of the SSP presented here detects all possible changes in linear trend of a time series in real time.

In the following, the parametric linear trend test for a sliding window of size w on the time series $\mathrm{Y_t}$, $\mathrm{t} = 1, \ldots, \mathrm{n}$, is presented. Thus, for the first window $[\mathrm{Y_1}, \ldots, \mathrm{Y_w}]^{\mathrm{T}}$ the least square estimator for the trend parameter β is obtained as $\hat{\beta} = \sum_{t-1}^{w}(t-\bar{t})Y_t / \sum_{t-1}^{w}(t-\bar{t})^2$ (1), where $\bar{t}$ is the average time. The standard error of $\hat{\beta}$ is estimated from the power spectrum $s_1(\hat{\beta}) = \left[2\int_0^{0.5} W(f)S(f)\right]^{1/2}$(2). In (2), $W(f) = \left|\sum_{t=1}^{w} b_t e^{-2\pi i f}\right|^2$ with $b_t = \frac{t-\bar{t}}{\sum_{t=1}^{w}(t-\bar{t})^2}$ and $S(f)$ denotes the sample power spectrum of ε_t given as $S(f) = \frac{1}{2\pi}(\hat{\gamma}_0 + 2\sum_{\kappa=1}^{w-1}\hat{\gamma}_\kappa \cos(2\pi f k)$. Parameter $\hat{\gamma}_\kappa$ denotes the estimate of the kth order autocovariance of ε_t, given as $\hat{\gamma}_\kappa = \frac{1}{w}\sum_{t=1}^{w-k}\hat{\varepsilon}_{t+k}\hat{\varepsilon}_t$ for $k > 0$, where $\hat{\varepsilon}_t = Y_t - \hat{a} - \hat{\beta}t$ are the estimated residuals ($\hat{a} = \overline{Y_t} - \hat{\beta}t$ and $\overline{Y_t}$ is the mean of the time series), and $\hat{\gamma}_0 = \frac{1}{w-2}\sum_{t=1}^{w}\hat{\varepsilon}_t{}^2$ for $k = 0$.

Figure 7 provides a demonstration of the linear trend profile of the fill level time series.

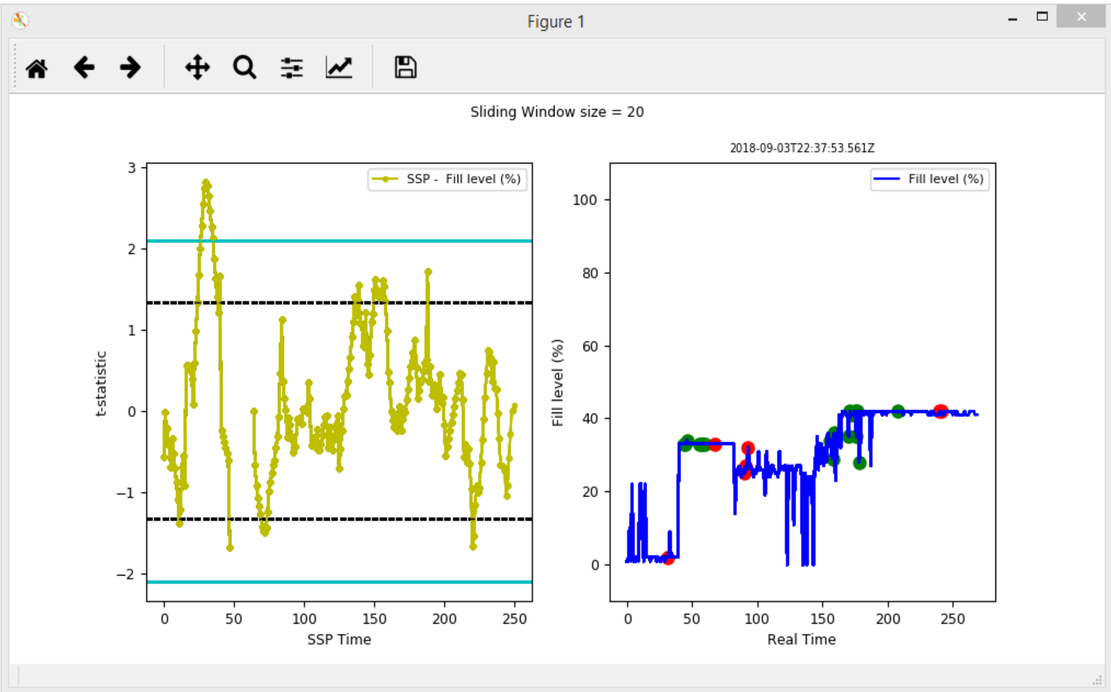


Figure 7: Linear trend profile (left subplot) and fill level time series (right subplot) during the detection of changes in linear trend. The cyan and dotted lines in negative denote the Lower$_1$ and Lower$_2$ thresholds, respectively, and the cyan and dotted lines in negative denote the Upper$_1$ and Upper$_2$ thresholds, respectively.

SSP method is applied on the time series of recordings (percentages) of fill level sensor described in Section 5.3. The interval between two recordings is 5 seconds (on average). The selected size of sliding window for SSP is 20 elements, based on intensive simulations and experiments. Figure 7(a) shows the linear trend profile (left subplot) and the fill level time series (right subplot). When the linear trend of fill level time series is steady, the linear trend profile points out that there are not significant changes indeed. In this test case, the time series has many ups and downs in small time intervals. This fact has as a result, the SSP methodology to detect many changes on the linear trend (see Figure 7(b), right subplot). In the case of fill level sensor, the SSP methodology, is set to detect only the time points when the linear trend has upward changes, a fact that indicates a state of bin fullness

In general, the real time approach of SSP method has the ability to detect multiple change points in linear trend of a time series and it is most effective in cases when the change(s) in a time series is(are) not abrupt, as happens in most natural variations.

6 Conclusions

In conclusion, the deliverable D3.7 – Computational Modelling, Simulation and Prediction in Logistics II describes the effort spent from M4 to M28 and represents the final status of Task3.3 – Simulation and Forecasting in Production and Logistics of WP3. This report documents the final results in the part of logistics similarly to production results which are documented in D3.5 – Computational Modelling, Simulation and Prediction in Production II (will be submitted in parallel). With this report, the work of Task 3.3 related to logistics is completed.

Based on the analysis of the applied methods, the Genetic Algorithm for ELDIA contractual recyclable material management indicates the optimum pair of routes and weights for waste and wood materials in order to enhance the decision making by the ELDIA purchase manager. Furthermore, the application of the Slope Statistic Profile methodology enhanced the COMPOSITION system with the ability to detect the waste bins fullness. By using this algorithm's estimations and notifications of the fill level containers, the scrap metal bidding process will be able to start and be completed earlier. The time series prediction engine provides a generalized tool for predictions and calculation of probabilities using state-of-art methodologies from the fields of statistics (auto-regression and moving average models) and probability theory (Markov chains).

7 List of Figures and Tables

7.1 Figures

Figure 1: COMPOSITION architecture functional view 7
Figure 2: Fill level sensor's position 8
Figure 3: Fill level sensor system overview 9
Figure 4: Ratios of routes per weights for all customers, materials and years 12
Figure 5: Predictions based on MA model from 2 to 5 months ahead. 14
Figure 6: Time series plots with the calculated transition probabilities for M=5 and states “Up” to “Up” (left plot) and M=5 and states “Up” to “Down” (right plot). 16
Figure 7: Linear trend profile (left subplot) and fill level time series (right subplot) during the detection of changes in linear trend. The cyan and dotted lines in negative denote the $Lower_1$ and $Lower_2$ thresholds, respectively, and the cyan and dotted lines in negative denote the $Upper_1$ and $Upper_2$ thresholds, respectively. 17

7.2 Tables

Table 1: Abbreviations and acronyms are used in this deliverable 5
Table 2: Contractual recyclable material management dataset 10
Table 3: Route/weights ratio 11
Table 4: Fitness function average value for material / customer for years 2015 and 2016. 11
Table 5: Objective bounds of routes and weights per material 11
Table 6: GA selections 12
Table 7: Customer’s route and tonnage statistics 13

8 References

(DeJong, 1975) K. A. DeJong, An Analysis of the Behaviour of a Class of Genetic Adaptive Systems, University of Michigan, 1975.

(Goldberg, 1989) D. E. Goldberg, Genetic Algorithms in Search, Optimization, Machine Learning, Addison Wesley, 1989.

(Holland, 1975) J. H. Holland, Adaptation in Natural and Artificial Systems, The University of Michigan Press, Ann Arbor, Mich, USA, 1975.

(Vafeiadis, 2011) Vafeiadis, T., Bora-Senta, E., Kugiumtzis, D., (2011). 'Estimation of linear trend onset in time series'. Simulation modelling – Practice and Theory, Vol. 19, Issue 5, p 1384 - 1398.