CAPTOR

Collective Awareness Platform for Tropospheric Ozone Pollution

Work package	WP3
Deliverable number	D3.3a:
Deliverable name	Report on data validation (a).M10
Deliverable type	R (Document report)
Dissemination level	PU (Public)
Estimated delivery date	M10
Actual delivery date	12/12/2016
Actual delivery date after EC review	15/09/2017
Version	2.0
Comments	Final version after EC review

Authors:
Mar Viana; Anna Ripoll; Marc Padrosa; Xavier Querol (CSIC)
Jorge García-Vidal; J.M. Barcelo-Ordinas (UPC)

Document History

Version	Date	Contributor	Comments
V0.1	28/11/2016	Mar Viana, Marc Padrosa, Anna Ripoll (CSIC)	Outline of contents, First Draft
V0.2	09/12/2016	Jose M. Barcelo-Ordinas (UPC)	Revision
V1.0	12/12/2016	Mar Viana, Marc Padrosa, Anna Ripoll (CSIC)	Final version submitted to EC
V2.0	15/09/2017	Mar Viana, Marc Padrosa, Anna Ripoll, Xavier Querol (CSIC); Jorge García-Vidal, J.M. Barcelo-Ordinas (UPC)	Final version after EC review

Table of Contents

List of figures

List of Tables

List of Abbreviations

DiY	Do it Yourself
EC	European Community
EU	European Union
MLRA	Multi-Linear Regression Analysis
NE	North-East
NOx	Nitrogen Oxides
SD card	Secure Digital card
VOCs	Organic Volatile Compounds

Executive Summary

Description of the work

The aim of deliverable D3.3a is to report on data validation during the summer 2016 monitoring campaign. This deliverable presents a review of the nodes deployed during the campaign and the data they reported, focusing on data quality and validation. It includes also the first quantitative results on ozone concentrations of the project. These results are the first real-world experience with the CAPTOR sensors, and are therefore subject to potentially frequent failures as discussed in deliverable 3.2a. Specifically, the work presented in this deliverable aims to understand these results and the reasons behind the positive and negative aspects of the 2016 summer campaign.

Objectives

In order to understand the results we present:

- The number of sensors deployed and how these work in order to assess data availability.
- Calibration of nodes by comparison with reference data, in order to obtain the most adequate correction factors (beta values).
- Quantification of actual ozone concentrations (in µg/m^3), including data treatment.
- Graphical representation of the final ozone concentration data.

Report on initial testing of sensor nodes

1. Introduction

The nature and impacts of air pollution effects on human health and ecosystems are relatively well known at present. It is for this reason that monitoring and quantifying the ambient concentrations of atmospheric pollutants is of major relevance. In particular, because of the high tropospheric ozone concentrations registered in rural areas, project CAPTOR aims to produce dense and high-quality network of sensor nodes to monitor the concentrations of this type of pollutant and to determine its spatial and temporal evolution. Tropospheric ozone is a secondary pollutant which originates from photochemical reactions linked to its gaseous precursors nitrogen oxides (NOx) and organic volatile compounds (VOCs), and solar radiation.

The aim of CAPTOR is to monitor ozone concentrations in rural areas, using low-cost and widely distributed sensors. To this end, a number of sensor nodes was deployed in the study area (Catalonia testbed) during summer 2016, which are able send data remotely and also to store them in an internal memory card. The 2016 summer campaign included three monitoring periods:

1) Calibration phase 1: prior to the monitoring campaign, this period covered the month of June 2016. The nodes were installed in the IDAEA-CSIC air quality monitoring station in Barcelona. During this period, the nodes were inter-compared with reference ozone data from the air quality monitoring station.
2) Monitoring campaign: between July and mid September. The nodes were deployed in the study area (Figure 1).
3) Calibration phase 2: after the monitoring campaign, this period covered the month of September 2016. The nodes were installed in the IDAEA-CSIC air quality monitoring station in Barcelona. During this period, the nodes were inter-compared with reference ozone data from the air quality monitoring station.

At the end of the summer campaign (July to September 2016), the nodes were collected and taken back to the lab for data analysis. The data collected were used to perform an initial study on node performance, data availability, and ozone concentrations. The data were also used as a basis for comparison with the results from the upcoming 2017 and 2018 summer campaigns within the project.

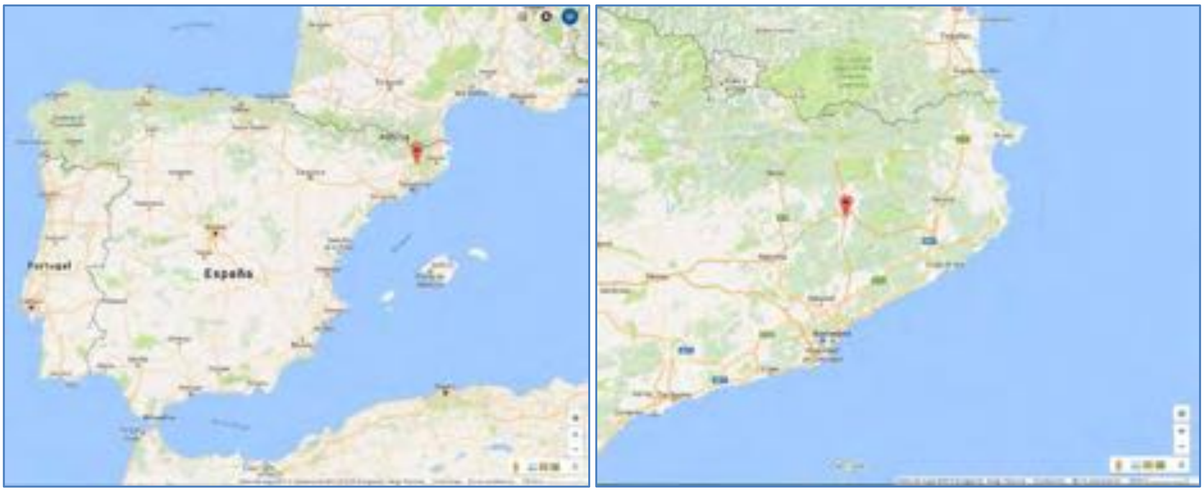


Figure 1. Location of the study area, between Barcelona and Vic, in Catalonia (NE Spain).

2. Results

2.1 Data collection

The first issue to be solved was data collection. The Captor nodes are designed to transfer data remotely to a server at the UPC in Barcelona (Commsensum). In addition, a second server was set up in France to be used as a backup, in such a way that all data were simultaneously sent to both servers. In addition, data were also stored internally in each Captor in an SD card. This system was designed to be redundant and minimise any potential for data loss.

For the 2016 campaign, the data presented in this deliverable were collected directly from the SD cards. The main reason for this was the failure to submit data remotely by a large number of nodes, as described in deliverable 3.2a. However, the adequate design of the data collection system, based on redundancy, ensured that the data were indeed not lost. One additional issue detected during this campaign was that when the data stored in the UPC and French servers was compared, it was observed that they were not exactly the same (as would have been expected). The differences detected were mainly missing or duplicate datapoints. As a result, the final data used for summer 2016 are those stored in the SD cards. The online submission of the data was in any case extremely useful during the campaign given that they helped us determine when the nodes did not work and therefore when they needed repair.

2.2 Data availability

The overall assessment of data availability is described in deliverable 3.2a. In short, during summer 2016 a total of 20 nodes was deployed in the study area. Out of these, the following was observed:

- 1/20 nodes did not submit any data.
- 7/20 nodes did not submit any data during the campaign (July-September period, in orange), although they did report data during the calibration periods (green).
- 10/20 nodes reported data during the calibration and the campaign periods, but the datasets are incomplete.
- 3/20 nodes produced full datasets for the entire period (calibrations and campaign).

In addition to these, 6 nodes were co-located at official air quality monitoring stations belonging to the regional network run by the Catalonia Government. These nodes aimed to provide data for onsite calibration outside Barcelona, where ozone concentrations were expected to be higher.

2.3 Calibration of nodes

Figure 2 shows the detailed results for each node and sensor deployed during the 2016 summer campaign. The following parameters are shown for each node:

- Top left: time series of electrical resistance data (raw output of the sensors) for the two calibration periods (before and after deployment in the field), for each ozone sensor in the node (3 sensors/node in most nodes). In parallel, the pink curve shows the ozone reference concentration (in µg/m^3) from the air quality monitoring station (Palau Reial, Barcelona).

- Top right: correlation plots between each individual sensor and the reference ozone concentrations.
- Bottom: time series of electrical resistance data (raw output of the sensors) for the campaign period, when the nodes were deployed in the field, for each ozone sensor in the node (3 sensors/node in most nodes). For this period no reference data are available for comparison, given that the nodes were deployed at volunteers’ homes.

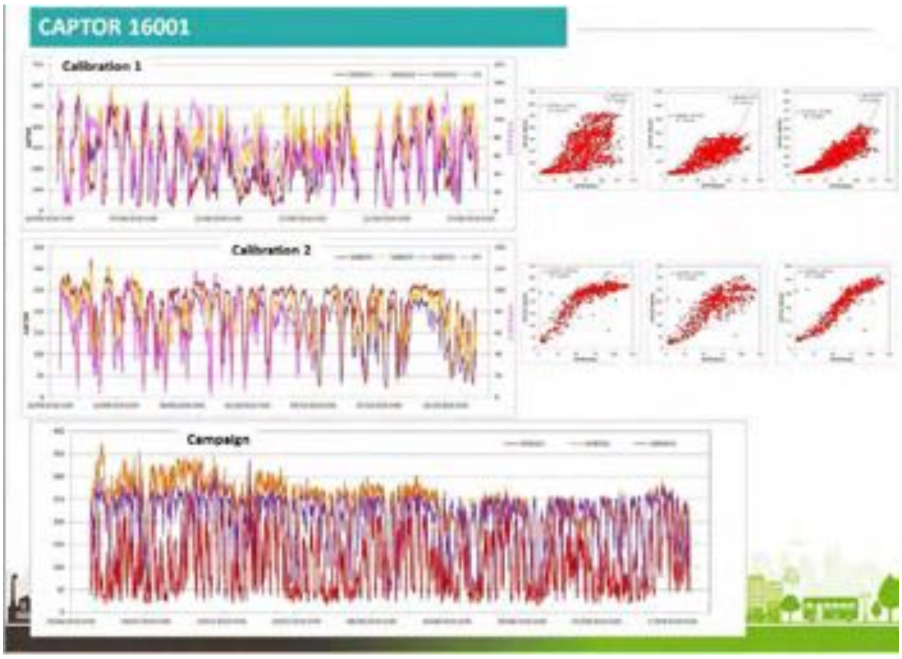


Figure 2. Summary of data performance and validation of the 20 nodes deployed at volunteers’ homes during the 2016 summer campaign.

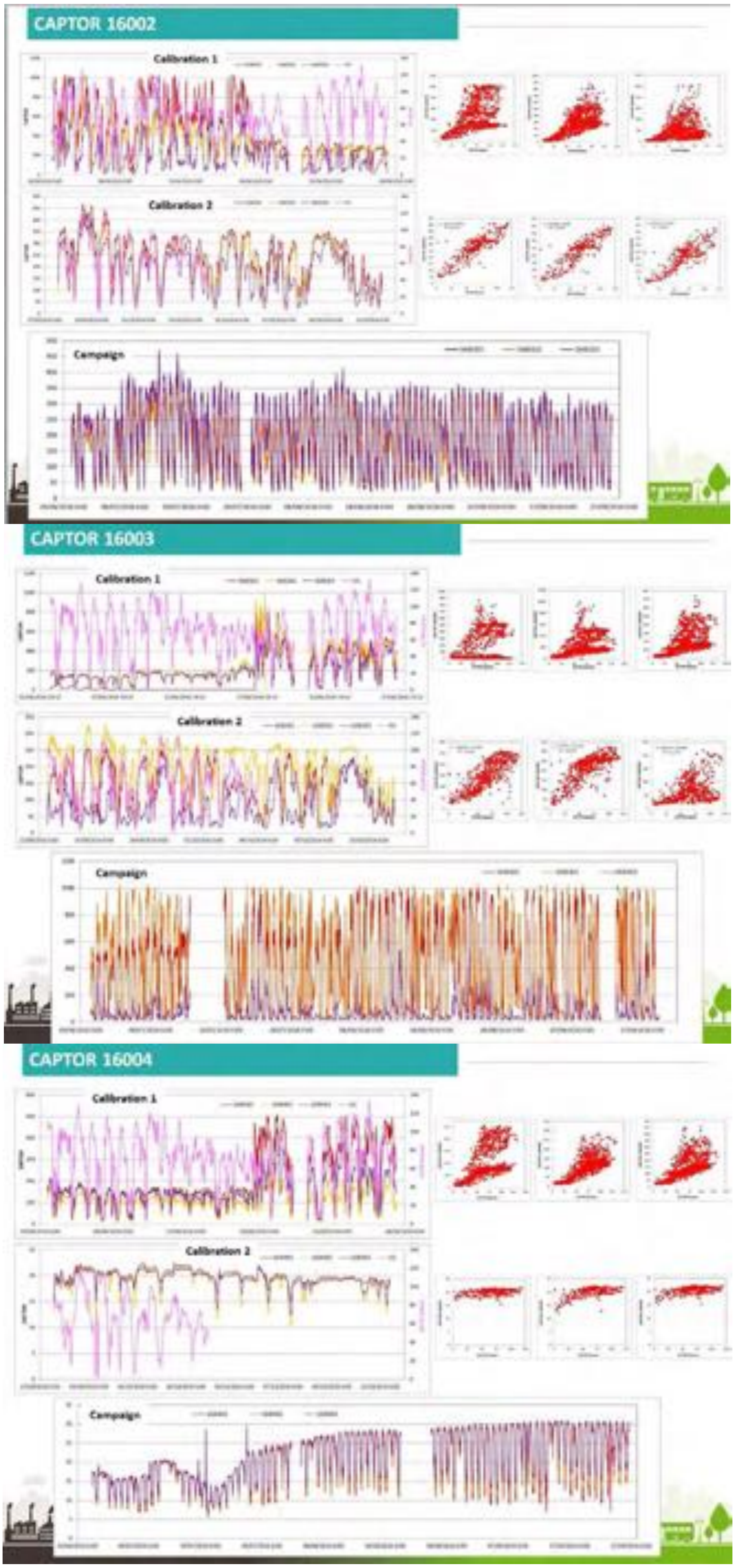


Figure 2. Continued.

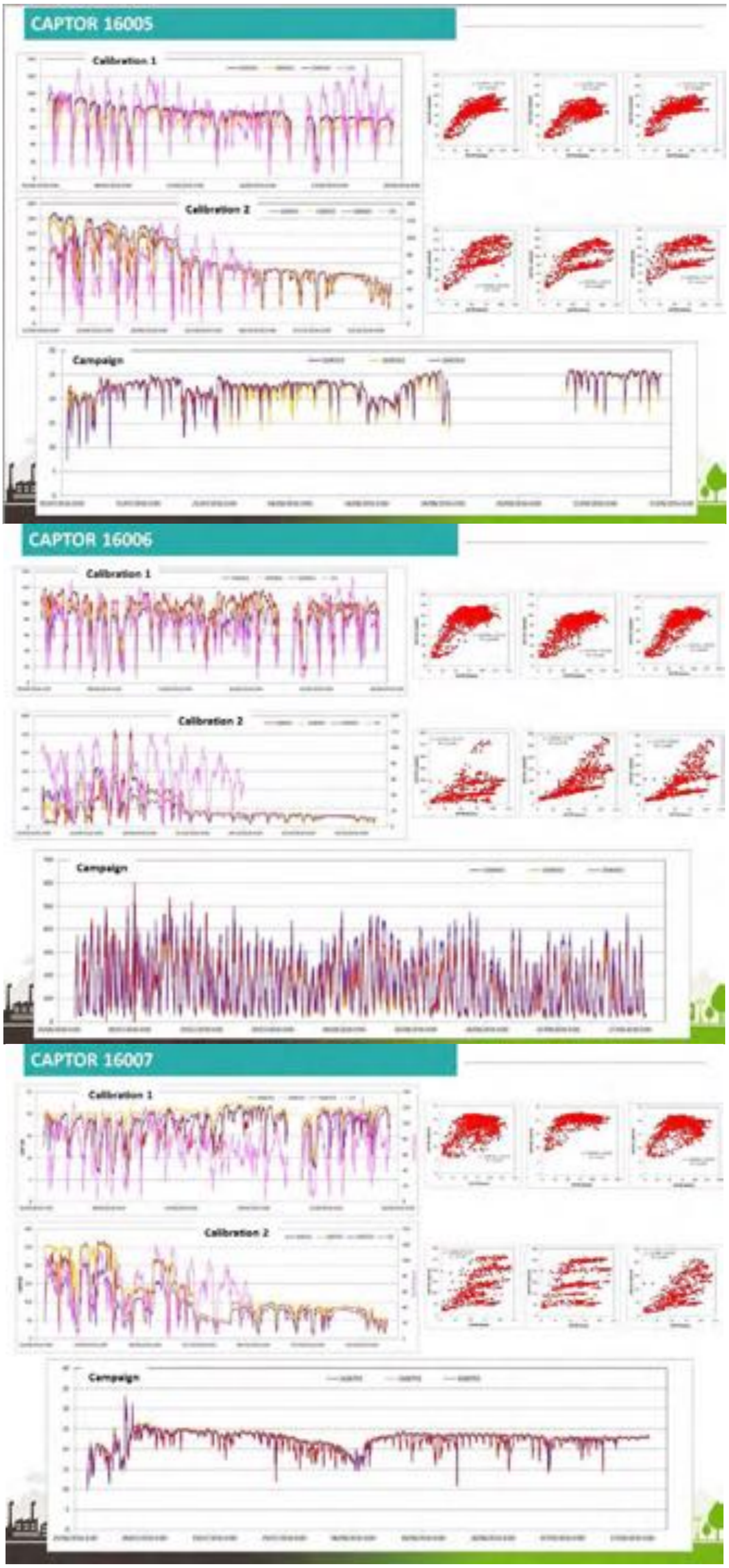


Figure 2. Continued.

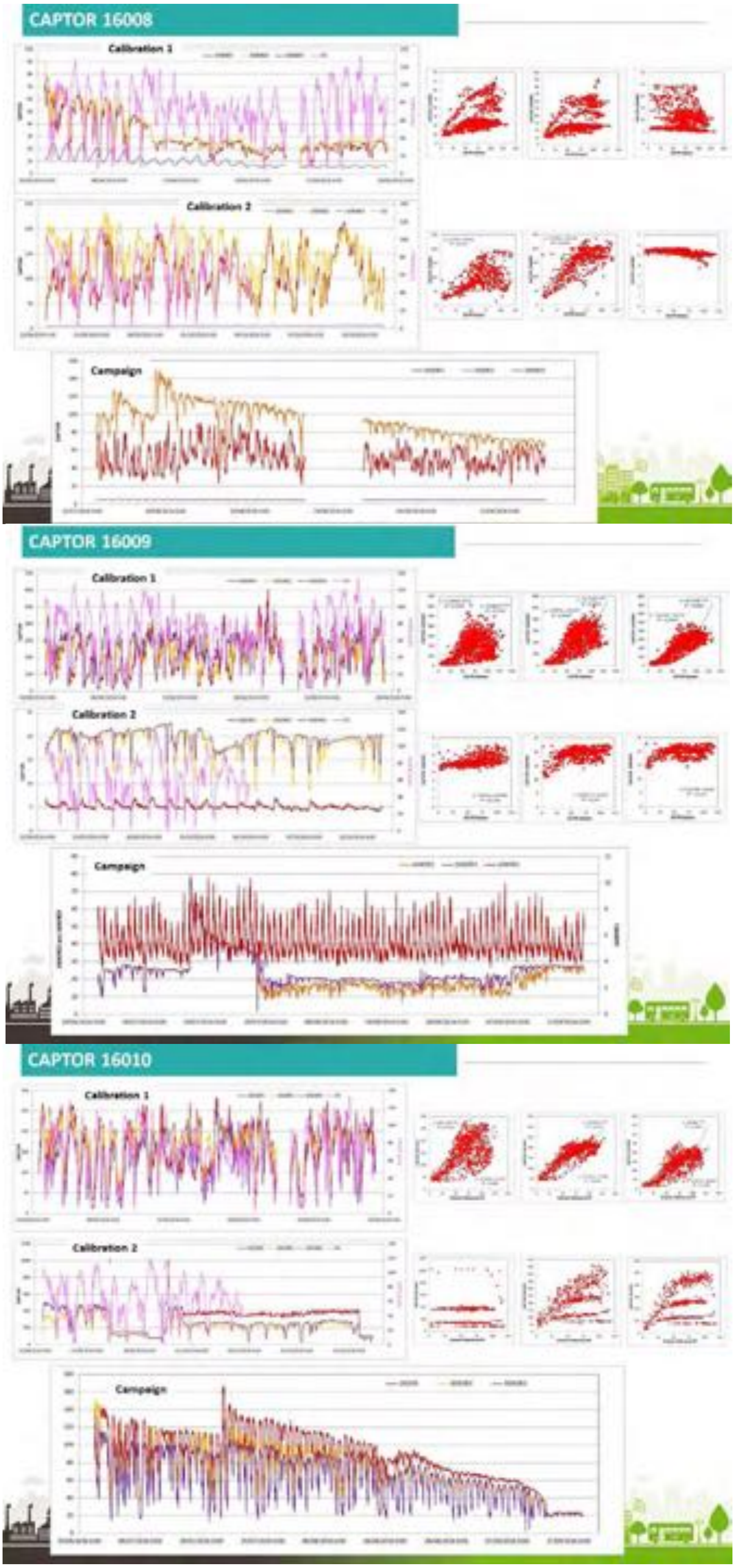


Figure 2. Continued.

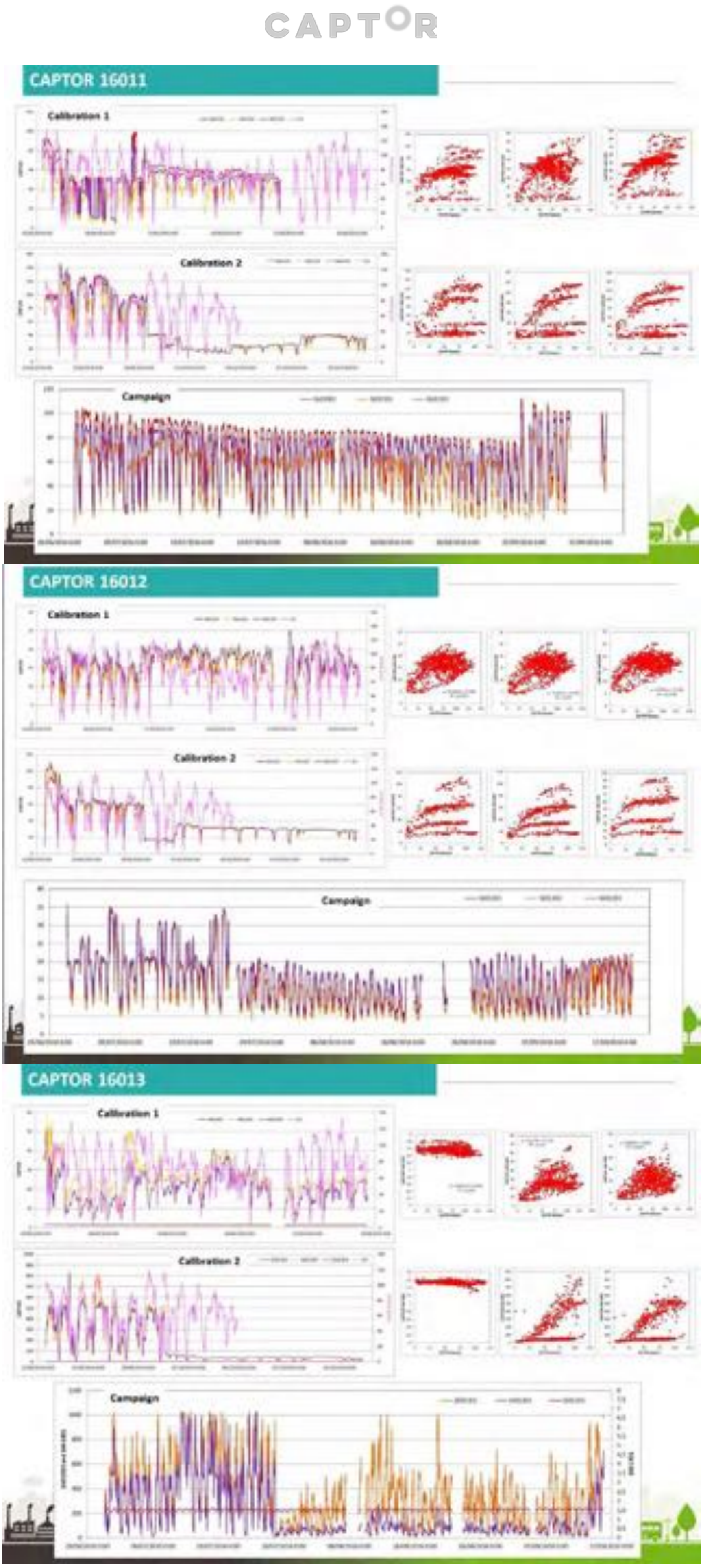


Figure 2. Continued.

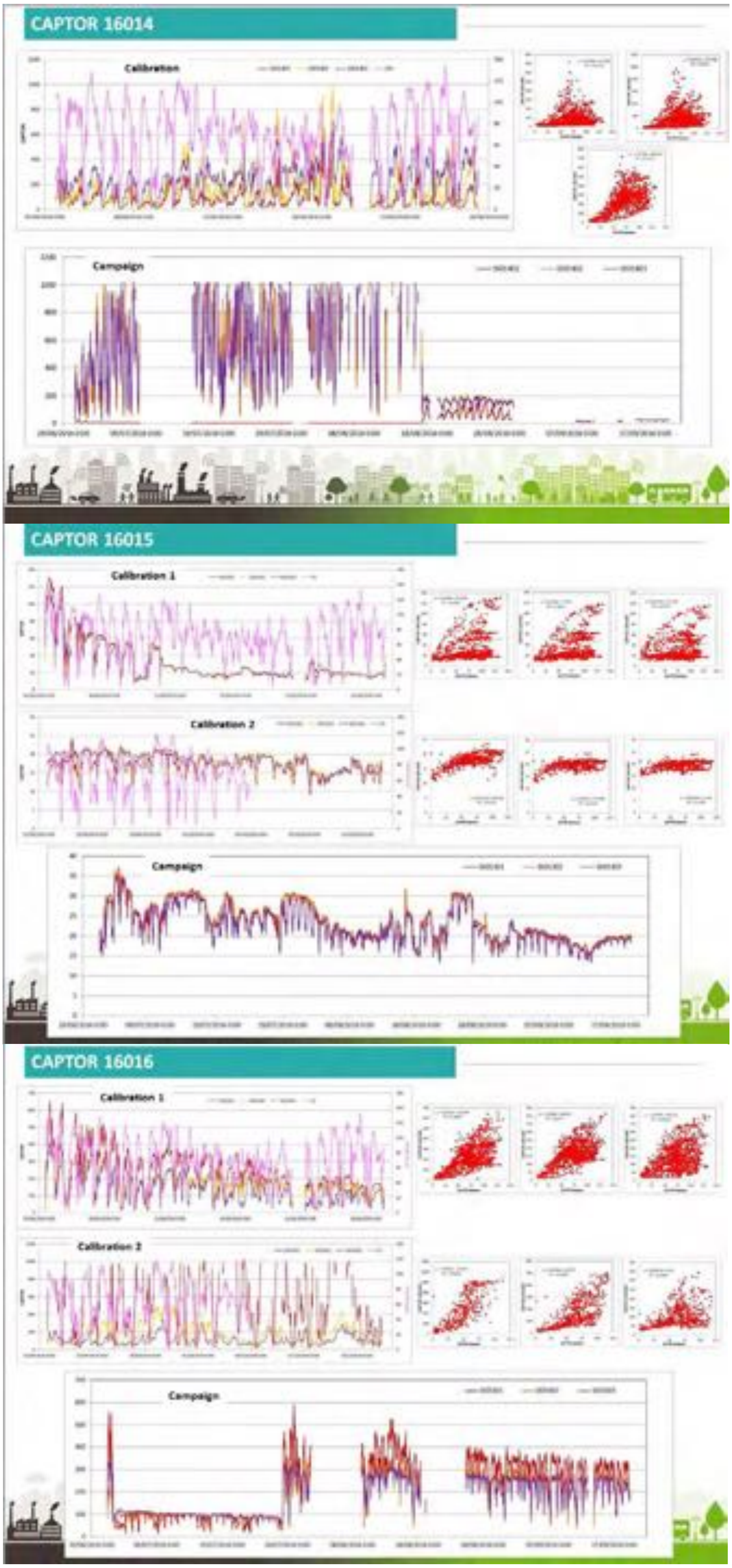


Figure 2. Continued.

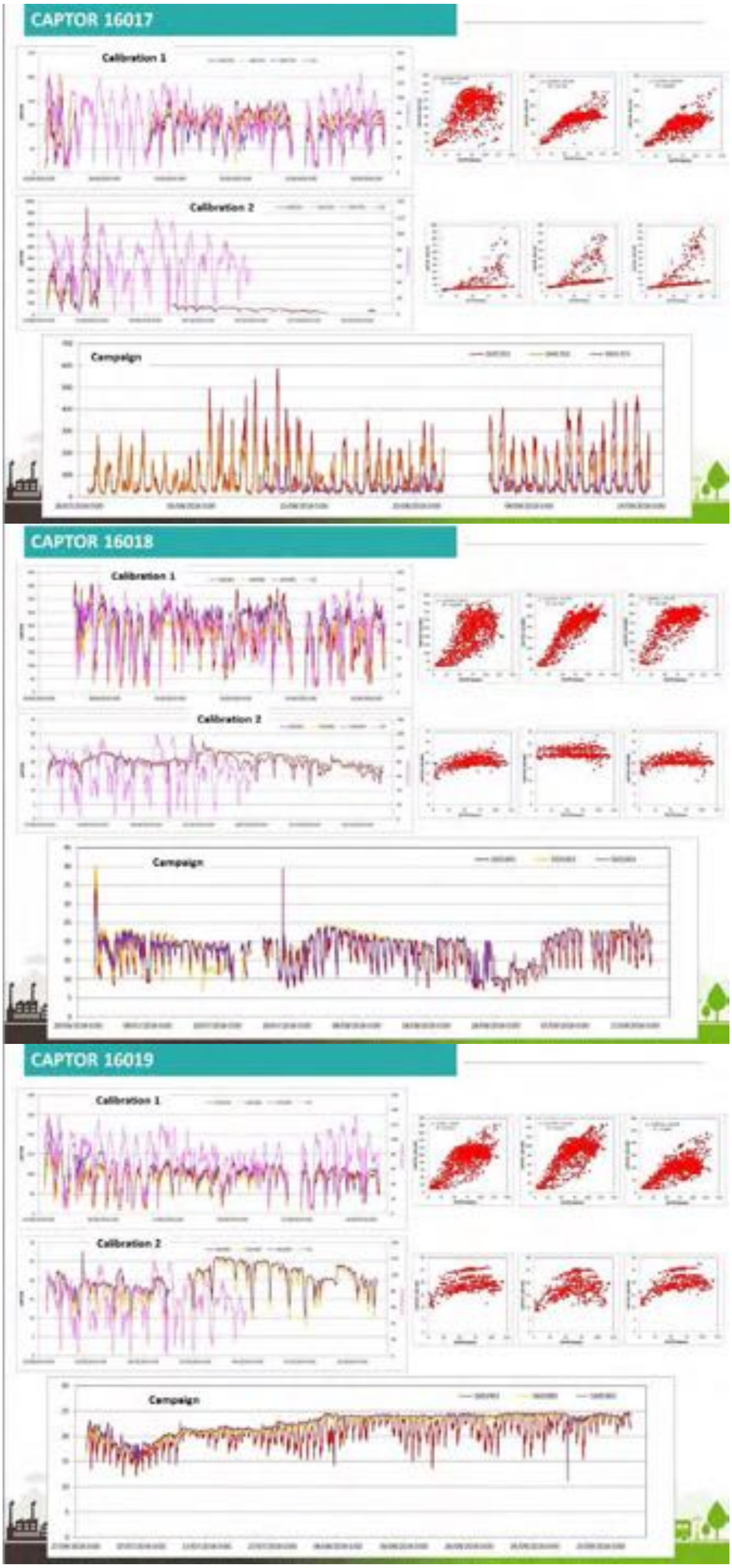


Figure 2. Continued.

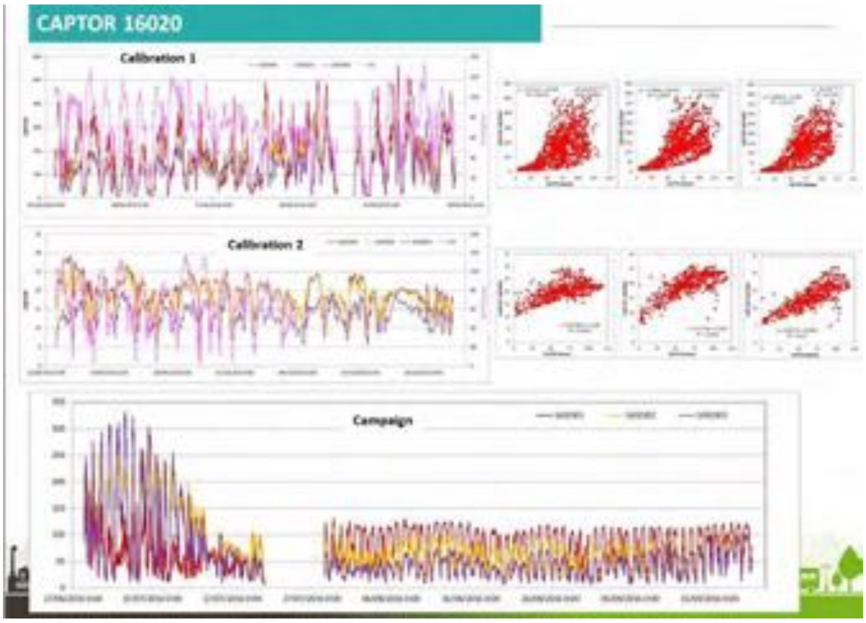


Figure 2. Continued.

The correlation between the individual sensors and the reference instrument is essential for this work, given that it is the only tool to convert the electrical signal produced by the sensors into ozone concentrations (in µg/m^3). This conversion is carried out by multi-linear regression between the ozone sensors, the humidity and the temperature sensors, and the reference data. Multi-linear regression produces correction factors referred to in this work as "beta coefficients".

As evidenced in Figure 2 and Figure 3 below, the performance varied largely as a function of the nodes. In short, three main behaviours were observed:

- Nodes classified as "Priority#1". There were 7 of these cases: 16001, 16002, 16003, 16005, 16016, 16020, and 16025. This kind of nodes provided sufficient data for both calibration periods, or at least one of them, as well as for the campaign period. This may mean that not all of the sensors performed correctly (e.g., sensor 1 in 16001, which showed a poor correlation with the reference).
- Nodes classified as "Priority#2". There were 11 of these cases: 16004, 16006, 16008, 16009, 16010, 16012, 16013, 16017, 16018, 16019, and 16028. This kind of node showed good correlation with the reference data for one of the calibration periods, but poor data for the other one. This implies that the beta coefficients may only be calculated with one of the calibration periods, and that thus we are unable to assess the validity of this beta coefficient, or potential drifts in the performance of the sensors. Given that this was the first time that Captor nodes were deployed widely in the field, no data are available yet regarding drifts over time. The purpose of carrying out two calibration campaigns, before and after the field work, was to assess the stability of the correction coefficients over time. This assessment could not be carried out for this kind of nodes.
- Nodes classified as "Priority#3". There were 4 of these cases 16007, 16011, 16015, and 16021. Correlations between the sensors and the reference were poor for both campaigns, thus rendering the data not valid. The absence of reliable correlations between sensor and reference data makes it impossible to obtain the necessary beta coefficients to calibrate the sensors.

- Nodes classified as “Priority#4”. There were 2 of these cases: 16022 and 16023. These nodes showed failures and require a longer calibration period. Therefore, they are still under assessment.
- Nodes classified as “Priority#5”. There were 2 of these cases: 16014 and 16024. The datasets for these nodes are still incomplete, but the cause is unclear. They require further assessment.
- Nodes classified as “Priority#6”. There were 2 of these cases: 16026 and 16027. Data available only for the first calibration campaign. The cause is unclear, and they require further assessment.

With regard to the nodes deployed at air quality monitoring stations (Figure 3), it should be noted that each node had 5 sensors (as opposed to 3 in the nodes deployed in volunteers’ homes). This was meant to increase accuracy of these nodes. Results showed a good performance for 3 of them (16023, 16024, and 16025) but poor performance for the rest.

Figure 3. Summary of data performance and validation of the 6 nodes deployed at official air quality monitoring stations during the 2016 summer campaign.

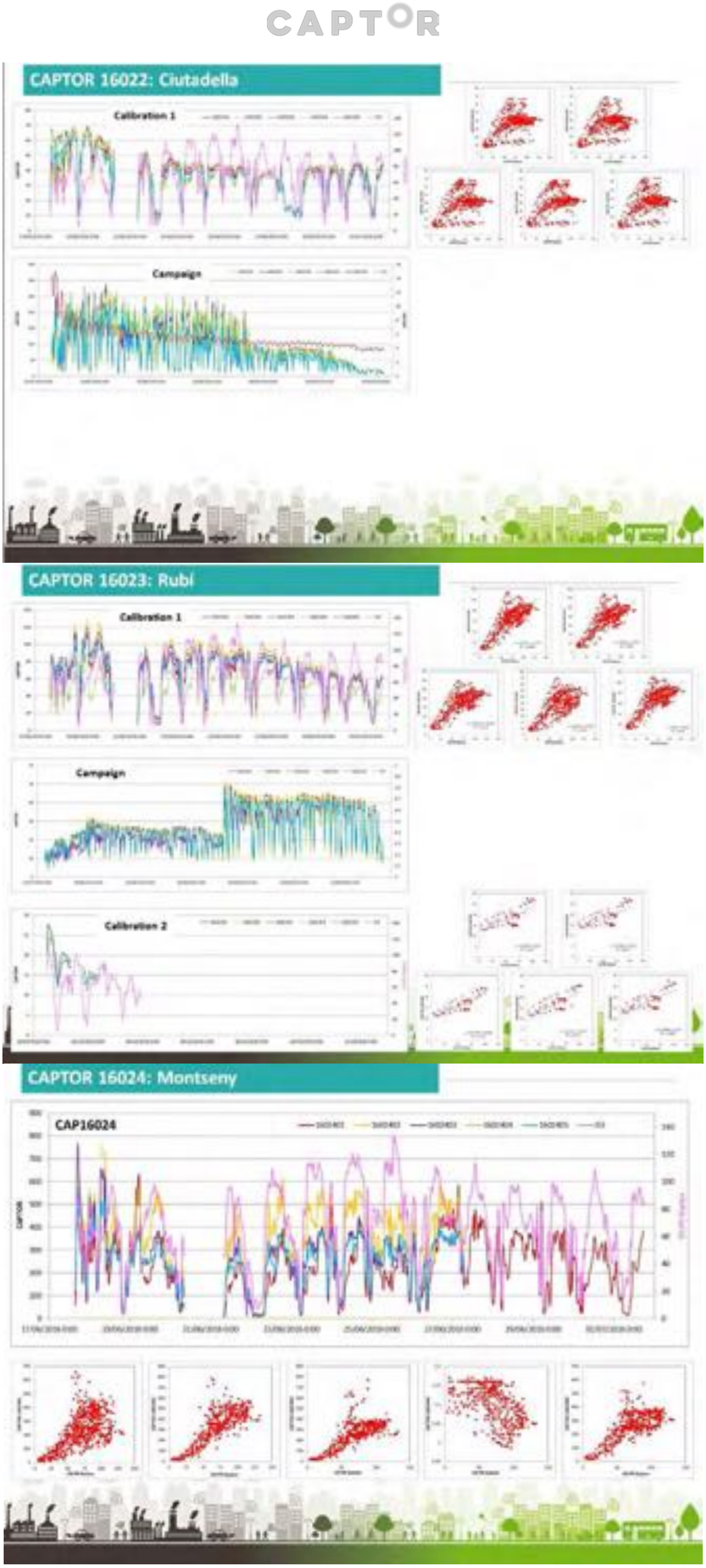


Figure 3. Continued.

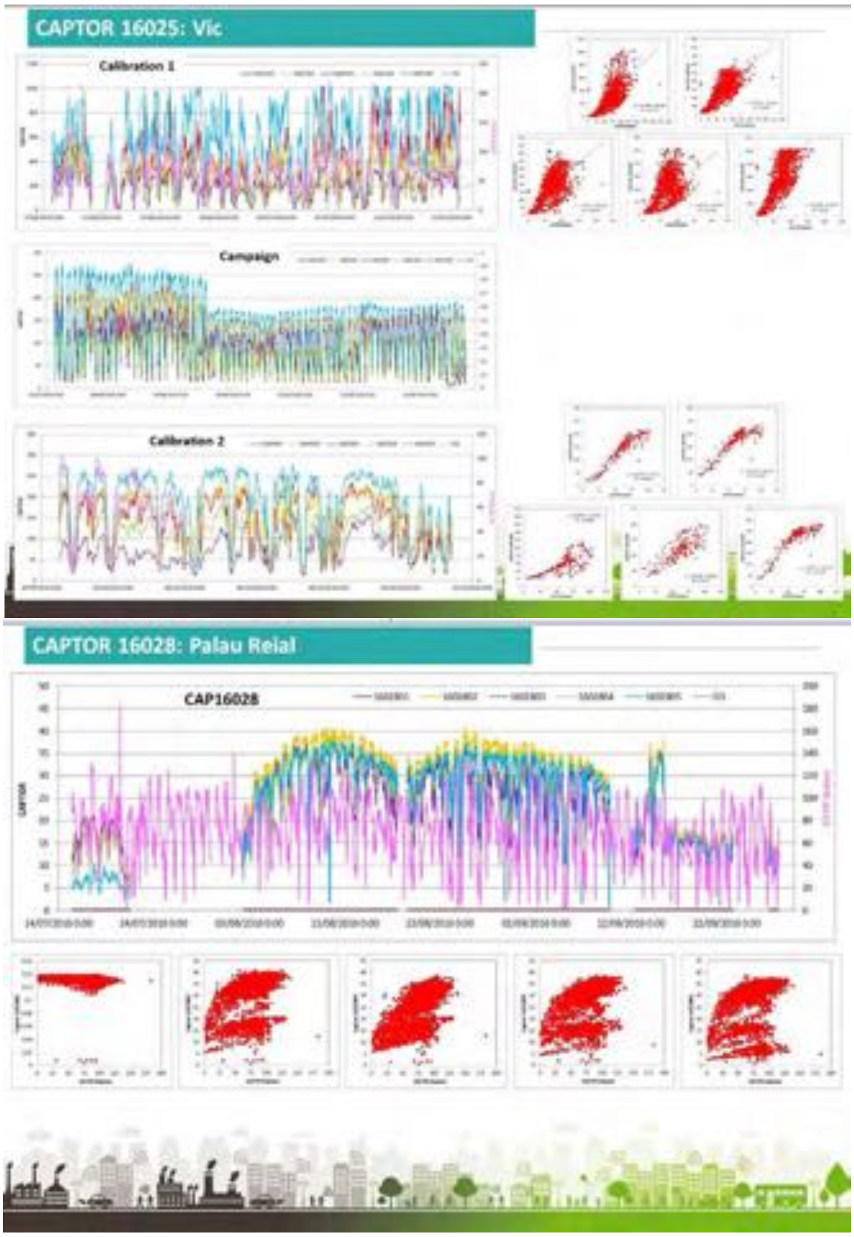


Figure 3. Continued.

Two additional issues were detected while assessing the data in Figure 2 and Figure 3, which complicated node calibration. The first was detected by analysing the time series of the reference and sensor data, when sudden increases in the baseline concentrations were observed (e.g., node 16023, Figure 3). These jumps in the baseline concentrations clearly seemed to result from instrumental issues, and not from the actual variability of ozone concentrations. The CAPTOR partners in charge of these issues are already assessing possible causes, working with a number of hypotheses. Their results will be reports in upcoming deliverables. The second main issue refers to the different performances observed for a number of nodes during the field campaign, when compared to the initial calibration period (e.g., nodes 16009 or 16010, Figure 2). In both cases the nodes showed good correlations during the calibration phase, but reported poor data during the field campaign. This decrease in performance could be related to the transport of the nodes to the volunteers' homes, whereby the sensors may have suffered or disconnected etc. However, other factors may have affected node performance, and the ultimate cause behind it is still unclear.

Finally, it should be noted that all of these experience have been collected and that the construction of new Captor prototypes is underway. These new prototypes will include improvements such as:

- Improved software for removal of outliers
- Improved software to avoid baseline changes, by isolating the readings of the individual sensors
- Improved hardware to prevent sensor detachment from the board
- Improved data processing, including input data scaling
- Improved deployment protocol, including exchange of malfunctioning sensors before deployment in the field.

The positive results from the implementation of some of the improvements described above may be observed in the results of Captor 16031 (Figure 4). It is expected that the results from the upcoming campaigns will resemble those from unit 16031 with regard to robustness and data comparability with respect to reference data.

Figure 4. Inter-comparison between sensor and reference data for one node including the improvements and lessons learnt from the 2016 summer campaign.

2.4. Data processing and calculation of beta coefficients

Data processing included careful filtering of the datasets, aligning of the sensor and reference time series, and selection of the specific time periods for each node to be used in the multi-linear regression analysis (MLRA). Time series were carefully screened to identify anomalous behaviours, which were excluded from the analysis. It also involved obtaining reference ozone concentrations from the local air quality monitoring network (station Palau Reial). The temporal resolution of the data was 30 minutes. Once this was carried out, MLRA was applied to obtain the correction factors (beta coefficients) to convert the electric signal produced by the sensors into ozone concentrations calibrated with respect to the ozone reference data from the air quality monitoring network.

Because two calibration periods were available for some of the nodes, it was decided that the most appropriate approach would be to apply separate MLRA to the data from each of them.

It should be noted that MLRA was applied in this case without prior scaling of the input data. This process (scaling) will be applied to the datasets from the 2017 and 2018 monitoring campaigns, but it was not completely implemented for the 2016 data. Therefore we have chosen to present the analysis in this deliverable without initial scaling. The impact of this may be observed in Table 1, where the beta coefficients for temperature are one order of magnitude higher than the rest, implying that the variability of the variable temperature will have a stronger impact on the final ozone concentrations than that of the rest of the variables. Scaling of the input data would avoid this artefact, providing equal relative weight to all of the variables. This will be improved for the 2017 and 2018 campaigns.

Table 1. Correction factors (beta coefficients) calculated by multi-linear regression analysis (MLRA) to calibrate the Captor node data (electrical signal) with respect to the reference ozone concentrations (in µg/m^3). Top (orange): Priority#1 nodes. Bottom (green): Priority#2 nodes.

		Sensor 1	Sensor 2	Sensor 3	Sensor 4	Sensor 5	Temperature	Humidity	intercept
16001	Cal 1	-0,09591	0,12653	0,13352			2,57589	0,38528	-58,05208
	Cal 2	0,08876	-0,03480	0,31464			1,33957	0,35530	-42,89418
16002	Cal 1	0,00438	0,10028	0,01545			2,50335	-0,03907	-25,54676
	Cal 2	0,16230	0,05042	0,00715			2,45662	0,18246	-57,06125
16003	Cal 1	0,06508	-0,01772	0,08572			3,52870	0,08542	-48,98216
	Cal 2	0,13785	0,15542	0,04925			3,02382	0,20225	-64,06390
16005	Cal 1	2,55484	-0,84446	-0,70218			2,45720	0,06508	-68,12839
	Cal 2	0,21144	1,93964	-1,56463			1,98852	-0,05301	-19,66666
16016	Cal 1	0,05712	0,19137	0,02518			3,17764	-0,01758	-55,37051
	Cal 2	x	x	x			x	x	x
16020	Cal 1	0,09240	-0,09665	0,21368			3,03467	0,04796	-34,38594
	Cal 2	0,62156	1,91168	3,17211			3,53245	0,09209	-110,82940
16025	Cal 1	0,03485	0,08342	0,18755	-0,25461	0,04646	x	x	28,42694
	Cal 2	0,12164	-0,08371	0,09369	0,10050	0,15871	3,08929	0,24328	-80,58510

		Sensor 1	Sensor 2	Sensor 3	Sensor 4	Sensor 5	Temperature	Humidity	intercept
16004	Cal 1	-0,02088	0,91819	-0,22283			2,83119	0,11074	-44,64942
	Cal 2	-15,91887	7,31508	10,08967			-0,44415	-5,41905	143,52008
16006	Cal 1	0,01483	0,32395	0,68392			2,63415	0,03467	-70,27904
	Cal 2	-0,01208	-0,09650	0,34590			0,65429	-0,45389	25,64485
16008	Cal 1	x	x	x			x	x	x
	Cal 2	0,14047	0,28738				3,87877	0,23815	-95,93529
16009	Cal 1	-0,00201	0,10542	0,10864			3,32786	0,32034	-76,90991
	Cal 2	x	x	x			x	x	x
16010	Cal 1	-0,00691	0,41406	0,06860			3,72029	0,37609	-116,39783
	Cal 2	x	x	x			x	x	x
16012	Cal 1	-8,63165	11,90772	3,21901			3,17723	0,46364	-114,36854
	Cal 2	-3,26348	4,50375	-0,19004			3,07714	0,26680	-62,48813
16013	Cal 1		-0,13777	2,16403			3,12081	0,18771	-59,63005
	Cal 2		0,02474	0,10834			2,73711	0,09326	-47,40284
16017	Cal 1	0,00448	0,35747	0,44760			2,13547	-0,21830	-47,56161
	Cal 2	-0,07806	0,19107	0,15194			1,15010	-0,15751	4,22973
16018	Cal 1	0,02734	0,36249	-0,08063			2,42649	0,09188	-47,38529
	Cal 2	x	x	x			x	x	x
16019	Cal 1	0,11668	0,36529	0,27306			3,59463	0,24232	-95,19755
	Cal 2	-6,35132	2,41631	10,61916			0,56960	-0,49104	-36,75602
16028	Cal 1		11,38995	-9,43254	3,13261	x	3,03862	-0,10026	-66,38308
	Cal 2	x	x	x	x	x	x	x	x

The results obtained from the MLRA analysis are summarised in Table 1. This analysis was applied to Priority#1 and Priority#2 nodes, given that they were the only ones for which the initial assessment showed sufficient data quality. Table 1 summarises the beta coefficients obtained for each node and calibration campaign (cal1 and cal2). Initially, the correction factors obtained from the two calibration periods should be similar: this would mean that we are observing a linear relation, and that there are no drifts over time (e.g., from ageing of the sensors). To the authors' knowledge, no ageing experiments under real-world conditions are available for this kind of sensors.

Aside from the different order of magnitude of the different variables, as described above, the results show that (once again) the different nodes had different behaviours. Whereas similar coefficients were obtained for both calibration periods for certain nodes (e.g., 16001), larger differences were obtained for others (e.g., 16004, 16005). Unfortunately, with the data available it is not possible to assess the potential influence of drifts over time given that the variability observed could be linked to actual drifts in sensor performance but also to other factors such as hardware problems, impact of transport, etc. Further studies including additional campaigns are necessary to assess this. There are numerous nodes for which calibration data are unavailable or showed poor performances. These have been removed from the analysis in order not to bias the results.

2.5. Quantification of ozone concentrations

The correction coefficients calculated in the previous sections were then applied to the sensor time series (3 ozone, 1 temperature and 1 relative humidity sensors) in order to obtain ozone concentrations in µg/m^3. As stated above, the beta coefficients were obtained without prior scaling of the input data, which may result in obtaining a slightly worse calibration of the nodes. The time series of ozone concentrations quantified for the field campaign period (July to September 2016) for each individual node are shown in Figure 5. When data were available from a nearby air quality monitoring station (within a <7 km radius), were available, these were plotted for visual comparison purposes (not for calibration). However, it must be noted that the station may be located at a different altitude or be influenced more or less by traffic emissions (which result in lower ozone concentrations) and therefore ozone concentrations may not be comparable.

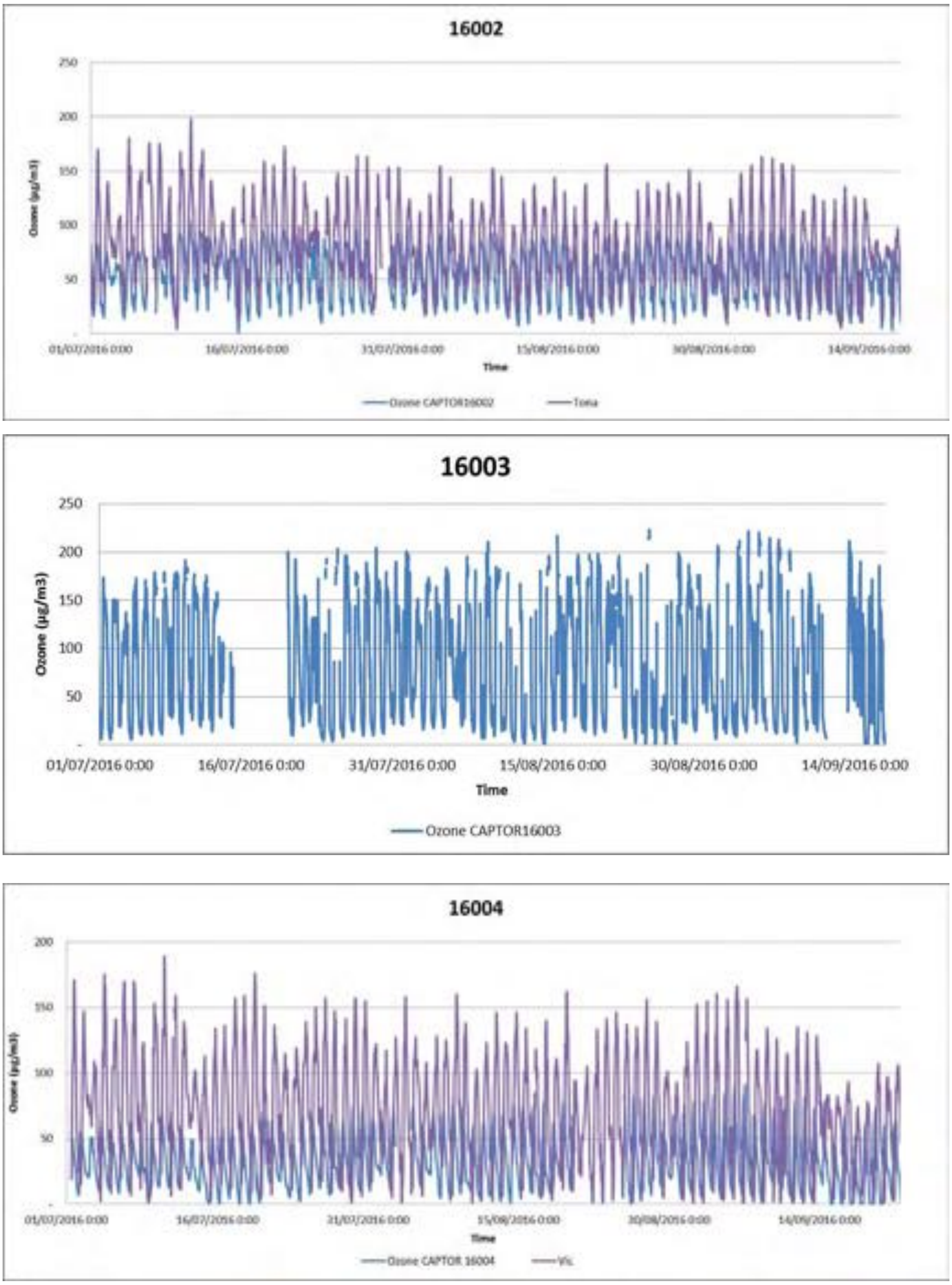


Figure 5. The time series of ozone concentrations quantified for the field campaign period (July to September 2016) for each individual node.

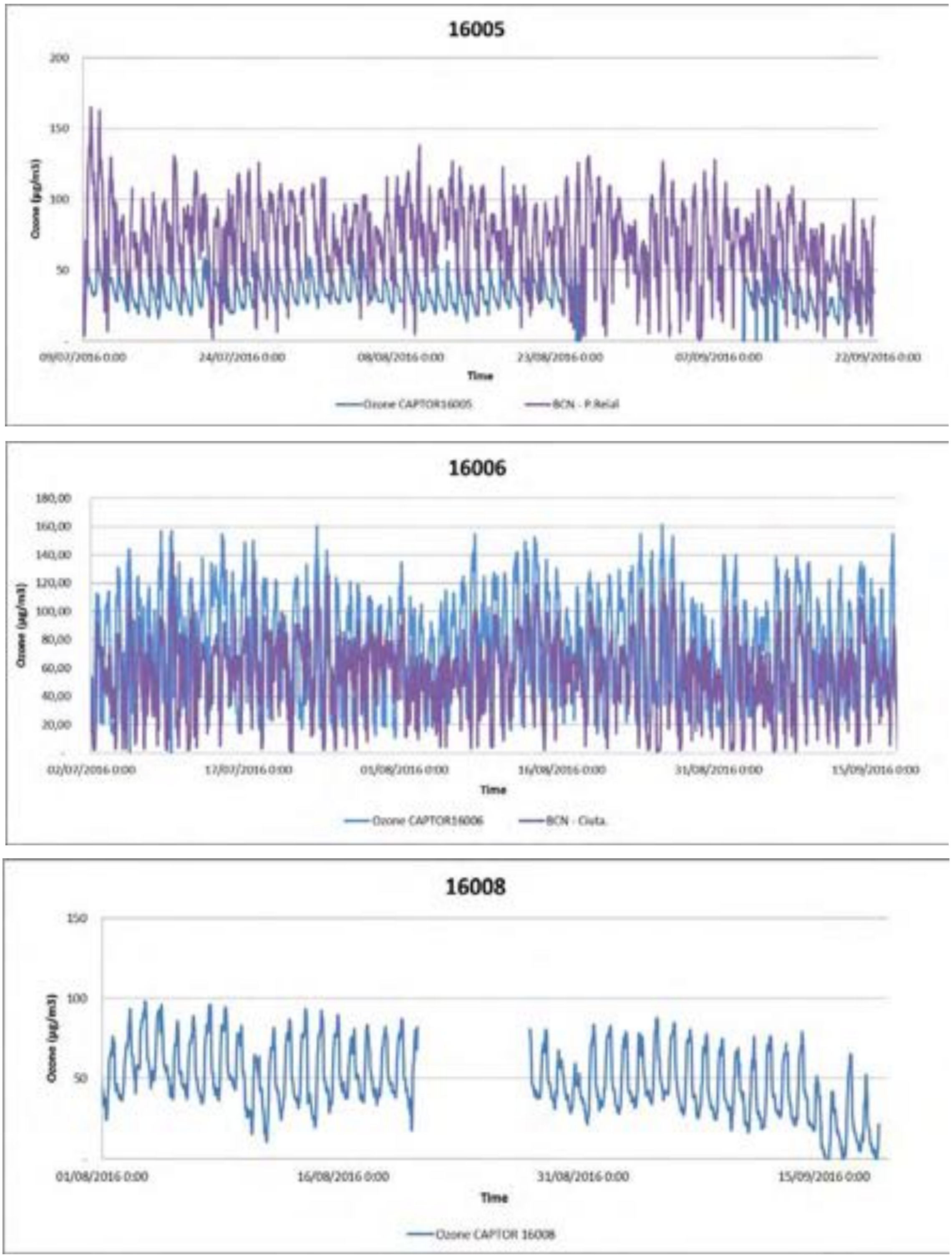


Figure 5. Continued

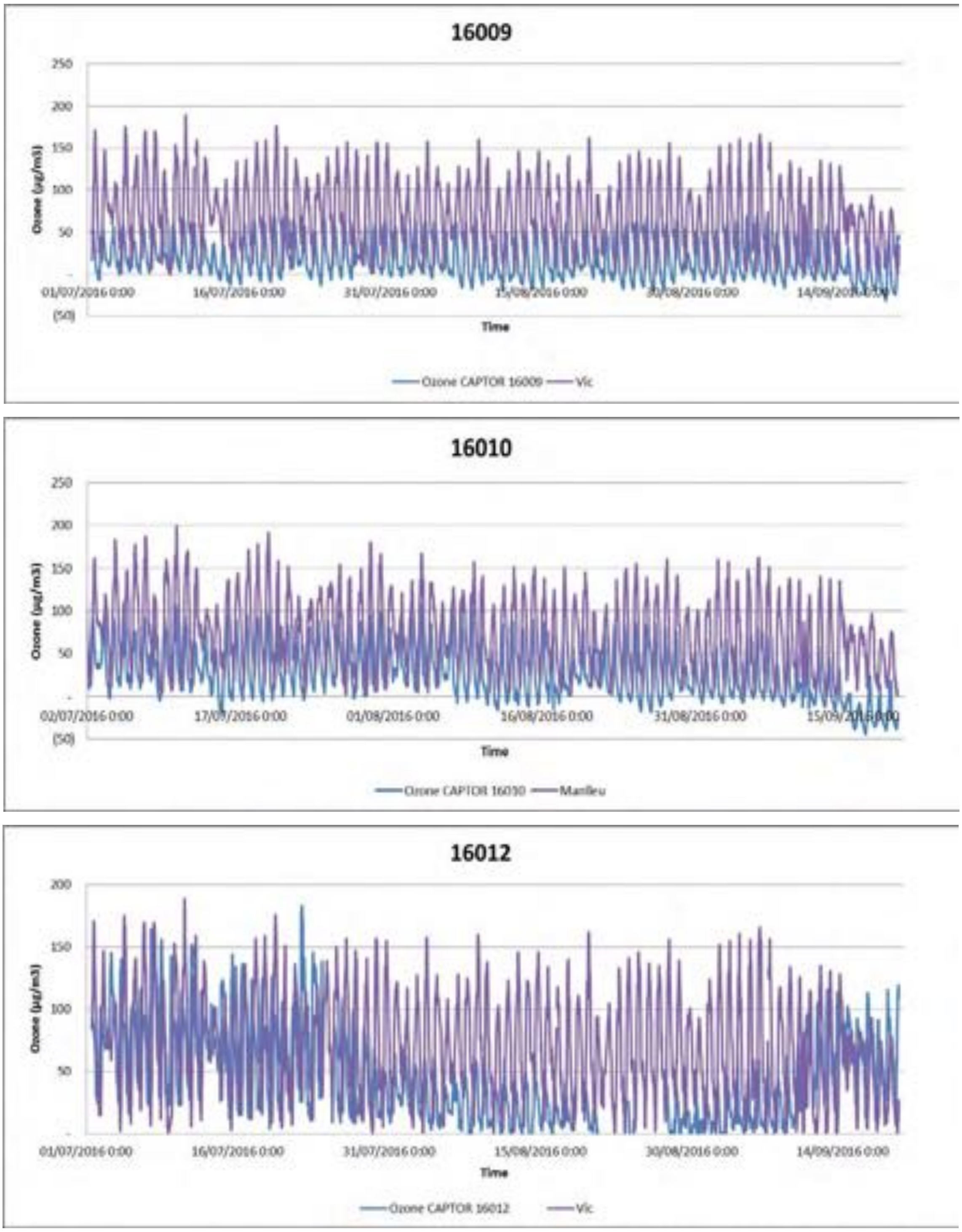


Figure 5. Continued

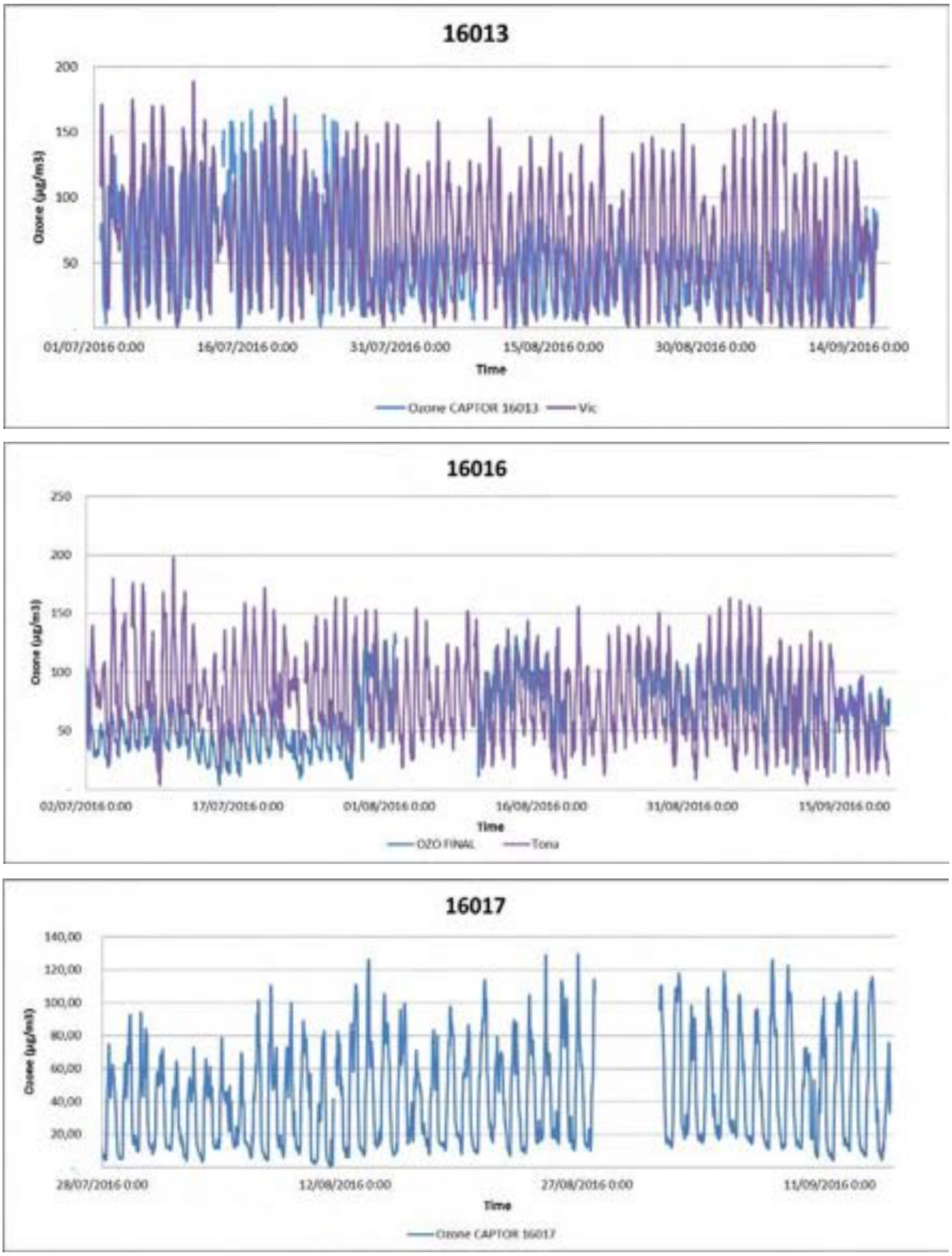


Figure 5. Continued

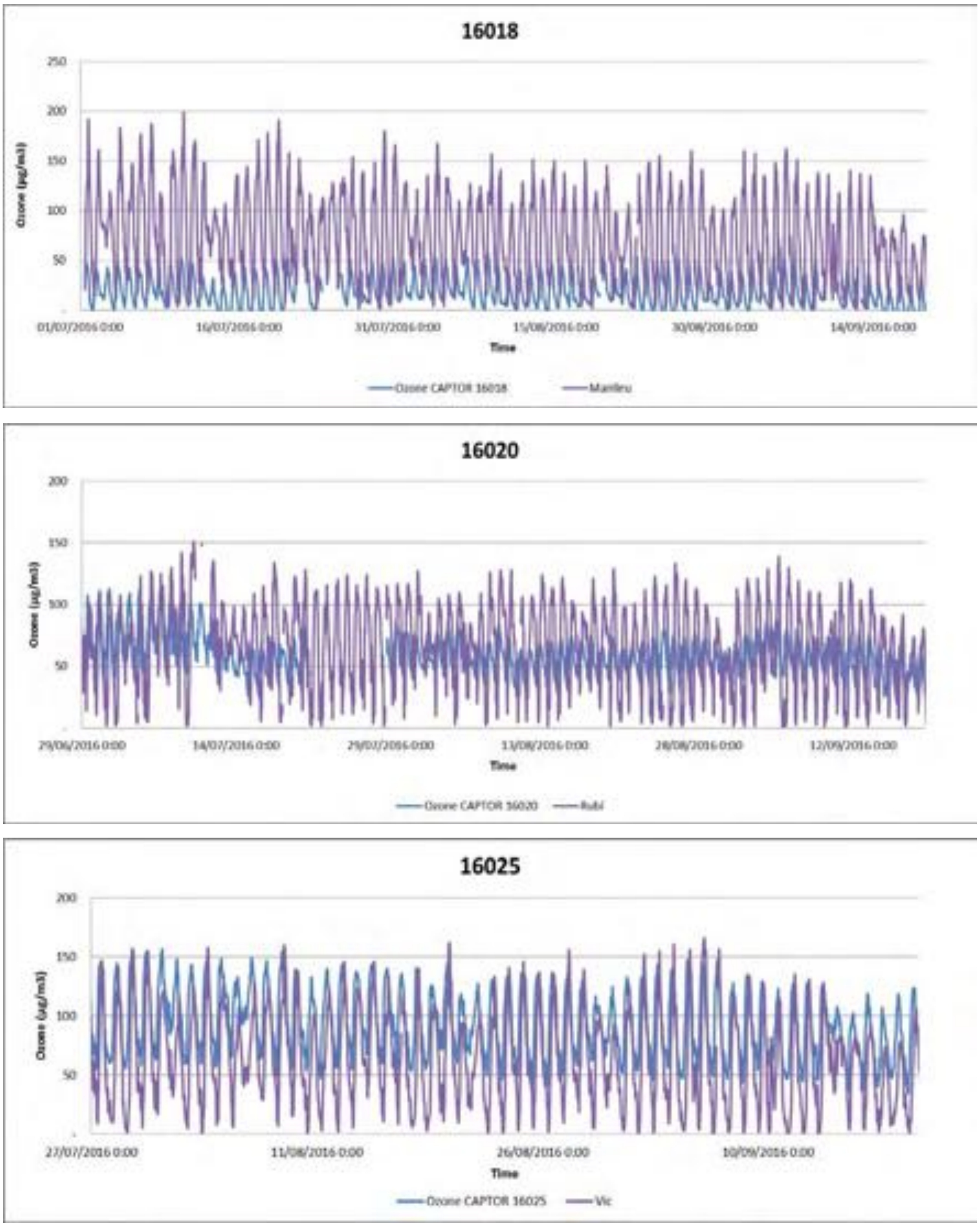


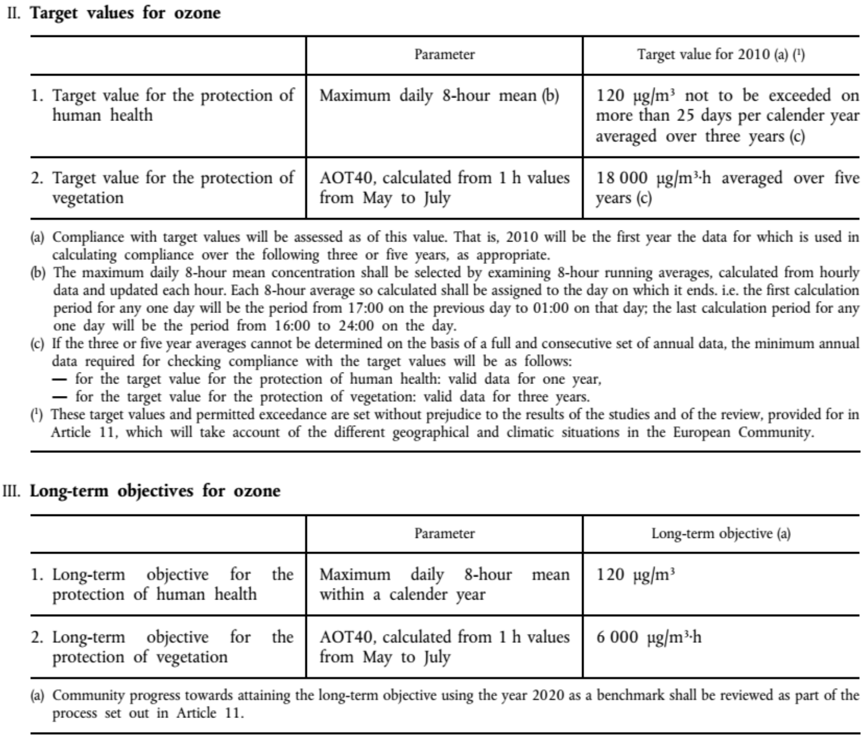
Figure 5. Continued

As expected, the results are node-dependent. All of the 16 time-series obtained (corresponding to 15 nodes at volunteers' homes and 1 located in the vicinity of an air quality monitoring station, node 16025) showed reasonable behaviours, following the characteristic daily cycles of ozone concentrations. Daily maxima were monitored around noon, and minimum values were recorded in the evenings and night. This is consistent with ozone chemistry. The order of magnitude of the concentrations measured was also within the usual ranges for this pollutant, although a detailed analysis is necessary for each node. This was expected given that the sensors were calibrated with respect to reference data. Over all, the data from five of the nodes (16001, 16002, 16004, 16006, and 16025) seemed to compare well with reference data from nearby stations (not directly co-located at these sites). The only node co-located at a reference site was 16025 (Vic), which provided very good results as shown in the Figure. The results in Figure 5 evidence that:

- The main limitation observed was the inability of certain nodes to capture the intensity of the daily ozone cycle, especially when results are compared to those from nearby reference stations. This problem was especially evident for nodes 16005, 16009, 16016, 16018, and 16020. This could indicate that the sensing nodes are unable to reproduce especially high or low hourly concentrations, but that they may generate valid data when looking at daily means.
- Re-basing issues were detected for 3 nodes (16010, 16013, and 16016), where significant changes in scale were detected. This kind of behaviour, improved in the upcoming versions of the nodes, requires careful filtering of the time series and assessment of the specific periods of time for which data are representative. One of the nodes (16012) showed a change in scale but which was not as abrupt as in the other re-basing cases; this may mean that more than one factor affected the node (negatively) at one same point in time.
- With regard to the absolute concentrations, 7 of the nodes (16002, 16004, 16005, 16009, 16012, 16016, and 16018) seemed to underestimate ozone concentrations when compared to reference data from nearby stations. It should be noted that it would not be scientifically robust to compare node with reference data when the instruments were not co-located, as is the case in Figure 5. The spatial variability of ozone, as a secondary pollutant, may not result in similar concentrations even when monitoring sites are located in the vicinity of each other. Differences in altitude above sea level may also impact ozone concentrations strongly. However, this could mean that the Captor node data could be higher or lower than the reference data, whereas in the case of these 7 nodes the result is consistently an underestimation of ozone concentrations. This suggests that there might be a bias in the sensing devices. This interpretation is supported by the fact that 6 of the 7 nodes (with the exception of 16005) were located in the vicinity of Vic, where the highest ozone concentrations were registered, thus suggesting that the performance of the Captor nodes may decrease at increasing ozone concentrations. Further assessments are necessary to confirm this behaviour.
- With regard to data availability, three nodes showed time series with frequent gaps, indicating technical issues with hardware or internet connectivity.

Subsequently to the time series analysis for each of the nodes, a statistical analysis was carried out based on the parameters requested by EU Directive 2002/3/EC for reporting of ozone by Member States in Europe. The Directive's long-term values and objectives for ozone are described in Table 2.

Table 2. Long-term values and objectives for ozone according to Directive 2002/3/EC.



II. **Target values for ozone**

	Parameter	Target value for 2010 (a) (¹)
1. Target value for the protection of human health	Maximum daily 8-hour mean (b)	120 µg/m³ not to be exceeded on more than 25 days per calender year averaged over three years (c)
2. Target value for the protection of vegetation	AOT40, calculated from 1 h values from May to July	18 000 µg/m³·h averaged over five years (c)

(a) Compliance with target values will be assessed as of this value. That is, 2010 will be the first year the data for which is used in calculating compliance over the following three or five years, as appropriate.

(b) The maximum daily 8-hour mean concentration shall be selected by examining 8-hour running averages, calculated from hourly data and updated each hour. Each 8-hour average so calculated shall be assigned to the day on which it ends. i.e. the first calculation period for any one day will be the period from 17:00 on the previous day to 01:00 on that day; the last calculation period for any one day will be the period from 16:00 to 24:00 on the day.

(c) If the three or five year averages cannot be determined on the basis of a full and consecutive set of annual data, the minimum annual data required for checking compliance with the target values will be as follows:

— for the target value for the protection of human health: valid data for one year,

— for the target value for the protection of vegetation: valid data for three years.

(¹) These target values and permitted exceedance are set without prejudice to the results of the studies and of the review, provided for in Article 11, which will take account of the different geographical and climatic situations in the European Community.

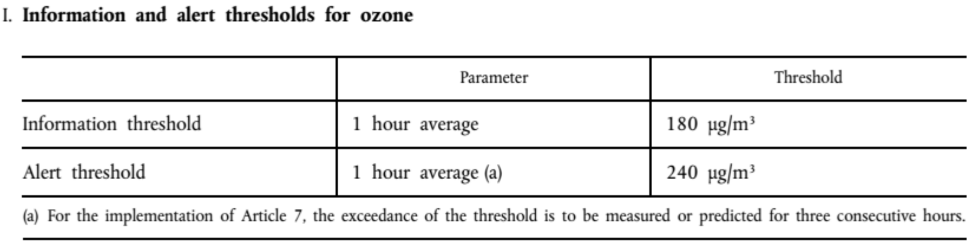
III. **Long-term objectives for ozone**

	Parameter	Long-term objective (a)
1. Long-term objective for the protection of human health	Maximum daily 8-hour mean within a calender year	120 µg/m³
2. Long-term objective for the protection of vegetation	AOT40, calculated from 1 h values from May to July	6 000 µg/m³·h

(a) Community progress towards attaining the long-term objective using the year 2020 as a benchmark shall be reviewed as part of the process set out in Article 11.

Information and alert thresholds are also provided by the Directive (Table 3).

Table 3. Information and alert thresholds for ozone provided by Directive 2002/3/EC.



I. **Information and alert thresholds for ozone**

	Parameter	Threshold
Information threshold	1 hour average	180 µg/m³
Alert threshold	1 hour average (a)	240 µg/m³

(a) For the implementation of Article 7, the exceedance of the threshold is to be measured or predicted for three consecutive hours.

The main statistics for the ozone concentrations monitored with the Captor nodes during the 2016 summer campaign are presented in Table 4. The indicators shown are the hourly average for the entire period, the 8-hour average, the number of exceedances of the long-term objective (120 µg/m^3), the number of exceedances of the alert threshold (240 µg/m^3), the 26th percentile highest value, the 93.2 percentile, and the average of the concentrations >120 µg/m^3 (every 8 hours).

Mean ozone concentrations for the entire field campaign from the Priority#1 nodes ranged between 34 and 91 µg/m^3, while they were lower in the Priority#2 nodes (19-75 µg/m^3). Even with the exception of node 16005 (33 µg/m^3), the remaining average concentrations for Priority#1 nodes (52-91 µg/m^3) were lower or close to the range expected for summer months in Spanish reference stations (Figure 6). It should be noted that these data correspond to the year 2015 and that the concentrations recorded in 2016 were slightly lower (Figure 7). In general, this could again indicate that the current version of the Captor nodes, with the current data processing (without scaling), could underestimate ozone concentrations when dealing with high levels. Local official data have been requested in order to carry out more locally-specific comparisons.

Table 4. Summary of ozone concentration results from the 2016 summer campaign.

Node	Hourly average for entire period	8-hour average for entire period	Number of exceedances of the long-term objective	Number of exceedances of the alert threshold	26th percentile highest value	93.2 percentile	Average of concentrations >120 µg/m3
16001	70.6	70.7	0	0	105	96	-
16002	51.7	51.8	0	0	90	81	-
16003	81.7	86.0	886	0	192	164	150
16005	33.7	33.8	0	0	53	47	-
16016	67.3	67.4	28	0	120	109	124
16020	59.7	59.4	0	0	100	81	-
16025	91.1	91.1	415	0	139	128	129
16004	34.0	34.0	0	0	71	61	-
16006	75.4	75.5	233	0	136	119	128
16008	49.2	49.3	0	0	85	76	-
16009	19.2	19.0	0	0	58	49	-
16010	30.8	30.7	0	0	80	68	-
16012	74.3	74.2	90	0	136	118	131
16013	52.3	52.3	69	0	128	108	126
16017	34.0	33.7	0	0	65	57	-
16018	19.6	19.4	0	0	46	39	-
16019	-	-	-	-	-	-	-
16028	-	-	-	-	-	-	-

Long-term objective: 120 µg/m^3; alert level: 240 µg/m^3.

Based on these data, ozone concentrations did not exceed the alert level in any of the study locations, whereas the long-term objective (120 µg/m^3) was exceeded by 3 of the 7 Priority#1 nodes (16003, 16016, 16025). These results are in line with the low concentrations reported above, which probably underestimate actual ozone concentrations.

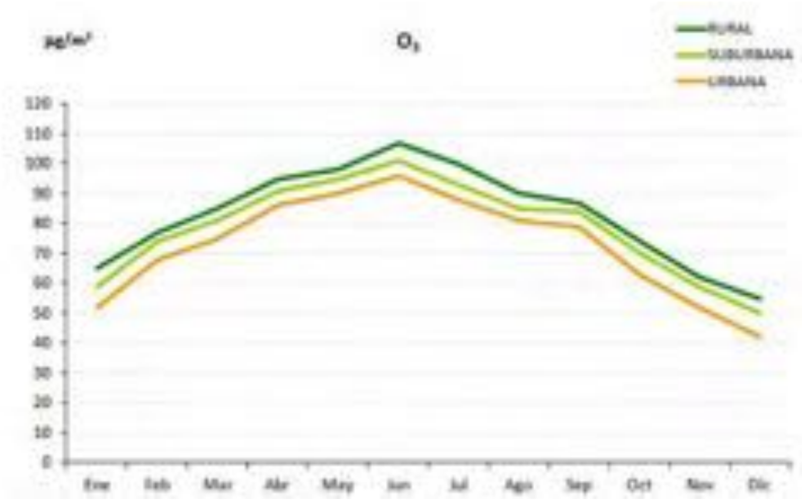


Figure 6. Monthly means of daily maxima for ozone in 2015, as a function of the type of location. Source: Spanish Ministry of the Environment.

Figure 7. Number of exceedances of the information threshold in Spain (April-September 2015, left; 2016, right). Source: Spanish Ministry of the Environment.

As a conclusion, on the basis of the results in Figure 5 and Table 4 a final selection of nodes was carried out to identify those with the most reliable data for the 2016 summer campaign. These were nodes 16001, 16006, 16009, 16010, and 16025. Taking into account that 20 nodes were initially deployed, this result evidences that the Captor nodes during the 2016 summer campaign provided relatively low data availability.

2.6. Spatial distribution of ozone concentrations

The spatial distribution of the mean ozone concentrations calculated for each node is shown in Figure 8. The location of each node is specified in Table 5. As discussed above, the nodes which are considered to have the highest data quality are 16001, 16006, 16009, 16010, and 16025. In general, the results suggest an increasing trend in ozone concentrations from the coastal area (Barcelona) towards the inland region (Vic), which is consistent with the trend described by the reference stations. The highest concentrations were registered in Vic (16025), Llinars del Vallès (16003) and Santa Eugènia de Berga (16012). Four of the nodes (16001, 16004, 16009, 16018) reported lower concentrations than would have been expected due to their location downwind of Barcelona and close to Vic (Sant Pere de Torellò, Gurb, Plà de Sant Sebastià, and Les Masies de Voltregà, respectively). However, these concentrations may be explained by the type of location (with higher contributions from traffic) or by the higher altitude of the monitoring sites (Sant Pere de Torellò, 621 m a.s.l; Gurb, 563 m a.s.l., Plà de Sant Sebastià, 480 m a.s.l.; and Les Masies de Voltregà, 533 m a.s.l., as opposed to 484 m a.s.l in Vic). Further assessments on mean daily ozone cycles and comparisons with locally-relevant reference stations are currently underway.

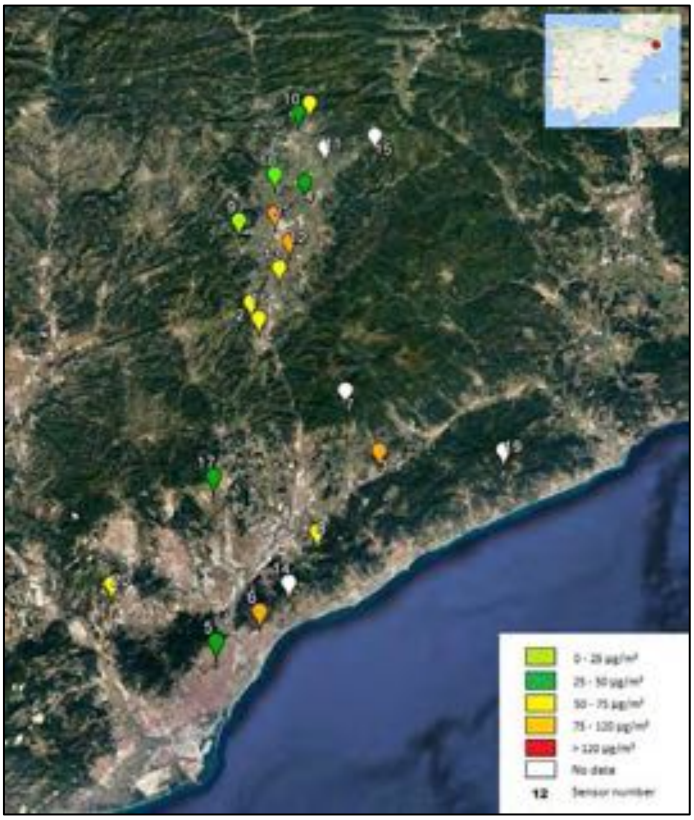


Figure 8. Spatial distribution of mean hourly ozone concentrations monitored with Captor nodes during the 2016 summer campaign.

Table 5. Location of the Captor nodes at volunteers' homes (top) and co-located with air quality monitoring reference stations.

ID captor	Coordinates	Location	ID captor	Coordinates	Location
16001	42° 4'49.22"N 2°17'33.72"E	Sant Pere de Torelló	16011	42° 1'19.78"N 2°19'5.83"E	Manlleu
16002	41°48'4.81"N 2°13'7.79"E	Centelles	16012	41°53'55.19"N 2°15'48.91"E	Santa Eugènia de Berga
16003	41°38'33.34"N 2°24'21.23"E	Llinars del Vallès	16013	41°51'54.81"N 2°14'55.05"E	Taradell
16004	41°58'34.86"N 2°17'12.70"E	Gurb	16014	41°29'29.20"N 2°16'18.95"E	Tiana
16005	41°25'33.97" 2°10'1.63"E	Barcelona	16015	42° 2'6.64"N 2°23'56.92"E	Cantonigròs
16006	41°27'33.71"N 2°13'46.95"E	Badalona	16016	41°49'16.85"N 2°12'19.96"E	Hostalets de Balenyà
16007	41°42'49.08"N 2°21'14.55"E	Vallfornes	16017	41°36'45.31"N 2° 9'20.36"E	Can Patxau
16008	41°32'56.97"N 2°18'47.97"E	Vilanova del Valles	16018	41°59'9.32"N 2°14'17.81"E	Les Masies de Voltregà
16009	41°55'30.02"N 2°10'58.90"E	Plà de Sant Sebastià	16019	41°38'34.40"N 2°35'40.83"E	Sant Cebrià de Vallalta
16010	42° 4'3.49"N 2°16'27.86"E	Sant Vicenç de Torelló	16020	41°29'24.54"N 2° 0'36.41"E	Rubí

Table 5. Continued.

ID captor	Coordinates	Location
16021	42° 0'12.12"N 2°17'13.53"E	Manlleu
16022	41°23'11.15"N 2°11'14.86"E	Ciutadella (Barcelona)
16023	41°29'31.62"N 2° 2'33.18"E	Rubí
16024	41°46'45.80"N 2°21'28.79"E	Montseny
16025	41°56'5.70"N 2°14'23.37"E	Vic

3. Conclusions

This deliverable aimed to review the quality of the datasets generated by the Captor nodes during the 2016 summer campaign. The main conclusions extracted may be summarized as follows:

- Performance was strongly node-dependent: this dependency affected calibration, how the nodes were (potentially) impacted by transport, potential drifts over time (assessed by means of the variability in the correction coefficients between both calibration campaigns), as well as other parameters. As a result, it is concluded that the nodes are not yet ready for deployment with a do-it-yourself (DIY) philosophy. This might be possible at later stages in the project, but at present the nodes require a major amount of specialised supervision in order to produce data which may be comparable to those generated by reference instrumentation. Further work must be carried out in order to validate performance and ensure data quality.
- A total of 5 nodes were selected as providing relatively reliable data, at least regarding daily mean concentrations. This is a positive result from the point of view of performance, given that this was the first real-world deployment of the nodes. However, in relation to the total number on nodes deployed (20), the percentage of data availability and comparability with reference should be considered low. As an example, the time series of one of the nodes providing reliable data, with the data from the reference station where it was co-located (Vic), is shown in Figure 9.

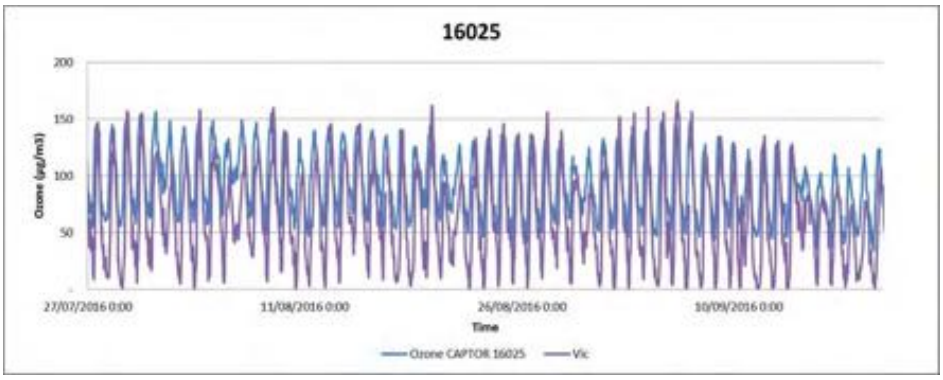


Figure 9. Times series of node 16025

- Drifts over time: unfortunately, the data available do not allow for an assessment of the potential influence of drifts over time given that the variability observed between both calibration periods could be linked to actual drifts in sensor performance but also to other factors such as hardware problems, impact of transport, etc. Further monitoring periods and data assessments are required to obtain robust conclusions.

- Intensity of the sensor signal: the data reported by the nodes followed clear daily cycles, as should be expected. However, the intensity of these cycles did not match that of the reference data, which might indicate that the sensing nodes are unable to reproduce especially high or low hourly concentrations. This would mean that the nodes may generate valid data when looking at daily means, although their performance might be limited when looking at hourly concentrations. This is of high importance, given that the main indicator when reporting ozone data is the number of hourly (or 8-hourly) exceedances of the 120 µg/m^3 threshold, which is a representation of high ozone concentrations. Therefore, the main parameter to be reported when looking at ozone, which is a measure of the highest hourly concentrations, cannot be monitored at present with the current version of the Captor nodes.
- Potential underestimation of mean concentrations for a limited number of nodes: 7 of the 10 Priority#1 nodes reported lower mean hourly concentrations than reference stations located in their vicinity (not co-located). The fact that they were not co-located should result in differences in concentrations, which could be higher or lower. However, these 7 nodes consistently underestimated ozone concentrations, thus suggesting that there might be a bias in the sensing devices. This interpretation is supported by the fact that 6 of the 7 nodes were located in the area where the highest ozone concentrations were registered. This result could imply that the performance of the Captor nodes may decrease with increasing ozone concentrations. Further assessments are necessary to confirm this behaviour.

Finally, as stated in deliverable 3.2a, the results from the 2016 summer campaign are considered highly useful and a good learning basis for the upcoming 2017 and 2018 summer campaigns. The experience gathered during 2016 regarding the construction of new Captor prototypes will add great value to the new campaigns. The main lessons learnt and improvements already underway are:

- Improved software for removal of outliers
- Improved software to avoid baseline changes, by isolating the readings of the individual sensors
- Improved hardware to prevent sensor detachment from the board
- Improved data processing, including input data scaling
- Improved deployment protocol, including exchange of malfunctioning sensors before deployment in the field.