



PONTIFICIA UNIVERSIDAD JAVERIANA  
MAESTRÍA EN HIDROSISTEMAS

# Assessment of Rainfall Influence Over Water Quality at the Effluent of an Urban Catchment by High Temporal Resolution Measurements

---

Santiago Sandoval (C.E.)

**Director:**

**Andrés Torres, PhD.**

**Evaluators:**

**Nicolas Caradot, MSc.**

**Guenter Gruber, PhD.**

**Peter A. Vanrolleghem, PhD.**

**Bogotá D.C.  
January 2013**

## TABLE OF CONTENTS

|       |                                                                                                                                       |    |
|-------|---------------------------------------------------------------------------------------------------------------------------------------|----|
| 1     | INTRODUCTION.....                                                                                                                     | 3  |
| 2     | LITERATURE REVIEW.....                                                                                                                | 6  |
| 2.1   | State of Art .....                                                                                                                    | 6  |
| 2.2   | Theoretical Framework .....                                                                                                           | 10 |
| 2.2.1 | Data Mining: .....                                                                                                                    | 10 |
| 2.2.2 | Principal Component Analysis.....                                                                                                     | 11 |
| 2.2.3 | Association Metrics.....                                                                                                              | 11 |
| 2.2.4 | Support Vector Machines.....                                                                                                          | 15 |
| 2.2.5 | Uncertainty Analysis .....                                                                                                            | 15 |
| 2.2.6 | Monte Carlo Simulation .....                                                                                                          | 17 |
| 2.2.7 | Signal Processing and Times Series Techniques.....                                                                                    | 17 |
| 3     | MATERIALS AND METHODS .....                                                                                                           | 20 |
| 3.1   | Experimental site.....                                                                                                                | 20 |
| 3.1.1 | Gibraltar Sanitary sub-basin.....                                                                                                     | 20 |
| 3.1.2 | Effluent of the catchment: Gibraltar Pumping Station (GPS).....                                                                       | 21 |
| 3.2   | Measurement Devices .....                                                                                                             | 24 |
| 3.2.1 | Flow Rate Measurement.....                                                                                                            | 24 |
| 3.2.2 | Water Quality Measurement .....                                                                                                       | 25 |
| 3.2.3 | Rain gauge stations .....                                                                                                             | 29 |
| 3.3   | Experimental Framework.....                                                                                                           | 30 |
| 3.3.1 | Flow Rate Measurement.....                                                                                                            | 30 |
| 3.3.2 | Water Quality Measurement .....                                                                                                       | 33 |
| 3.4   | Data Description and Computational Tools .....                                                                                        | 34 |
| 3.4.1 | Data Description.....                                                                                                                 | 34 |
| 3.4.2 | Computational Tools .....                                                                                                             | 37 |
| 3.5   | Developed Methods.....                                                                                                                | 39 |
| 3.5.1 | Dry weather characterization and rainy weather identification.....                                                                    | 39 |
| 3.5.2 | Recurrent relations between rainfall pulses and water quality's time windows<br>ranging outside of typical dry weather behavior ..... | 46 |
| 3.5.3 | Association of recurrent relations with characteristics of the urban catchment .....                                                  | 50 |

|       |                                                                                                                                       |     |
|-------|---------------------------------------------------------------------------------------------------------------------------------------|-----|
| 4     | RESULTS AND DISCUSSION .....                                                                                                          | 52  |
| 4.1   | Dry weather characterization and rainy weather identification.....                                                                    | 52  |
| 4.1.1 | Differences between dry and rainy days in terms of water quality parameters.....                                                      | 52  |
| 4.1.2 | Dry weather probabilistic characterization.....                                                                                       | 59  |
| 4.1.3 | Identification of rainfall pulses and water quality's time windows ranging outside<br>typical dry weather behavior .....              | 77  |
| 4.2   | Recurrent relations between rainfall pulses and water quality's time windows ranging<br>outside of typical dry weather behavior ..... | 109 |
| 4.2.1 | Echo Profile Flowmeter .....                                                                                                          | 109 |
| 4.2.2 | Uni-parametric sensors.....                                                                                                           | 113 |
| 4.3   | Association of recurrent relations with physical characteristics of the urban catchment                                               | 128 |
| 4.4   | Future Work .....                                                                                                                     | 130 |
| 5     | CONCLUSIONS .....                                                                                                                     | 137 |
| 6     | RECOMMENDATIONS .....                                                                                                                 | 141 |
| 7     | REFERENCES .....                                                                                                                      | 143 |
| 8     | APPENDICES .....                                                                                                                      | 149 |

## LIST OF FIGURES

|                                                                                                                                                                                                                                              |    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 3.1 map of the <i>Gibraltar</i> 's Sanitary sub-basin. Source: adapted by author from Google Earth.                                                                                                                                   | 20 |
| Figure 3.2 a) Location of the three EAAB rain gauge stations used for rainfall measurements in the catchment; b) The study zone in Bogotá.....                                                                                               | 21 |
| Figure 3.3 Pumping station <i>Estación Elevadora Gibraltar</i> .....                                                                                                                                                                         | 22 |
| Figure 3.4 Description of the study zone. Sub-basins 461, 462, 464, 465, 467, 470 and 501, “Gibraltar” urban catchment. Source: Legal Contract PUJ – EAAB No. 9-07-25100-0763-2010...                                                        | 22 |
| Figure 3.5 Echo Profile Flowmeter device a) installation, b) data transmitter and c) velocity sensors. (Source: <a href="http://www.schillig.com.ar/nivus_fr.htm">www.schillig.com.ar/nivus_fr.htm</a> , December 7 <sup>th</sup> 2012)..... | 24 |
| Figure 3.6 Visoturb 700 IQ sensor uni-parametric; a) measurment's angles and b) parts of the sensor. Source: WTW <i>on-line</i> , 2010. ....                                                                                                 | 25 |
| Figure 3.7 Sensolyt 700 IQ sensor uni-parametric; a) dimensions and b) parts of the sensor. Source: WTW <i>on-line</i> , 2010. ....                                                                                                          | 26 |
| Figure 3.8 Visolid 700 IQ sensor uni-parametric; a) dimensions and b) parts of the sensor. Source: WTW <i>on-line</i> , 2010. ....                                                                                                           | 26 |
| Figure 3.9 FDO 700 IQ sensor uni-parametric; parts of the sensor. Source: WTW <i>on-line</i> , 2010...                                                                                                                                       | 27 |
| Figure 3.10 Electromagnetic spectra- frequency and wavelength (Source: <a href="http://www.s-can.at/parameter.php?langcode">www.s-can.at/parameter.php?langcode</a> , December 7th 2012) .....                                               | 28 |
| Figure 3.11 Water quality parameters obtained from spectrometer measurements (Source: <a href="http://www.s-can.at/text.php?kat=11&amp;id=59&amp;langcode">www.s-can.at/text.php?kat=11&amp;id=59&amp;langcode</a> , December 7th 2012)..... | 28 |
| Figure 3.12 Sensor's dimensions (s::can, 2007).....                                                                                                                                                                                          | 29 |
| Figure 3.13 Sensor's components (s::can, 2007) .....                                                                                                                                                                                         | 29 |
| Figure 3.14 Rainfall raw data registered by the studied rain gauge station .....                                                                                                                                                             | 30 |
| Figure 3.15 Hydraulic system. a) Upstream and downstream section of the discharge tunnel; b) entry to the discharge tunnel .....                                                                                                             | 31 |
| Figure 3.16 Discharge tunnel .....                                                                                                                                                                                                           | 31 |
| Figure 3.17 Transmitter safeguard .....                                                                                                                                                                                                      | 32 |
| Figure 3.18 Structure 1: transmitter safeguard in field.....                                                                                                                                                                                 | 32 |
| Figure 3.19 Structure 2 scheme: sensors installation inside the discharge tunnel.....                                                                                                                                                        | 32 |
| Figure 3.20 Structure 2: sensors installation inside the discharge tunnel.....                                                                                                                                                               | 33 |
| Figure 3.21 Installation boat for infra-red, ion sensitive and UV-VIS spectrometry sensors .....                                                                                                                                             | 34 |
| Figure 3.22 Example of PCA execution, a) separation of groups by rainfall registered at a given rain gauge station; b) variables correlated with the principal component 1 or 2 (x and y axis). ....                                         | 40 |
| Figure 3.23 Example of the probabilistic behavior of TSS for a typical dry weather day. ....                                                                                                                                                 | 41 |
| Figure 3.24 Estimation of a probability x, by the use of the PDF for the random variable X.....                                                                                                                                              | 42 |
| Figure 3.25 Evaluation of each time-step's probability of being part of typical dry weather behavior for a rainy day signal, highlighting probabilities of dry weather behavior below 5 %. ....                                              | 43 |
| Figure 3.26 TSS measured by uni-parametric sensors autocorrelation function for the whole measurement campaign.....                                                                                                                          | 44 |
| Figure 3.27 Comparison between: a) original signal b) Moving Average Filter c) Median Filter....                                                                                                                                             | 45 |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |    |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 3.28 Identification of time windows $[a, b]$ in which $P_{dry}(a_j < t < b_j)_{filt} < 0.05$ ; where $a$ , $b$ are change points obtained by the Step Detection calculations over the signal $P_{dry}(t)_{filt}$ . ....                                                                                                                                                                                                                                                                                                          | 45 |
| Figure 3.29 Identification of rainfall pulses for a given rain gauge station. ....                                                                                                                                                                                                                                                                                                                                                                                                                                                      | 46 |
| Figure 3.30 Two rainfall pulses (rainfall) affecting the same time window that ranged outside of the typical dry weather behavior (output). ....                                                                                                                                                                                                                                                                                                                                                                                        | 47 |
| Figure 3.31 Assessment of recurrent relation between a rainfall and a water quantity characteristic (Pearson Correlation's test, Linear Model Regression, Average Mutual Information). ....                                                                                                                                                                                                                                                                                                                                             | 49 |
| Figure 3.32 Assessment of recurrent relation between a rainfall and water quality characteristic (PLS approach).....                                                                                                                                                                                                                                                                                                                                                                                                                    | 50 |
| Figure 4.1 PCA analysis for the <i>Bosa</i> rain gauge station (rainy days' data blue group, dry days' data green group); a), b) separation given by principal components 1 (x-axis) and 2 (y-axis); b), c) separation given by principal components 1 (x-axis) and 3 (y-axis). ....                                                                                                                                                                                                                                                    | 54 |
| Figure 4.2 PCA analysis for the <i>Bosa</i> rain gauge station (rainy days' data blue group, dry days' data green group); a), b) separation given by principal components 1 (x- axis) and 2 (y- axis). ....                                                                                                                                                                                                                                                                                                                             | 56 |
| Figure 4.3 PCA analysis for the <i>Bosa</i> (a,b) and <i>Casablanca</i> (c,d) rain gauge stations (rainy days' data blue group, dry days' data green group); a), b) separation given by principal components 1 (x-axis) and 2 (y-axis). ....                                                                                                                                                                                                                                                                                            | 58 |
| Figure 4.4 Dry weather probabilistic characterization for flow rate measured by Echo Profile Flowmeter sensor, using rainfall registrations at different rain gauge stations ( <i>Fontibon</i> (left), <i>Bosa</i> (center) and <i>Casablanca</i> (right)); a, b, c) Probability density function (PDF) for each time-step; d, e, f) Cumulative probability function (CPF) for each time-step; g, h, i) Mode, Median and Expectation for each time-step; j, k, l) Interquartile distance and standard deviation for each time-step..... | 60 |
| Figure 4.5 Dry weather probabilistic characterization for DO measured by uni-parametric sensor, using rainfall registrations at different rain gauge stations ( <i>Fontibon</i> (left), <i>Bosa</i> (center) and <i>Casablanca</i> (right)); a, b, c) Probability density function (PDF) for each time-step; d, e, f) Cumulative probability function (CPF) for each time-step; g, h, i) Mode, Median and Expectation for each time-step; j, k, l) Interquartile distance and standard deviation for each time-step.....                | 62 |
| Figure 4.6 Dry weather probabilistic characterization for TSS measured by uni-parametric sensors, using rainfall registrations at different rain gauge stations ( <i>Fontibon</i> (left), <i>Bosa</i> (center) and <i>Casablanca</i> (right)); a, b, c) Probability density function (PDF) for each time-step; d, e, f) Cumulative probability function (CPF) for each time-step; g, h, i) Mode, Median and Expectation for each time-step; j, k, l) Interquartile distance and standard deviation for each time-step.....              | 63 |
| Figure 4.7 Dry weather probabilistic characterization for Turbidity measured by uni-parametric sensors, using rainfall registrations at different rain gauge stations ( <i>Fontibon</i> (left), <i>Bosa</i> (center) and <i>Casablanca</i> (right)); a, b, c) Probability density function (PDF) for each time-step; d, e, f) Cumulative probability function (CPF) for each time-step; g, h, i) Mode, Median and Expectation for each time-step; j, k, l) Interquartile distance and standard deviation for each time-step.....        | 65 |
| Figure 4.8 Dry weather probabilistic characterization for ORP measured by uni-parametric sensors, using rainfall registrations at different rain gauge stations ( <i>Fontibon</i> (left), <i>Bosa</i> (center) and <i>Casablanca</i> (right)); a, b, c) Probability density function (PDF) for each time-step; d, e, f) Cumulative probability function (CPF) for each time-step; g, h, i) Mode, Median and Expectation for each time-step; j, k, l) Interquartile distance and standard deviation for each time-step.....              | 67 |
| Figure 4.9 Dry weather probabilistic characterization for Temperature measured by uni-parametric sensors, using rainfall registrations at different rain gauge stations ( <i>Fontibon</i> (left), <i>Bosa</i> (center)                                                                                                                                                                                                                                                                                                                  |    |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |    |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| and <i>Casablanca</i> (right)); a, b, c) Probability density function (PDF) for each time-step; d, e, f) Cumulative probability function (CPF) for each time-step; g, h, i) Mode, Median and Expectation for each time-step; j, k, l) Interquartile distance and standard deviation for each time-step.....                                                                                                                                                                                                                                                                                                  | 68 |
| Figure 4.10 Dry weather probabilistic characterization for pH measured by uni-parametric sensors, using rainfall registrations at different rain gauge stations ( <i>Fontibon</i> (left), <i>Bosa</i> (center) and <i>Casablanca</i> (right)); a, b, c) Probability density function (PDF) for each time-step; d, e, f) Cumulative probability function (CPF) for each time-step; g, h, i) Mode, Median and Expectation for each time-step; j, k, l) Interquartile distance and standard deviation for each time-step.....                                                                                   | 69 |
| Figure 4.11 Correlation value among the different absorbance values registered during the measurement campaign, for each wavelength.....                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | 70 |
| Figure 4.12 Dry weather probabilistic characterization for absorbance values registered in wavelength 205 nm (Nitrites-Nitrates range) by means of the UV-VIS spectrometry sensor, using rainfall registrations at different rain gauge stations ( <i>Fontibon</i> (left), <i>Bosa</i> (center) and <i>Casablanca</i> (right)); a, b, c) Probability density function (PDF) for each time-step; d, e, f) Cumulative probability function (CPF) for each time-step; g, h, i) Mode, Median and Expectation for each time-step; j, k, l) Interquartile distance and standard deviation for each time-step ..... | 72 |
| Figure 4.13 Dry weather probabilistic characterization for absorbance values registered in wavelength 745 nm (VIS range) by means of the UV-VIS spectrometry sensor, using rainfall registrations at different rain gauge stations ( <i>Fontibon</i> (left), <i>Bosa</i> (center) and <i>Casablanca</i> (right)); a, b, c) Probability density function (PDF) for each time-step; d, e, f) Cumulative probability function (CPF) for each time-step; g, h, i) Mode, Median and Expectation for each time-step; j, k, l) Interquartile distance and standard deviation for each time-step .....               | 73 |
| Figure 4.14 Dry weather probabilistic characterization for TSS values registered by means of the UV-VIS spectrometry sensor, using rainfall registrations at different rain gauge stations ( <i>Fontibon</i> (left), <i>Bosa</i> (center) and <i>Casablanca</i> (right)); a, b, c) Probability density function (PDF) for each time-step; d, e, f) Cumulative probability function (CPF) for each time-step; g, h, i) Mode, Median and Expectation for each time-step; j, k, l) Interquartile distance and standard deviation for each time-step .....                                                       | 74 |
| Figure 4.15 Dry weather probabilistic characterization for COD values registered by means of the UV-VIS spectrometry sensor, using rainfall registrations at different rain gauge stations ( <i>Fontibon</i> (left), <i>Bosa</i> (center) and <i>Casablanca</i> (right)); a, b, c) Probability density function (PDF) for each time-step; d, e, f) Cumulative probability function (CPF) for each time-step; g, h, i) Mode, Median and Expectation for each time-step; j, k, l) Interquartile distance and standard deviation for each time-step.....                                                        | 76 |
| Figure 4.16 Autocorrelation functions for rainfall registered at the <i>Fontibon</i> , <i>Bosa</i> and <i>Casablanca</i> rain gauge station .....                                                                                                                                                                                                                                                                                                                                                                                                                                                            | 78 |
| Figure 4.17 Step detection over the filtered signal for rainfall registered at the <i>Fontibon</i> , <i>Bosa</i> and <i>Casablanca</i> rain gauge stations.....                                                                                                                                                                                                                                                                                                                                                                                                                                              | 79 |
| Figure 4.18 Autocorrelation function of flow rate, considering the complete measurement period.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | 80 |
| Figure 4.19 Probability of dry weather signal filtered by a Median Filter, for Flow rate signal .....                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | 81 |
| Figure 4.20 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding Flow rate (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the <i>Bosa</i> rain gauge station.....                                                                                                                                                                                                                                                                                                                      | 83 |

|                                                                                                                                                                                                                                                                                                                  |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 4.21 Autocorrelation functions for DO, Total Suspended Solids, Turbidity, ORP, Temperature and pH; measured by uni-parametric (considering the complete measurement period)                                                                                                                               | 84  |
| Figure 4.22 Probability of dry weather signal filtered by a Median Filter for DO, TSS, Turbidity, ORP, Temperature and pH signals                                                                                                                                                                                | 86  |
| Figure 4.23 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding DO (uni-parametric) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the <i>Bosa</i> rain gauge station                     | 88  |
| Figure 4.24 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding TSS (uni-parametric) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the <i>Bosa</i> rain gauge station                    | 90  |
| Figure 4.25 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding Turbidity (uni-parametric) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the <i>Bosa</i> rain gauge station              | 92  |
| Figure 4.26 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding ORP (uni-parametric) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the <i>Bosa</i> rain gauge station                    | 94  |
| Figure 4.27 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding Temperature (uni-parametric) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the <i>Bosa</i> rain gauge station            | 95  |
| Figure 4.28 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding pH (uni-parametric) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the <i>Bosa</i> rain gauge station                     | 96  |
| Figure 4.29 Autocorrelation function for wavelength 205 nm (left) and 745 nm (right); measured by UV-VIS spectrometry (considering the complete measurement period)                                                                                                                                              | 98  |
| Figure 4.30 Probability of dry weather signal filtered by a Median Filter, for wavelength 205 nm and wavelength 745 nm signals                                                                                                                                                                                   | 99  |
| Figure 4.31 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding wavelength 205 nm (UV-VIS spectrometry) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the <i>Bosa</i> rain gauge station | 100 |
| Figure 4.32 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding wavelength 745 nm (UV-VIS spectrometry) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the <i>Bosa</i> rain gauge station | 101 |
| Figure 4.33 Autocorrelation function for COD (left) and TSS (right); measured by UV-VIS spectrometry (considering the complete measurement period)                                                                                                                                                               | 102 |
| Figure 4.34 Probability of dry weather signal filtered by a Median Filter, for TSS and COD signals                                                                                                                                                                                                               | 103 |
| Figure 4.35 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding TSS (UV-VIS spectrometry) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the <i>Bosa</i> rain gauge station               | 104 |

|                                                                                                                                                                                                                                                                                                         |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 4.36 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding COD (UV-VIS spectrometry) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the <i>Bosa</i> rain gauge station..... | 105 |
| Figure 4.37 Lag times for measured water quality parameters at the effluent of the catchment, by rainfall registered at the <i>Fontibon</i> , <i>Bosa</i> and <i>Casablanca</i> rain gauge stations (grey boxes: no lag times reported) .....                                                           | 107 |
| Figure 4.38 Recurrent relations between flow rate and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the <i>Fontibon</i> rain gauge station.....                                                                                                 | 110 |
| Figure 4.39 Recurrent relations between flow rate and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the <i>Bosa</i> rain gauge station .....                                                                                                    | 111 |
| Figure 4.40 Recurrent relations between flow rate and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the <i>Bosa</i> rain gauge station (PLS model).....                                                                                         | 112 |
| Figure 4.41 Recurrent relations between the DO and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the <i>Fontibon</i> rain gauge station.....                                                                                                    | 114 |
| Figure 4.42 Recurrent relations between DO and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the <i>Fontibon</i> rain gauge station (PLS model). ..                                                                                             | 114 |
| Figure 4.43 Recurrent relations between the DO and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the <i>Bosa</i> rain gauge station .....                                                                                                       | 116 |
| Figure 4.44 Recurrent relations between DO and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the <i>Bosa</i> rain gauge station (PLS model).....                                                                                                | 117 |
| Figure 4.45 Recurrent relations between the DO and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the <i>Casablanca</i> rain gauge station .....                                                                                                 | 118 |
| Figure 4.46 Recurrent relations between DO and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the <i>Casablanca</i> rain gauge station (PLS model). .....                                                                                        | 118 |
| Figure 4.47 Recurrent relations between the Total Suspended Solids and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the <i>Fontibon</i> rain gauge station .....                                                                               | 119 |
| Figure 4.48 Recurrent relations between the Total Suspended Solids and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the <i>Bosa</i> rain gauge station....                                                                                     | 120 |
| Figure 4.49 Recurrent relations between Total Suspended Solids and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the <i>Bosa</i> rain gauge station (PLS model). .....                                                                          | 121 |
| Figure 4.50 Recurrent relations between the Turbidity and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the <i>Fontibon</i> rain gauge station .....                                                                                            | 122 |
| Figure 4.51 Recurrent relations between the Turbidity and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the <i>Bosa</i> rain gauge station.....                                                                                                 | 123 |
| Figure 4.52 Recurrent relations between the ORP and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the <i>Fontibon</i> rain gauge station.....                                                                                                   | 124 |
| Figure 4.53 Recurrent relations between the ORP and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the <i>Bosa</i> rain gauge station .....                                                                                                      | 125 |
| Figure 4.54 Recurrent relations between ORP and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the <i>Bosa</i> rain gauge station (PLS model).....                                                                                               | 126 |

|                                                                                                                                                                                                                                                                                                                                                                                                |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figure 4.55 Recurrent relations between ORP and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the <i>Casablanca</i> rain gauge station (PLS model)                                                                                                                                                                                     | 127 |
| Figure 4.56 PCA analysis regarding UGA's characteristics and its influence from rainfall, given by their nearest rain gauge station (component 1: x-axis, component 2: y-axis).....                                                                                                                                                                                                            | 129 |
| Figure 4.57 Dry weather probabilistic characterization for TSS loads, using rainfall registrations at the <i>Bosa</i> rain gauge stations; a) Probability density function (PDF) for each time-step; b) Cumulative probability function (CPF) for each time-step; c) Mode, Median and Expectation for each time-step; d) Interquartile distance and standard deviation for each time-step..... | 130 |
| Figure 4.58 Autocorrelation function for TSS loads .....                                                                                                                                                                                                                                                                                                                                       | 131 |
| Figure 4.59 Probability of dry weather signal filtered by a Median Filter, for TSS loads signal ...                                                                                                                                                                                                                                                                                            | 131 |
| Figure 4.60 Rainfall pulses ("rainfall"), time windows raining outside of typical dry weather behavior regarding TSS loads ("output") and the overlap between the rainfall pulse and the influenced output ("both"), for rainfall registered at the <i>Bosa</i> rain gauge station.....                                                                                                        | 133 |
| Figure 4.61 Recurrent relations between the TSS loads and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the <i>Bosa</i> rain gauge station.....                                                                                                                                                                                        | 134 |
| Figure 4.62 Modeling results for TSS loads regarding the a) calibration rainfall event, b) validation rainfall event, c)observed-simulated values (calibration rainfall event) .....                                                                                                                                                                                                           | 136 |

## TABLES

|                                                                                                                         |    |
|-------------------------------------------------------------------------------------------------------------------------|----|
| Table 2.1 Summary of the state of art .....                                                                             | 8  |
| Table 3.1 Summary table with physical characteristics of the UGAs of <i>Gibraltar</i> urban catchment                   | 23 |
| Table 3.2 Data base of the measurements at the effluent of the catchment .....                                          | 36 |
| Table 3.3 Daily rainfall registered at the <i>Fontibon</i> , <i>Bosa</i> and <i>Casablanca</i> rain gauge stations..... | 37 |

## LIST OF APPENDICES

|                                                                                                                                                                                                                                                                                                                 |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Appendix 8.1 Data base of the measurements undertaken at the effluent of the studied catchment .....                                                                                                                                                                                                            | 149 |
| Appendix 8.2 Implementation of the proposed methods in R-project .....                                                                                                                                                                                                                                          | 149 |
| Appendix 8.3 PCA analysis for the <i>Fontibon</i> rain gauge station (rainy days' data blue group, dry days' data green group); a), b) separation given by principal components 1 (x-axis) and 2 (y-axis); b), c) separation given by principal components 1 (x-axis) and 3 (y-axis).....                       | 149 |
| Appendix 8.4 PCA analysis for the <i>Casablanca</i> rain gauge station (rainy days' data blue group, dry days' data green group); a), b) separation given by principal components 1 (x-axis) and 2 (y-axis); b), c) separation given by principal components 1 (x-axis) and 3 (y-axis).....                     | 150 |
| Appendix 8.5 PCA analysis for the <i>Fontibon</i> rain gauge station (rainy days' data blue group, dry days' data green group); a), b) separation given by principal components 1 (x- axis) and 2 (y- axis). .....                                                                                              | 151 |
| Appendix 8.6 PCA analysis for the <i>Casablanca</i> rain gauge station (rainy days' data blue group, dry days' data green group); a), b) separation given by principal components 1 (x-axis) and 2(y-axis). .....                                                                                               | 151 |
| Appendix 8.7. PCA analysis for the <i>Fontibon</i> rain gauge station (rainy days' data blue group, dry days' data green group); a), b) separation given by principal components 1 (x-axis) and 2 (y-axis). .....                                                                                               | 152 |
| Appendix 8.8 Rainfall pulses ("rainfall"), time windows raining outside of typical dry weather behavior regarding Flow rate ("output") and the overlap between the rainfall pulse and the influenced output ("both"), for rainfall registered at the <i>Fontibon</i> rain gauge station .....                   | 153 |
| Appendix 8.9 Rainfall pulses ("rainfall"), time windows raining outside of typical dry weather behavior regarding Flow rate ("output") and the overlap between the rainfall pulse and the influenced output ("both"), for rainfall registered at the <i>Casablanca</i> rain gauge station. ....                 | 154 |
| Appendix 8.10 Rainfall pulses ("rainfall"), time windows raining outside of typical dry weather behavior regarding DO (uni-parametric) ("output") and the overlap between the rainfall pulse and the influenced output ("both"), for rainfall registered at the <i>Fontibon</i> rain gauge station .....        | 155 |
| Appendix 8.11 Rainfall pulses ("rainfall"), time windows raining outside of typical dry weather behavior regarding DO (uni-parametric) ("output") and the overlap between the rainfall pulse and the influenced output ("both"), for rainfall registered at the <i>Casablanca</i> rain gauge station.....       | 156 |
| Appendix 8.12 Rainfall pulses ("rainfall"), time windows raining outside of typical dry weather behavior regarding TSS (uni-parametric) ("output") and the overlap between the rainfall pulse and the influenced output ("both"), for rainfall registered at the <i>Fontibon</i> rain gauge station .....       | 158 |
| Appendix 8.13 Rainfall pulses ("rainfall"), time windows raining outside of typical dry weather behavior regarding TSS (uni-parametric) ("output") and the overlap between the rainfall pulse and the influenced output ("both"), for rainfall registered at the <i>Casablanca</i> rain gauge station.....      | 158 |
| Appendix 8.14 Rainfall pulses ("rainfall"), time windows raining outside of typical dry weather behavior regarding Turbidity (uni-parametric) ("output") and the overlap between the rainfall pulse and the influenced output ("both"), for rainfall registered at the <i>Fontibon</i> rain gauge station. .... | 159 |

|                                                                                                                                                                                                                                                                                                                               |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Appendix 8.15 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding ORP (uni-parametric) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the <i>Fontibon</i> rain gauge station. ....                     | 160 |
| Appendix 8.16 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding ORP (uni-parametric) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the <i>Casablanca</i> rain gauge station.....                    | 161 |
| Appendix 8.17 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding wavelength 745 nm (UV-VIS spectrometry) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the <i>Fontibon</i> rain gauge station.....   | 162 |
| Appendix 8.18 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding wavelength 745 nm (UV-VIS spectrometry) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the <i>Casablanca</i> rain gauge station..... | 162 |
| Appendix 8.19 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding TSS (UV-VIS spectrometry) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the <i>Fontibon</i> rain gauge station. ....                | 163 |
| Appendix 8.20 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding COD (UV-VIS spectrometry) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the <i>Fontibon</i> rain gauge station. ....                | 164 |

*The execution of this thesis was the result of the author's work, with the support of other professionals. The author wishes to express his gratitude to all the people and institutions that made this accomplishment possible.*

*To my family and my relatives, your unconditional support has been essential for my development as a researcher. Thank you for being there, encouraging me to keep on pursuing my dreams.*

*To my professors, for being a permanent guide and a role model not only during the development of this project, but also for giving me invaluable tools for facing my future challenges, my eternal gratitude.*

*My sincere thanks to The Water Utilities of Bogotá Empresa de Acueducto y Alcantarillado de Bogotá (EAAB) which under the Administrative Contract No. 9-07-25100-0763-2010 provided all the necessary means for the data collection and the field work, in addition to their unconditional collaboration. Finally, special thanks to all the hydrology team of this company, whom supported me in the most difficult moments through the execution of this project.*

## ABSTRACT

The present study aimed to establish the rainfall influence over the water quality at the effluent of an urban catchment by the use of temporal high resolution measurements. The water quantity and quality parameters at the effluent of the studied urban catchment (*Gibraltar*: Bogotá, Colombia) were measured by means of infra-red, ion sensitive, UV-VIS spectrometry (water quality) and ultrasonic (water quantity: flow rate) sensors. Hence, the rainfall pulses were registered at three rain gauge stations (*Fontibon*, *Bosa* and *Casablanca*) in proximities of the studied urban catchment. A methodology for identifying periods in which water quality and quantity seemed to be ranging outside of their typical dry weather behavior was designed, by the use of signal processing and time series techniques (Median Filter, Step Detection and Autocorrelation functions-ACF-). A Principal Components Analysis (PCA) brought up a preliminary insight of which water quality and quantity parameters seem to explain mostly the differences between the variability of dry and wet weather data. The most explicative parameters measured by means of uni-parametric and ultrasonic sensors were: (i) the Dissolved Oxygen (DO), (ii) the Flow Rate, (iii) the Oxidation Reduction Potential (ORP), (iv) the Turbidity and (v) the Total Suspended Solids (TSS). The pH and the Temperature seemed not to be altered from dry to rainy weather. For the case of measurements undertaken by UV-VIS spectrometry, the absorbance values registered by the boundary wavelengths of the UV-VIS range (200 nm to 750 nm) seems to explain better the differences between variability of dry and wet weather days' data, rather than for the wavelengths closer to the center of the UV-VIS range (for about 400 nm). Hence, considering the most explicative water quality and quantity parameters from PCA, the designed methodology was then applied to the time series registered during rainy days. A temporal delimitation of the rainfall events was obtained, which exposed an appropriate consistency. Regarding the identified rainfall events, association metrics (Pearson's Correlation Test, Average Mutual Information, Linear Models and Partial Least Squares Model) between different rainfall and water quality and quantity statistical characteristics per event (*e.g.* mean, standard deviation, entropy and duration) were undertaken, with the purpose of detecting recurrent relations between rainfall and water quality and quantity (for rainfall registered at different rain gauge stations). The results showed that: (i) Turbidity seems to be related with the duration of the rainfall pulses and with the mean rainfall height, (ii) DO seems to have an association with total volume of rainfall and with the rainfall pulse's entropy, (iii) flow rate was found to be related with mean rainfall height, standard deviation of rainfall heights and the mean rainfall intensity, (iv) TSS behavior was found to be linked with characteristics of rainfall such as the duration, the mean rainfall height, the standard deviation of the rainfall heights and the memory of the process (evaluated by the ACF of each specific rainfall pulse), (v) ORP seemed to be connected with the characteristics of rainfall events such as the duration, the Antecedent Dry Weather Period (ADWP), the total rainfall volume, the entropy and the standard deviation of the rainfall pulse(s). Rainfall registered at the *Bosa* rain gauge station seems to be more correlated with characteristics of the effluent in terms of quality and quantity, rather than *Fontibon* and *Casablanca* rain gauge stations. Preliminary insights showed that the principal sources of variability between the spatial influence of a given rain gauge station (in terms of the amount of recurrent) with different sub basins of the urban catchment are: (i) the distance to the rain gauge station which is mostly influencing a specific sub basin, (ii) the variability of the pipes' slope at a given sub basin, (iii) the industrial land use regarding a sub basin, (iv) the mean pipes' slope for a given sub basin and (v) the variability of pipes' length at a sub basin.

## RESUMEN

El presente estudio se trazó como objetivo establecer la influencia de la lluvia sobre la calidad del agua en el efluente de una cuenca urbana mediante el uso de medición en alta resolución temporal. Los parámetros de calidad y cantidad de agua en el efluente de la cuenca urbana estudiada (Gibraltar: Bogotá, Colombia) se midieron mediante captosres con tecnología uni-paramétrica, espectrometría UV-VIS (en cuanto a calidad) y medición ultrasónica (en cuanto a cantidad). Los registros de lluvia se efectuaron a través de tres estaciones de lluvia (Fontibon, Bosa y Casablanca) en proximidades de la cuenca urbana estudiada. Mediante el uso de técnicas de procesamiento de señales y análisis de series de tiempo (filtro de medianas, Step Detection y funciones de autocorrelación-ACF-), se propuso una metodología orientada a identificar aquellos periodos durante los cuales un parámetro dado de calidad o cantidad de agua se encuentra fluctuando por fuera de su comportamiento típico durante el tiempo seco. Mediante un Análisis de Componentes Principales (PCA) se estableció de manera preliminar qué parámetros de calidad y cantidad de agua explican en mayor medida las diferencias entre la variabilidad de tiempo seco y tiempo de lluvia. Los parámetros más explicativos medidos a través de captosres con tecnología uni-paramétrica y ultrasónica fueron: (i) oxígeno disuelto (OD), (ii) el caudal, (iii) el potencial de oxidación reducción (ORP), (iv) la turbiedad y los sólidos suspendidos totales (SST). El pH y la temperatura mostraron no ser diferentes para tiempo seco o tiempo de lluvia. En cuanto a la espectrometría UV-VIS, los valores de absorbancia registrados en las longitudes de onda borde del rango UV-VIS (200 nm a 750 nm) mostraron explicar en mayor medida las diferencias en variabilidad entre tiempo seco y tiempo de lluvia, que aquellas longitudes de onda más cercanas al centro del rango UV-VIS (e.g. 400 nm). Se aplicó la metodología propuesta a las series de tiempo de parámetros explicativos (del PCA), registradas durante días lluviosos. Se obtuvo una delimitación temporal de diferentes eventos lluviosos, la cual mostró una consistencia apropiada. Utilizando dichos eventos, se implementaron métricas de asociación (Pearson's Correlation Test, Average Mutual Information, Linear Models and Partial Least Squares Model) entre estadísticos de la lluvia y la calidad y cantidad de agua (e.g. media, desviación estándar, entropía y duración), con el fin de detectar relaciones recurrentes entre la lluvia (de tres estaciones) y la calidad y cantidad de agua. Se obtuvieron los siguientes resultados: (i) la turbiedad parece estar relacionada con la duración de los eventos de lluvia y con la altura promedio de agua precipitada, (ii) el OD parece estar asociado con el volumen total de agua precipitada y con la entropía de los pulsos de lluvia, (iii) el caudal se encontró relacionado con la altura media de agua precipitada, la desviación estándar de las alturas de agua precipitadas y con la intensidad promedio del evento, (iv) los SST parecen estar ligados con características de la lluvia tales como la duración del evento, la altura de lluvia promedio, la desviación estándar de las alturas de lluvia y con la memoria de la lluvia (medida a través de la AFC de la lluvia, durante un evento), (v) el ORP parece estar relacionado con características de la lluvia como la duración, el periodo seco antecedente, el volumen total de lluvia, la entropía de la lluvia y la desviación estándar de las alturas de lluvia, durante un evento lluvioso. La lluvia registrada en la estación de Bosa parece estar más correlacionada con características de calidad y cantidad de agua, que para el caso de las estaciones de Fontibón y Casablanca. Preliminarmente, las principales fuentes de variabilidad entre la influencia espacial de una estación de lluvia sobre las diferentes sub cuencas son: (i) la distancia a la estación de lluvia, (ii) la variabilidad de las pendientes de las tuberías, (iii) el porcentaje de uso industrial, (iv) la pendiente promedio de las tuberías y (v) la variabilidad de las longitudes de tuberías.

# 1 INTRODUCTION

Experience has shown an important influence of rainstorm water on the arrival of pollutant loads and concentrations into the receiving water bodies (Warith et al., 1998). For several flow rate values, wastewater composition varies due to dilution phenomena caused by rainfall water (Warith et al., 1998). Under this perspective, rainstorm events might represent a significant variability in water quality for receipt water bodies with important anthropic influences.

During the past years, advances towards the understanding of pollutants concentrations have been performed based mainly in data acquired from monitoring campaigns, samples collection and laboratory analysis (see. Athayde *et al.*, 1983; Saget, 1994; Duncan, 1999; Pitt *et al.*, 2004; Brombach *et al.*, 2005; Ellis *et al.*, 2006). Hence, some significant restrictions can be mentioned from this approach: (i) campaigns with limited number of rainstorm events and operational high costs, (ii) limited observations of pollutographs dynamics, (iii) low representativeness of phenomena's variability and (iv) difficulties for statistical analysis due to a limited number of data (Métadier & Bertrand-Krajewski, 2012).

Since 90's, with the development of continuous and reliable measurements in the sewer system, water quality monitoring during rainy weather has been implemented with several purposes. Although the mentioned studies had considered the behavior of various pollutants in urban drainage systems, nowadays, a detailed relationship between rainfall characteristics and water quality parameter's variability in the effluent of urban catchments seems to be not clearly established.

Furthermore, the variability of the pollutographs produced by the runoff during rainfall events at the effluent of an urban catchment requires detailed analyses of the variability and complexity of processes and interactions involved: for identical flow rate values, different pollutants concentrations have been observed. The mentioned variability has been addressed by multiple authors (*e.g.* Barraud *et al.*, 2002; Veldkamp *et al.*, 2002; Grüning and Orth, 2002; Gruber *et al.*, 2005; Hochedlinger *et al.*, 2006; Bertrand-Krajewski *et al.*, 2007; Aumond and Joannis, 2008; Lacour, 2009; Lacour *et al.*, 2009; Schilperoort, 2011; Bertrand-Krajewski, 2012).

Studies like Lee *et al.* (2002) showed relations among the total magnitudes of runoff volumes from rainstorm events and concentrations of urban pollutants constituents such as Total Suspended Solids (TSS). These experiences can be described as approaches focused on event mean concentrations (EMC) of the pollutants constituents during the event. It must be pointed out that this type of analysis may not include other characteristics of rainfall and water quality parameters during a specific rainfall event (*e.g.* rainfall maximum heights, maximum concentration, maximum concentration's variation and rainfall event's duration). Under this perspective, analyses related to other characteristics of rainfall and pollutants during a specific rainfall event (*e.g.* rainfall maximum intensity, maximum concentration, maximum concentration's variation and rainfall event's duration) may be omitted.

Studies have shown the relevance of rainfall dynamics over the transport of pollutants in urban catchments (Rosenzweig et al., 2008), but most of these do not inquire into the spatial relationship of rainfall and pollutants' concentrations and loads (Lee et al., 2002). Studies like Tiefenthaler & Schiff (2002) and Gnecco et al. (2005) showed influences between rainfall event characteristics (particularly total heights, intensities and durations) over pollutant concentrations (specifically TSS,

metals and organic compounds) in a test environment. This fact makes results described above hardly generalizable to large and non-homogeneous urban catchments. Measurements of water quality parameters inside the catchment could be implemented with the purpose of taking into account their variability due to the interaction of the runoff and the urban surfaces (like roads or roofs: see for example Tiefenthaler & Schiff, 2002; Gnecco et al., 2005). However, the purpose of this document is to figure out these interactions as an integrated phenomenon, which is observable in the total behaviour of the effluent, in terms of water quality. Additional recommendations can be pointed (see for example Tiefenthaler & Schiff, 2002; Gnecco et al., 2005) to monitoring water quality parameters in the effluent of the catchment. These practices can bring up the possibility of studying the interaction between pollutant concentrations and the transport of water flow through the sewer system.

In Colombia, studies related with the relation between rainfall and water quality draining from an urban catchment have been addressed under a modeling approach (Díaz Granados & Camacho, 2003; Estupiñán et al., 2009). Although these kinds of physically-based numerical models can become supporting tools in drainage systems management, their main objective was not to inquire in the relationship between rainfall characteristics (intensity, rainfall height, spatial distribution, antecedent dry weather period) and the water quality parameters of the effluent. For most of these models rainfall is set as an input variable, without detailing into the possible relation between rainfall and pollutographs in terms of detailed recurrent characteristics and variability among rainfall events (generally due to absence of continuous measurements).

Therefore, taking into account the identified knowledge gaps, study of water quality of wastewater seems to be fundamental for urban sanitation water systems' management (*e.g.* sewer network, Wastewater Treatment Plants, overflow structures). In this sense, necessity of scientific advances towards the understanding of pollutants behavior, not just for the sanitary flow rates, but also of rainwater and runoff and its relation with rainfall characteristics within or in proximities of a specific urban catchment arises. In addition, linkages detected between rainfall and water quality by the use of statistical and mathematical techniques, without detailing into the followed physical process, brings up the possibility of defining: (i) regulation and control, operation, planning and rehabilitation requirements in sewer systems, (ii) real time control mechanisms, (iii) integrated urban drainage modeling tools (Métadier & Bertrand-Krajewski, 2012). The contribution to the knowledge given by the present thesis is related with the following aspects: (i) as previously exposed, the research question has not been addressed, which primarily regards with the influence of rainfall over water quality effluent of an urban catchment, by the use of continuous measurements (1 min time step), (ii) dealing with large data bases product of continuous measurements is a challenge which demand the use of innovative methodologies for the acquisition as well as the processing and validation of reliable data.

Specifically, the aim of this thesis was to identify influences of pluvial characteristics over variables related to water quality parameters, registered with continuous measurements, by the use of statistical tools and mathematical classification methods. In order to accomplish this objective, several rainfall events registered over a study zone (*Estación Elevadora Gibraltar - Cuenca el Tintal* - , Bogotá, Colombia) were analyzed.

The present document is composed by the following sections: (i) Introduction, which justifies and briefly explain the question, the approach and the scope of the present research; (ii) literature review, including the state of art and the definition of theoretical concepts that were addressed in the development of the present thesis; (iii) materials and methods section, including a description of the study zone, the equipment and data used and the developed methodologies for addressing the scientific problem; (iv) the results obtained from applying the proposed methodologies to the experimental data; (v) conclusions and recommendations, regarding the results obtained.

## 2 LITERATURE REVIEW

In the following paragraphs the literature review is described. The first part consists in the state of art, regarding the addressed scientific problem. The state of art includes researches towards related topics, under a national and international context. A second section, proposes a theoretical framework with a definition of the concepts and tools implemented in this research.

### 2.1 State of Art

Jacobsen & Madsen (1996) measured the concentration of dissolved oxygen in an urban creek located in the suburbs of Copenhagen, Denmark, during a period of three months. With the collected data, statistical modeling techniques were used in order to explain the dissolved oxygen concentration's variability, in the mentioned creek. This study, brought up relevant contributions, but did not analyze other water quality parameters commonly used as Chemical Oxygen Demand (COD), Total Suspended Solids (TSS) or Conductivity (K). Although environment temperature was used as an input for the model, water temperature in the creek was not object of analysis.

A study executed by Tiefenthaler & Schiff (2003) sought to establish a relationship between rainfall intensity and its duration with runoff's water quality. A physical rainfall simulator was used under controlled conditions. Samples of several simulated rainfall events with different durations and rainfall intensities were analyzed. It was found that TSS concentrations, metals and organic compounds were considerably higher at the beginning of the rainfall events. An inverse correlation between these constituents and studied rainfall characteristics (intensity and duration) was found. On the other hand, it was determined that total runoff volumes produced similar pollutants concentrations. Advantages of the obtained results can be described in a more detailed analysis of rainfall specific characteristics during an event (for this case duration and intensity). Previously exposed results can be described as hardly generalizable to non-homogeneous and large urban catchments, due to their test environment nature.

Rosenzweig et al. (2008) described the effect of rainfall dynamics of various rainfall events over the transport of dissolved inorganic nitrogen from highland areas to downstream in an urban catchment located in the Northeastern region of the United States. The general conclusion of this study was that the flow rate seemed to be the most influent fact in the transport of nitrogen, following a direct relation: for higher flow rate values, higher nitrogen concentrations were observed. Hence, the highest nitrogen concentrations were detected afterwards of the hydrograph's peak, for all analyzed rainfall events. Although this study brought up new insights concerning the research question, other water quality parameters of common interest for urban drainage were not discussed (*e.g.* COD, TSS, K, SS).

Gnecco *et al.* (2005) studied the behavior of water constituents in urban environments, by distinguishing two types of runoff: roofs' runoff and roads' runoff. A small urban catchment located in Villa Cambiaso, Genova (Italy) was selected as the study zone. By analyzing the results of a monitoring campaign, they found in both types of runoff higher concentrations of TSS at the beginning of the event, fading with the passage of time. Detailing in the influence of various hydrological variables over event mean concentration (EMC), it was observed a correlation just for EMC of TSS and rainfall's maximum intensity, during a given event. Heavy metals' concentrations seem to be influenced by rainfall intensity and frequency. This study focused in several water

quality parameters (TSS, COD, and heavy metals), fact that gave it enough representativeness of physical and biochemical pollution of analyzed water systems. As identified gaps regarding this study, two aspects can be pointed: (i) with the purpose of including the interaction of the runoff with the sewer system, measurements at the effluent of the catchment can be implemented; (ii) as the selected catchment have a strictly residential land use, considering a study zone with more land uses will bring up more general results.

Lee *et al.* (2002) studied 13 urban catchments in the city of Chongju (South Korea), with several characteristics: (i) combined or separated sewer network, (ii) commercial or residential land use, (iii) different drainage areas, (iv) different surface's permeability; among others. By registering 38 rainstorm events, urban pollutant's runoff was studied. Wash off magnitudes were assessed through a data analysis technique oriented to EMC. As most relevant results, it was detected that constituents such as SS seem to be influenced by rainfall, oppositely to pollutants such as COD. No correlation between the pollutants runoff and the antecedent dry weather period (ADWP) was observed. An increasing of pollutants towards the beginning of the event was observed mainly for small catchments. This research provided a more general vision of the research question by having different catchment's characteristics like land use and area. As an identified gap of this study, it can be mentioned that rainfall and pollutographs characteristics (such as maximum intensity, maximum concentration, maximum gradient of concentration, events duration) for a given rainfall event were not included in the analysis.

Regarding local studies, the exposed problem has been addressed under modeling approaches. Díaz Granados & Camacho (2003) proposed a water quality simplified model, which was implemented in *cuenca del Virrey* in Bogotá (sub basin with an area of 93 Ha, located in Bogota's Northeastern zone). The result of this experience brought up advances towards the understanding of rainfall-runoff and water quality processes inside the studied catchment, by obtaining model's optimal parameters, with an appropriate predictability. By the use of water quality measurements in multiple points of the sewer system, water quality was modeled along the sewer pipes, for each time step. This study brought up new insights of real time operation for a Bogota's basin, based on mathematical modeling with parameters calibration. However, a relation between calibrated temporal parameters and rainfall characteristics (intensity, rainfall volumes, spatial distribution, ADWP) for the events registered during the simulations was not considered.

Estupiñán *et al.* (2009), by the use of rigorous mathematical modeling, represented urban runoff's behavior and its pollution, in terms of TSS. The variables measured were the atmospheric deposition and the sediments wash-off. The main result of this research was that models with a simpler structure showed the better performance. A strong relation between the ADWP and the amount of wash-off sediments was detected. Although exposed results were promising, the methodology needs measurements (including atmospheric depositing and wash-off sediments), which are commonly difficult to carry out. Thus, catchments with limited instrumentation might not be appropriated for implementing the model. In general terms, concerning local experiences, the phenomenon has been analyzed following an impulse-response approach, without inquiring into possible recurrent relations between rainfall characteristics and variability, regarding independent rainfall events.

With the purpose of giving a brief summary of the state of art, Table 2.1 describes for each article: (i) the objective of the research, (ii) the implemented tool, (iii) the strength of the proposed methodology, (iv) the scientific contribution and (v) the weaknesses identified in terms of knowledge gaps. This table is organized by researches under an international and national context.

**Table 2.1 Summary of the state of art**

| <b>Reference</b>             | <b>Objective</b>                               | <b>Tool</b>                                          | <b>Strength</b>                  | <b>Contribution</b>                                                                                                                                                   | <b>Weakness</b>                                                           |
|------------------------------|------------------------------------------------|------------------------------------------------------|----------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------|
| Jacobsen & Madsen (1996)     | Modeling DO                                    | Statistical modeling techniques                      | Detailing variability            | Variability of DO                                                                                                                                                     | Other water quality parameters not analyzed                               |
| Tiefenthaler & Schiff (2003) | Rainfall characteristics<br>Vrs. Water quality | Physical model, ANOVA, correlation tests             | Several rainfall characteristics | High TSS, heavy metals and organic compounds at the beginning of the event                                                                                            | Test of environment                                                       |
| Rosenzweig et al. (2008)     | Inorganic Nitrogen transport                   | M/V curves, descriptive statistic                    | Nitrogen                         | Nitrogen as a function of flow rate                                                                                                                                   | Other water quality parameters not analyzed                               |
| Gnecco et al. (2005)         | Pollution in roofs and streets' urban runoff   | M/V curves, descriptive statistic, correlation tests | Detailed measurements            | Greater concentrations of TSS at the beginning of the event. TSS as a function of maximum intensity, heavy metals' variability and rainfall's intensity and frequency | test environment, Non taking into account travel through the sewer system |

| Reference                      | Objective                                                           | Tool                                                 | Strength                                   | Contribution                                                                     | Weakness                                                          |
|--------------------------------|---------------------------------------------------------------------|------------------------------------------------------|--------------------------------------------|----------------------------------------------------------------------------------|-------------------------------------------------------------------|
| Lee et al. (2002)              | Relation rainfall-pollutants, including catchments' characteristics | M/V curves, descriptive statistic, correlation tests | Detailed measurements, 13 urban catchments | rainfall total volumes produces higher TSS concentrations, independently of ADWP | just focused on EMC, few hydrological variables                   |
| Díaz Granados & Camacho (2003) | Water quality modeling                                              | Physically based equations, parameters tuning        | Detailed description                       | Understanding of rainfall-runoff and water quality processes                     | Relation between model's parameters with rainfall characteristics |
| Estupiñán <i>et al.</i> (2009) | Mathematical modeling                                               | Physically based equations, parameters tuning        | Detailed description                       | Relation AWPD and the amount of washed sediments                                 | Measurements of sediments wash off and atmospheric deposition     |

## 2.2 Theoretical Framework

Study of the relation between the pollutant's variability at the effluent of an urban catchment and rainfall characteristics requires more detailed analyses due to variability and complexity of additional processes and interactions involved (Métadier & Bertrand-Krajewski, 2012). In this sense, necessity of using more flexible tools, which can figure out the relation between rainfall and pollutants at the effluent of the catchment from continuous measurements, arises. The information extracted from data bases product of mentioned measurements, should have the following properties: (i) assure representativeness in agreement with the spatial and temporal scales, (ii) reliability and precision of data and (iii) being sufficiently detailed, according to the proposed scopes (Fletcher & Deletic, 2007).

With the purpose of introducing some relevant concepts that were addressed for the present document, following paragraphs are proposed. In this theoretical framework, concepts such as data mining definition, statistical and black-box techniques and signal processing concepts are contextualized. Hence, keeping in mind that this thesis is closely related to variability, a brief explanation of this concept becomes relevant.

Time series usually exposes periods in which fluctuations of independent variable increases rather than in other periods, as outbreaks of agitation and uncertainty, followed by relatively calm periods. An isolated observation with a high magnitude does not represent strictly variability. Given the fact that these behaviors are very common in any stochastic process, variability can be defined just when a consecutive series of observations report higher or lower average values, dynamically over time. Commonly used estimators of variability can be stated as follows: standard deviation, variance, variation coefficient, interquartile distance, interdecile distance; non necessarily constant with the passage of time (volatile stochastic process or heteroscedasticity) (Montenegro, 2010).

Therefore, one of the most recently identified challenges for urban drainage is to deal with extended time series and data bases, which expose non-linear and hardly predictable behaviors (as described above), obtained by means of technologies such as online measurements of different variables of the system (adapted from Fletcher & Deletic, 2007). In this sense, techniques like data mining and statistical models seems appropriate for facing this new challenges.

### 2.2.1 Data Mining:

Data mining implementation in urban drainage studies arises from the necessity of identifying patterns, tendencies and cause-effect relations between the different variables, coming from measurements and monitoring, registered in data bases. Specifically, data mining gives a high flexibility in the study of multidimensional problems and hidden rules. In this sense, its implementation seems to be pertinent for the research question addressed (Fletcher & Deletic, 2007).

Throughout the years an important amount of technologies and tools have been developed for data processing, with the aim of analyzing complex and high-variability problems. These tools, that are generally used altogether, can be classified as follows: (i) tools for pre-processing, aggregation and data sub-sampling, (ii) data mining for recognizing patterns in order to validate theoretical relations between the variables (Fletcher & Deletic, 2007).

The extraction of statistical information is probably the first step for studying a data base. In general, the following concepts can be mentioned in a statistical data analysis: (i) extraction of characteristics (*e.g.* mean, median, mode, harmonic mean and geometric mean), (ii) extraction of variability of single values (*e.g.* standard deviation, variance, coefficient of variation, asymmetry) and (iii) transforming the domain of studied phenomenon (time-space), to equivalent domains such as frequency-time or Wavelet. Previously exposed techniques can be used, either as inference tools or also as pre-processing tools in order to reduce the dimensionality of the studied problem (Fletcher & Deletic, 2007).

Data mining or machine learning can be classified in the following categories: (i) “*data-driven*”, which aims to recognize hidden patterns embedded in the data, without any hypothesis established by the user, (ii) “*theory-driven*”, with a previously established cause-effect relations between the variables, (iii) evaluation and polishing of hypotheses (Fletcher & Deletic, 2007). Accordingly, two main objectives are sought with the implementation of data mining: (i) Prediction, including classification, regression and time series analysis and (ii) observation and description, characterizing deviations, variability, segmentation, grouping, associations, rules, synthesis, visualization and text mining (Fletcher & Deletic, 2007).

### 2.2.2 Principal Component Analysis

The principal component analysis (PCA) is a widely used statistical technique in urban drainage studies. The objective is basically to extract information of measured data, identifying the variability’s direction by the use of linear combinations among the variables (Mohamed-Faouzi et al., 2006). The concept “linear combination” is referred to the fact that the scores in a component are created by the sum of the scores in the observed variables. The observed variables are averaged, seeking that the resulting components are capable to represent the greater amount of variance of the data set. For achieving a reliable PCA analysis, a data set greater than 100 or five times the number of analyzed variables is required. The PCA analysis can be used also as a grouping tool, accordingly with the influence of each observation over a common feature (Hernández, 1998). For the present research, the PCA was used with two objectives: (i) identifying which water quality parameters seem to explain the differences in the variability of dry weather and wet weather data and (ii) detecting the most relevant factors regarding the physical characteristics of the basin which better explain the variability in the influence of rainfall over water quality, at the effluent of the catchment.

### 2.2.3 Association Metrics

In the following section some association metrics addressed for the present research are described. These metrics include linear and non-linear, univariate and multivariate approaches. For the proposed metrics, their significance calculation procedure is then given, whenever it could be formulated. For the present research, the association metrics were implemented for establishing correlations between rainfall and water quality characteristics, for different rainfall events.

#### 2.2.3.1 Correlation Tests

Identifying the degree of association between two variables seems fundamental in order to answer the proposed research questions. The correlation test is a statistical tool widely used for this task in related studies (*e.g.* Gnecco et al., 2005; Lee et al., 2002; Tiefenthaler & Schiff, 2003; Métadier & Bertrand-Krajewski, 2012).

The Pearson's Correlation Coefficient of a sample is a measurement of the linear association between two variables  $X$  and  $Y$ . In case that a normal bivariate distribution cannot be applied for  $X$  and  $Y$ , the Spearman's Correlation Coefficient  $r_s$ , is another (non-parametric) measurement of association, which can be appropriate for these cases (Alvarado, 2008).

The magnitude of the calculated statistic estimator is comparable with tabulated values, in order to determine if the differences are "statistically significant" (Alvarado 2008). For this, an arbitrary significance level ( $\alpha$ ) is chosen.  $\alpha = 5\%$  or  $1\%$ , means that it exists a probability of  $5\%$  or  $1\%$  that the differences detected can be due to randomness, when no difference between the treatments effectively exists. If the  $t$  estimator is greater than the critical value tabulated for  $\alpha = 5\%$ , it can be stated that  $p < 0.05$  (Alvarado 2008). This statement implies that it exists a chance less than  $5\%$  that the observed differences can be explained just by randomness and less than  $1\%$  in case of using  $\alpha = 1\%$  and  $p < 0.01$ . Thus, when observed differences are truly due to random causes, an error of  $5\%$  is committed, whenever is stated that groups are different when they are actually different. This error is called type 1 and the probability of committing it is measured by the  $p$  value; therefore a test with  $p < 0.01$  is more restrictive than other with  $p < 0.05$  (Alvarado, 2008).

#### 2.2.3.2 Average Mutual Information (AMI)

Another relevant concept for the present thesis, which is a metric of the association degree between two variables, is the AMI statistic (Average Mutual Information). This metric corresponds to the average mutual information  $I(X;Y)$  between two random variables  $X$  and  $Y$  and can be interpreted as the average amount of information that will be obtained of  $X$ , from knowing  $Y$ . Mathematically, it can be expressed as follows (Charniak, 1993):

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right) \quad \text{Eq. 2.1}$$

were:

$I(X;Y)$ : average mutual information between  $X$  and  $Y$  (degree of association between the two variables).

$P(X,Y)$ : joint probability of  $X$  and  $Y$ .

$P(X)$ : marginal probability of  $X$ .

$P(Y)$ : marginal probability of  $Y$ .

For testing the significance of the AMI estimation, the null hypothesis that the datasets  $X$  and  $Y$  are independent is established. Only if the observed mutual information is not consistent with this null hypothesis, it can be claimed that the null hypothesis is wrong, and the data thus contains significant correlation. However, the rejection of a given null hypothesis is never absolute, but always with respect to a certain significance level. In order to test the null hypothesis it can be constructed an ensemble of surrogate datasets  $\{X_s, Y_s\}$ , consistent with the null hypothesis: the approach consists in creating random permutations of the original data  $X$  and  $Y$  as surrogate datasets  $\{X_s, Y_s\}$ . The values are constrained to take on the same values as the data, just in random order. Therefore, the

mean and the standard deviation  $\sigma$  of the observed mutual information using the ensemble of surrogate data is estimated. The significance  $S$  is given by:

$$S = \frac{I(X, Y)^{data} - \langle I(X, Y)^{data} \rangle}{\sigma_{surr}} \quad \text{Eq. 2.2}$$

Since this approach implicitly assumes knowledge about the distribution of the mutual information (e.g.  $|S| \geq 2.6$  can be treated as significant at a level of 99 % assuming a *Gaussian* distribution), it is suggested to use a more general reasoning (Schreiber & Schmitz, 2000; Theiler et al., 1992): First a probability  $\alpha$  of false rejection is selected. Then, for a sided test,  $M^S = \frac{1}{\alpha} - 1$  surrogates are generated. Including the original data, an ensemble of  $1/\alpha$  datasets can be proposed. Thus, the probability that the data has the largest mutual information within the whole ensemble merely be coincidence is exactly  $\alpha$ . The null hypothesis could therefore be rejected at a significance level  $(1-\alpha)*100$  % (Steuer et al., 2002).

### 2.2.3.3 Simple Linear Regression

To describe the linear association between quantitative variables, a statistical procedure called regression often is used to construct a model. Regression is used to assess the contribution of one or more “explanatory” variables (called independent variables) to one “response” (or dependent) variable. It also can be used to predict the value of one variable based on the values of others. When there is only one independent variable and when the relationship can be expressed as a straight line, the procedure is called simple linear regression. Any straight line in two-dimensional space can be represented by the following equation:

$$y = a + bx \quad \text{Eq. 2.3}$$

where  $y$  is the variable on the vertical axis,  $x$  is the variable on the horizontal axis,  $a$  is the  $y$ -value where the line crosses the vertical axis (often called the intercept), and  $b$  is the amount of change in  $y$  corresponding to a one-unit increase in  $x$  (often called the slope).

Regression is an inferential procedure, meaning that it can be used to draw conclusions about populations based on samples randomly drawn from those populations. With the purpose of testing significance, it can be established a null hypothesis where the value of the slope  $b$  is equal to zero; hence the alternative hypothesis is determined by any value of  $b$  different from zero. For the proposed test is required that: (i) dependent variable  $y$  has a linear relationship with the independent variable  $x$ , (ii) for each value of  $x$ , the probability distribution of  $y$  has the same standard deviation (variance homogeneity), (iii) for any given value of  $x$ , the  $y$  values are normally distributed (Shapiro Wilk Test). This prove can be assessed by means of a *t-test*. The test statistic can be defined by the following equation:

$$t = \frac{b_1}{SE} \quad \text{Eq. 2.4}$$

where  $b_1$  is the slope of the sample regression line, and  $SE$  is the standard error of the slope, calculated as follows:

$$SE = \frac{\sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{(n-2)}}}{\sqrt{\sum (x_i - \bar{x})^2}} \quad \text{Eq. 2.5}$$

where  $y_i$  are the observed values of the dependent variable,  $\hat{y}_i$  are the estimated values of the dependent variable,  $x_i$  are the observed values of the independent variable,  $\bar{x}$  is the mean of the independent variable and  $n$  is the number of observations. From the test statistic and by knowing the degrees of freedom ( $n-2$ , for linear models) it can be calculated the p-value (probability of observing a sample statistic as extreme as the test statistic). This estimation, compared to a pre-established significance level, can be considered as evidence for accepting or rejecting the null hypothesis (Mevik & Wehrens, 2007).

#### 2.2.3.4 Partial Least Squares Regression (PLS)

Multivariate regression methods like principal components regression (PCR) and partial least squares regression (PLSR) are widely used in a wide set of urban drainage applications (*e.g.* Winkler *et al.*, 2008). The main reason is that they have been designed to confront the situation that there are many, possibly correlated, predictor variables, and relatively few samples.

In the usual multiple linear regression (MLR) context, the least-square solution for

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \varepsilon \quad \text{Eq. 2.6}$$

is given by

$$\mathbf{B} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} \quad \text{Eq. 2.7}$$

The problem is that  $\mathbf{X}^t\mathbf{X}$  is singular, either because the number of variables (columns) in  $\mathbf{X}$  exceeds the number of objects (rows), or because of collinearities. Both PCR and PLSR circumvent this by decomposing  $\mathbf{X}$  into orthogonal scores  $\mathbf{T}$  and loadings  $\mathbf{P}$ .

$$\mathbf{X} = \mathbf{T}\mathbf{P} \quad \text{Eq. 2.8}$$

and regression  $\mathbf{Y}$  is implemented not on  $\mathbf{X}$  itself but on the first  $a$  columns of the scores  $\mathbf{T}$ . In PCR, the scores are given by the left singular vector  $\mathbf{X}$ , multiplied with the corresponding singular values, and the loadings are the right singular vectors of  $\mathbf{X}$ . This, however, only takes into account information about  $\mathbf{X}$ , and therefore may be suboptimal for prediction purposes. PLSR aims to incorporate information of both  $\mathbf{X}$  and  $\mathbf{Y}$  in the definition of the scores and loadings. Scores and loadings are selected in such way that to describe as much as possible of the *covariance* between  $\mathbf{X}$  and  $\mathbf{Y}$ , where PCR concentrates on the *variance* of  $\mathbf{X}$ . For this reason PLS can be selected as an appropriate multivariate regression technique in order to identify the most important variables from  $\mathbf{X}$  in terms of explaining  $\mathbf{Y}$ . The PLS can be understood as a Principal Component Analysis (PCA), including a multiple regression over the principal components (Mevik & Wehrens, 2007).

#### 2.2.4 Support Vector Machines

The Support Vector Machines are a classificatory linear application between two groups for a set of input individuals. Its principle is based on the representation of each of the training individuals (with its membership to one of the sets previously defined) as a point in a given space. From this representation, a plane with a maximum distance to the individuals of each group is defined (Schölkopf & Smola, 2002). For the present research, the Support Vector Machines and the Kernel Methods were used for assessing probability density functions of the dry weather behavior, without using the traditional approach of histograms due to its high sensibility to the bins' length selection (adapted from Schölkopf & Smola, 2002).

Another application of Kernel Methods is for estimating the frequency histograms of a data set. The traditional method for estimating this data characteristic depends on a selected class mark, which highly influences the shape of the histogram. For this thesis Kernel Methods were also used for estimating the continuous probability density function (PDF) of the observed data. This technique achieves continuous and smooth density functions, independently of the selected class mark.

Formally, given a sample with  $n$  observations  $x_1 \dots x_i$ , a Kernel type function is defined as follows:

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad \text{Eq. 2.9}$$

Where  $K(x)$  is a Kernel function,  $h_n$  is a sequence of positive values known as the bandwidth:  $h_n$  will establish the smoothness of the estimation. The distribution  $K(x)$  reallocates the probability mass  $1/n$  in the neighborhood of each sample point (see details in Schölkopf & Smola, 2002).

#### 2.2.5 Uncertainty Analysis

Uncertainty can be defined as a property which characterizes the dispersion of the values that can be reasonably attributed to a mesurand (term related with the magnitude to the measurement) (CEM, 2008).

Assessing the uncertainty is an important measure of how reliable is our data. This estimation is difficult because of the unpredictability, complexity and nonlinearity of urban water systems. Predictions made from measured data will involve errors in all cases (adapted from Fletcher & Deletic, 2007).

Uncertainty analysis is a rigorous methodology for estimating uncertainties in measurements and in the results calculated from them. Uncertainty assessed by the use of statistical methods is often called Type A uncertainty; in contrast, instruments' resolution can be assigned to a Type B uncertainty. For the present work uncertainties will be estimated by means of Type A and Type B approaches (adapted from Fletcher & Deletic, 2007).

Generally, uncertainty sources are assumed to be normally distributed and thus can be estimated by means of the standard deviation of the measurement population. On the other hand, when data is used in order to express indirect measurements by means of direct measurements (e.g. flow rate can be derived from the measurement of the water level and the mean velocity in a sewer system),

uncertainty of dependent variable will be greater than the original uncertainty of independent variables (combined standard uncertainty). This propagation of elemental errors can be estimated by at least two methodologies: (i) first-order Taylor series approximation and (ii) Monte Carlo simulation (adapted from Fletcher & Deletic, 2007).

#### 2.2.5.1 Type A Evaluation of Uncertainty

A practical and reliable method for estimating the expected value  $\mu_q$  of a random variable  $q$  measured by  $n$  independent observations is an arithmetic average of all of the  $q_k$  observations of the measured variable. This value can be computed as follows:

$$\bar{q} = \frac{1}{n} \sum_{k=1}^n q_k \quad \text{Eq. 2.10}$$

Thus, the individual observations of  $q$ , expressed by  $q_k$  can be different among them due to random effects in the quantities. Therefore, the experimental variance of the observations can be estimated by the use of the following equation:

$$s^2(q_k) = \frac{1}{n-1} \sum_{k=1}^n (q_k - \bar{q})^2 \quad \text{Eq. 2.11}$$

This statistic characterizes the variability of the observed values  $q_k$ ; how far is the data from the mean  $\bar{q}$ , in average. Hence, the best estimate of the variance of the mean is given by:

$$s^2(\bar{q}) = \frac{s^2(q_k)}{n} \quad \text{Eq. 2.12}$$

The experimental variance of the mean or, standard deviation of the mean (as  $s(\bar{q}) = \sqrt{s^2(\bar{q})}$ ), quantifies how well  $\bar{q}$  estimates the expectation  $\mu_q$  of  $q$ , which can be understand as an assessment of the uncertainty of  $\bar{q}$  (adapted from Fletcher & Deletic, 2007).

#### 2.2.5.2 Type B Evaluation of Uncertainty

In a certain case in which  $x$  is estimated from an input quantity  $X_i$  that has not been obtained from repeated measurements, the associated estimated variance or standard deviation  $\mu(x_i)$  which represents the type A uncertainty of the measurand  $X_i$  cannot be defined. In these cases, uncertainty should be assessed by terms of scientific judgment, manufacturer's specification, knowledge of instruments, or as values taken from handbooks. This type of uncertainty is often called type B uncertainty (adapted from Fletcher & Deletic, 2007).

#### 2.2.5.3 Combined Standard Uncertainty

Suppose  $y$  is an estimate of the measurand  $Y$ , which is a result of a mathematical model with inputs  $x_1, x_2, x_3, \dots, x_N$ . Whenever input quantities are not correlated, combined standard uncertainty  $\mu_c^2(y)$  of a measurement  $y$ , assessed from measurement of independent variables  $x_1, x_2, x_3, \dots, x_N$  and their uncertainties  $\mu(x_1), \mu(x_2), \mu(x_3), \dots, \mu(x_N)$ , can be estimated by the following equation:

$$\mu_c^2(y) = \sum_{i=1}^N \left( \frac{\partial f}{\partial x_i} \right)^2 \mu^2(x_i) \quad \text{Eq. 2.13}$$

This proposed expression is based on a first order Taylor approximation. Another approach is by means of a Monte Carlo simulation (see section 2.2.6 for details). This methodology consists in the generation of random simulations of a set of given values which belong to the uncertainty domain  $\mu(x_1), \mu(x_2), \mu(x_3) \dots \mu(x_N)$  of independent variables  $x_1, x_2, x_3 \dots x_N$ . For a given simulation, the function  $y$  is calculated for the values of  $x_1, x_2, x_3 \dots x_N$ , ranging between  $-\mu(x_i)$  to  $+\mu(x_i)$ . Hence, the uncertainty is estimated *via* the type A procedure as described above (section 2.2.5.1).

### 2.2.6 Monte Carlo Simulation

The Monte Carlo method is a numerical technique which can be used for the solution of mathematical problems by the random sampling of random variables. This method consists in the development of an analytical model by the use of computational tools in order to predict the behavior of the system. Additional considerations are undertaken in order to ensure the completely randomized selection of system's parameters, in agreement with their probability density function (PDF) (Ayyub & Mccuen, 2003). By the use of this methodology non exact, but approximate solutions can be found with several iterations. Precision of the final solution will be leaded by the number of simulations performed (A.H.E.P.E., 2006).

### 2.2.7 Signal Processing and Times Series Techniques

For the present work some signal processing and time series techniques were implemented. Those techniques were used mainly for filtering the signals in order to assess average behaviors (Median Filter) and detect abrupt changes in the mean level of the time series (Step Detection). In addition, the ACF concept was implemented with the purpose of estimating an appropriate length of the mobile time window used for the Median Filter. This concept brought up insights of the periodical behavior of water quality time series and of the “inertia” given by the different measurement devices implemented.

#### 2.2.7.1 Median Filtering

In signal processing and time series analysis, it is often desirable to reduce the presence of noise in a given signal. This pre-processing technique is commonly used before applying a step detection algorithm, as is strongly recommended for highlighting abrupt changes of the signal. Such filter class is based in data ordering. Let  $x_i, i=1, \dots, n$ , be  $n$  observations. Thus, this data can be ordered according to their magnitude as follows:

$$x(1) < x(2) < \dots < x(n) \quad \text{Eq. 2.14}$$

where  $x(i)$  denotes the  $i$ -th order statistic and  $x(1)$  and  $x(n)$  are the minimum and maximum observations respectively, of a given time window. The observation  $x(v+1)$  lies in the middle and is called the median of the observations, also denoted by  $\text{med}(x_i)$ . The algorithm used for the present research consisted basically in replacing the value of the signal  $x(t)$  in the time  $t$  by the value of  $\text{med}(x_i)$ , defined by a central time window over the value  $x(t)$ , by the consecutive registers

regarding the past and future observations. This technique, rather than the traditional moving average processing, gives three advantages in terms of the signal obtained after processing: (i) signal will not lose abrupt changes from its original configuration; (ii) the technique is less sensitive to outlier presence; (iii) groups of high or low data will keep their original values (Arce, 2005).

### 2.2.7.2 Step Detection

Non-stacionarity or changes in generative parameters is often a key aspect of time series. Change point detection (CPD) aims to recognize regime abrupt changes in the distribution of a given signal at unknown time-steps. The problem can be stated as follows: there is a sequence consisting of  $n$  observations from the random variables  $X_1, \dots, X_n$ , and it is required to test whether this sequence contains any change points. Whenever the problem is addressed by means of a sequential detection (observations are processed sequentially over time), several approaches can be implemented for solving the statement (*e.g.* control charts CUSUM (Page 1954), Exponential Weighted Moving Averages (Roberts 1959), or sequential Bayesian methods (Chib 1998; Fearnhead and Lai 2007)).

This type of change detection is not retrospective, meaning that the decision whether a change has occurred at a particular point in the sequence is made using only the observations from the past at a given time-step, fact that makes it computationally feasible and also appropriate for real time operation.

Let denote  $x_t$  as the  $t^{\text{th}}$  observation that has been received, where  $t \in \{1, 2, \dots\}$ . Every time that a new observation  $x_t$  is received, the CPM approach treats  $x_1, \dots, x_t$  as being a fixed length sequence sample, computes the statistic  $D_t$  (statistic which evaluates all possible ways of splitting the data into two contiguous subsequences) as:

$$D_t = \max_k \frac{D_{k,t} - \mu D_{k,t}}{\sigma D_{k,t}} \quad 1 > k > t \quad \text{Eq. 2.15}$$

Were the null hypothesis of no change is then rejected if  $D_t > h_n$  for some appropriate threshold  $h_n$ . This threshold is selected based on the Mann Whitney statistic test (Pettitt, 1980). For every  $t^{\text{th}}$  observation were the null hypothesis is rejected the observation is received, then  $D_{t+1}$  is compared to  $h_{t+1}$  (Gordon, 2012).

### 2.2.7.3 Autocorrelation Function

The covariance between two observations  $X_n$  and  $x_{n+k}$  of a stationary stochastic process is defined as:

$$r(k) = \text{cov}(x_n, x_{n+k}) = E[(x_n - \mu)(x_{n+k} - \mu)] \quad \text{Eq. 2.16}$$

The quantity  $r(k)$  is defined for all integral values of  $k$ , and together it is called *the autocovariance function* of  $X_n$ . It measures the covariance between pairs at a distance or lag  $k$ . The value for the autocorrelation at lag 0 is 1. A long *autocovariance function* indicates that the data vary slowly (Priestley, 1981). A short covariance function indicates that the data at short distances are not related or correlated. From the above equation it can be conclude that:

$$r(0) = cov(x_n, x_{n+k}) = E[(x_n - \mu)^2] = \sigma_x^2 \quad \text{Eq. 2.17}$$

Hence, the *autocovariance function*  $r(k)$  can be normalized in order to obtain the *autocorrelation function*:

$$\rho(k) = \frac{r(k)}{r(0)} = \frac{r(k)}{\sigma_x^2} \quad \text{Eq. 2.18}$$

By the evaluation of the ACF of a given water quality signal  $X_i$ , it can be estimated how influenced is any value at  $X_n$  by their past observations, indicating how long can be the memory of a given physical process. The ACF, generally, has the characteristic of decaying with the number of lags until the correlation's value is less than the value expected for a Gaussian noise process. The first lag in which the ACF is lower than a Gaussian Noise correlation value is the point in which a present observation  $X_n$  starts to forget its past (adapted from Priestley, 1981).

## 3 MATERIALS AND METHODS

### 3.1 Experimental site

In the following lines the experimental site and the study zone is introduced. This section includes a detailed description of the urban catchment (*Gibraltar* Sanitary Sub-basin) and its effluent (*Estación Elevadora Gibraltar*), place where the *in situ* water quality measurements were undertaken.

#### 3.1.1 Gibraltar Sanitary sub-basin

Río Bogotá's principal sub basins can be listed as follows: (i) *Fucha*, (ii) *El Salitre*, (iii) *Tunjuelo* (principal sub basins), (iv) *Jaboque* and (v) *El Tintal* (secondary sub-basins). The study zone for the present research was a specific sub basin located at the Northwestern area of Bogotá (Colombia), called *Gibraltar*, which constitutes about 90 % of *El Tintal* sub-basin's total area. The total area of *Gibraltar* that has a wastewater and stormwater contribution to the effluent is estimated in about 2226 ha. Regarding year 2007, *Gibraltar* had an afferent population of 423 216 Hab. and a population density of 127 hab/ha (HMV, 2007). Therefore, Figure 3.1 of the urban catchment *Gibraltar* is shown:



Figure 3.1 map of the *Gibraltar*'s Sanitary sub-basin. Source: adapted by author from Google Earth.

It is important to discuss about the type of the basin (separate or combined) that was addressed for the present research. Although the *Gibraltar* sub-basin was designed mainly as a separate system, due to presence of: (i) misconnections and (ii) an inappropriate separation of wastewaters in houses' facilities (roofs and sewage waters), this sewer system can be considered mainly as combined.

From the total area afferent to this catchment, wastewater can be classified as: (i) 83.13 % of domestic activities; (ii) 0.31 % of commercial activities and (iv) 3.18 % of industrial dumps. Hence, 1.03 % of the total area is protected land and 12.35 % is not yet established (HMV, 2007).

In terms of rainfall, three rain gauge stations are located within and in proximity of the studied urban catchment. For illustrative purposes, Figure 3.2 shows a map with the rain gauge stations located in the study zone:

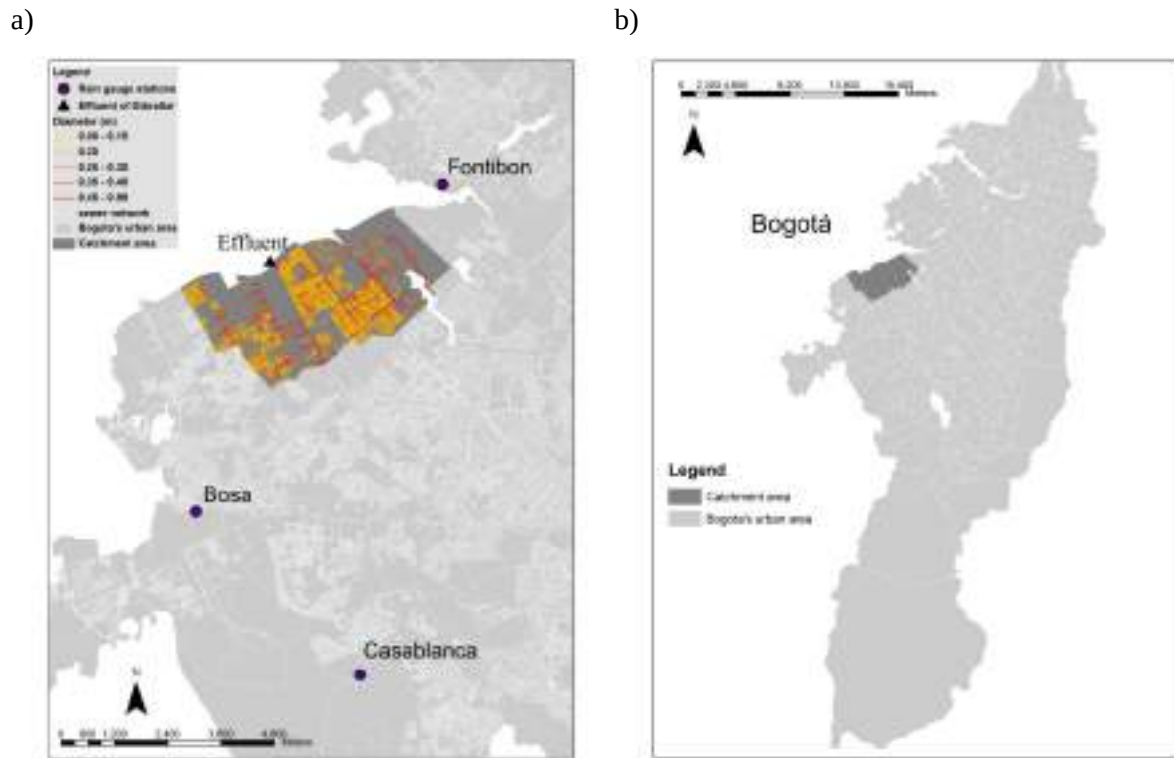


**Figure 3.2 a) Location of the three EAAB rain gauge stations used for rainfall measurements in the catchment; b) The study zone in Bogotá**

The studied catchment can be considered as representative enough for this study due to its heterogeneity in terms of: (i) land uses (residential, commercial and industrial), (ii) important drainage area (>1500 Ha), (iii) sewer system with a part which works as combined and another as separate, (iv) closeness to rain gauge stations with reliable measurements, (v) high variability of pollutants in the effluent of the catchment.

### 3.1.2 Effluent of the catchment: Gibraltar Pumping Station (GPS)

The present research was executed by the use of *in situ* measurements at the effluent of the studied urban catchment (*Gibraltar*). The effluent of this basin consists mainly in a pumping station (*Estación Elevadora Gibraltar*, cord: 105769 N, 88418.7 E), the flow of which is discharged to Bogotá's river, by the use of four Archimedes pumps (with a capacity of 1.27 m<sup>3</sup>/s, each one), see Figure 3.3.



**Figure 3.3 Pumping station *Estación Elevadora Gibraltar***

This pumping station collects wastewater from Sub-basins, 461, 462, 464, 465, 467, 470 and 501. The connectivity among the different Sub-basins is shown (Figure 3.4):



**Figure 3.4 Description of the study zone. Sub-basins 461, 462, 464, 465, 467, 470 and 501, “Gibraltar” urban catchment. Source: Legal Contract PUJ – EAAB No. 9-07-25100-0763-2010.**

It can be appreciated that all UGAs of the sub basin discharge their wastewaters to *Interceptor Cundinamarca*; a principal sewer pipe which is connected to the GPS. The Table 3.1 is a summary table including physical characteristics of each sub-basin.

**Table 3.1 Summary table with physical characteristics of the UGAs of *Gibraltar* urban catchment**

| Sub-basin | Nearest rain gauge station | Hab.    | Users | Pop. density (Hab/km <sup>2</sup> ) | Comm. use (%) | Ind. use (%) | Resid. use (%) | Runoff coefficient "C" | Lag time "k" | Mean length (m) | Area (m <sup>2</sup> ) | Dist. To the nearest rain gauge station (m) | Distance to the effluent (m) | Std. pipes' slope (%) | Mean pipes' slope (%) | Std. pipes' diameter (m) | Mean pipes' diameter (m) | Std. pipes' length (m) | Mean pipes' length (m) | # of pipes | Total pipes' length (m) | Pipes density (m/km <sup>2</sup> ) |
|-----------|----------------------------|---------|-------|-------------------------------------|---------------|--------------|----------------|------------------------|--------------|-----------------|------------------------|---------------------------------------------|------------------------------|-----------------------|-----------------------|--------------------------|--------------------------|------------------------|------------------------|------------|-------------------------|------------------------------------|
| 467       | BOSA                       | 26628.5 | 5455  | 0.03                                | 0.9           | 0            | 0.1            | 0.619                  | 2044         | 4265.7          | 9E+05                  | 4436.886                                    | 1980.746                     | 2.738                 | 0.723                 | 0.174                    | 0.396                    | 28.9                   | 53.841                 | 151        | 8130                    | 171.56                             |
| 501       | BOSA                       | 23802.4 | 4569  | 0.02                                | 0             | 0            | 1              | 0.619                  | 2507         | 5156.1          | 1E+06                  | 4618.212                                    | 1635.881                     | 1.852                 | 0.344                 | 0.102                    | 0.247                    | 27.55                  | 48.601                 | 303        | 14726                   | 232.61                             |
| 470       | BOSA                       | 52336.1 | 10307 | 0.03                                | 0.02          | 0.01         | 0.97           | 0.62                   | 3337         | 6939.1          | 2E+06                  | 4262.636                                    | 1640.802                     | 2.296                 | 0.597                 | 0.122                    | 0.271                    | 26.11                  | 47.177                 | 741        | 34958                   | 367.81                             |
| 461       | FONTIBON                   | 42027.8 | 8289  | 0.03                                | 0.1           | 0.02         | 0.88           | 0.622                  | 2156         | 6090.8          | 1E+06                  | 1621.553                                    | 3150.288                     | 0.693                 | 0.134                 | 0.24                     | 0.528                    | 22.01                  | 56.944                 | 54         | 3075                    | 38.94                              |
| 462       | FONTIBON                   | 46972.6 | 9264  | 0.03                                | 0.1           | 0.02         | 0.88           | 0.619                  | 2603         | 6810.8          | 2E+06                  | 2301.104                                    | 2582.894                     | 1.301                 | 0.207                 | 0.127                    | 0.355                    | 25.2                   | 68.687                 | 147        | 10097                   | 95.05                              |
| 465       | FONTIBON                   | 90666.5 | 16551 | 0.04                                | 0.05          | 0            | 0.95           | 0.62                   | 3800         | 7743.1          | 2E+06                  | 3880.772                                    | 1450.137                     | 2.167                 | 0.499                 | 0.099                    | 0.254                    | 27.75                  | 44.555                 | 1505       | 67056                   | 617.71                             |
| 464       | FONTIBON                   | 68777.9 | 13565 | 0.03                                | 0.1           | 0.02         | 0.88           | 0.419                  | 3605         | 7096.4          | 2E+06                  | 3020.198                                    | 2037.275                     | 0.749                 | 0.403                 | 0.137                    | 0.286                    | 27.82                  | 50.033                 | 995        | 49783                   | 432.39                             |

This table reports several characteristics regarding: (i) the land use, (ii) the sewer pipes network, (iii) the hydraulic characteristics, (iv) the closeness to the effluent and to rainfall rain gauge stations and (v) the population of each UGA. The distance to the effluent and to the nearest rain gauge station are measured from the centroid of the UGA.

### 3.2 Measurement Devices

For the present research several measurement devices were installed at the effluent of the catchment. For the case of flow rate, a NIVUS OCM Pro CF flow rate measurement equipment (Echo Profile Flowmeter) was used. With the purpose of measuring water quality two types of sensors were selected: uni-parametric sensors (infra-red and ion sensitive) and multi-parametric sensors (UV-VIS spectrometry). The WTW manufacturer offers several sensors, which works using different technologies, practical for measuring various water quality parameters. Depending on which water quality parameter is desirable to measure, the technologies used varies from infra-red light (wavelength about 860 nm) to ion sensitive electron technology (see Cammann and Schroeder, 1979). For the case of the UV-VIS spectrometry, the measurement of another group of water quality parameters can be undertaken. Hence, the absorbance of light by different wavelengths, in the analyzed sample, can be registered (spectral analysis). All the measurement devices are capable of registering data with a time-step of one minute. This was the time-step selected for registering the water quality data at the effluent of the catchment.

#### 3.2.1 Flow Rate Measurement

An advanced water/wastewater flow meter from NIVUS UK was used for the present research. The NIVUS OCM Pro CF (incorporating POA Sensor) Echo Profile Flowmeter is an area velocity flow meter which employs the cross correlation principle to deliver highly accurate open channel flow measurement. The NIVUS flow meter meets strict specification for flow monitoring in open channels between 0 m/s to 2 m/s velocity and 0.1 m to 1 m depth. Typical applications for the flow meter are as diverse as abstraction, wastewater treatment, stormwater, industrial wastewater, cooling water, and sluice stages in rivers and irrigation (NIVUS, 2012).

This flow measurement device reports an accuracy of 1 % in its measurements (NIVUS, 2012). The device measures and indicates the flow velocity profile with the spatial allocation of measured single velocities. The NIVUS OCM Pro CF measures in all channels partly filled and full pipes and flumes, including heavily polluted and abrasive media. The reported data can be stored by the use of a Compact Flash card or with an online connection/data transmission via internet. Illustrative figures of the measurement device are presented in Figure 3.5.



**Figure 3.5 Echo Profile Flowmeter device a) installation, b) data transmitter and c) velocity sensors. (Source: [www.schillig.com.ar/nivus\\_fr.htm](http://www.schillig.com.ar/nivus_fr.htm), December 7<sup>th</sup> 2012)**

The installation of the device b) shown in Figure 3.1 (a). The mentioned equipment has a Transmitter in which data is registered and stored (see Figure 3.1 b). The velocity sensors are

installed in the bed of the channel, see (Figure 3.1 a, c). The device is capable of measuring with a temporal resolution of one minute time-step.

### 3.2.2 Water Quality Measurement

Water quality measurements were undertaken by means of different technologies. The sensors implemented for the present research were the uni-parametric sensors and the UV-VIS spectro::lyser sensors manufactured by WTW and s:can, respectively. Some water quality parameters were measured by means of uni-parametric sensors and UV-VIS spectrometry (*e.g.* TSS) simultaneously. On the other hand, parameters like Oxidation Reduction Potential (ORP) and pH could be measured by the use of an ion sensitive technology, given by a uni-parametric sensor. Parameters related to the ultraviolet range of the light spectra such as COD, could be measured by means of the UV-VIS spectrometry sensor. Additional details regarding the measurement devices used for the present research will be described in the following sections.

#### 3.2.2.1 Uni-parametric sensors

For the present research, three uni-parametric sensors were used. The turbidity measurements were undertaken by means of the device Visoturb 700 IQ, capable to measure up to 4000 FNU (for a wavelength of 860 nm, with a measurement angle of 90°). The resolution of measurement is reported as 1 NTU (WTW *on-line*, 2010). A scheme of the device is shown (including the measurement angles and the device structure) in Figure 3.6.

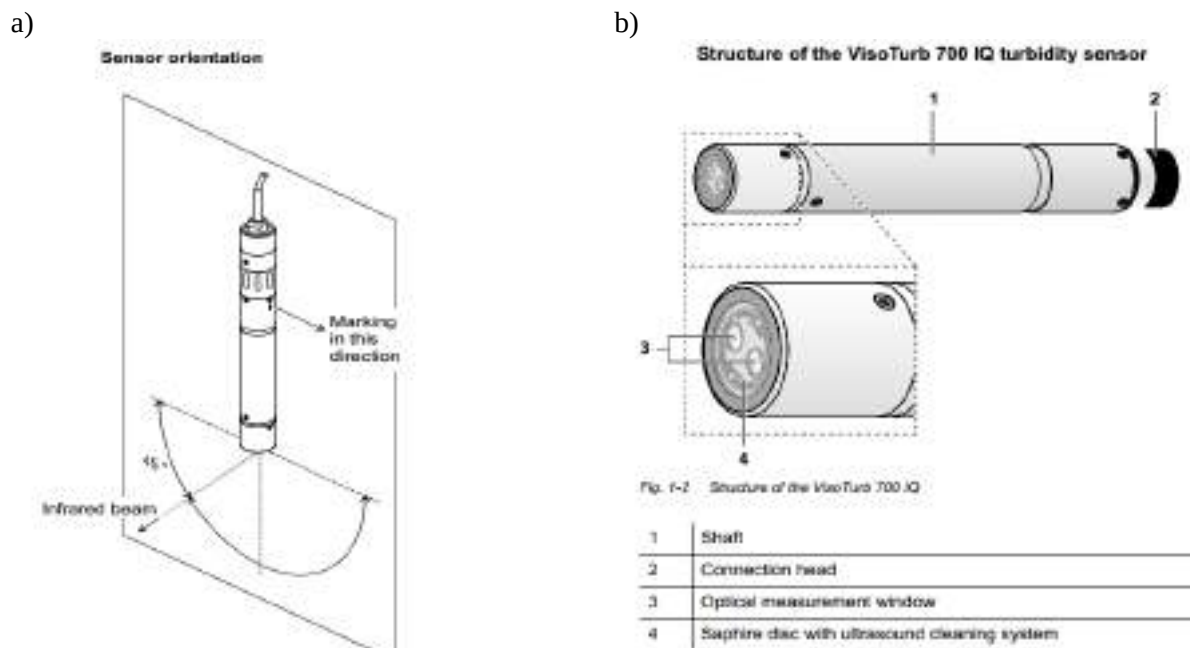
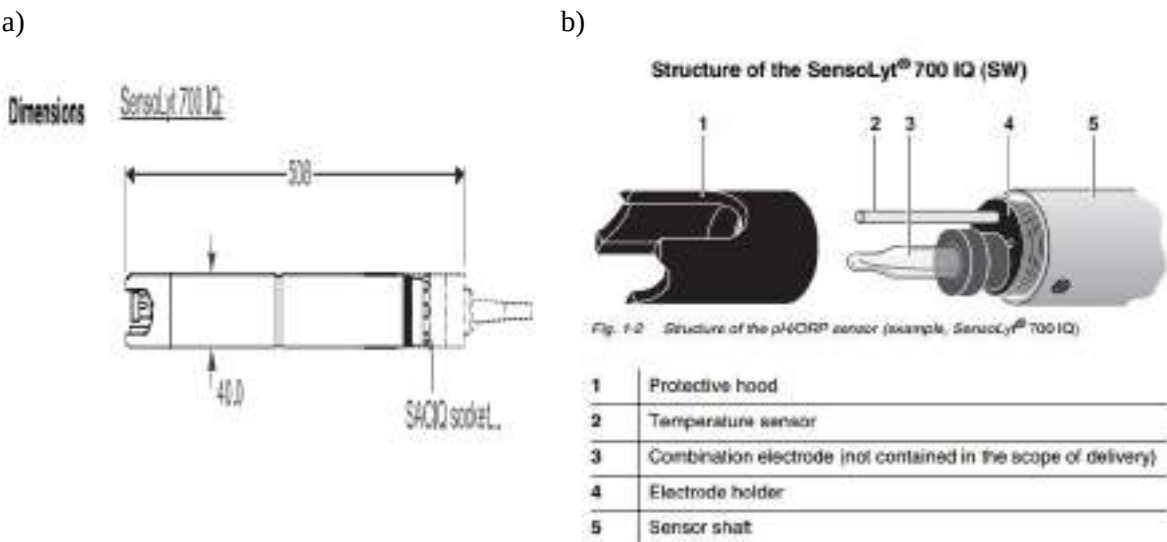


**Figure 3.6** Visoturb 700 IQ sensor uni-parametric; a) measurement's angles and b) parts of the sensor. Source: WTW *on-line*, 2010.

Measurements of Oxidation Reduction Potential (ORP), pH and temperature were undertaken by means of a Sensolyt 700 IQ sensor, which implements an ion sensitive electrode (ISE), based on the potentiometric principle (Cammann and Schroeder, 1979). The measurement range for ORP is from -2000 mV to 2000 mV, with a measurement resolution of 1 mV. For the case of pH, the measurement range is reported from 0.00 pH units to 14.00 pH units, including a measurement

resolution of 0.01 pH units. Regarding the temperature, the measurement range is established from 0° to 60°, with a resolution of 0.01° (WTW *on-line*, 2010). Some schemes of the Sensolyt 700 IQ sensor are shown in Figure 3.7.



This technology works under the principle of ion sensitive. The sensor is composed by a shaft, a connection head and an optical measurement window made of sapphire. The infra-red beam has a range of operation angles as seen in Figure 3.8. The DO was measured by means of the following sensor:

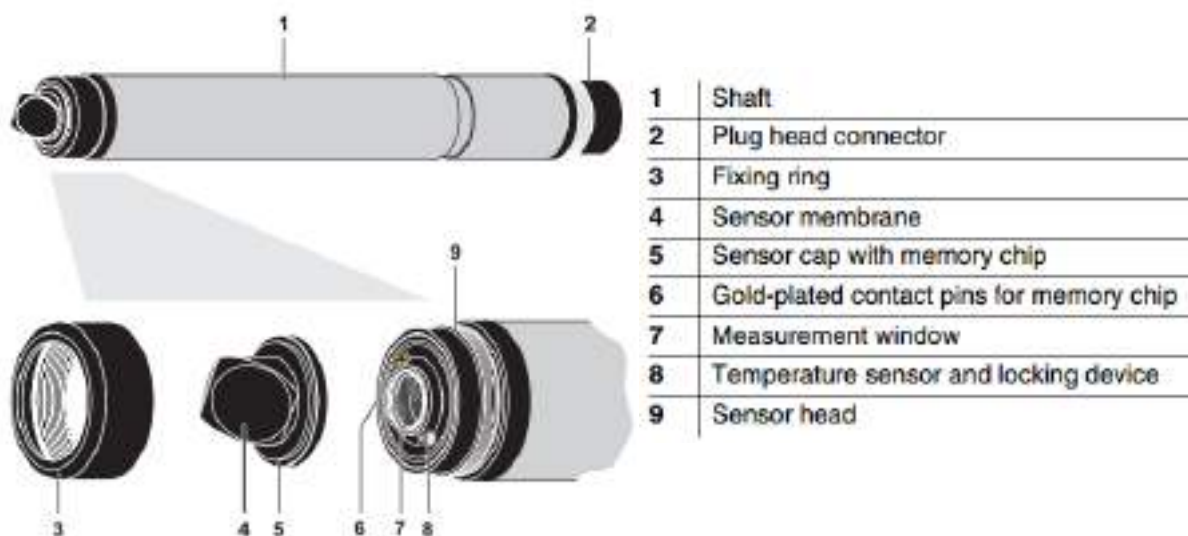


**Figure 3.9 FDO 700 IQ sensor uni-parametric; parts of the sensor. Source: WTW *on-line*, 2010.**

The exposed sensor works under the optical principle: a fluorescent dye is stimulated in the membrane of the FDO® 700 IQ by a short wave length lightsource. By falling back into the passive state, long wave light is emitted, which is recorded as a measurement signal. If oxygen contacts the dye by diffusing through the membrane the period of back scattering light is shortened according to the oxygen concentration of the sample. In principle the measurement of the fluorescent signal come back to a highly precise time measurement (WTW *on-line*, 2010).

Depending on the user's requirements, different sensors can be implemented. With the purpose of registering continuous time series for mentioned water quality parameters, the three sensors were operated simultaneously.

### 3.2.2.2 UV-VIS spectrometry

Several spectrometric techniques are used for the measurement of ions derived from molecules. This identification can lead to identify sample's compounds. The techniques are divided in agreement with the type of radiation used, as follows: (i) atomic absorption spectroscopy (AAS), (ii) infra-red (IR), (iii) images by magnetic resonance (IMR), (iv) mass spectroscopy (MS), (v) nuclear magnetic resonance (NMR), (vi) RAMAN, (vii) ultraviolet (UV), (viii) ultraviolet and visible (UV/Visible) and (ix) x-ray (X). A scheme relating the different types of radiation with different wavelengths is shown in Figure 3.10.

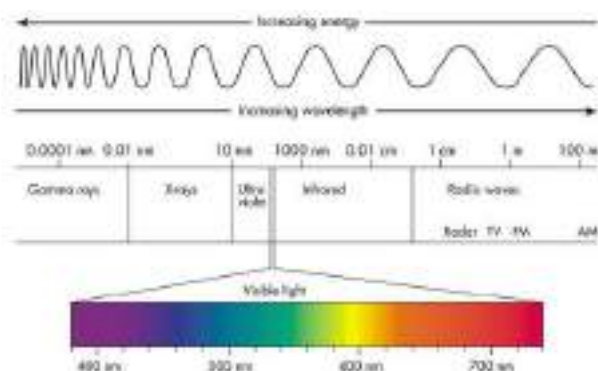


**Figure 3.10** Electromagnetic spectra- frequency and wavelength (Source: [www.s-can.at/parameter.php?langcode](http://www.s-can.at/parameter.php?langcode), December 7th 2012)

For each substance, the absorption rate depends of the molecule's structure and will occur at (a) particular wavelength(s). The absorption of the light depends not only on the concentration of a given substance, but also on the distance that the light should travel. For this reason, in the case of extremely contaminated substances, a shorter measuring window of the sensor is used (s::can, 2007).

In Figure 3.10, approximate wavelength ranges in which the absorbance values correlate with common water quality parameters analyzed in wastewater and stormwater are exposed:

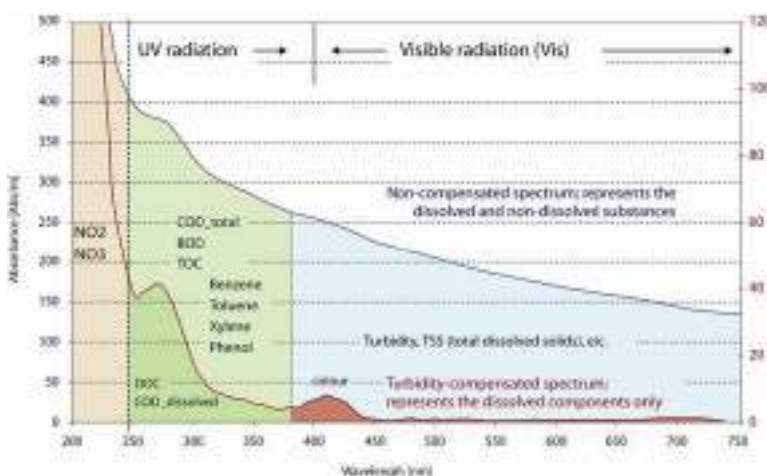


**Figure 3.11** Water quality parameters obtained from spectrometer measurements (Source: [www.s-can.at/text.php?kat=11&id=59&langcode](http://www.s-can.at/text.php?kat=11&id=59&langcode), December 7th 2012)

It is important to clarify that there are compounds which can absorb light from more than one wavelength, for example the COD. The UV-VIS spectrometry sensor can be defined as a submersible spectrometer, capable of measuring on-line the absorbance spectra (UV-VIS) directly from a liquid media (*in situ*). A scheme of the sensor can be proposed in Figure 3.12.

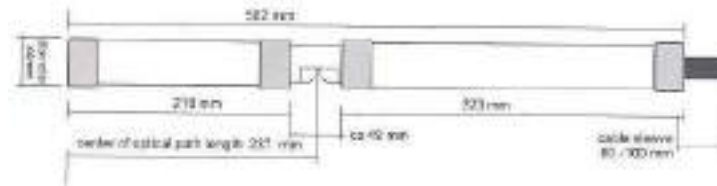


**Figure 3.12 Sensor's dimensions (s::can, 2007)**

By the described procedure, the UV-VIS spectrometry sensor can detect the concentration of a given substance, by measuring the absorbed radiation in certain wavelengths of a ray of light, emitted by the equipment.

The sensor is composed basically by three components: (i) the transmitter, (ii) the measurement cell and (iii) the detection unit (see Figure 3.13). The transmitter's central element is a source of light projected by a xenon lamp (s::can, 2007).



**Figure 3.13 Sensor's components (s::can, 2007)**

Water quality parameters that can be measured by means of the UV-VIS spectrometry sensor are various (*e.g.* BOD, true color, DOC, TOC, COD, TSS and Turbidity), from the whole spectra of 220 wavelengths. For the present research, two water quality parameters were addressed by means of this technology: COD and TSS (equivalent water quality parameters); in addition with an analysis of the complete spectra. This selection was undertaken with the purpose of having a water quality parameter more representative of the ultraviolet range (COD) and another of the visible range (TSS).

### 3.2.3 Rain gauge stations

The three rain gauge stations (*Fontibon, Bosa and Casablanca*) used for this research were installed by EAAB (*Empresa de Acueducto y Alcantarillado de Bogotá*) in the following years: 1955 (*Fontibon*), 1958 (*Bosa*) and 1976 (*Casablanca*). The three rain gauges were manufactured by Lambrecht®. The rain gauges' technical requirements of installation, calibration and data processing were executed by following the recommendations of the Guide to Hydrological Practices (Vol. 1) (WMO, 1996). Regarding the rainfall location, the three rain gauges were located in places with an appropriate exposure, in order to mitigate obstacles or wind's effect over the final measurement. The precipitation is measured by the use of a graduated cylinder, for the three rain gauge stations. The rainfall register by a given rain gauge station is reported in a format in Figure 3.14.

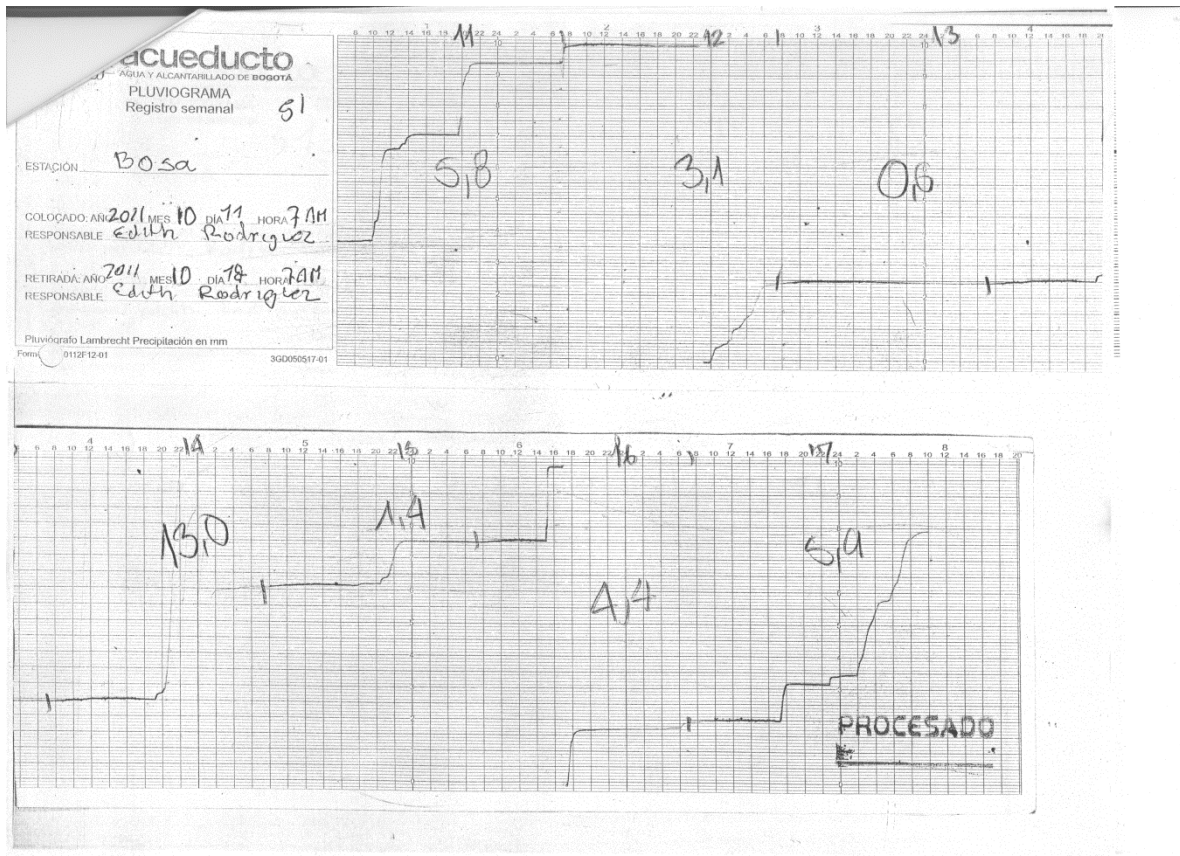


**Figure 3.14 Rainfall raw data registered by the studied rain gauge station**

The raw data is reported in the format exposed in Figure 3.14. The measurement procedure can be described as follows: the rainfall fills a recipient with the course of time and a pen is coupled with a float inside the recipient. Thus, the water level that is filling the recipient is correlated with the lectures reported by the pen, which is writing constantly on the paper. Whenever the pen reaches the maximum level for writing in the paper, the lecture is continued in the lower part. The resolution of the measurements was reported as 0.1 mm, for all of the studied rain gauge stations.

### 3.3 Experimental Framework

The experimental framework was undertaken at the pumping station GPS. In the following paragraphs, a detailed description of the installation of the measurement devices in the pumping station is specified. Technical details of the installation such as the constructive process and the specific location of each of the measurement devices (Echo Profile Flowmeter, uni-parametric, and UV-VIS spectrometry sensors) are described.

#### 3.3.1 Flow Rate Measurement

For the measurement of flow rates in the effluent of the catchment (GPS), a flow measurement device (Echo Profile Flowmeter) was installed with the purpose of storing the flow rate data, reported for each minute. Measurements were undertaken from October 18<sup>th</sup> to November 11<sup>th</sup> 2011. In order to install the measurement device, two structures had to be constructed (structure 1: transmitter safeguard; structure 2: sensors location inside the discharge tunnel). In Figure 3.15 location of the measurement device in the E.E. Gibraltar is shown.



**Figure 3.15 Hydraulic system. a) Upstream and downstream section of the discharge tunnel; b) entry to the discharge tunnel**

The Figure 3.15 presents the hydraulic system in the GPS. The wastewater and stormwater effluent from the studied urban catchment passes through a discharge tunnel. For measuring the flow rate with the online technology proposed, the device was installed inside this discharge tunnel, which is described by the following scheme:

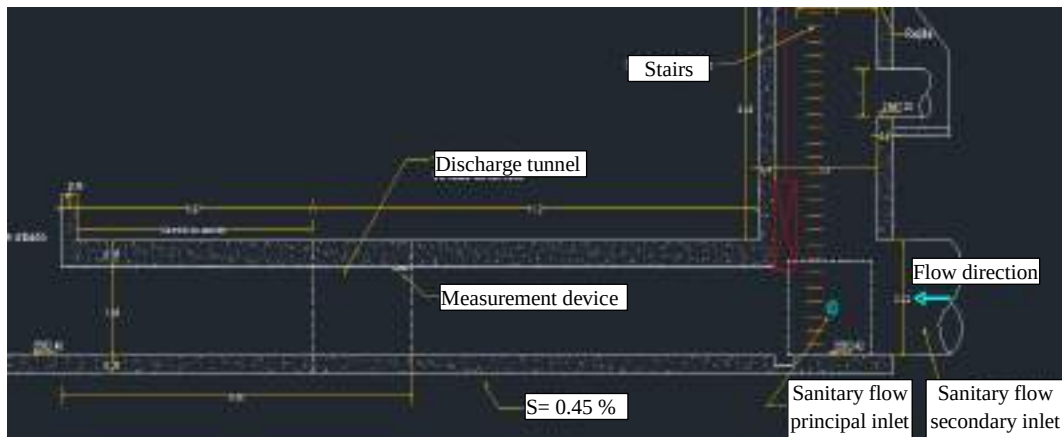


**Figure 3.16 Discharge tunnel**

The Figure 3.16 gives a detailed description of the tunnel and the place where the device was installed. The open gate is located in the upstream direction and the Archimedes pumps are located towards the downstream direction. The first structure built was the transmitter safeguard, which is illustrated in Figure 3.17.

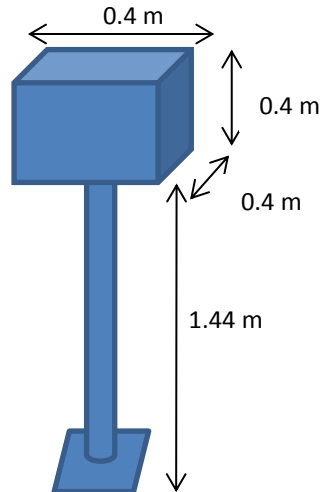


**Figure 3.17 Transmitter safeguard**

The base was fixed to the concrete surface of the floor by the use of two bolts of  $\frac{3}{4}$ "X 3". The transmitter (device where the data is stored) was kept in a metallic box, in order to protect it from outdoor conditions. Photos of this structure are shown in Figure 3.18.



**Figure 3.18 Structure 1: transmitter safeguard in field**

The structure was protected with a plastic cover and cables were tied to the yellow rail. Each three days field inspections were undertaken to check the installation and the operation of the equipment. A scheme of the second structure is shown in Figure 3.19.

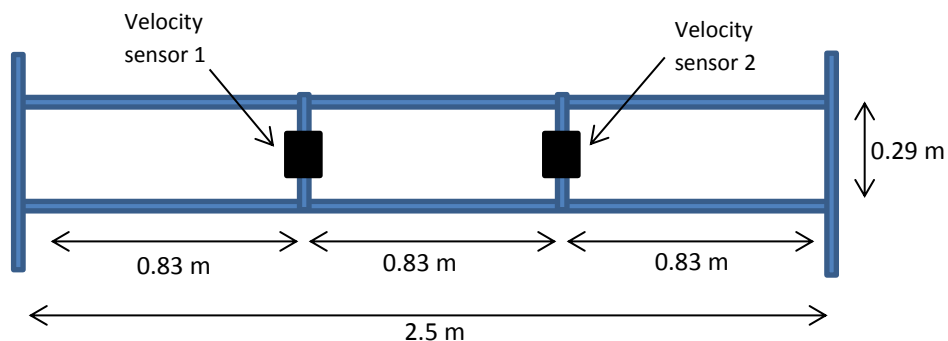


**Figure 3.19 Structure 2 scheme: sensors installation inside the discharge tunnel.**

For installing this structure, it was necessary to completely close the gate located in the upstream section, in order to block the flow through the discharge tunnel. The structure was fixed to the tunnel's ceiling with  $\frac{3}{4}$ "x 3" bolts. The structure could not be fixed to the tunnel's floor due to a remaining water depth of about 60 cms, regardless the gate was closed. The structure was constructed with 2"x  $\frac{3}{16}$ " stainless steel bars. The cables of the sensors were fixed to the wall by the use of plastic staples. The structure was installed 5.80 m inside the tunnel, from the downstream direction, while the gate was completely closed.



**Figure 3.20 Structure 2: sensors installation inside the discharge tunnel.**

The structure was entered and installed inside the discharge tunnel with the sensors already coupled. For this procedure, the working men should use protection masks and security clothing in order to protect themselves appropriately from the abrasive and toxic environment inside the tunnel (wastewater and stormwater discharge).

### **3.3.2 Water Quality Measurement**

The water quality sensors implemented for the measurement of water quality at the effluent of the studied urban catchment (*Gibraltar*), required an additional installation process in terms of location of the devices inside the pumping station and of the structures necessary for protecting and operating the equipment. The uni-parametric and the UV-VIS spectrometry sensors were installed in a stainless steel boat as is shown in Figure 3.21.



**Figure 3.21** Installation boat for infra-red, ion sensitive and UV-VIS spectrometry sensors

This boat was located at the wastewater and stormwater inlet to the pumping station, considering an appropriate distance from the Archimedes pumps in order to affect the water quality online measurements as like as possible. The boat was installed by the use of metal cables tied to fixed supports. The data were registered in a computer sheltered from outdoor conditions by an additional structure, placed in the upstream direction of the inlet of the pumping station.

### 3.4 Data Description and Computational Tools

In this section, a description of the data base obtained from the measurement campaign is given. The data measured at the effluent of the catchment were obtained by the use of the measurement technologies described in section 3.2. Hence, the process for obtaining the rainfall data registered by rain gauge stations which were located in section 3.1.1 is also explained. In addition, computational tools used for the present research are mentioned in this section. These tools were mainly based on R Project Statistical Software.

#### 3.4.1 Data Description

By the use of the Echo Profile Flowmeter equipment (section 3.2.1), the flow rate was measured at the effluent of the catchment. In terms of water quality, two technologies were used: (i) uni-parametric and (ii) UV-VIS spectrometry sensors. For the uni-parametric sensors, the following water quality parameters were measured: (i) Dissolved Oxygen, (ii) Turbidity, (iii) Total Suspended Solids, (iv) Temperature, (v) pH and (vi) ORP. Regarding the UV-VIS spectrometry sensors, two main aspects were registered: (i) the spectra produced by wavelengths from 200 nm to 750 nm (UV-VIS range) and (ii) the concentrations of TSS and COD reported by the global calibration.

Rainfall raw data were registered by three rain gauge stations within or in proximities of the studied catchment (see Figure 3.2), by the use of pluviographs. This information was processed with the *GetData Digitalizer* software, obtaining a rainfall time series for each rain gauge station with a

temporal resolution of one minute. Data obtained from the mentioned water quality and rainfall measurements was compiled in a data base (see details in Appendix 8.1).

A multivariate time series of different water quality parameters and rainfall registered at different rain gauge stations was achieved. This time series had a total length of about 34000 minutes corresponding to 25 days of measurements (October 18<sup>th</sup> to November 11<sup>th</sup> of 2011), for 229 variables (including each wavelength of the spectra and the TSS and COD series of the UV-VIS spectrometry, in addition to the uni-parametric sensors' registers). For illustrative purposes, a specific part of the time series, from 1h52 to 2h18 in October 19<sup>th</sup> is shown in Table 3.2.

**Table 3.2 Data base of the measurements at the effluent of the catchment**

| Time       |          | Measurements on the effluent of the catchment (uniparametric) |            |           |      |          |                             |          | rainfall (mm) |            |              |
|------------|----------|---------------------------------------------------------------|------------|-----------|------|----------|-----------------------------|----------|---------------|------------|--------------|
| day        | hour     | flowrate (m <sup>3</sup> /s)                                  | Turb (NTU) | DO (mg/l) | pH   | ORP (mV) | TSS (mg/l) SiO <sub>2</sub> | Temp (C) | p_fontibon    | p_bosa     | p_casablanca |
| 19.10.2011 | 01:52:00 | 1.748                                                         | 450.3      | 1.425     | 7.53 | -86.97   | 1968.5                      | 17.38    | 0.002452002   | 0.00100252 | 0.000136532  |
| 19.10.2011 | 01:53:00 | 1.774                                                         | 449.4      | 1.697     | 7.52 | -80.69   | 1970.4                      | 17.37    | 0.002452002   | 0.00100252 | 0.000136532  |
| 19.10.2011 | 01:54:00 | 1.766                                                         | 443.6      | 1.933     | 7.51 | -76.46   | 1981                        | 17.37    | 0.002452002   | 0.00100252 | 0.000136532  |
| 19.10.2011 | 01:55:00 | 1.712                                                         | 442.4      | 2.066     | 7.53 | -75.05   | 1988.8                      | 17.4     | 0.002452002   | 0.00100252 | 0.000136532  |
| 19.10.2011 | 01:56:00 | 1.627                                                         | 444.1      | 2.077     | 7.53 | -74.6    | 1990.2                      | 17.37    | 0.002452002   | 0.00100252 | 0.000136532  |
| 19.10.2011 | 01:57:00 | 1.609                                                         | 443.8      | 1.984     | 7.53 | -75.56   | 1979.5                      | 17.38    | 0.002452002   | 0.00100252 | 0.000136532  |
| 19.10.2011 | 01:58:00 | 1.636                                                         | 447.6      | 1.821     | 7.53 | -78.45   | 2011.2                      | 17.37    | 0.002452002   | 0.00100252 | 0.000136532  |
| 19.10.2011 | 01:59:00 | 1.653                                                         | 453.4      | 1.654     | 7.53 | -79.35   | 2036.7                      | 17.37    | 0.002452002   | 0.00100252 | 0.000136532  |
| 19.10.2011 | 02:00:00 | 1.634                                                         | 460.7      | 1.513     | 7.54 | -79.54   | 2054.1                      | 17.37    | 0.002452002   | 0.00100252 | 0.000136532  |
| 19.10.2011 | 02:01:00 | 1.613                                                         | 466.1      | 1.427     | 7.54 | -80.12   | 2094.4                      | 17.37    | 0.002452002   | 0.00100252 | 0.000136532  |
| 19.10.2011 | 02:02:00 | 1.615                                                         | 479.5      | 1.392     | 7.54 | -80.63   | 2129.5                      | 17.37    | 0.002452002   | 0.00100252 | 0.000136532  |
| 19.10.2011 | 02:03:00 | 1.610                                                         | 487        | 1.4       | 7.56 | -80.12   | 2138.5                      | 17.37    | 0.002452002   | 0.00100252 | 0.000136532  |
| 19.10.2011 | 02:04:00 | 1.585                                                         | 491.3      | 1.433     | 7.55 | -79.35   | 2141.3                      | 17.37    | 0.002452002   | 0.00100252 | 0.000136532  |
| 19.10.2011 | 02:05:00 | 1.605                                                         | 485.8      | 1.461     | 7.54 | -78.64   | 2376.5                      | 17.38    | 0.002452002   | 0.00100252 | 0.000136532  |
| 19.10.2011 | 02:06:00 | 1.665                                                         | 476.8      | 1.452     | 7.53 | -79.99   | 2112.5                      | 17.37    | 0.002452002   | 0.00100252 | 0.000136532  |
| 19.10.2011 | 02:07:00 | 1.694                                                         | 468.2      | 1.4       | 7.52 | -79.6    | 2068.9                      | 17.37    | 0.002452002   | 0.00100252 | 0.000136532  |
| 19.10.2011 | 02:08:00 | 1.684                                                         | 458.3      | 1.298     | 7.5  | -81.01   | 2040.6                      | 17.37    | 0.002452002   | 0.00100252 | 0.000136532  |
| 19.10.2011 | 02:09:00 | 1.650                                                         | 451.8      | 1.18      | 7.5  | -81.2    | 2025.4                      | 17.37    | 0.002452002   | 0.00100252 | 0.000136532  |
| 19.10.2011 | 02:10:00 | 1.604                                                         | 446.1      | 1.08      | 7.49 | -81.14   | 2017.2                      | 17.37    | 0.002452002   | 0.00100252 | 0.000136532  |
| 19.10.2011 | 02:11:00 | 1.583                                                         | 442.2      | 0.999     | 7.48 | -81.33   | 2014.5                      | 17.37    | 0.002452002   | 0.00100252 | 0.000136532  |
| 19.10.2011 | 02:12:00 | 1.579                                                         | 442.4      | 0.946     | 7.47 | -82.87   | 2014                        | 17.38    | 0.002452002   | 0.00100252 | 0.000136532  |
| 19.10.2011 | 02:13:00 | 1.638                                                         | 442.4      | 0.911     | 7.49 | -84.02   | 2015.1                      | 17.39    | 0.002452002   | 0.00100252 | 0.000136532  |
| 19.10.2011 | 02:14:00 | 1.662                                                         | 441.3      | 0.894     | 7.5  | -84.6    | 2018.1                      | 17.39    | 0.002452002   | 0.00072996 | 0.000136532  |
| 19.10.2011 | 02:15:00 | 1.620                                                         | 440.6      | 0.895     | 7.53 | -86.72   | 2004.6                      | 17.39    | 0.002452002   | 0.00067784 | 0.000136532  |
| 19.10.2011 | 02:16:00 | 1.553                                                         | 438.7      | 0.898     | 7.54 | -87.1    | 2003.3                      | 17.38    | 0.002452002   | 0.00067784 | 0.000136532  |
| 19.10.2011 | 02:17:00 | 1.520                                                         | 432.7      | 0.887     | 7.54 | -88.25   | 1995.5                      | 17.39    | 0.002452002   | 0.00067784 | 0.000136532  |
| 19.10.2011 | 02:18:00 | 1.524                                                         | 428.5      | 0.859     | 7.54 | -88.77   | 1977                        | 17.39    | 0.002452002   | 0.00067784 | 0.000136532  |

This data base had different missing values (“NA”), for each water quality parameter. Those unsuccessful registrations by the measurement devices can be explained by obstructions in the equipment and particularly, for flow rate, whenever the discharge tunnel was not under a full pipe condition (about 20 % of the total measurement time). The uncertainty in the reported data was assessed as the measurement’s resolution of the devices used (see details in section 3.2). A summary of the rainfall data registered at the three studied rain gauge stations is given in Table 3.3.

**Table 3.3 Daily rainfall registered at the *Fontibon*, *Bosa* and *Casablanca* rain gauge stations**

| Date       | Daily rainfall (mm) |             |                   |
|------------|---------------------|-------------|-------------------|
|            | <i>Fontibon</i>     | <i>Bosa</i> | <i>Casablanca</i> |
| 18.10.2011 | 0.3                 | 2.6         | 0.9               |
| 19.10.2011 | 6.7                 | 10.8        | 6.5               |
| 20.10.2011 | 3.4                 | 6.9         | 13.0              |
| 21.10.2011 | 0.2                 | 0.1         | 0.0               |
| 22.10.2011 | 0.4                 | 0.3         | 0.3               |
| 23.10.2011 | 3.3                 | 0.0         | 0.1               |
| 24.10.2011 | 29.0                | 20.3        | 15.4              |
| 25.10.2011 | 0.2                 | 3.4         | 18.2              |
| 26.10.2011 | 0.0                 | 0.0         | 0.0               |
| 27.10.2011 | 1.2                 | 0.0         | 0.0               |
| 28.10.2011 | 0.4                 | 0.0         | 0.2               |
| 29.10.2011 | 0.0                 | 0.0         | 0.0               |
| 30.10.2011 | 1.5                 | 0.4         | 0.4               |
| 31.10.2011 | 0.0                 | 0.0         | 0.0               |
| 01.11.2011 | 0.1                 | 0.0         | 0.0               |
| 02.11.2011 | 0.0                 | 0.0         | 0.0               |
| 03.11.2011 | 0.0                 | 0.0         | 0.0               |
| 04.11.2011 | 0.0                 | 0.0         | 0.0               |
| 05.11.2011 | 0.4                 | 4.1         | 8.9               |
| 06.11.2011 | 0.1                 | 0.1         | 0.1               |
| 07.11.2011 | 0.0                 | 0.0         | 0.0               |
| 08.11.2011 | 12.1                | 10.5        | 1.5               |
| 09.11.2011 | 6.1                 | 10.3        | 24.8              |
| 10.11.2011 | 19.2                | 12.4        | 23.5              |
| 11.11.2011 | 0.1                 | 0.0         | 0.1               |

The following table exposes the daily rainfall registered in each of the studied rain gauge stations. A completely dry time was registered just for six days. Therefore, the dry weather characterization needed to be executed considering the registers in each rain gauge station independently, classifying days with a total rainfall height lower or equal to 0.1 mm as dry days (reporting a total of 11 dry days for *Fontibon*, 14 for *Bosa* and 13 for *Casablanca*).

### 3.4.2 Computational Tools

A brief description of the used computational tools is provided. The first section contains a description of the Software implemented for the proposed methodology (R-Project). In the following chapters, a description of the libraries and the functions used for the proposed implementation is given.

#### 3.4.2.1 R-Project

R is a programming language for graphic and statistical analyses (linear and non-linear modeling, statistical tests, time series analysis, classification, etc.) (R Development Core Team, 2012). This software is for free distribution under the requirements of Free Software Foundation's GNU General PublicLicence in source code format. It can be compiled in the UNIX, Windows and MacOS platforms (R Development Core Team, 2012).

### 3.4.2.2 R libraries and functions (R Development Core Team, 2012).

In the following lines, a description of the libraries and functions implemented for the proposed methods is given.

`library(gplots)`: Functions for plotting graphical shapes such as ellipses, circles, cylinders, arrows, *etc* (Warnes *et al.*, 2012).

`barplot2()`: An enhancement of the standard `barplot()` function. Creates a bar plot with vertical or horizontal bars. Can plot confidence intervals for each bar, a lined grid behind the bars, change plot area color and logarithmic axes may be used.

`library(fBasics)`: The R metrics "fBasics" package is a collection of functions to explore and to investigate basic properties of financial returns and related quantities. The covered fields include techniques of explorative data analysis and the investigation of distributional properties, including parameter estimation and hypothesis testing. Even more, there are several utility functions for data handling and management (Wuertz *et al.*, 2012).

`library(entropy)`: `entropy` estimates the Shannon entropy  $H$  of the random variable  $Y$  from the corresponding observed counts  $y$  (Hausser & Strimmer, 2012).

`library(pls)`: Functions to perform partial least squares regression (PLSR), canonical powered partial least squares (CPPLS) or principal component regression (PCR), with a formula interface. Cross-validation can be used. Prediction, model extraction, plot, print and summary methods exist (Mevik *et al.*, 2012).

`pls()`: The functions fit PLSR, CPPLS or PCR models with 1, ..., `ncomp` number of components. Multi-response models are fully supported. The type of model to fit is specified with the `method` argument. Four PLSR algorithms are available: the kernel algorithm ("`kernelpls`"), the wide kernel algorithm ("`widekernelpls`"), SIMPLS ("`simpls`") and the classical orthogonal scores algorithm ("`oscorespls`"). One CPPLS algorithm is available ("`cppls`") providing several extensions to PLS. One PCR algorithm is available: using the singular value decomposition ("`svdpc`"). If `method` is "`model.frame`", the model frame is returned. The functions `pcr`, `pls` and `cppls` are wrappers for `mvr`, with different values for `method`.

`library(ade4)`: Multivariate data analysis and graphical display (Dray & Dufour, 2007).

`dudi.pca()`: performs a principal component analysis of a data frame and returns the results as objects of class `pca` and `dudi`.

`library(cpm)`: An implementation of several different change point models (CPMs) for performing both parametric and nonparametric change detection on univariate data streams (Gordon & Ross, 2007).

`processStream()`: This function is used to detect a multiple change points in a sequence of observations using the Change Point Model (CPM) framework for sequential (Phase II) change detection. The observations are processed in order, starting with the first, and a

decision is made after each observation whether a change point has occurred. Unlike the `detectChange` function, `processStream` does not terminate and return when a change point is encountered. Instead, a new CPM is initialized immediately following the change point, with all previous observations being discarded. The monitoring then continues, starting from the first observation after the change point. If more change points are discovered later in the sequence, the CPM is again reinitialized after each one. In this way, the whole sequence of observations will be processed and multiple change points may be detected.

`library(zoo)`: `zoo` is the creator for an S3 class of indexed totally ordered observations which includes irregular time series. `Zoo` provides infrastructure for ordered observations which are stored internally in a vector or matrix with an index attribute (of arbitrary class, see below). The index must have the same length as `NROW(x)` except in the case of a zero length numeric vector in which case the index length can be any length. Emphasis has been given to make all methods independent of the index/time class (given in `order.by`). In principle, the data `x` could also be arbitrary, but currently there is only support for vectors and matrices and partial support for factors. `Zoo` is particularly aimed at irregular time series of numeric vectors/matrices, but it also supports regular time series (i.e., series with a certain frequency). `Zoo`'s key design goals are independence of a particular index/date/time class and consistency with `ts` and base R by providing methods to standard generics. Therefore, standard functions can be used to work with "zoo" objects and memorization of new commands is reduced (Zeileis & Grothendieck, 2005).

`na.approx`: Generic functions for replacing each NA with interpolated values.

### 3.5 Developed Methods

A methodology for identifying time windows in which water quality parameters' output signals are influenced by rainfall is proposed, based on a typical dry weather behavior characterization at the effluent of the catchment. A Principal Components Analysis (PCA) is proposed regarding a preliminary insight of which water quality and quantity parameters seem to explain mostly the differences between the variability of dry and wet weather data. Hence, considering the most explicative water quality and quantity parameters from PCA, the designed methodology is then applied to the time series registered during rainy days. Hence, rainfall pulses characteristics are proposed to be correlated with rainfall events characteristics (in terms of trend and variability), with the purpose of detecting recurrent relations between rainfall and water quality characteristics, for a given set of rainfall events. During a rainfall event, water quality data influenced by rainfall can be considered the output signal, as well as the rainfall can be understood as the input signal of the system, under a causality hypothesis.

#### 3.5.1 Dry weather characterization and rainy weather identification

In order to define a rainfall event, identifying null precipitation periods, the beginning of the event, the end of the event, and the duration of null precipitation period that separates two independent rainfall events becomes necessary. In general terms, two rainfall events can be considered mutually independent if the resulting effects of the first event have ceased before the beginning of the second one (adapted from Chocat, 1997).

### 3.5.1.1 Differences between dry and rainy days in terms of water quality parameters

The first part is proposed with the aim of identifying which water quality parameters seem to be more influenced by rainfall registered at a given rain gauge station and its typical behavior during a dry day, by the use of a Principal Component Analysis (PCA). This tool seeks to explain the differences between each multivariate time-step by separating the sample into orthogonal principal components composed by linear combinations of the variables. About 75 % of the sample's variability must be collected with the selected components in order to have a representative PCA analysis. The observations contained in the data base can be grouped by hypotheses of separation with the purpose of identifying generalizable, instead of particular differences. The data base can be separated into two groups based on the rainfall registered during each of the analyzed days of the measuring campaign: (i) all time-steps of multivariate data of a day in which no rainfall was registered at a given rain gauge station as "DIAS\_SECO"; (ii) otherwise as "DIAS\_LLUVIA". An additional fact should be considered: the "DIAS\_SECO" group might not include data of days in which some rainfall was registered a short enough time before (less than the response time of the system) the midnight. In those cases the data of mentioned days can be better classified as "DIAS\_LLUVIA". With an appropriate separation between the group 1 and 2 plotted in the principal components domain, it can be verified which independent variables (water quality parameters) better explain the differences between rainy and totally dry days data. As an example the Figure 3.22 is shown.

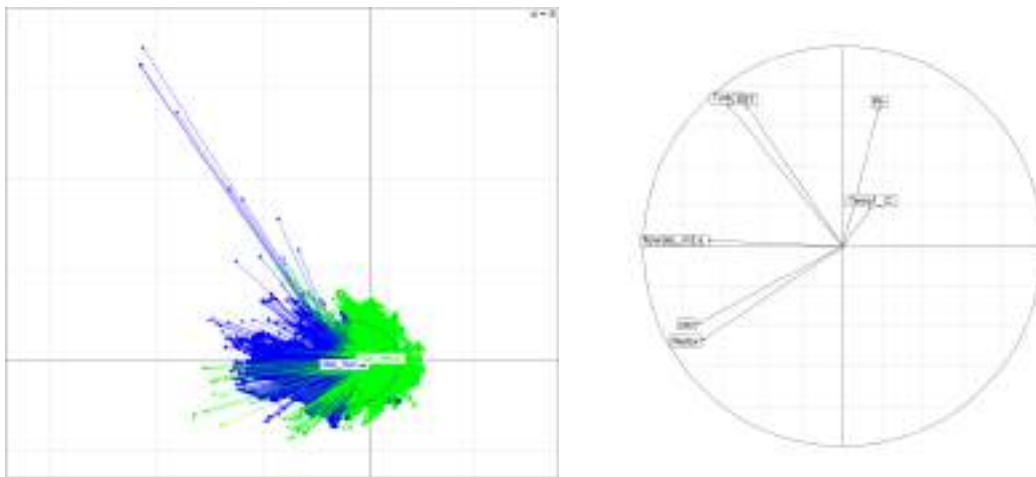


**Figure 3.22 Example of PCA execution, a) separation of groups by rainfall registered at a given rain gauge station; b) variables correlated with the principal component 1 or 2 (x and y axis).**

The interpretation of the PCA is based on two aspects: (i) a separation between groups of data and (ii) the variables that constitute each of the principal components that separates the groups. Regarding the example proposed in Figure 3.22, a separation between the dry and rainy weather data (blue and green groups in Figure 3.22 a) could be established by means of the principal component 1 (x axis). On the other hand, the vectors graph (Figure 3.22 b) shows the importance of each variable over the studied principal components (1 and 2, for this case). Hence, the projection of each vector over the x axis represents the importance that each variable has over the separation between the dry and wet weather data (given by principal component 1). Therefore, it can be detected that the DO variable has the most important incidence in the separation of rainy and dry days. However, Flow rate, TSS, Turbidity, and ORP has also influence in the explanation of why

multivariate data registered at rainy days is different from multivariate data reported during dry days.

By the use of this preliminary identification, a more detail analyses of the relationship between rainstorm and most relevant variables in the output (detected from the PCA), can be performed. Therefore, it is proposed in order to identify, for the rainy days (group 2 established by the PCA), the specific time-steps that are influenced by rainfall pulses over the output signal that were previously established as influenced by rainfall. The definition of a rainstorm event must be kept in mind: the beginning and the ending are attached to the behavior of the analyzed water quality parameter; the influence of rainstorm will always stop when this parameter returns to its dry weather behavior. From this definition arises the importance of characterizing the behavior of influenced-water quality parameters during a typical dry day.

### 3.5.1.2 Dry weather probabilistic characterization

With the purpose of assessing the dry weather typical behavior, all signals of output variables registered during dry days (group 1, PCA analysis), are superposed for generating an envelope in which, for each time-step of a typical dry day, multiple realizations are registered (about 12 realizations in each time-step due to the registration of about 12 completely dry days, for the case study). This curve can be useful to evaluate, for a given rainy day, how far a value of any water quality parameter, registered at a specific time-step, is from its typical dry weather behavior.

With the purpose of performing an appropriate probabilistic characterization of each time-step for a typical dry weather day as a stochastic process with scarce realizations, a probabilistic density function PDF estimation technique can be used (Kernel Density Estimators). The objective is to estimate the PDF for each time-step, by the use of the mentioned 12 realizations, during a typical dry day. In this sense, a figure can be constructed for representing the typical dry weather behavior (see Figure 3.23).

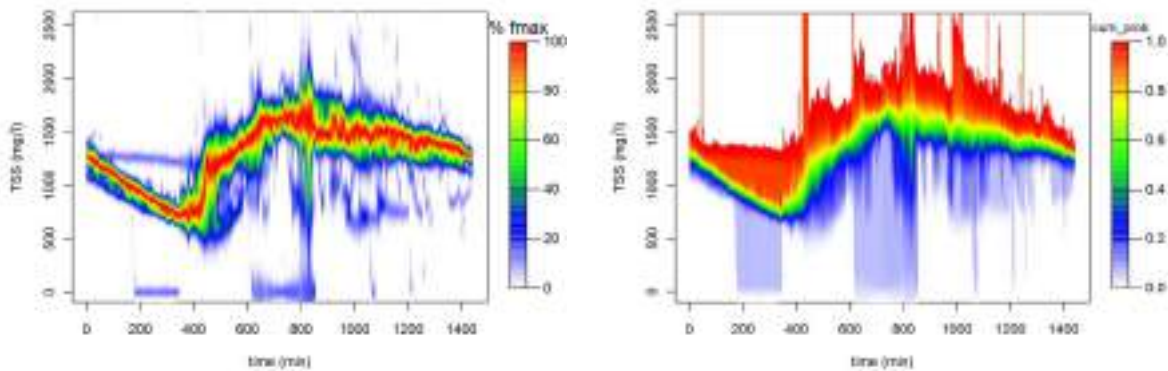


**Figure 3.23 Example of the probabilistic behavior of TSS for a typical dry weather day.**

By the use of KDE and data classified as group 1 for the PCA analysis (dry weather data), the PDF and the CDF at any given time-step  $t_i$ , with  $t_i$  belonging to  $[1,1440]_{\min}$  (length of a day), can be estimated (Figure 3.23). These curves are useful for calculating, at a given time-step, the probability of being part of the typical dry weather behavior, for any given TSS value. This computation can be executed as follows:

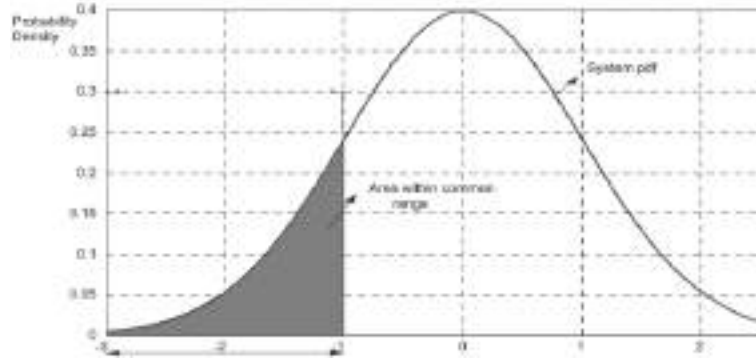


**Figure 3.24 Estimation of a probability  $x$ , by the use of the PDF for the random variable  $X$ .**

For a random TSS value  $x$ , the probability that this value is “near” the typical weather behavior will be expressed by the grey area multiplied by two (due to condition of a two tailed PDF). It is important to remember that the exposed PDF curve at Figure 3.24 will be different for any given time-step  $t_i$ , ranging from 1 to 1440 (length of the typical dry weather day).

### ***3.5.1.3 Identification of water quality’s time windows ranging outside of typical dry weather behavior***

Once the probability density function for  $t_i$  is estimated, the probability that a water quality data  $X_i(t)$  registered at time  $t$ , is part of the typical weather behavior can be computed. Therefore, all days of the time series which were classified into group 2 in PCA analysis (rainy days data) are evaluated with the purpose of detecting which time windows of the total rainy days’ time series have a low enough probability of being part of typical dry weather behavior. As a statistical proposal, each data part of a rainy day, with a probability equal or lower than 5 % of being part of dry weather behavior can be considered part of a rainstorm event.

Due to the phenomena’s natural variability and possible outlier’s presence, data with a high temporal closeness to rainfall pulses can report a high probability of membership to the dry weather envelope. Furthermore, data very far from rainfall pulses can report a low probability of membership to the dry weather envelope. In this sense, the exposed indicator presents discontinuities that should be treated. See for example the results in Figure 3.25.

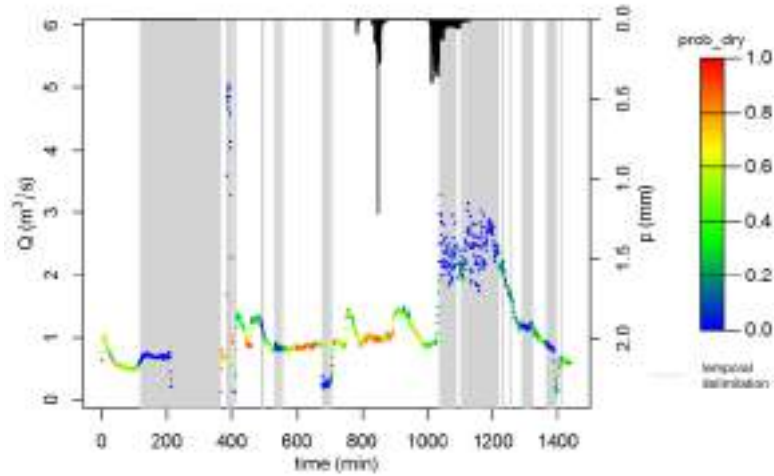


**Figure 3.25 Evaluation of each time-step's probability of being part of typical dry weather behavior for a rainy day signal, highlighting probabilities of dry weather behavior below 5 %.**

With the purpose of dealing with discontinuities given by the above reasons, average behavior of the probability evolution over time, for a rainy day signal, should be assessed. Whenever the output signal  $X_i$  is ranging outside of its typical dry weather behavior during a long enough time period, it was considered that rainfall had an influence over  $X_i$ . In order to treat these noisy discontinuities by estimating in which time windows the probability is in average  $< 5\%$ , a Median Filtering technique is proposed over the probability of being part of dry weather signal  $P_{dry}(t)$ .

However, some important questions arise from the definition proposed above. How many neighbor values of the evaluated time-step should be included in the mobile window for filtering  $P_{dry}(t)$ ? Values from the past, future or from both should be included? The answer to these queries is not easy. The author proposes for the present study to support the selection of a time window based on the autocorrelation function (ACF) concept.

By the evaluation of the ACF of a given water quality signal  $X_i$ , it can be estimated how any value at  $X_i(t)$  is influenced by their past observations, indicating how long the memory of the physical process can be. As an example, a water quality signals' *autocorrelation function* is in Figure 3.26.

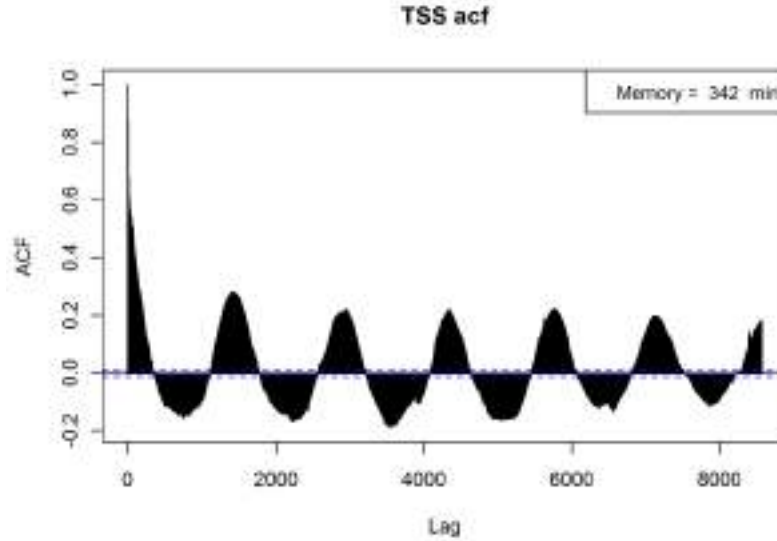


**Figure 3.26 TSS measured by uni-parametric sensors autocorrelation function for the whole measurement campaign**

Analyzing this figure, it can be stated that the TSS water quality parameter exhibits a daily periodical behavior. This fact can be an indication that the typical weather behavior of TSS should not expose a dramatic variability. Hence, the dashed line shows the maximum correlation value that the signal should have for not being considered as Gaussian Noise (significant correlation). It can be noticed that the correlation is still strong for the first lags and it starts to decay for longer lag values. The first lag in which the ACF is lower than the dashed line (correlation value for Gaussian Noise) is the point in which a present observation  $X_i(t)$  starts to forget its past. For this case, it can be reported to be approximately 342 minutes.

As a first approach, this value can be used as the length of the mobile time window in the Median Filter, for evaluating the probability of being part of dry weather time series  $P_{dry}(t)$ . It is important to clarify that the ACF criteria are proposed over the water quality signal  $X_i(t)$ , instead of over  $P_{dry}(t)$ . This decision had the purpose of achieving a more physically-based support to the proposed methodology, based in the memory of the physical process rather than on the memory of a set of evaluated probabilities (given by  $P_{dry}(t)$ ).

This new filtered signal  $P_{dry}(t)_{filt}$ , obtained by the use of a time window length estimated from the ACF of the  $X_i(t)$  signal, will expose more clearly abrupt changes in time and groups of high or low values ( $P_{dry}(t)_{filt} < 5\%$ ), by removing also isolated registered values. These advantages are linked more to the median filtering technique rather than to the traditional moving average approach. In order to explain this statement, differences between the expected results from a Moving Average Filtering and a Median Filtering are given in Figure 3.27.

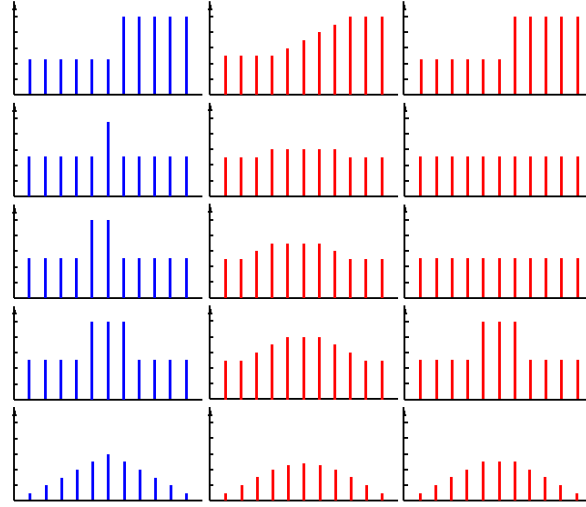


**Figure 3.27 Comparison between: a) original signal b) Moving Average Filter c) Median Filter.**

It can be appreciated that by the use of median filtering the original signal is not smoothed and abrupt changes are highlighted. Hence, isolated values (possibly outliers) will be removed from the median filtered signal, without affecting the signal values as in a moving average filter.

Therefore, it can be detected in which part of the signal  $P_{dry}(t)_{filt}$  abrupt changes occurred, by means of a Step Detection technique. This technique will lead to identify when the probability of dry weather behavior changes noticeably. Therefore, it can be analyzed how are the  $P_{dry}(a < t < b)_{filt}$  values between a couple of Step Detection points  $a$  and  $b$ . Regarding the above proposals, whenever the mean value  $\overline{P_{dry}(a < t < b)_{filt}} < 0.05$  (over the time window  $[a, b]$ ), it can be considered that the water quality signal  $X_i(t)$  was “far” enough from the typical dry weather behavior during the time window  $[a, b]$ . This fact makes the time window  $[a, b]$  a continuous gap in which the rainfall seemed to have an influence over the water quality output signal  $X_i(t)$ , at the effluent of the catchment. With the purpose of illustrating the mentioned procedure, the Figure 3.28 is shown:

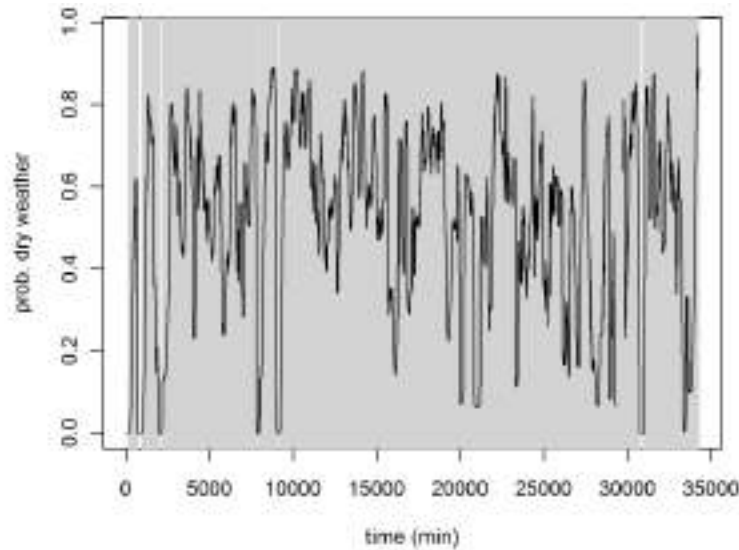


**Figure 3.28 Identification of time windows  $[a, b]_j$  in which  $\overline{P_{dry}(a_j < t < b_j)_{filt}} < 0.05$ ; where  $a, b$  are change points obtained by the Step Detection calculations over the signal  $P_{dry}(t)_{filt}$ .**

From this figure, it can be concluded that among the group of identified step detection points, just for six time windows  $[a, b]_k$  (where  $k = 1, \dots, 6$  are the number of windows and  $a, b$  are the change detection points), the  $\overline{P_{dry}(a_k < t < b_k)_{filt}}$  was estimated to be below 0.05 (white stripes in Figure 3.28). Thus, for the output signal of the water quality parameter  $X_i(t)$ , a total of six rainfall events were detected, delimited by the time windows  $[a, b]_k$ , during the measuring campaign (from 0 to 34627 minutes). In this sense, six time windows were identified in which  $X_i(t)$  had an “unusual” behavior, possibly caused by rainfall (due to the presence of rainfall during these days).

#### 3.5.1.4 Rainfall pulses identification

Based on the methods proposed in section 3.5.1.3, intervals  $[c, d]_k$  for the filtered input rainfall signal  $Y_j(t)_{filt}$  (for a  $j$  rain gauge station) in which abrupt changes occur, can be detected (by means of the Step Detection technique). In this sense, the Step Detection technique for identifying rainfall pulses can be applied as in Figure 3.29.

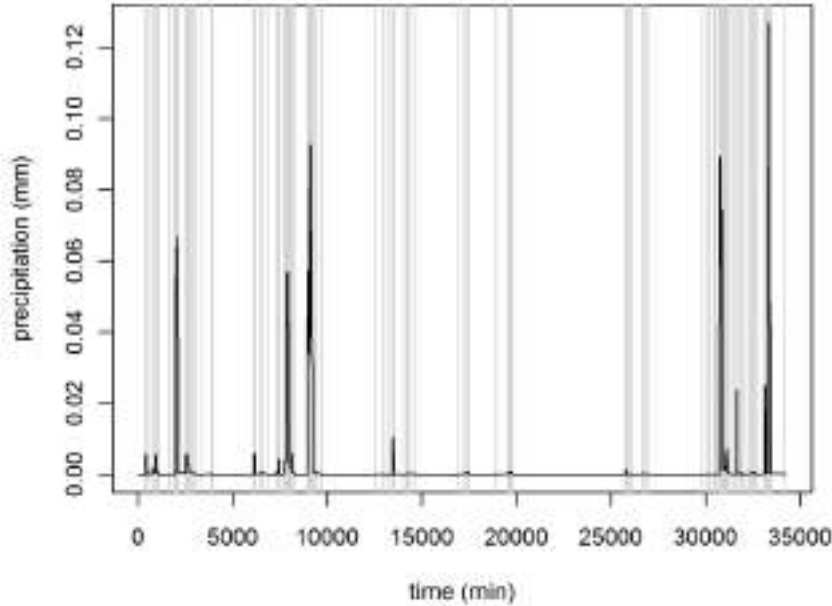


Figure 3.29 Identification of rainfall pulses for a given rain gauge station.

The grey lines identify where the possible rainfall pulses started and ended  $[c, d]_k$ , during the measurement campaign, for  $Y_j(t)_{filt}$ . Intervals  $[c, d]_k$  where the mean value  $\overline{Y(c_k < t < d_k)_{filt}} > 0.1$  mm/h (a high enough intensity for being considered a significant rainfall pulse, which may disturb the output signal  $X_i(t)$ ) can be identified. Those intervals  $[c, d]_k$  can be understood as the length of a given rainfall pulse, with average rainfall intensity greater than 0.1 mm/h.

#### 3.5.2 Recurrent relations between rainfall pulses and water quality's time windows ranging outside of typical dry weather behavior

Regarding the identified rainfall events, association metrics (Pearson's Correlation Test, Average Mutual Information, Linear Models and Partial Least Squares Model) are proposed between statistical characteristics of rainfall signal  $Y_j(t)$  and water quality and quantity output signal  $X_i(t)$  per event (*e.g.* mean, standard deviation, entropy and duration), with the purpose of detecting recurrent relations between rainfall and water quality and quantity (for rainfall registered at different rain gauge stations). Hence, uncertainty was assessed in rainfall events characteristics, as

well as in the recurrent relations between rainfall and water quality parameters, in order to know the reliability of the reported results.

By the identification of rainfall pulses in the input of the catchment  $Y_j(t)$  (section 3.5.1.4) and rainfall events in the water quality parameter signal  $X_i(t)$  (output or effluent of the catchment) (section 3.5.1.3), recurrent relations between the rainfall pulses and a water quality parameter measurements can be assessed. For the complete time series, the number of rainfall pulses should be equal to the number of time windows that ranges outside the typical dry weather behavior. If the number of rainfall pulses is greater, this will mean that more than one rainfall pulse is affecting a given time window  $[a, b]_k$  of  $X_i(t)$ , in which  $\overline{P_{dry}(a_k < t < b_k)filt} < 0.05$ . The author proposes to join the nearest rainfall pulses until the number of rainfall pulses is equal to the number of time windows outside the typical dry weather behavior. This proposal can be supported by the fact that, the more frequent that two or more continued rainfall pulses occur, the lower is the probability that the effect over the output  $X_i(t)$  has completely ceased from the beginning of the rainfall event. With the purpose of exposing this situation, the Figure 3.30 is shown.

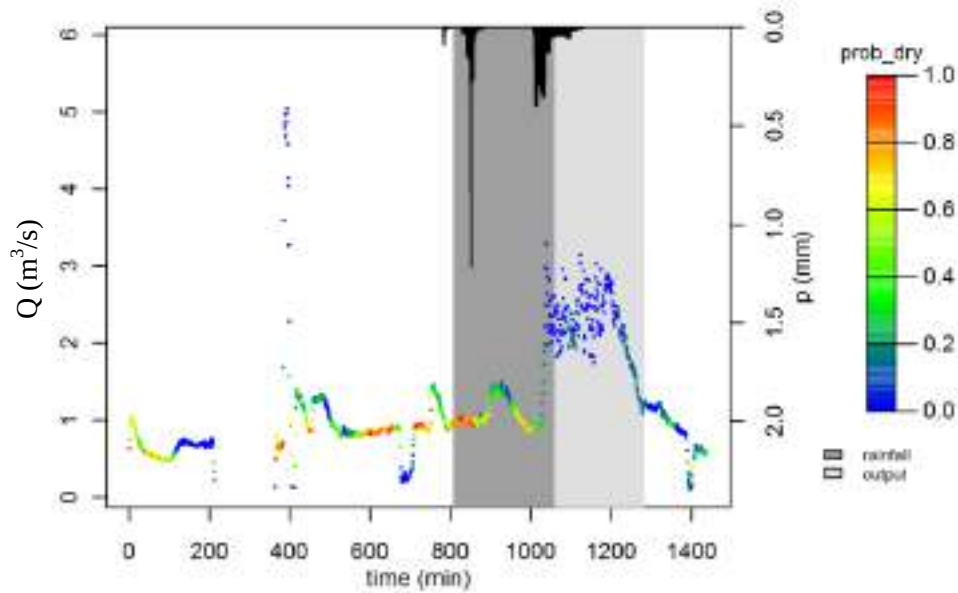


**Figure 3.30 Two rainfall pulses (rainfall) affecting the same time window that ranged outside of the typical dry weather behavior (output).**

In this figure, two rainfall pulses were identified as separated by the proposed methodology (section 3.5.1.4). However, as the number of pulses was greater than the number of time windows that grouped low probabilities of being part of the typical dry weather behavior (continuous blue reports of the signal  $X_i(t)$ ), the rainfall pulses were joined in order to make one entire rainfall pulse assignable to the unique time window  $[a, b]_k$  ( $\overline{P_{dry}(a_k < t < b_k)filt} < 0.05$ ) of the output.

Whenever the number of time windows  $[a, b]_k$  ( $\overline{P_{dry}(a_k < t < b_k)filt} < 0.05$ ) is greater than the number of rainfall pulses, the remaining time windows  $[a, b]_k$  ( $\overline{P_{dry}(a_k < t < b_k)filt} < 0.05$ ) possibly are not caused by rainfall and can be removed from the analysis.

By defining appropriately  $[a, b]_k$  ( $\overline{P_{dry}(a_k < t < b_k)}_{filt} < 0.05$ ) and their possible causing rainfall pulses  $[c, d]_k$  ( $\overline{Y(c_k < t < d_k)}_{filt} > 0.1$  mm/h), individual characteristics of both,  $X_i[a_k < t < b_k]$  and  $Y_i[c_k < t < d_k]$  can be inferred (e.g.  $\overline{X[a_k < t < b_k]}$ ,  $sd(X[a_k < t < b_k])$ ,  $\frac{dX[a_k < t < b_k]}{dt}$ ,  $sd\left(\frac{dX[a_k < t < b_k]}{dt}\right)$ ,  $\overline{Y[c_k < t < d_k]}$ ,  $sd(Y[c_k < t < d_k])$ ,  $\frac{dY[c_k < t < d_k]}{dt}$ ,  $sd\left(\frac{dY[c_k < t < d_k]}{dt}\right)$ , for each  $k$ ).

The correlation can be estimated between rainfall characteristics and water quality parameters among the  $k$  rainfall events detected. These recurrent relations can be explored by means of at least four approaches: (i) a Pearson's correlation test, by verifying statistical significance of the relation found ( $p$ -value) - this test can only be executed whenever normality and homogeneity of variance of both independent variables  $X$  and  $Y$  is proven, by the use of Shapiro Wilk Test (for normality) and the Bartlett Test (homogeneity of variance) - ; (ii) a linear model, evaluating the null hypothesis that the slope  $m$  of a linear relation between  $X$  and  $Y$  is greater than zero, with an acceptable significance ( $p$ -value)-also requirements about normality and homogeneity of variance need to be verified-; (iii) Average Mutual Information (AMI), verifying the significance of a non-linear relation between independent variables  $X$  and  $Y$ ; (iv) Partial Least Squares (PLS), evaluating the importance of each variable  $X_i$  from a multivariate input, in terms of prediction of the dependent variable  $Y$ .

### 3.5.2.1 Assessing uncertainty of rainfall events' characteristics, for the input (rainfall) and output (water quality) of the system

Each rainfall's or water quality's characteristic of a given  $k$  rainfall event (e.g.  $\overline{X[a_k < t < b_k]}$ ,  $sd(X[a_k < t < b_k])$ ,  $\frac{dX[a_k < t < b_k]}{dt}$ ,  $sd\left(\frac{dX[a_k < t < b_k]}{dt}\right)$ ,  $\overline{Y[c_k < t < d_k]}$ ,  $sd(Y[c_k < t < d_k])$ ,  $\frac{dY[c_k < t < d_k]}{dt}$ ,  $sd\left(\frac{dY[c_k < t < d_k]}{dt}\right)$ ) has an uncertainty in its estimation, due to presence of uncertainty in the original data  $X_i$  and  $Y_j$ . The uncertainty of the original data  $\mu(X_i)$  and  $\mu(Y_j)$  can be estimated as a Type B uncertainty, based on the resolution of the measurement device, given by the manual of the equipment and experts' judgment. Any statistic  $S$  estimated from a given rainfall event  $S(X[a_k < t < b_k])$  or  $S(Y[c_k < t < d_k])$  will have additional uncertainty rather than just the individual uncertainty values of  $\mu(X_i)$  and  $\mu(Y_j)$ . For this reason, it is necessary to assess uncertainties  $\mu(S(X[a_k < t < b_k]))$  and  $\mu(S(Y[c_k < t < d_k]))$ , for any statistic  $S$  to be used.

By knowing the equation from which  $S$  is calculated (e.g. the mean water quality value during a rainfall event  $k$  can be computed as:  $\sum \frac{X[a_k < t < b_k]}{N}$ ) two approaches can be implemented: (i) a first order Taylor approximation, estimating how the uncertainty of a set of independent observations  $X[a_k < t < b_k]$  with independent uncertainties  $\mu(X_i)$  is then propagated, (ii) A Monte Carlo simulation, in which for each value of  $X[a_k < t < b_k]$  a random perturbation  $x'$  is added (were  $x'$  represents a vector of random numbers:  $-\mu(X_i) < x' < \mu(X_i)$ ), for calculating the statistic  $S$  over the set expressed by  $X[a_k < t < b_k] + x'$ ; this procedure is performed several times (named as the number of simulations), giving that, for each simulation, a different value of  $S(x)$  is going to be reported. Therefore, the variability of  $S(x)$  can be computed as the standard deviation  $sd(S(x))$ , equivalent to the uncertainty in the estimation of  $S$ , from the sample  $X[a_k < t < b_k]$ , with independent uncertainty  $\mu(X_i)$ , denoted as  $\mu(S(X[a_k < t < b_k]))$ . For this research, the author

will use the Monte Carlo approach for assessing the uncertainties  $\mu(S(X[a_k < t < b_k]))$  and  $\mu(S(Y[c_k < t < d_k]))$ , for any statistic  $S$ .

### 3.5.2.2 Assessing uncertainty in recurrent relations between rainfall pulses and water quality's time windows ranging outside of typical dry weather behavior

As explained in section 3.5.2, the estimation of the correlation between rainfall characteristics and the water quality parameters' characteristics among the  $k$  detected rainfall events can be performed by different statistical techniques; but uncertainty in the measured variables  $X$  and  $Y$  has not been included yet. As described in 3.5.2.1, uncertainty in any statistic  $S$  can be assessed by means of a Monte Carlo approach. By knowing the uncertainty of  $\mu(S(X[a_k < t < b_k]))$  and  $\mu(S(Y[c_k < t < d_k]))$ , it is important to determine the probability that the relation between the variables  $X$  and  $Y$  is significant, regarding the results computed by methodologies mentioned in 3.5.2. This assessment can also be done by means of a Monte Carlo approach, generating random values as  $(S(X[a_k < t < b_k])) + x'$ , where  $x'$  belongs to the interval  $-\mu(S(X[a_k < t < b_k])) < x' < +\mu(S(X[a_k < t < b_k]))$ . For each generated random number, the criteria exposed in section 3.5.2 for Pearson Correlation Test, Linear Model approach, AMI significance and PLS can be evaluated among all rainfall events. Figure 3.31 shows an example of the results.

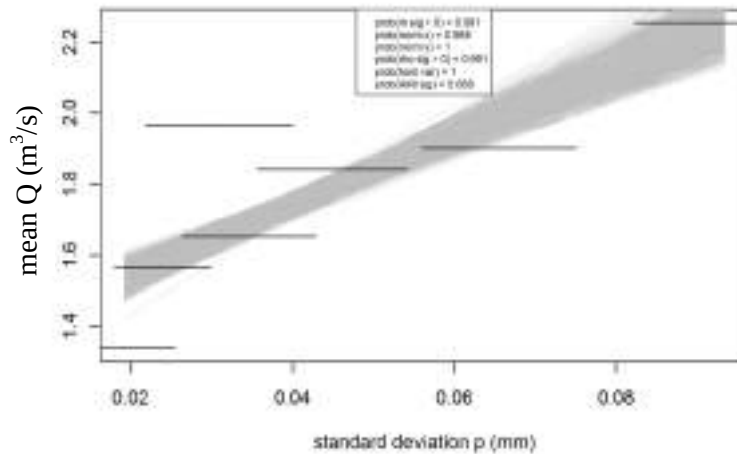


**Figure 3.31 Assessment of recurrent relation between a rainfall and a water quantity characteristic (Pearson Correlation's test, Linear Model Regression, Average Mutual Information).**

For this case, seven rainfall events were reported. For each event, the uncertainty in the estimation of the standard deviation (statistic  $S$ ) of the rainfall  $\mu(S(Y[c_k < t < d_k]))$  was computed (uncertainties shown as horizontal bars). Furthermore, the uncertainties of the flow rate mean value (statistic  $S$ ) during each event's uncertainty  $\mu(S(X[a_k < t < b_k]))$  was assessed by means of the same approach (uncertainties shown as vertical bars): this uncertainty was much lower than the uncertainty associated to rainfall, reason for not being appreciable in Figure 3.31.

Therefore, several Monte Carlo simulations were executed and the number of times in which each of the four relationship criteria were significant were counted and divided by the total number of simulations. In other words, the probability that each of the explored criteria ensures that a significant relation between the standard deviation of the rainfall and the mean flow rate value

exists was estimated. For example, for Figure 3.31, it was assessed that, with a probability of 99 %, the relation between the standard deviation of the rainfall and the mean flow rate value can be represented by a linear relation (Pearson's Correlation Test and Linear Model evaluation). Hence, with a probability of about 67 %, it can be concluded that the relationship between this variables follows a non-linear behavior (AMI results). In terms of a multivariate analysis, an example of the PLS results is shown in Figure 3.32.

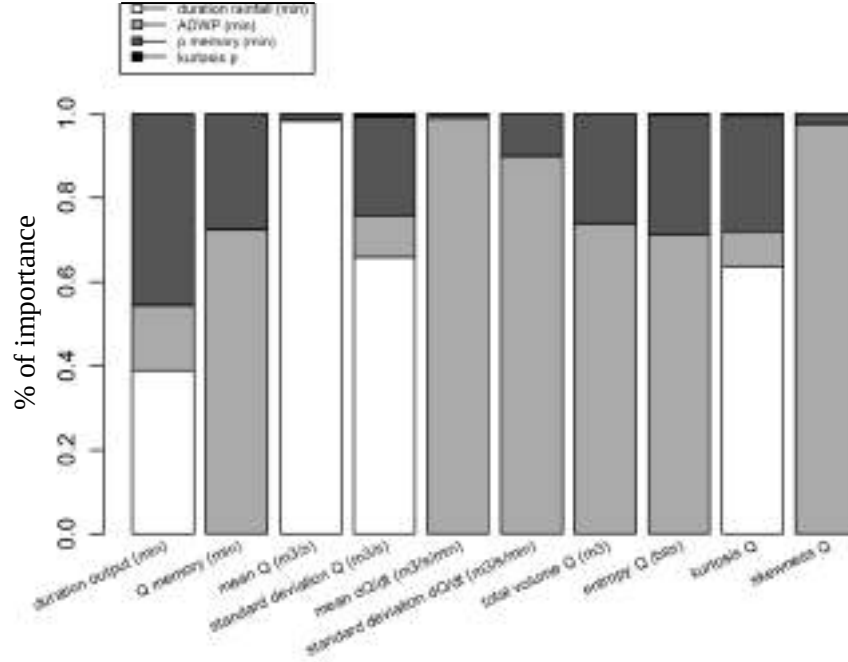


**Figure 3.32 Assessment of recurrent relation between a rainfall and water quality characteristic (PLS approach).**

This figure exposes several relations among different rainfall characteristics and different characteristics of the flow rate measured at the effluent of the catchment. For mean flow rate values (mean Q in Figure 3.32), it was detected that in about 99 % of the occasions (for several values of  $\mu(S(X[a_k < t < b_k]))$  and  $\mu(S(Y[c_k < t < d_k]))$  obtained from a Monte Carlo approach), the most important explanatory variable is the duration of the rainfall. The PLS analysis does not have a significance test formulation, which makes results hardly generalizable, just bringing insights of local comparisons among the output's evaluated characteristics (*e.g.* flow rate mean, standard deviation, total volume, entropy). On the other hand, for this technique, although significance level cannot be evaluated (*p*-value), a multivariate analysis with scarce observations (*e.g.* seven rainfall events) can be undertaken.

### 3.5.3 Association of recurrent relations with characteristics of the urban catchment

With the purpose of supporting the recurrent relations found, a physically based method is proposed. The methodology consists in assigning to each of the studied rain gauge stations a categorical range of “influence”, over a given water quality parameter. This assignation should be supported by the recurrent relations found from the previously described exercise: the rain gauge station which reported, with a higher frequency, significant relations regarding different characteristics of the rainfall pulses and of the water quality registers in time windows ranging outside of their typical dry weather behavior, will be a “High Influence” rain gauge station. Hence,

rain gauge stations can also be classified to have “Medium” or “Low” influence, depending on the amount of significant recurrent relations observed, for different comparisons between water quality parameters and rainfall characteristics. The number of ranges can be established by different criteria such as: (i) the number of rain gauge stations and (ii) the available data of physical characteristics of the catchment. Therefore, the physical characteristics of each spatial point within the urban catchment (*e.g.* average sewer pipelines features, land use, permeability and nearness to the effluent) can be related with a studied rain gauge station by means of spatial criteria (*e.g.* the nearest rain gauge station will be assigned as representative of the rainfall influence, for a given point within the catchment). Considering that for each of the studied rain gauge stations a range of influence is assigned; a range of “High”, “Medium” or “Low” influence can also be specified for each point of the catchment, by means of the afferent rain gauge station (which was also labeled as “High”, “Medium” or “Low”). In order to explain the differences between the variability of the physical characteristics of each point inside the catchment with a “High”, “Medium” or “Low” influence, a PCA is proposed. This analysis will lead to visualize the most explanatory factors (physical characteristics representative of each point), between the variability of a “Low”, “Medium” or “High” amount of recurrent relations given by the afferent rain gauge station to each point within the catchment.

## 4 RESULTS AND DISCUSSION

The proposed methods described in section 3.5 were applied to a case of study (urban catchment described in section 3.1) for several water quality and quantity parameters measured by means of different technologies (Echo Profile Flowmeter, uni-parametric and UV-VIS spectrometry). Measurements at the effluent of the catchment were joined with rainfall registrations given by three rain gauge stations (*Fontibon*, *Bosa* and *Casablanca*) in a whole dataset described in section 3.4.1. Hence, the proposed methods were implemented in R-project (see details Appendix 8.2).

### 4.1 Dry weather characterization and rainy weather identification

The dry weather characterization and the rainy weather identification for the studied water quality parameters were executed. It is important to clarify that the dry weather characterization was performed with the information of each separated rain gauge station (*Fontibon*, *Bosa* or *Casablanca*). This division had the purpose of analyzing if a given rain gauge station had more influence over some water quality parameter, rather than the other studied stations. Whenever this is the case, the mentioned approach will lead to perceive if a given rain gauge station has more influence over water quality at the effluent of the catchment than on other rain gauge station (possibly due to their spatial differences and zone location's characteristics).

#### 4.1.1 Differences between dry and rainy days in terms of water quality parameters

Differences between dry and rainy weather were identified by means of a multivariate statistical technique: the Principal Component Analysis (PCA). These calculations led to identify which water quality parameters measured by means of different technologies seem to be influenced by rainfall rather than the other studied parameters. The exposed results just give relative insights; in order to find recurrent and generalizable relations further analyses are proposed (*e.g.* Correlation Tests). However, this preliminary identification will be an input for this mentioned analysis. On the other hand, it seems reasonable to expect that the flow rate values will show an important influence over the variability of rainy and dry weather data, at least more than the other analyzed water quality parameters. In this sense, with the purpose of having a reference, flow rate values were included in all PCA calculations. The data was separated into two groups: the dry weather data (regarding days in which no rainfall was registered at a given rain gauge station) and the rainy days data (including data that was reported during rainy days).

##### 4.1.1.1 Uni-parametric sensors

Regarding the water quality parameters values measured by uni-parametric sensors, a PCA analysis was executed with the rainfall registered in each of the studied rain gauge stations. These calculations were performed in order to detect the most relevant water quality parameters considered (DO, ORP, pH, Temperature, Turbidity and TSS), over the differences of the variability of dry and wet weather data. First calculations were undertaken for the *Fontibon* rain gauge station, representing the data by three principal components (comp. 1: 30 %, comp. 2: 58 % and comp. 3: 73 % of the cumulative variance).

The projection of each vector over the x axis represents the importance that each variable has over the separation between the dry and wet weather data (given by principal component 1, Appendix 8.3). Therefore, it can be detected that the DO variable has the most important incidence in the separation of rainy and dry days. From the above result, it can be concluded that the DO seems to

be the parameter that has most influence over the variability of rainy and dry weather data, regarding the Appendix 8.3 (a) and (b) (inclusive more than Flow rate, due to the length of the projection over component 1). It must be kept in mind that the classification of data between dry or rainy was assessed by the use of rainfall registered at the *Fontibon* rain gauge station. Furthermore, Flow rate, ORP, Turbidity and TSS seem to also have an important influence in the explanation of why rainy weather data is different from dry weather data. On the other hand, Temperature and pH seem not to be related with the split-up of the two groups, due to their separation from the principal component 1 (horizontal axis). In Appendix 8.3 (c) and (d) (separation of the groups by effect of components 1 and 3), similar results were obtained. Differences between the two separation groups (dry and rainy weather) were mostly explained by uni-parametric water quality measurements (except Temperature and slightly pH) and the flow rate. However, the separation between the groups is qualitative. With the purpose of verifying the hypothesis established by means of these results, further analyses are proposed.

For the case of the rainfall registered at the *Bosa* rain gauge station separated by three components (comp. 1: 30 %, comp. 2: 58 % and comp. 3: 73 % of the cumulative variance), the following results were obtained:

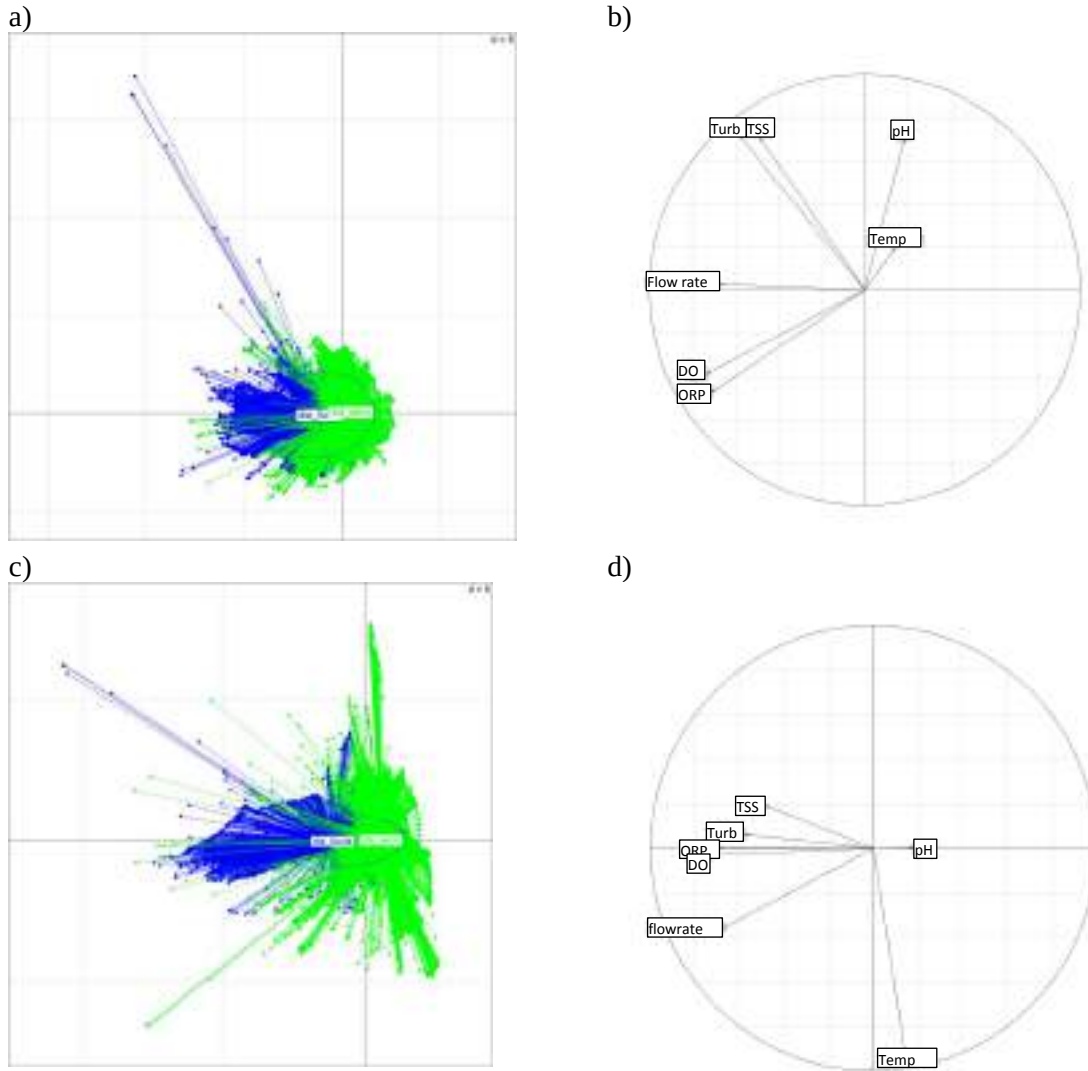


**Figure 4.1** PCA analysis for the *Bosa* rain gauge station (rainy days' data blue group, dry days' data green group); a), b) separation given by principal components 1 (x-axis) and 2 (y-axis); b), c) separation given by principal components 1 (x-axis) and 3 (y-axis).

In terms of the principal components 1 and 2 (Figure 4.1 a, b), the results were equivalent to the results reported for *Fontibon* rain gauge station. Regarding the results described by Figure 4.1 (b) and (c), although the principal components 2 and 3 represent a separation between the groups of rainy and dry weather data, this separation is not so clear as in case of *Fontibon* rain gauge station. This fact may lead to conclude that probably the *Bosa* rain gauge station has a slightly more detectable influence over water quality than the *Fontibon* rain gauge station. However, this is a preliminary conclusion because the results reported for both rain gauge stations (*Bosa* and *Fontibon*) are quite similar.

A slight influence of pH was detected in the Figure 4.1 (b) and (c) between the separation of the two groups. The difference with the other variables (Flow rate, TSS, Turbidity, DO and ORP) is that the pH vector points in the positive direction of the component 3 (right side of the circle). This fact can

lead to state that the greater pH values were registered during dry days, rather than in rainy days. On the other hand, this influence is probably insignificant, due to the short norm of the pH vector.

Analyzing the results obtained from rainfall registered at the *Casablanca* rain gauge station, regarding three principal components (comp. 1: 30 %, comp. 2: 58 % and comp. 3: 73 % of the cumulative variance), the Appendix 8.4 results were obtained.

The results reported for the *Casablanca* rain gauge station are equivalent to the results reported for the *Fontibon* rain gauge station. This fact can be explained because for the *Fontibon* and *Casablanca* rain gauge stations, the dry days were the same. In this sense, dry weather characterization for the mentioned rain gauge stations will be equivalent. Possible differences in terms of rainfall influence over water quality between these stations were addressed in further analysis. On the other hand, the initial hypothesis that Flow rate will show an important separation between the dry and rainy groups could be verified, although DO showed more influence over the separation of groups than the Flow rate. However, none of the analysis undertaken exposed a clear separation of the “dry” and “wet” weather data groups, due to the presence of overlaps. This fact may lead to state that the exposed results are just preliminary insights: more general conclusions will be sought in further sections.

#### 4.1.1.2 UV-VIS spectrometry

As water quality parameters were also measured by means of a UV-VIS spectrometry device, the proposed PCA was also undertaken for data reported by this sensor. The analysis was divided into: (i) a spectral analysis, including each wavelength as an input variable; (ii) the default water quality values reported by the global calibration of the equipment. The flow rate measurements were also included in all of the three proposed approaches (as a reference variable, equivalent to the uniparametric sensor approach).

##### 4.1.1.2.1 Spectral analysis

For these paragraphs, the PCA analysis regarding the spectra of the UV-VIS measurements (wavelengths from 250 nm to 750 nm) was undertaken. The first analyzed rain gauge station was *Fontibon*. Its data was represented by means of three principal components (comp. 1: 90 %, comp. 2: 98 % and comp. 3: 99 % of the cumulative variance). However, the only appreciable separation between the “dry” and “rainy” data was for the principal components 1 and 2 (principal comp. 2-3 or 1-3 do not exposed a clear separation). The obtained results regarding these components are exposed in the Appendix 8.5.

A separation between the blue and the green group can be explained principally due to component 2 Appendix 8.5 (a) and (b). Hence, the flow rate value is the variable that has a more important influence over the differences of the variability between these two groups. In terms of wavelengths, it can be appreciated that UV range (from wavelength 250 nm to 380 nm) seems to have a positive (+) incidence over the rainy days values, as they point to the positive part of the component 2 (up direction). Thus, the VIS part of the spectra (from wavelength 380 nm to 750 nm) seems to have a negative (-) incidence over the variability of the dry weather data, as they point to the negative part of the component 2 (down direction). Hence, the most extreme wavelengths of the UV-VIS range (250 nm and 750 nm wavelengths) seems to be the most significant variables over the variability of water quality measured during dry and rainy days, in a positive and negative mathematical sense,

respectively. For wavelengths in the VIS range (380 nm to 750 nm), the mathematical relation was detected as direct: the higher absorbance's values seem to be registered during rainy days. On the other hand, for wavelengths in the UV range (250 nm to 380 nm), the relation with measured water quality might be inverse: the lower absorbance's values are registered during dry days. Wavelengths in the middle of the spectra (about 400 nm) seem not to be influenced by rainfall (see Appendix 8.5 a and b).

Results obtained from the rainfall registered at the *Bosa* rain gauge station (comp. 1: 90 %, comp. 2: 98 % and comp. 3: 99 % of the cumulative variance), regarding the spectra registered by the UV-VIS spectrometry sensor are exposed for the principal components 1 and 2 (remaining components do not exposed an appreciable separation between the “dry” and “rainy” groups):

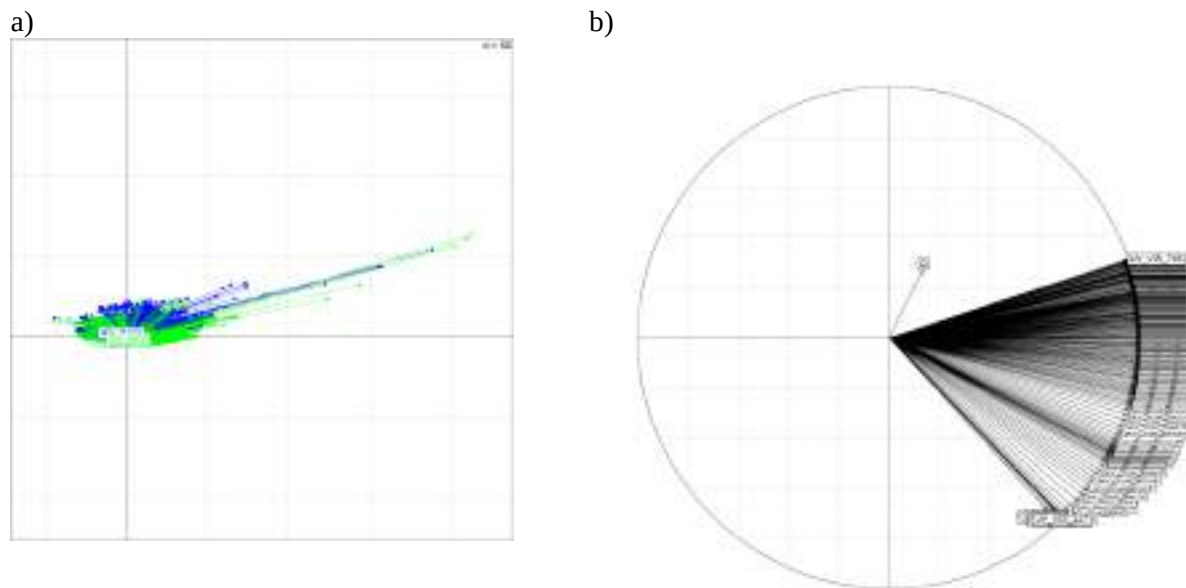


**Figure 4.2 PCA analysis for the *Bosa* rain gauge station (rainy days' data blue group, dry days' data green group); a), b) separation given by principal components 1 (x- axis) and 2 (y- axis).**

From Figure 4.2 (a) and (b), the proposed statements for results reported by rainfall registered at the *Fontibon* rain gauge station are equivalent. Therefore, in terms of the spectra registered by the UV-VIS spectrometry sensor for measured water quality parameters, there was not a detectable difference between dry weather defined by registrations in *Fontibon* or *Bosa* rain gauge station. This statement seems to be in contradiction with the results obtained for uni-parametric sensors, where slightly differences from the PCA for *Fontibon* and *Bosa* rain gauge stations were detected (see Figure 4.2 a and b). Thus, two possible explanations to this circumstance are formulated: (i) non-registered values “NA” in the UV-VIS spectrometry sensor were more frequent than for the uni-parametric sensors, (ii) differences detected between aggrupation performed by the use of *Bosa* rain gauge station's registers are so slightly that can be neglected.

For the *Casablanca* rain gauge station, the same analysis in terms of the spectra registered by means of the UV-VIS spectrometry sensor were undertaken. The data was represented by three principal components (comp. 1: 90 %, comp. 2: 98 % and comp. 3: 99 % of the cumulative variance).The

results were representative for components 1 and 2, which exhibited a separation between the “dry” and “rainy” groups (see Appendix 8.6).

It can be stated that, in terms of the spectra measured by the UV-VIS sensor, the results obtained from the PCA execution by using the rainfall registered at the *Bosa*, *Fontibon* or *Casablanca* rain gauge station are equivalent. This fact may lead to conclude that, probably, for all days in which rainfall was registered at a given rain gauge station, rainfall was also registered at the other studied stations. Regarding the Spectral analysis, none of the studied cases reported a full separation between the “dry” and “rainy” groups, due to the presence of overlaps. However, these results bring up preliminary statements of the influence of rainfall over water quality, which will be an input for further analysis in the following sections.

#### 4.1.1.2.2 Equivalent water quality parameters

By the use of the UV-VIS spectrometry sensor and equipment’s global calibration, equivalent TSS and COD values were reported. These values were classified into groups of “dry” and “rainy” data in agreement with the rainfall registered by different rain gauge stations (*Fontibon*, *Bosa* and *Casablanca*). Whenever a rainfall pulse was registered (at a given rain gauge station), data reported during this day was assigned in the “rainy” group; otherwise in the “dry” group. The PCA was undertaken considering three principal components (comp. 1: 60 %, comp. 2: 93 % and comp. 3: 99 % of the cumulative variance). However, the principal components 1 and 2 were the only couple of components which represented a separation between the “dry” and “rainy” groups (although this separation exhibits a substantial overlap). The results obtained regarding the rainfall registrations at the *Fontibon* rain gauge station are shown in Appendix 8.7.

For this case, a significant difference between dry and rainy days, by any variable besides the flow rate could not be appreciated. On the other hand, the equivalent TSS concentration seems to be more related to rainy days’ variability and COD concentration to dry days’ group. Calculations were executed for the rainfall registers of *Bosa* and *Casablanca* rain gauge stations. The data was represented by means of three principal components (comp. 1: 60 %, comp. 2: 93 % and comp. 3: 99 % of the cumulative variance; for *Bosa* and also *Casablanca*). The separation between “dry” and “rainy” groups was based on the rainfall registers: whenever no rainfall was registered at the *Bosa* or *Casablanca* rain gauge stations during a complete day, the water quality data reported during that day was classified as “dry”; otherwise as “rainy”. The results regarding *Bosa* and *Casablanca* rain gauge stations are presented in Figure 4.3.

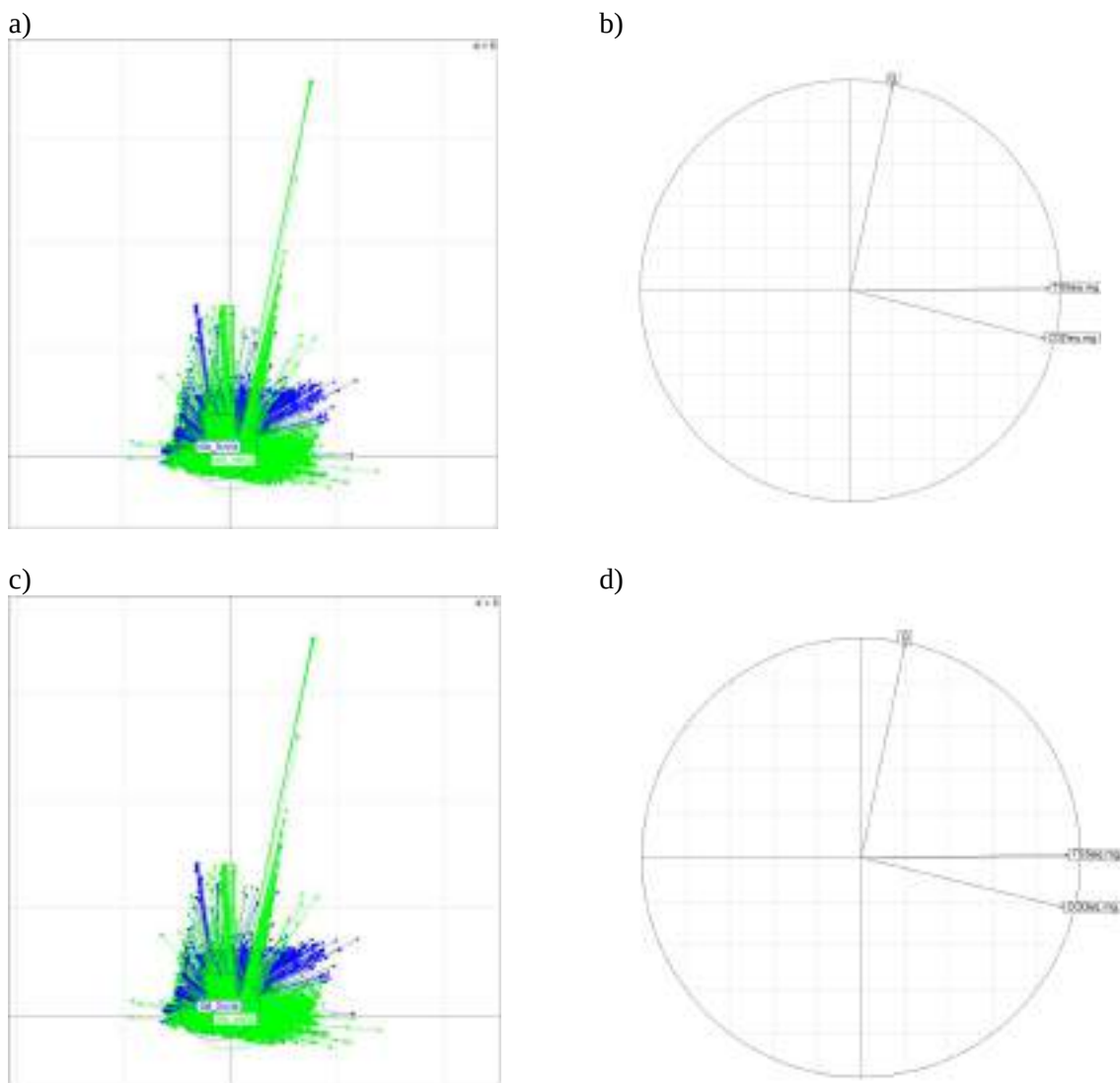


**Figure 4.3 PCA analysis for the Bosa (a,b) and Casablanca (c,d) rain gauge stations (rainy days' data blue group, dry days' data green group); a), b) separation given by principal components 1 (x-axis) and 2 (y-axis).**

The results for all three rain gauge stations seem to be equivalent. This fact means that, for the measurements undertaken by the use of the UV-VIS spectrometry sensor, no differences among the studied rain gauge stations were detected. Possible explanations for this result, as discussed in section 4.1.1.2.1, can be the presence of an important amount of “NA” reports. Another possibility is that all rainfall events during the measuring campaign were registered by the three studied rain gauge stations simultaneously.

From the PCA, it can be concluded that the water quality parameters measured by means of the uni-parametric sensors: DO, ORP, Turbidity, TSS seem to be more influenced by rainfall registrations, than pH and Temperature. Therefore, it can be stated that parameters not directly related with solid particles (*e.g.* DO, flow rate and ORP) seem to be more related with rainfall, rather than parameters that describe particulate pollution (TSS and Turbidity). This fact may lead to conclude that resuspension and settling phenomena in the wastewater (including stormwater) effluent of the studied urban catchment exhibits a complex behavior, with variations that can be more difficult to

relate to dry or wet weather conditions than for water characteristics non-related exclusively with solid particles (e.g. DO, flow rate and ORP). For the UV-VIS spectrometry equipment, the VIS range of the spectra (wavelengths 380 nm to 750 nm) seems to more appropriately explain the variability of rainy days' data than the UV range (wavelengths 250 nm to 380 nm), which seems to have a more important influence over the variability of dry weather data. Furthermore, the more extreme wavelengths in the UV-VIS range (250 and 750 respectively), seems to have the most important influence over the variability of water quality data. While the wavelength value gets nearer to the center of the spectra (about 400 nm), its representativeness in the variability of dry or rainy days data gradually fades. The Flow rate value was an appropriate reference, as it appeared as an important explanatory variable of the variability between dry and wet weather groups for all analyzed cases. Therefore, the PCA execution (as a multivariate analysis) seems to be more appropriate for a preliminary identification of the most influenced water quality parameters rather than the others, without concerning generalizable behaviors (if for the PCA the DO seemed to be influenced by rainfall more than the other parameters, it would not mean that variability of DO can be explained just by means of rainfall characteristics). In addition, considering that separation between "dry" and "rainy" data is characterized by overlaps in all PCA undertaken, the exposed results are even rougher. With the purpose of establishing more generalizable results, further analyses are proposed.

#### **4.1.2 Dry weather probabilistic characterization**

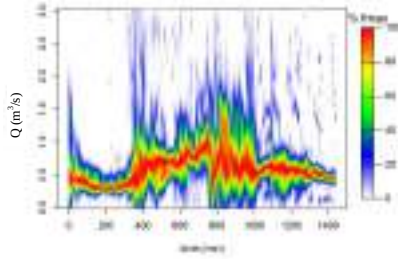
Throughout these paragraphs the probabilistic characterization of the typical dry weather behavior of the water quality parameters measured by means of uni-parametric and UV-VIS spectrometry sensors was assessed. Although the variability of the dry and rainy weather's behavior was not quite different by an analysis carried out with rainfall registered by different rain gauge stations (*Fontibon, Bosa, Casablanca*), the dry weather characterization was undertaken with the individual records, for the case study. This exercise was executed with the purpose of having the most available data from the dry weather, for each of the water quality and quantity parameters. If the rainfall registered by a given rain gauge station does not have an influence over a water quality or quantity parameter, the data from this day will be discarded unnecessarily. Hereafter, the Flow rate and each of the water quality parameters measured by means of the uni-parametric and UV-VIS spectrometry sensors were analyzed independently.

##### **4.1.2.1 Echo Profile Flowmeter**

The dry weather characterization obtained for the flow rate measurements, by analyzing rainfall registrations of each rain gauge station independently, is described in Figure 4.4.

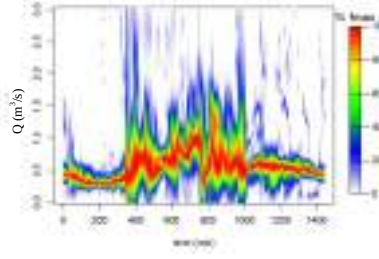
*Fontibon*

a)



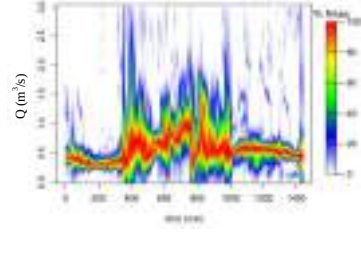
*Bosa*

b)

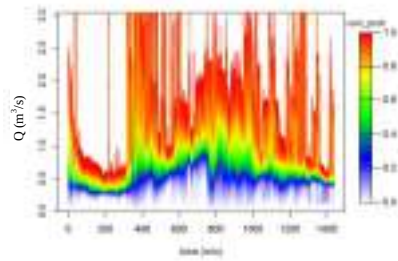


*Casablanca*

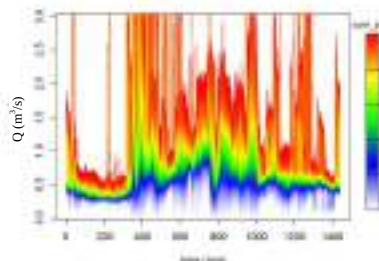
c)



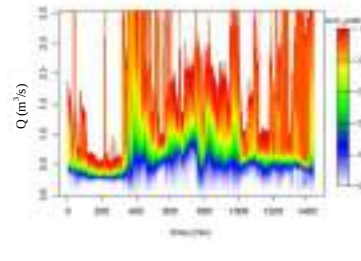
d)



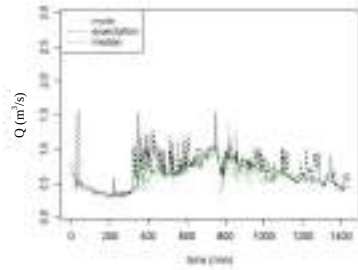
e)



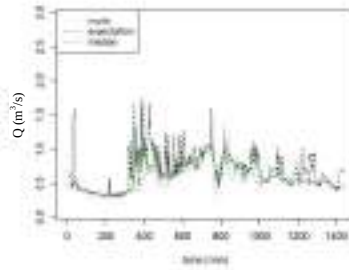
f)



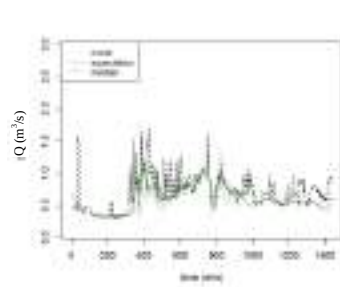
g)



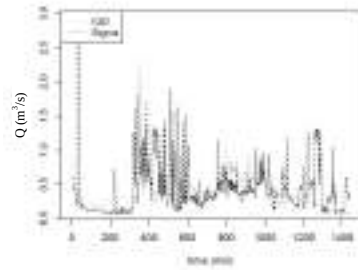
h)



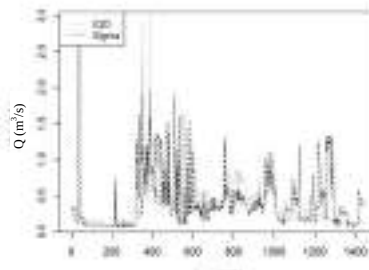
i)



j)



k)



l)

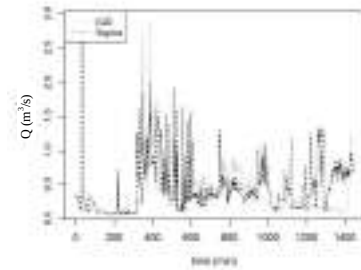


**Figure 4.4** Dry weather probabilistic characterization for flow rate measured by Echo Profile Flowmeter sensor, using rainfall registrations at different rain gauge stations (*Fontibon* (left), *Bosa* (center) and *Casablanca* (right)); a, b, c) Probability density function (PDF) for each time-step; d, e, f) Cumulative probability function (CPF) for each time-step; g, h, i) Mode, Median and Expectation for each time-step; j, k, l) Interquartile distance and standard deviation for each time-step

Dry weather characterization for the *Bosa* and *Casablanca* rain gauge stations seems to be equivalent; for the case of the *Fontibon* rain gauge station slight differences at midnight of the typical dry weather day's characterization (about 00h00) can be noticed. For this characterization,

the variability during this period of time seems to be greater than for the other studied stations (see Figure 4.4 j, k, l). This fact can be due to the registration of spurious data when some rainfall was reported just at the *Fontibon* rain gauge station, without being reported at the other stations.

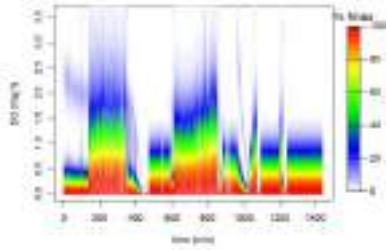
The variability of Flow rate values seems to be greater on approximately 300 min to 600 min (5h00 to 10h00) and lower from 0 min to 300 min (12h00 to 5h00) than for the rest of the typical dry weather day, for all analyzed cases in terms of rainfall registrations (see Figure 4.4 j, k, l). On the other hand, flow rate mean values seemed to be lower (tendency curves) during the same time window in which variability was also lower (12 pm to 5 am) (see Figure 4.4 g, h, i).

#### **4.1.2.2 Uni-parametric sensors**

By the use of the uni-parametric sensors, different water quality parameters were measured: Dissolved Oxygen, Total Suspended Solids, Temperature, Turbidity, pH and ORP. The typical dry weather behavior of the mentioned water quality was characterized by means of a probabilistic approach.

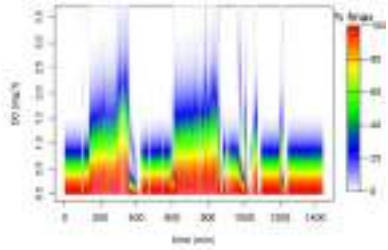
*Fontibon*

a)



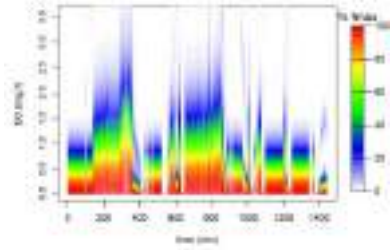
*Bosa*

b)

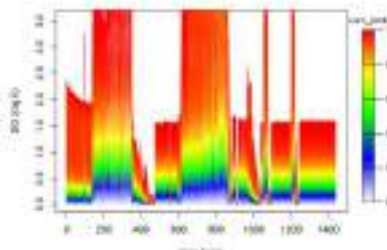


*Casablanca*

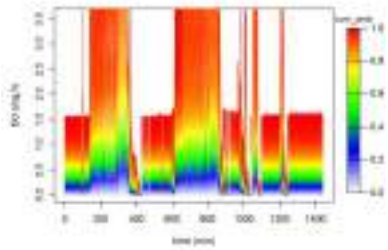
c)



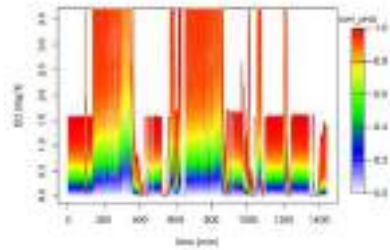
d)



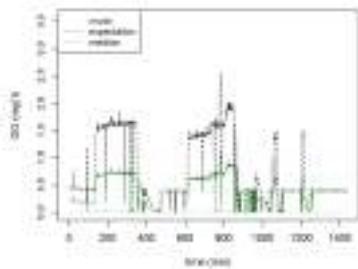
e)



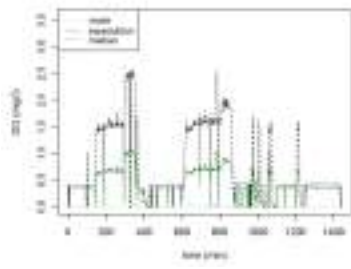
f)



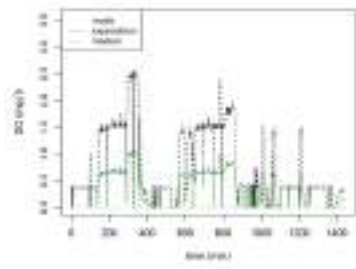
g)



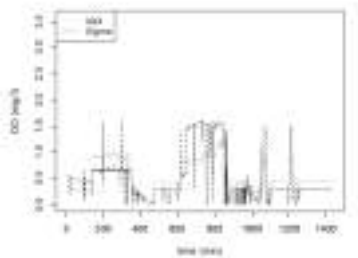
h)



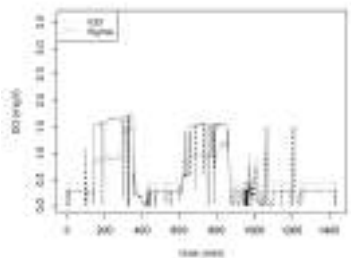
i)



j)



k)



l)

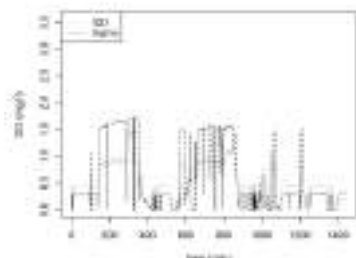


**Figure 4.5** Dry weather probabilistic characterization for DO measured by uni-parametric sensor, using rainfall registrations at different rain gauge stations (*Fontibon* (left), *Bosa* (center) and *Casablanca* (right)); a, b, c) Probability density function (PDF) for each time-step; d, e, f) Cumulative probability function (CPF) for each time-step; g, h, i) Mode, Median and Expectation for each time-step; j, k, l) Interquartile distance and standard deviation for each time-step

From the results obtained regarding the DO measured by uni-parametric sensors and the rainfall registered by the *Fontibon*, *Bosa* and *Casablanca* rain gauge stations, the typical dry weather characterization was slightly different for the three studied rain gauge stations, especially at night

time (see Figure 4.5). As a generalizable conclusion, it can be stated that this distribution was one side tailed, with the highest probability values closer to 0.0 mg/l.

The next water quality parameter analyzed was the Total Suspended Solids (TSS). The results can be shown in Figure 4.6.

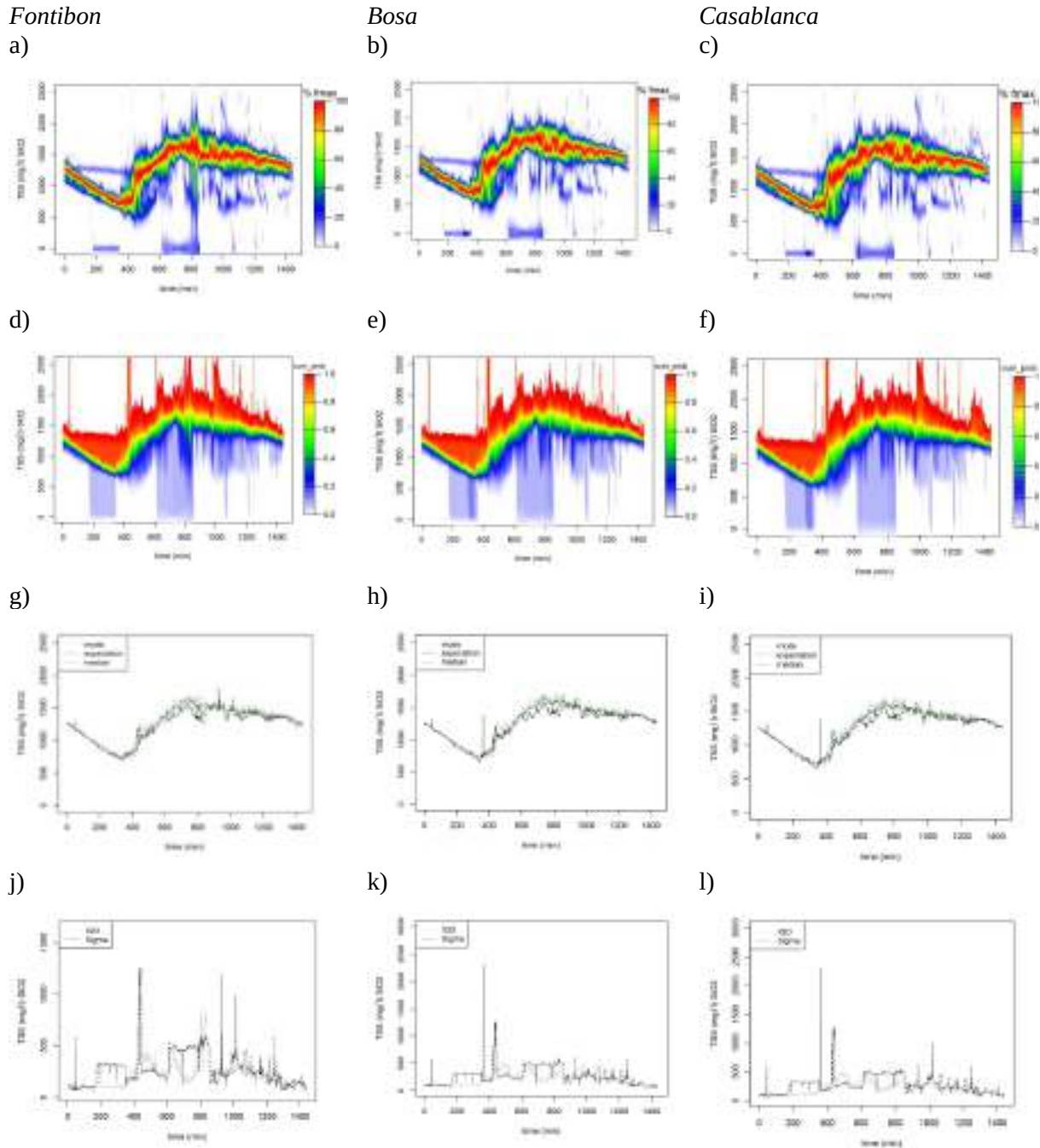


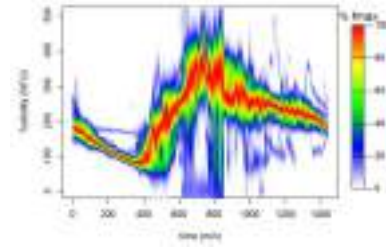
**Figure 4.6** Dry weather probabilistic characterization for TSS measured by uni-parametric sensors, using rainfall registrations at different rain gauge stations (*Fontibon* (left), *Bosa* (center) and *Casablanca* (right)); a, b, c) Probability density function (PDF) for each time-step; d, e, f) Cumulative probability function (CPF) for each time-step; g, h, i) Mode, Median and Expectation for each time-step; j, k, l) Interquartile distance and standard deviation for each time-step

From the above results, it can be stated that very small differences between the dry weather characterizations of *Fontibon*, *Bosa* and *Casablanca* rain gauge stations were observed. Concerning the typical dry weather characterization undertaken by the use of the rainfall registrations of the *Fontibon*, *Bosa* and *Casablanca* rain gauge station it can be stated that the PDF, the CPF, the tendency and the variability were very similar among them (Figure 4.6). It can be detected that there are no particular windows, during the typical dry weather, in which the variability seems higher than for others, independently of the station selected for undertaking the dry weather characterization. On the other hand, it can be concluded that the minimum TSS values are about 600 mg/l SiO<sub>2</sub> and occurs at 380 min, approximately. The maximum TSS values (about 1500 mg/l SiO<sub>2</sub>) were reported at 700 min, approximately. The TSS units were reported in mg/L SiO<sub>2</sub> because this was the measurement unit that the uni-parametric sensor has by default.

The Turbidity registered by means of the uni-parametric sensors was analyzed, in terms of its typical dry weather behavior, defined by rainfall information reported by studies rain gauge stations (*Fontibon*, *Bosa* and *Casablanca*).

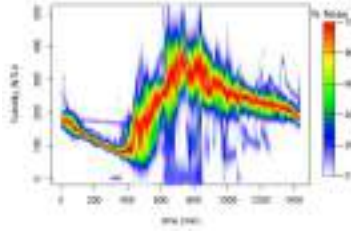
Fontibon

a)



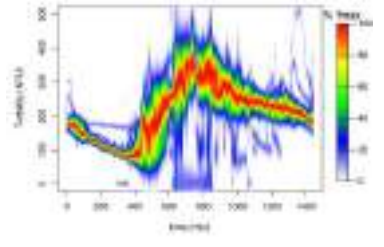
Bosa

b)

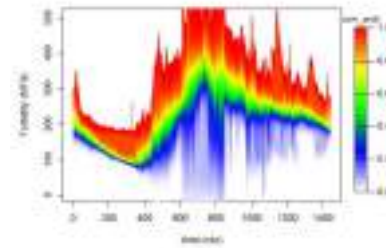


Casablanca

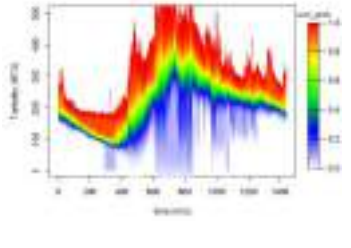
c)



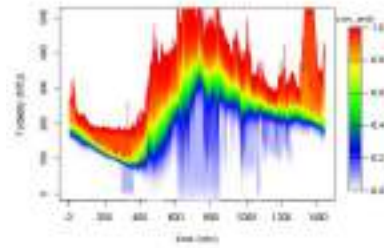
d)



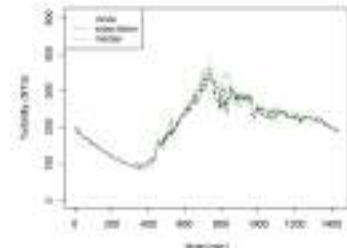
e)



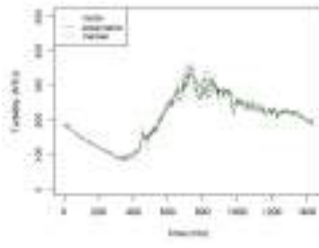
f)



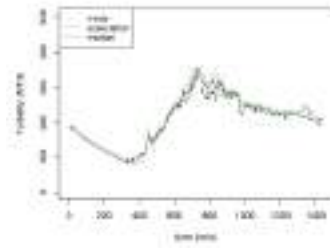
g)



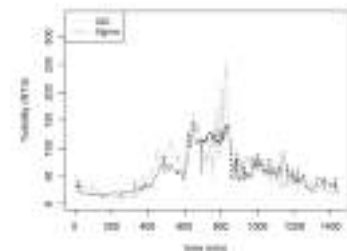
h)



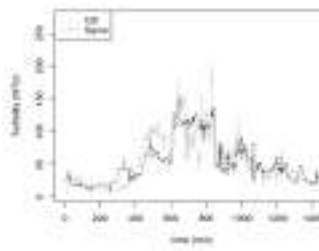
i)



j)



k)



l)

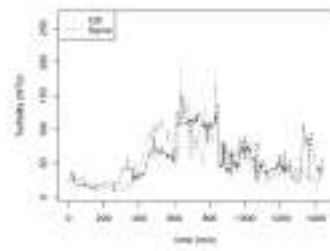


**Figure 4.7 Dry weather probabilistic characterization for Turbidity measured by uni-parametric sensors, using rainfall registrations at different rain gauge stations (*Fontibon* (left), *Bosa* (center) and *Casablanca* (right)); a, b, c) Probability density function (PDF) for each time-step; d, e, f) Cumulative probability function (CPF) for each time-step; g, h, i) Mode, Median and Expectation for each time-step; j, k, l) Interquartile distance and standard deviation for each time-step**

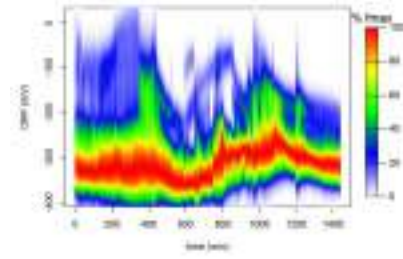
From the results in Figure 4.7, it can be observed a very similar behavior between the Turbidity and the TSS, in terms of means, medians and modes (see Figure 4.6 and Figure 4.7 g, h, i). This relationship can suggest a strong correlation between the Turbidity and the TSS (*e.g.* Mishendo *et al.*, 1999). The variability's behavior (see Figure 4.7 j, k, l) seems to be different between Turbidity and TSS: an increasing variability of turbidity can be noticed from about 400 min to 1000 min, different than for the case of TSS (Figure 4.7 j, k, l). Thus, it can also be stated that Turbidity exposes its highest variability between 600 min and 800 min rather than for other time windows of

the typical dry weather behavior; independently from rainfall registers used for undertaking the dry weather probabilistic characterization.

The results obtained for an analysis of ORP are shown in Figure 4.8. The ORP's typical dry weather behavior exposes two slight peaks at about 400 min and 1000 min, respectively. Furthermore, as for Turbidity, it was detected that the windows in which the tendency values are maximum (Figure 4.8 g, h, i), the variability is also maximum (for the first peak of ORP, which was reported at 400 min, the expectation is about -240 mV, with an interquartile distance (IQD) of 160 mV; the second peak of ORP, reported at approximately 1000 min, had an expectation of -260 mV, with an IQD of -140 mV). The behavior of the ORP, as for the other studied water quality parameters (e.g. DO, TSS, Turbidity), seems to be independent of the selection of the rain gauge station for undertaking the dry typical weather day's classification. This fact can lead to conclude that there is not a particular rain gauge station which shows a very different rainfall patterns from the others. Therefore, the dry weather characterization can be undertaken without detailing into spatial variability of the rainfall, regarding the ORP.

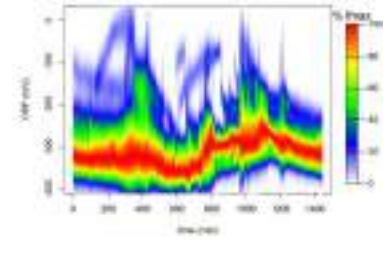
*Fontibon*

a)



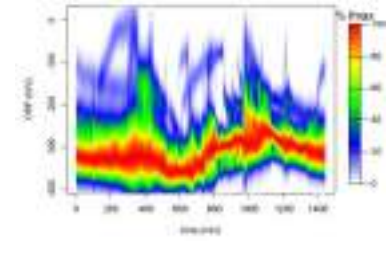
*Bosa*

b)

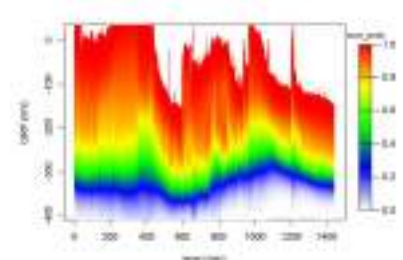


*Casablanca*

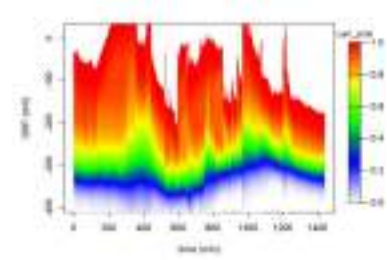
c)



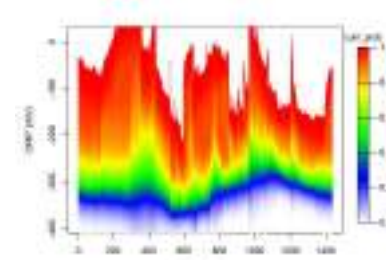
d)



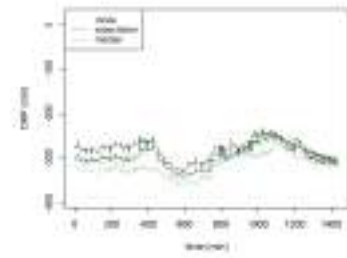
e)



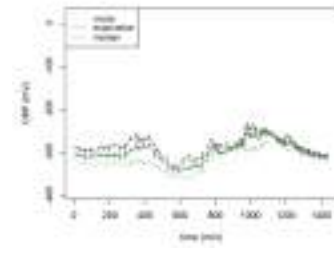
f)



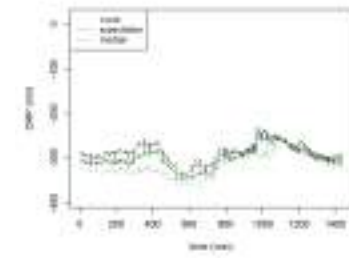
g)



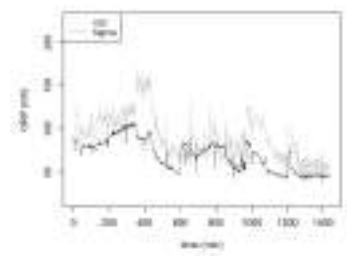
h)



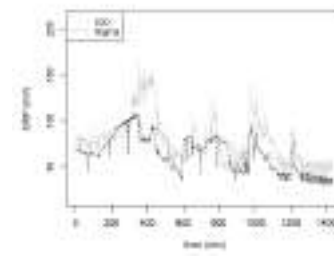
i)



j)



k)



l)

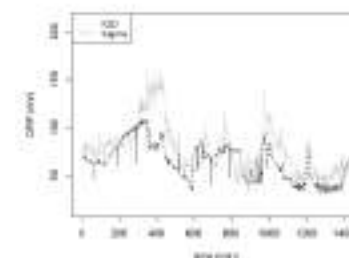


**Figure 4.8 Dry weather probabilistic characterization for ORP measured by uni-parametric sensors, using rainfall registrations at different rain gauge stations (*Fontibon* (left), *Bosa* (center) and *Casablanca* (right)); a, b, c) Probability density function (PDF) for each time-step; d, e, f) Cumulative probability function (CPF) for each time-step; g, h, i) Mode, Median and Expectation for each time-step; j, k, l) Interquartile distance and standard deviation for each time-step**

For the case of Temperature, the obtained results are exposed in Figure 4.9.

## Temperature

*Fontibon*

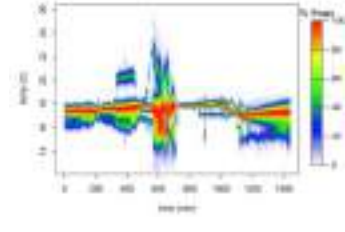
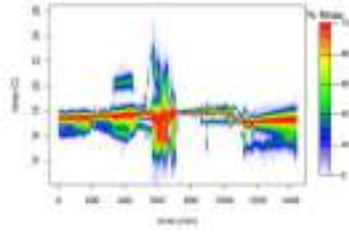
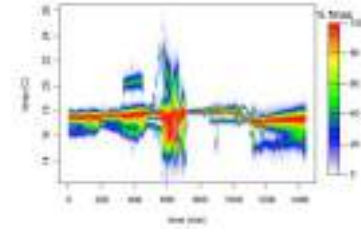
*Bosa*

*Casablanca*

a)

b)

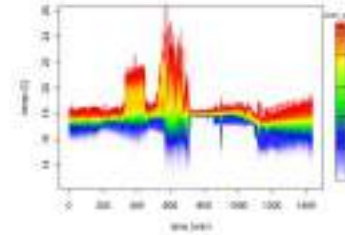
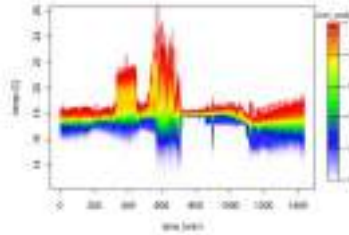
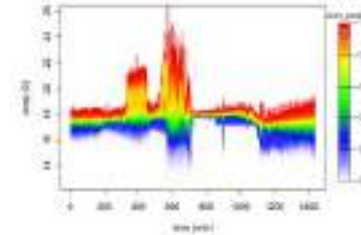
c)



d)

e)

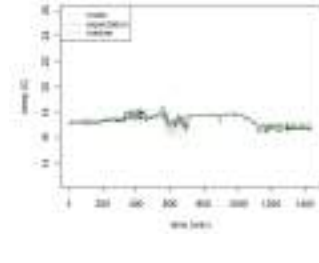
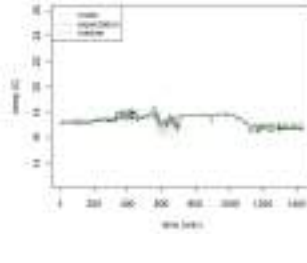
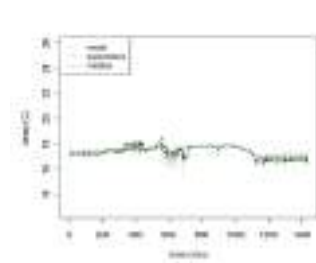
f)



g)

h)

i)



j)

k)

l)

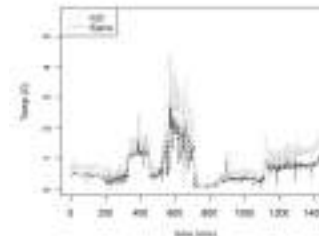
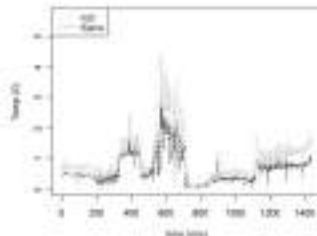
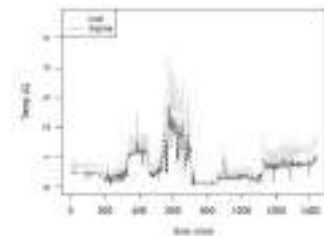


**Figure 4.9** Dry weather probabilistic characterization for Temperature measured by uni-parametric sensors, using rainfall registrations at different rain gauge stations (*Fontibon* (left), *Bosa* (center) and *Casablanca* (right)); a, b, c) Probability density function (PDF) for each time-step; d, e, f) Cumulative probability function (CPF) for each time-step; g, h, i) Mode, Median and Expectation for each time-step; j, k, l) Interquartile distance and standard deviation for each time-step

From the above results it can be concluded that Temperature seems to have a constant behavior over time, during dry weather. An important peak of variability was detected at about 600 min (Figure 4.9 j, k, l), although the tendency values were not higher (Figure 4.9 g, h, i) for this time-step. Once again, the dry weather behavior seems to be independent of the rainfall registers used for delimitating the dry days data (*Fontibon*, *Bosa* or *Casablanca*).

Regarding the pH measurements undertaken by means of the uni-parametric sensors, the following results presented in Figure 4.10 were obtained. The pH values during a typical dry day seem to have a maximum value at about 550 min, of 8.2 pH units. The variability of the pH, during a dry day, seems to be greater for the time window ranging from 400 min to 1000 min, but without a detectable peak (as for the cases of Temperature and Turbidity). The lowest values of pH (about 7.4 pH units) were detected at the beginning and ending of the peak.

*Fontibon*

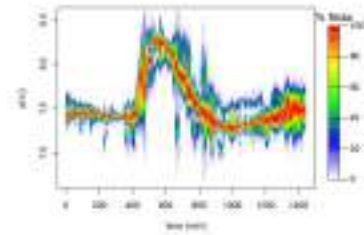
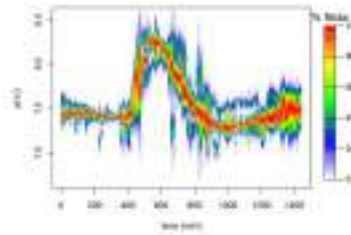
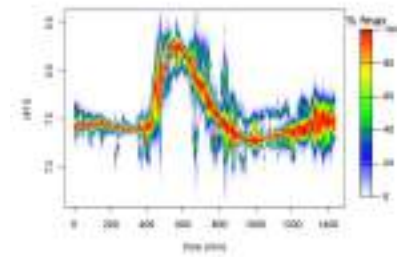
*Bosa*

*Casablanca*

a)

b)

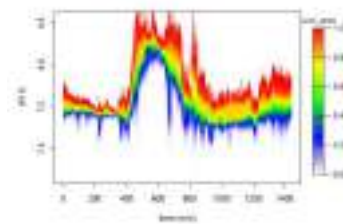
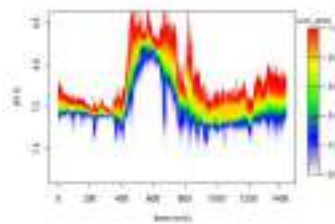
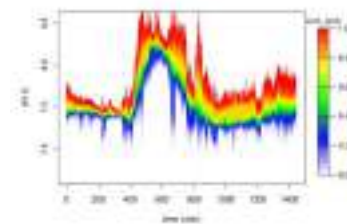
c)



d)

e)

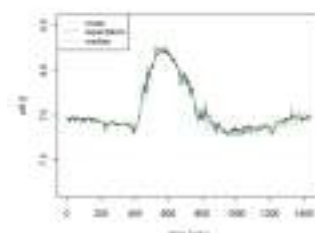
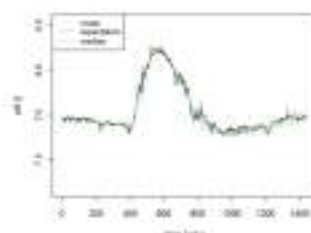
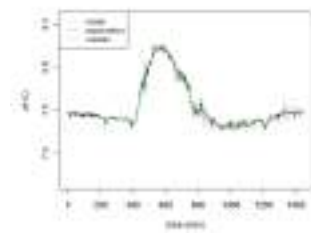
f)



g)

h)

i)



j)

k)

l)

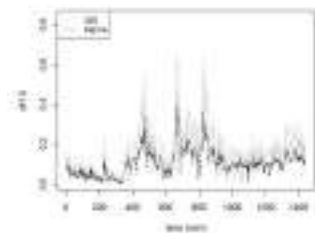
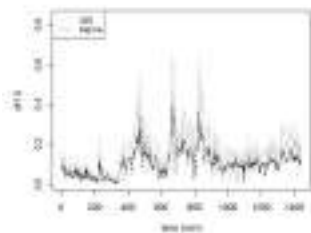
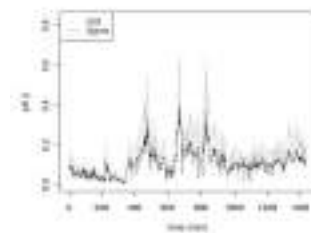


**Figure 4.10** Dry weather probabilistic characterization for pH measured by uni-parametric sensors, using rainfall registrations at different rain gauge stations (*Fontibon* (left), *Bosa* (center) and *Casablanca* (right)); a, b, c) Probability density function (PDF) for each time-step; d, e, f) Cumulative probability function (CPF) for each time-step; g, h, i) Mode, Median and Expectation for each time-step; j, k, l) Interquartile distance and standard deviation for each time-step

An overall dry weather characterization for different water quality parameters measured by means of uni-parametric sensors was undertaken. Parameters such as DO and TSS did not report a

particular structure over the behavior of the variability. On the other hand, water quality parameters such as ORP, Turbidity and Temperature exposed important and well-defined peaks of variability during a typical dry weather curve. In terms of tendency statistics (expectation, mode, median), the turbidity and the TSS exhibit very similar behaviors during a dry day. This fact can be explained due to the correlation between Turbidity and TSS reported by different studies (*e.g.* Mishendo *et al.*, 1999). A decrement of values just before an increasing tendency with maximum value at about half of the day was observed, for both water quality parameters. The ORP exposed two slight peaks during a typical day, different than for the pH values, which registered just one high peak at the half of the day. For all analyzed water quality parameters, the selection of which rain gauge station's registers should be used for defining dry days showed to be almost equivalent for all of the three studied rain gauge stations (*Fontibon, Bosa and Casablanca*).

#### 4.1.2.3 UV-VIS spectrometry

The typical dry weather behavior was also assessed for the measurements undertaken by means of the UV-VIS spectrometry sensor. As for the PCA, the dry weather probabilistic characterization was divided into two groups: (i) a spectral analysis, including the wavelengths from 250 to 750 individually; (ii) an analysis for some equivalent water quality parameters obtained from the spectra (equivalent TSS and COD concentrations), by the use of the global calibration's model of the equipment. In both cases, the characterization was executed independently for the three different rain gauge stations (*Fontibon, Bosa and Casablanca*).

##### 4.1.2.3.1 Spectral Analysis

Regarding the spectra reported by the UV-VIS spectrometry, the results were obtained by classifying the days as “dry weather” by the use of the rainfall registered at the *Fontibon* rain gauge station. Furthermore, the typical dry weather characterization of the spectra was divided into three parts: (i) the nitrate-nitrites range (wavelengths from 200 nm to 250 nm), (ii) the UV range (wavelengths from 250 nm to 380 nm) and (iii) the VIS range (wavelengths from 380 nm to 750 nm). The Spectral Analysis has a particularity in terms of the number of analyzed variables (in contrast to the Echo Profile Flowmeter, the uni-parametric sensors and the equivalent water quality parameters): as each wavelength can be proposed as an independent variable, the total number of independent variables will be of about 220 (from wavelength 200 nm to 750 nm, measured each 2.5 nm). With the purpose of detecting for which variables it seemed to be more appropriate to undertake the dry weather characterization, additional calculations were assessed. As a first approach, a Correlation Test was proposed in order to appreciate the relation among the different wavelengths.

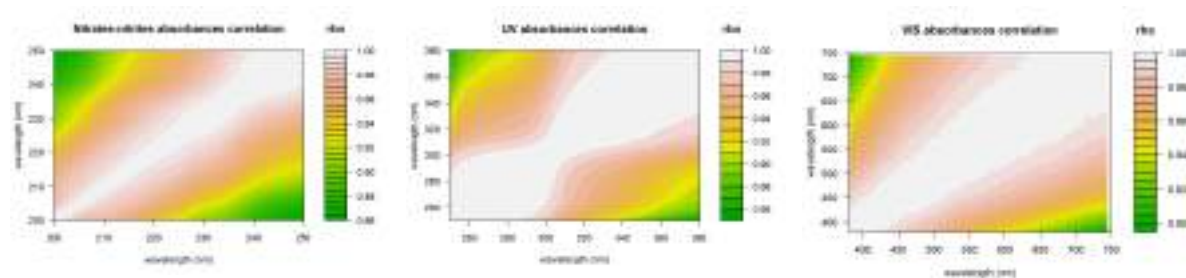


**Figure 4.11** Correlation value among the different absorbance values registered during the measurement campaign, for each wavelength.

From Figure 4.11, it can be observed that the closer the wavelength value is, the more correlated their respective absorbance's data are. This behavior is more evident for the VIS range than for the UV and Nitrites-Nitrates ranges. Some discontinuities for wavelengths about 310 nm in the correlation trend over the different wavelengths can be noticed. Therefore, it can be concluded that for wavelengths about 310 nm the correlation with neighbors is less strong than for other wavelengths of the UV spectra.

High correlation values computed between neighbor wavelengths may lead to conclude that by analyzing the appropriate wavelengths, the influence of rainfall over a given wavelength will be equivalent for its neighbor wavelengths. As exposed in the PCA (see section 4.1.1.2.1), the extreme wavelengths in the whole UV-VIS range (*e.g.* 250 nm and 750 nm) seems to have the most relevant influence over the differences detected in the variability of dry and wet weather data. Therefore, by undertaking the dry weather probabilistic characterization over the most extreme wavelength values, the result can be considered as representative enough of the dry weather behavior of the spectra, measured by means of the UV-VIS spectrometry sensor.

In this sense, the dry weather characterization is exposed for the wavelength 205 nm and for the wavelength 745 nm. It is important to clarify that the extreme wavelengths (200 nm and 750 nm) were not chosen due to a higher presence of not registered values "NA". First, results of dry weather characterization for wavelength 205 nm are presented (Figure 4.12). The dry weather characterization was alike to *Fontibon*, *Bosa* and *Casablanca* rain gauge stations.

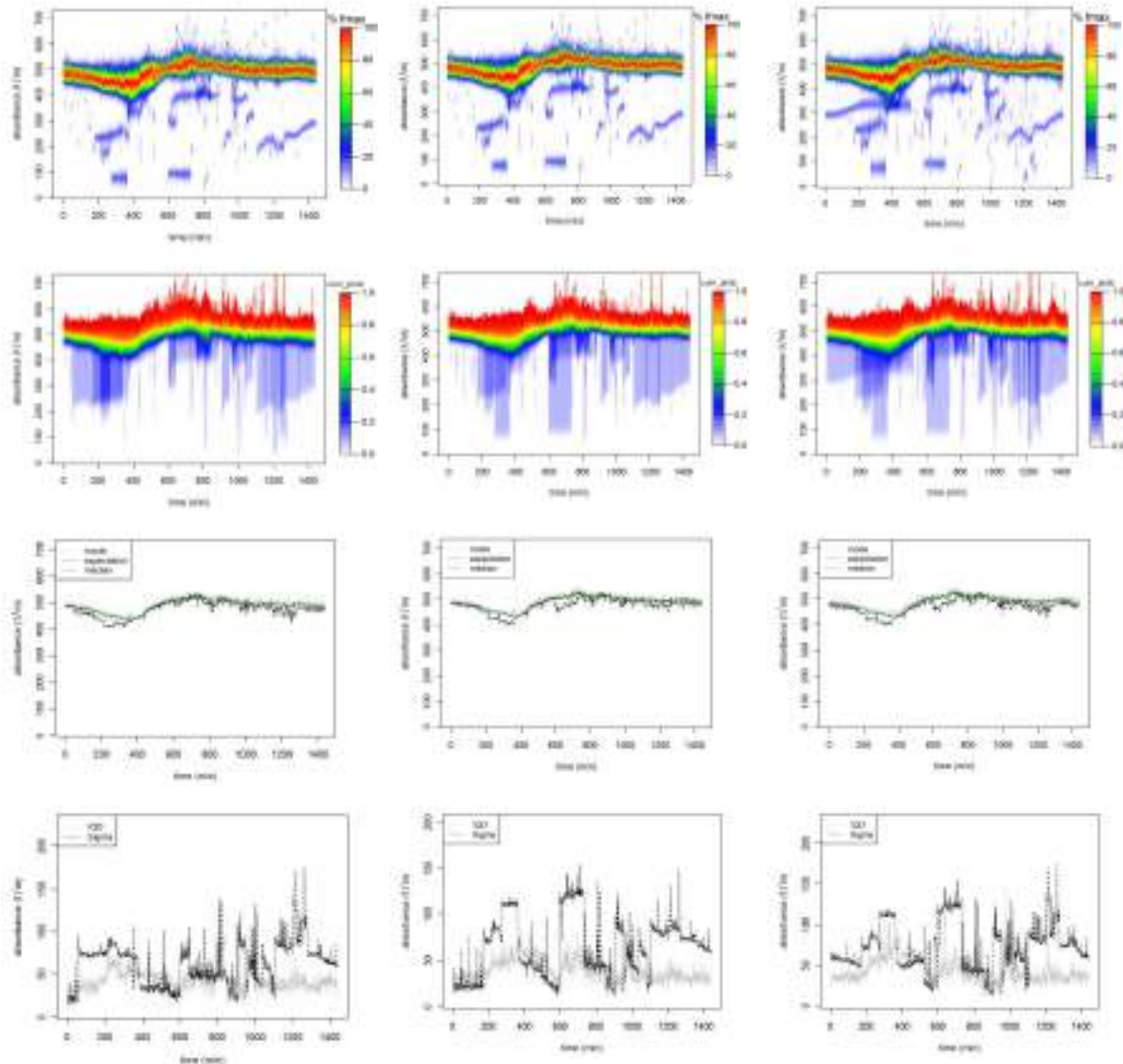


Figure 4.12 Dry weather probabilistic characterization for absorbance values registered in wavelength 205 nm (Nitrites-Nitrates range) by means of the UV-VIS spectrometry sensor, using rainfall registrations at different rain gauge stations (*Fontibon* (left), *Bosa* (center) and *Casablanca* (right)); a, b, c) Probability density function (PDF) for each time-step; d, e, f) Cumulative probability function (CPF) for each time-step; g, h, i) Mode, Median and Expectation for each time-step; j, k, l) Interquartile distance and standard deviation for each time-step

It can be concluded from the tendency values that, over the typical dry weather day curve, the only window in which the slope was positive (increasing of absorbance over time) seems to happen in a window from 400 min to 800 min. However, these changes of slope in the curve are small. There were no easily detectable variability peaks. No important variations from the behavior computed by dry weather registered at different rain gauge stations (*Bosa*, *Fontibon* and *Casablanca*) were detected.

The results obtained for the dry weather characterization regarding wavelength 745 nm are shown in Figure 4.13.

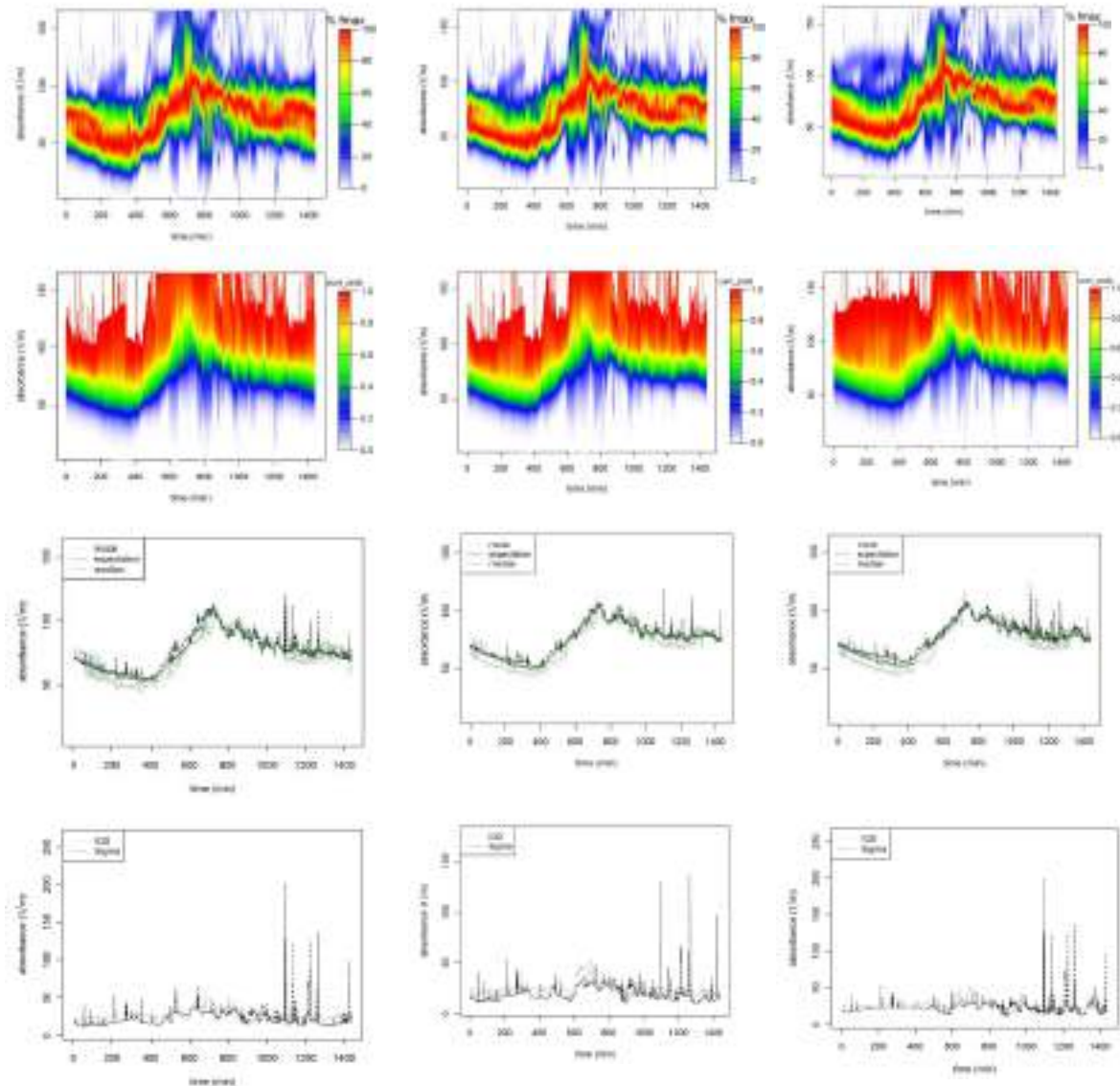


**Figure 4.13** Dry weather probabilistic characterization for absorbance values registered in wavelength 745 nm (VIS range) by means of the UV-VIS spectrometry sensor, using rainfall registrations at different rain gauge stations (*Fontibon* (left), *Bosa* (center) and *Casablanca* (right)); a, b, c) Probability density function (PDF) for each time-step; d, e, f) Cumulative probability function (CPF) for each time-step; g, h, i) Mode, Median and Expectation for each time-step; j, k, l) Interquartile distance and standard deviation for each time-step

The peak near the middle of the typical dry day was more pronounced than for the case of wavelength 205 nm. Hence, the variability statistics were more constant over time than for the case of wavelength 205 nm. This curve seems very similar to the TSS and Turbidity results exposed in section 0. (uni-parametric sensors) shown in Figure 4.6. This fact suggests a relation between high wavelengths (e.g. 745 nm) and the Turbidity and TSS values measured by the uni-parametric sensors. No significant differences were appreciated in terms of dry weather characterization among the studied rain gauge stations (*Fontibon*, *Bosa* and *Casablanca*).

#### 4.1.2.3.2 Equivalent water quality parameters

By the use of the UV-VIS sensor global's calibration, equivalent TSS and COD concentration values can be registered. Hence, the same approach can be implemented for characterizing a typical dry day.

*Fontibon*

*Bosa*

*Casablanca*

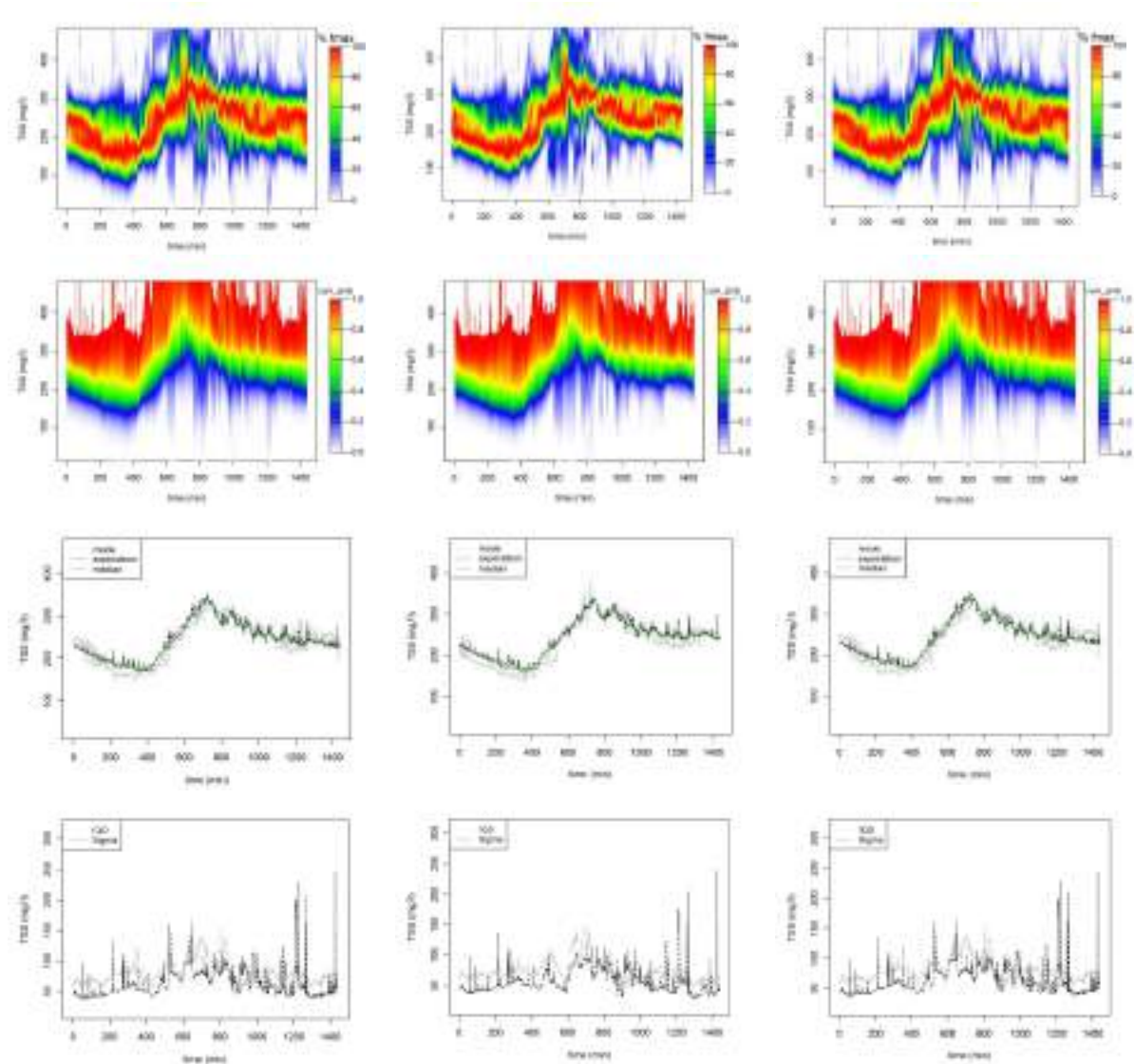


**Figure 4.14** Dry weather probabilistic characterization for TSS values registered by means of the UV-VIS spectrometry sensor, using rainfall registrations at different rain gauge stations (*Fontibon* (left), *Bosa* (center) and *Casablanca* (right)); a, b, c) Probability density function (PDF) for each time-step; d, e, f) Cumulative probability function (CPF) for each time-step; g, h, i) Mode, Median and Expectation for each time-step; j, k, l) Interquartile distance and standard deviation for each time-step

Similarities between the dry weather characterizations undertaken by the use of uni-parametric sensors (section 0) and by means of the UV-VIS spectrometry sensor (Figure 4.14) can be detected. Both curves report comparable behaviors in terms of the occurrence of a peak over the half of the dry day, although the peak seems more noticeable for the case of UV-VIS spectrometry equivalent TSS concentration values' computations. However, two additional differences can be observed: (i)

for the TSS values reported by the uni-parametric sensors, variability seems lower than for the case of UV-VIS spectrometry (Figure 4.6 and Figure 4.14 j, k, l), (ii) the reported TSS values seems to be meaningfully higher for the uni-parametric sensors than for the UV-VIS spectrometry measurements (about 350 mg/l to 1500 mg/l). This fact can be explained by the use of the default calibrations of both equipments: the UV-VIS spectrometry sensor seems to report more reasonable TSS values than the uni-parametric sensors, without a calibration for the specific wastewater (previous studies has established a mean TSS concentration of 233 mg/l, for the studied zone (HNV, 2007)). On the other hand, for this research, the absolute concentration's values were definitely not relevant. As the aim of these paragraphs was mainly to characterize relative behaviors during a random dry day (without rainfall influence), without absolute and generalizable values (calibration's scope), both equipments (uni-parametric and UV-VIS spectrometry sensors) were used with their default calibrations without additional calibration analyses including samples and laboratory tests. As for uni-parametric sensors, the equivalent TSS concentration values reported by means of the UV-VIS exposed a similar behavior with the spectral analysis, for wavelength 745 nm (see Figure 4.13). No noteworthy differences in the dry weather probabilistic characterization were identified by the use of the different registers reported in the studied rain gauge stations (*Fontibon, Bosa and Casablanca*).

The Chemical Oxygen Demand (COD) registered by UV-VIS Spectrolyer's global calibration was analyzed, regarding its probabilistic behavior during dry weather. Results are given in Figure 4.15.

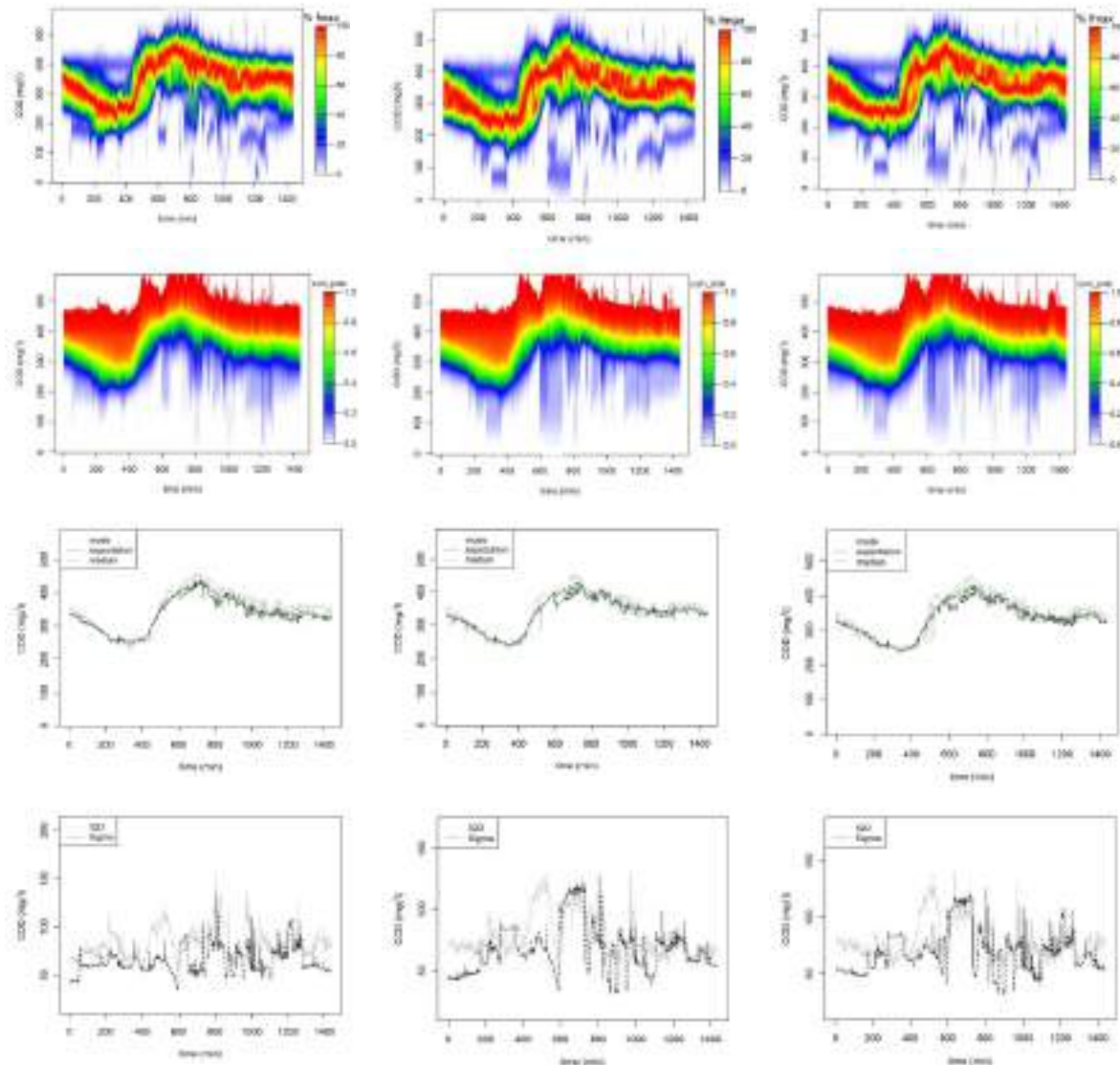


**Figure 4.15** Dry weather probabilistic characterization for COD values registered by means of the UV-VIS spectrometry sensor, using rainfall registrations at different rain gauge stations (*Fontibon* (left), *Bosa* (center) and *Casablanca* (right)); a, b, c) Probability density function (PDF) for each time-step; d, e, f) Cumulative probability function (CPF) for each time-step; g, h, i) Mode, Median and Expectation for each time-step; j, k, l) Interquartile distance and standard deviation for each time-step

From the above results, it can be stated that the COD exposed a certain similarity with the TSS, independently of the measurement sensors (UV-VIS spectrometry or uni-parametric sensors), in terms of reporting a peak closer to the half of the typical dry weather day (12h00) (result that seems to be in agreement with other studies such as Rodríguez *et al.*, 2011). However, it can be noticed, from the TSS's dry weather probabilistic characterization (Figure 4.15), that TSS behavior seems to be more alike to the wavelength 745 nm's absorbance values, during dry days. Therefore, for TSS and wavelength 745 nm, the crest of the probabilistic curve seems sharper than for the COD curve, in which a kind of "S", with a smoother peak, can be appreciated. However, the probabilistic dry weather characterization for COD seems not as similar as the similarity detected between TSS and

wavelength 745 nm. Therefore, it can be stated that the COD probably is not well represented by means of a single wavelength. Hence, for calculating the COD from the wavelengths' spectra, several wavelengths (including the VIS range) should be studied.

In general terms, from the results obtained by means of the UV-VIS spectrometry sensor, it can be concluded that the dry weather characterization for extreme wavelengths of the registered spectra (regarding the PCA results, most relevant wavelengths in terms of explaining dry and rainy weather variability) could be undertaken. Hence, a probabilistic dry weather characterization of the COD and TSS values reported by means of the equipment's global calibration was assessed. From these calculations, in addition with results obtained for uni-parametric sensors, it could be concluded that: (i) the TSS dry weather's probabilistic behavior was similar for the uni-parametric and the UV-VIS spectrometry sensors, (ii) the dry weather characterization could be assessed for uni-parametric and the UV-VIS spectrometry sensors by means of the equipment's default calibration, (iii) water quality parameters such as TSS and Turbidity exposed a sharper curve during a typical dry day than parameters such as COD, in which their typical dry day's curve seemed to be smoother, (iv) the variability for a same water quality parameter (*e.g.* TSS) seems to be different for measurements undertaken by means of the uni-parametric or the UV-VIS spectrometry sensor, (v) most of the water quality parameters seemed to expose a two tailed PDF for each time step, except for the DO (which was registered as approximately 0.0 mg/l for most of the dry weather time), (vi) the dry weather probabilistic characterization for wavelengths at the end of the VIS range (about 745 nm) was similar to the TSS dry weather characterization, (vii) no substantial differences in characterizing the probabilistic behavior of water quality measurements were detected between the rainfall registered in *Fontibon*, *Casablanca* or *Bosa* rain gauge stations. This fact means all the studied rain gauge stations have an influence over water quality. Whenever it is the case, if the dry weather characterization is highly different for a given rain gauge station, it could be concluded that during the rainy days the water quality/quantity signal will not probably be detected as ranging outside of the typical dry weather behavior. The results obtained for the TSS (measured by uni-parametric and UV-VIS) and the COD seems to have similarities with related studies (Rodríguez *et al.*, 2011) in terms of the mean value behavior during a typical dry day. However, the variability was assessed in the present document as very volatile during the typical dry weather day (see Figure 4.6, Figure 4.14 and Figure 4.15), in contrast to this previous experiences (in which the variability was reported almost constant during the typical dry day).

#### **4.1.3 Identification of rainfall pulses and water quality's time windows ranging outside typical dry weather behavior**

Based on Median Filtering and Step Detection techniques, time windows in which rainfall pulses occur and water quality measurements ranged outside their typical dry weather behavior, were identified. Before estimating when a given water quality time series ranges outside its typical behavior, additional calculations are required. By the use of the dry weather probabilistic characterization results, the probability that a given water quality measurement, at a specified time-step, is part of typical dry weather behavior was computed. Whenever this probability value is lower than 0.05, the analyzed water quality measurement can be considered part of a rainfall event.

#### 4.1.3.1 Rainfall Pulses identification

The rainfall pulses identification consisted in delimitating the time windows in which considerable rainfall intensity was registered (rainfall mean intensity greater than 0.1 mm/h). With the purpose of undertaking this identification, the first step was to filter the rainfall signal registered at a certain rain gauge station, by means of a Median Filter. This filtering technique will expose abrupt changes of the signal, throughout the time; also isolated noisy data “outlier” will then be removed. Therefore, by the use of a Step Detection algorithm (based in Mann-Whitney statistical test), time steps in which abrupt changes took place over the rainfall time series are detected. The described methods were applied to rainfall time series registered at each of the studied rain gauge stations (*Fontibon*, *Bosa* or *Casablanca*). As a key parameter, the time window length used in the Median Filter must be defined. With the purpose of identifying an appropriate length of the time window used in the median filtering, the autocorrelation functions ACF for each time series can be computed. Results regarding the ACF for the *Fontibon*, *Bosa* and *Casablanca* rain gauge stations are presented in Figure 4.16.

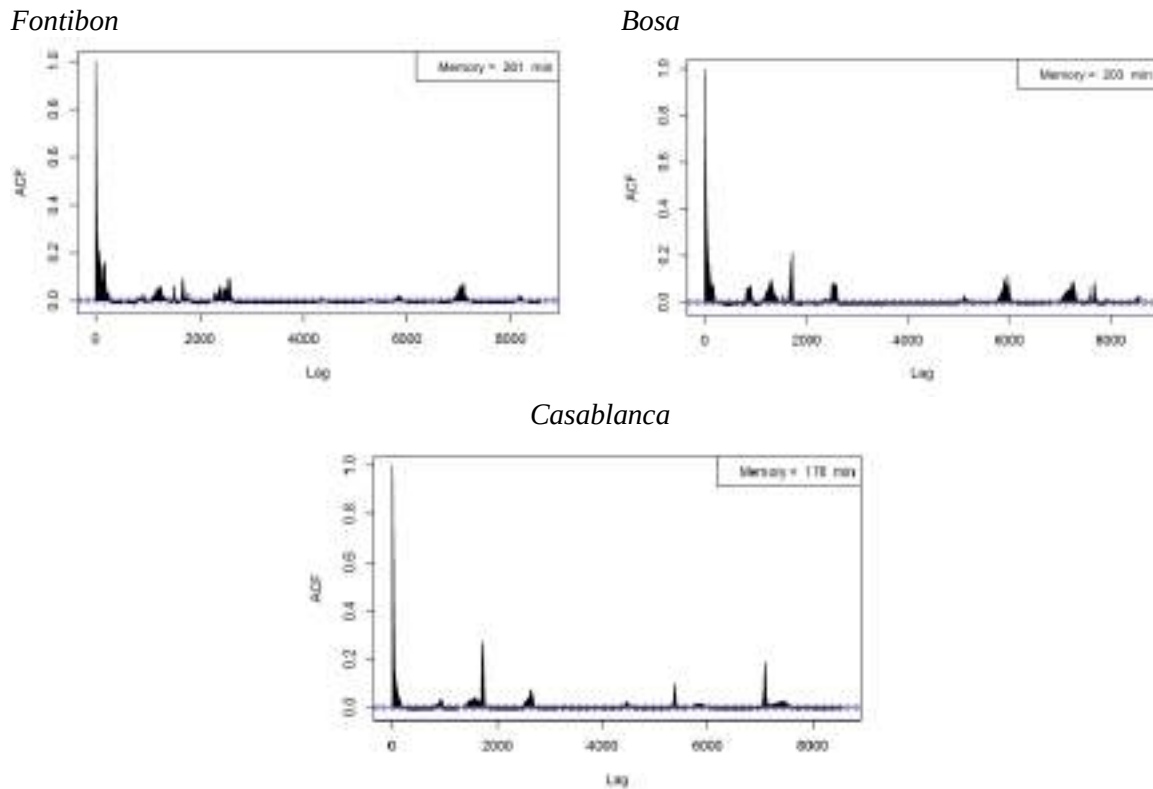


**Figure 4.16** Autocorrelation functions for rainfall registered at the *Fontibon*, *Bosa* and *Casablanca* rain gauge station

The ACF generally has the characteristic of decaying with the number of lags until the correlation's value is less than the value expected for a Gaussian noise process (dashed line, Figure 4.16). In this sense, the lags for which the ACF decayed as described were: (i) *Fontibon* 261 min, (ii) *Bosa* 203 min and (iii) *Casablanca* 176 min. These values can be interpreted as estimates of the physical process' memory. In other words, a rainfall registered at time-step  $t$ , at a studied rain gauge station, starts to “forget” the observations from the past when the above reported times had lapsed, respectively. Therefore, the length of the windows used for applying the Median Filter over the

rainfall registered at each of the studied rain gauge stations was the memory of the process (lags for which the ACF has decayed enough). Once the rainfall signal related to each of the studied rain gauge stations was filtered by means of a Median Filter, the proposed Step Detection algorithm was applied over the mentioned time series. Thus, the points in which an abrupt change in the filtered rainfall signals occurred are shown by the grey lines in Figure 4.17.

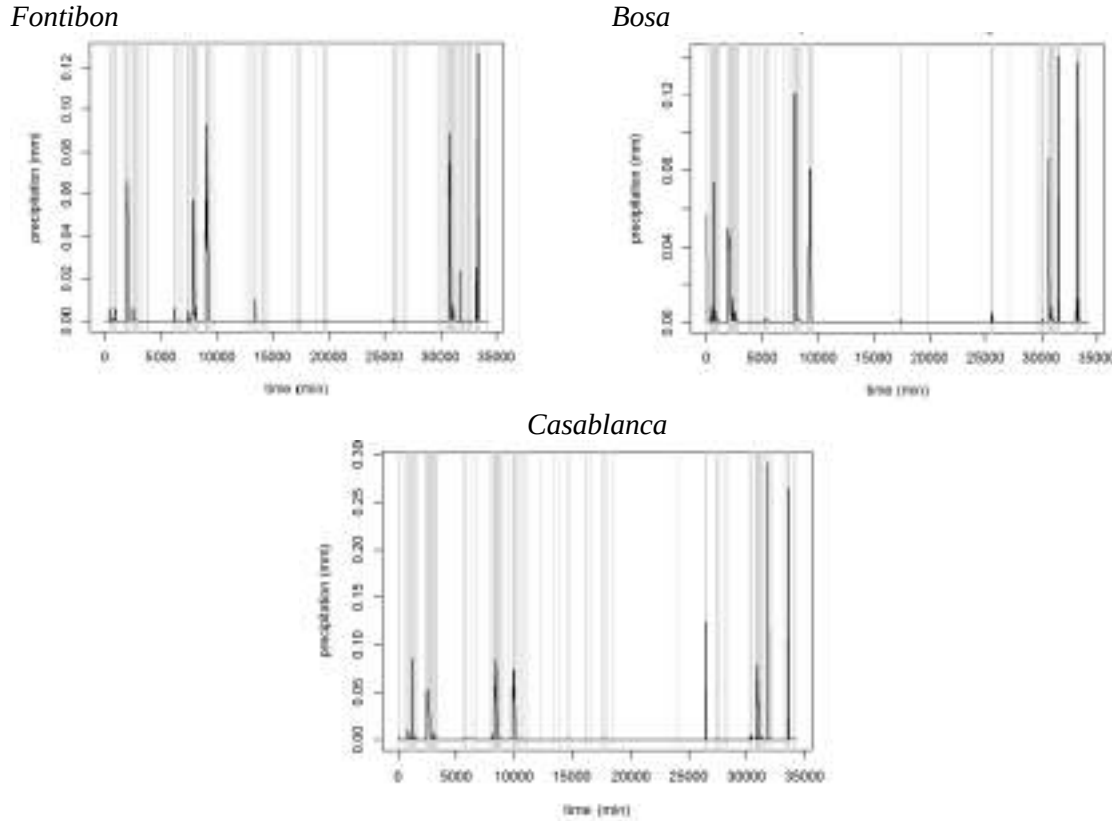


**Figure 4.17** Step detection over the filtered signal for rainfall registered at the *Fontibon*, *Bosa* and *Casablanca* rain gauge stations

In all intervals composed by two neighbor points (grey lines) in which an abrupt change of the signal was detected, the condition that the mean rainfall intensity reported during these intervals was greater than 0.1 mm/h, was then evaluated. By means of these calculations the rainfall pulses were identified, for the rainfall registered at the *Fontibon*, *Bosa* and *Casablanca* rain gauge stations. The results are exposed coupled with the identification of the time windows in which water quality parameters were ranging outside typical dry weather behavior (see Figure 4.17).

#### 4.1.3.2 *Water quality's time windows ranging outside typical dry weather behavior*

For rainy days, a specified water quality or quantity parameter measured by means of an online sensor will present a different behavior compared to a dry day. Hence, by means of the probabilistic dry weather probabilistic characterization, an evaluation of the time-steps in which water quality or quantity measurements ranges outside of their typical behavior, with an associate probability, can be assessed. In order to report complete time windows in which a given water quality or quantity signal ranges outside its typical dry weather behavior (instead of isolated observations), a Median

Filter and a Step Detection algorithm were applied. For each neighbor points in which the Step Detection algorithm detected abrupt changes over the Median Filtered “probability of being dry” signal, the mean probability of being part of dry weather behavior was calculated for all data reported during this time interval. Whenever this probability value was lower than 0.05 (average probability of being part of the typical weather behavior), the water quality or quantity signal registered during the analyzed time window was considered to be influenced by rainfall.

#### 4.1.3.2.1 Echo Profile Flowmeter

For the flow rate measurements, the results obtained are exposed in this section. Hence, the rainfall pulses’ identification is included. As rainfall pulses depend on rainfall registrations and the dry weather characterization was assessed by rainfall registered at different rain gauge stations, the results are presented independently, in terms of each rain gauge station considered (*Fontibon*, *Bosa* or *Casablanca*). Therefore, for every time-step the probability of being part of dry weather was calculated by means of each dry weather probabilistic characterization. Thus, the resulting signal (probability of dry weather signal) was filtered by means of a Median Filter. Afterwards, the Step Detection algorithm was applied over the filtered probability of being part of a dry weather signal, in order to identify abrupt changes.

A fundamental parameter for the Median Filter is the length of the mobile window; it was assessed based on the ACF of the flow rate signal, measured at the effluent of the catchment. The results of the ACF are shown in Figure 4.18.

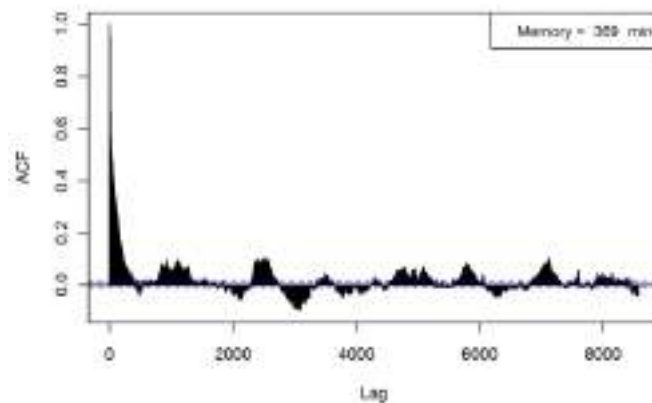


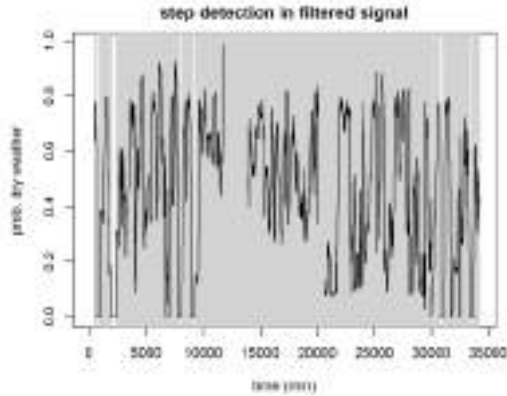
**Figure 4.18 Autocorrelation function of flow rate, considering the complete measurement period**

From the exposed results, it can be concluded that the process’ memory is about 370 min. Thus, an appropriate recommendation for implementing the Median Filtering technique over the flow rate signal consists in using a mobile window which includes approximately 370 observations, independently of the rain gauge station which is being analyzed.

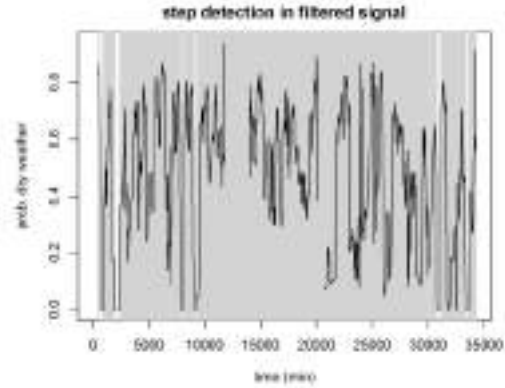
Once the step detection algorithm is applied over the Median Filtered signal, leading to identify points in which abrupt changes over the dry weather probability signal occurred, the mean value of probability of dry weather between the neighbor abrupt changed points (from Step Detection algorithm) was calculated. Whenever this value was lower than 0.05 and the analyzed window showed to be close to a rainfall pulse, it was considered that this time window was ranging outside

typical dry weather behavior. Results for rainfall registered at the *Fontibon*, *Bosa* and *Casablanca* rain gauge stations are summarized in Figure 4.19.

*Fontibon*



*Bosa*



*Casablanca*

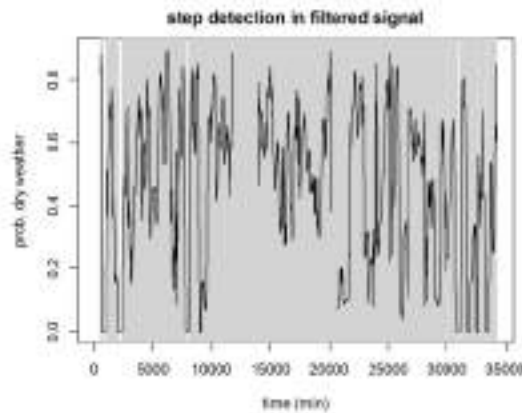
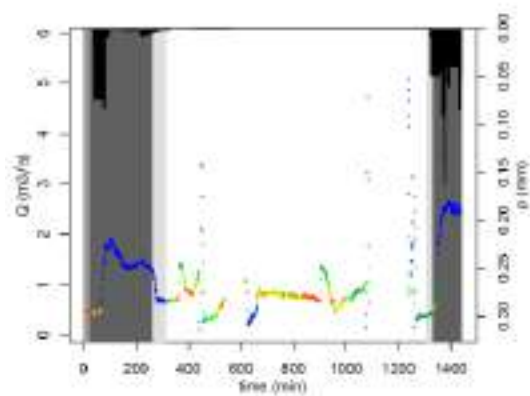


**Figure 4.19 Probability of dry weather signal filtered by a Median Filter, for Flow rate signal**

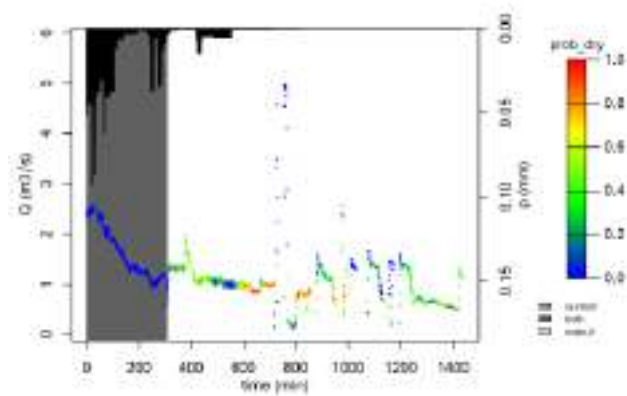
The Figure 4.19 exposes the identified rainfall events in the output signal (flow rate), for each of the studied rain gauge stations (*Fontibon*, *Bosa* and *Casablanca*) (white stripes). Two conditions are accomplished by the identified time windows: (i) the mean value of probability of dry weather (filtered) is in average lower than 0.05, (ii) abrupt changes in probability signal can be observed at the beginning and ending of the time window -inside the white stripes no significant change over the signal was detected (Step Detection algorithm, Mann Whitney test)-. For the case of *Fontibon* rain gauge station, the events identified in terms of rainfall and/or the output signal, are presented in Appendix 8.8.

It can be appreciated that a total of six rainfall events (Appendix 8.8: a-b, c-d, d, e-f, f, g) could be identified, including the rainfall pulses and their consequences over the output (flow rate). As exposed, groups of values with a low probability of being part of dry weather behavior (blue points) were detected by means of the proposed methods. The first hydrograph reported in the exposed time series was not analyzed, due to its distance of a rainfall pulse. This fact may lead to conclude that the mentioned hydrograph could be caused by rainfall registered at a different station than the *Fontibon* rain gauge station. The same procedure was then applied, but considering the rainfall registered at the *Bosa* rain gauge station. The obtained results are given in Figure 4.20.

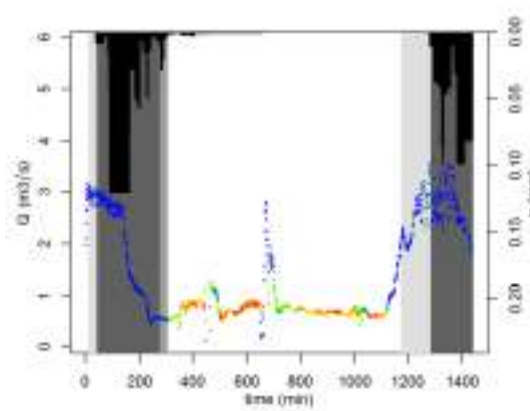
a)



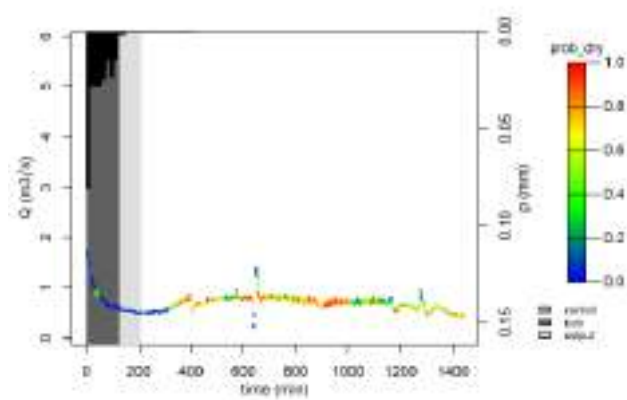
b)



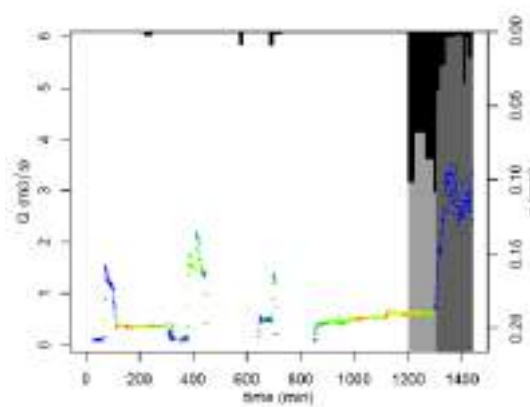
c)



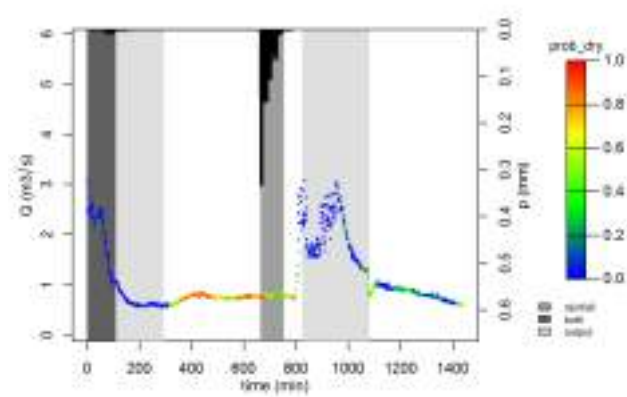
d)



e)



f)



g)

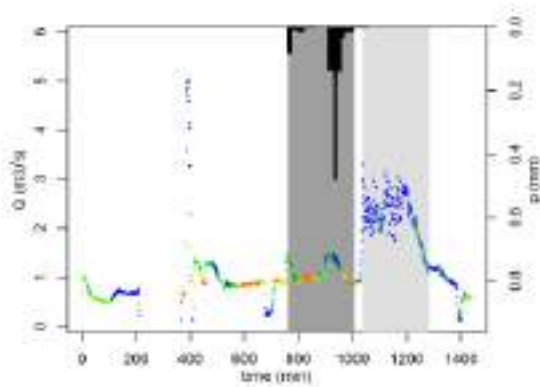


**Figure 4.20** Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding Flow rate (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Bosa* rain gauge station.

From the above results, it can be stated that the identification of rainfall events in terms of rainfall pulses and the output of the catchment (flow rate), by the use of rainfall registered at the *Bosa* rain gauge station, could be assessed for a total of seven rainfall events (Figure 4.20: a, a-b, c, c-d, e-f, f, g). It can be noticed that here the first hydrograph registered for the rainy days’ data was then identified as a rainfall event. This fact can be explained by the temporal closeness to a rainfall pulse. Another aspect to be distinguished is that isolated blue points (low probability values) were not included as time windows in which flow rate was ranging outside its typical dry weather behavior. This situation can be explained by means of the median filtering technique’s properties, applied to the “probability of dry weather” time series. The described procedure was finally also applied to the flow rate signal, compared with rainfall pulses detected at the *Casablanca* rain gauge station. Results are described in Appendix 8.9.

For the case of the *Casablanca* rain gauge station, only three rainfall events (Appendix 8.9: a, c-d, d) could be identified by means of the proposed methodology. Due to their temporal farness, rainfall pulses isolated from time windows identified at the output were discarded.

#### 4.1.3.2.2 Uni-parametric sensors

For the case of water quality measurements undertaken by means of the uni-parametric sensors (DO, TSS, Turbidity, ORP, Temperature and pH), the methodology described in section 3.5 was applied as for section 4.1.3.2.1 (for flow rate measurements). Regarding the results, for implementing the Median Filter over the dry weather probability signal, the memory of the process was estimated by the use of the autocorrelation function (ACF), for each of the analyzed water quality parameters. The results are shown in Figure 4.21:

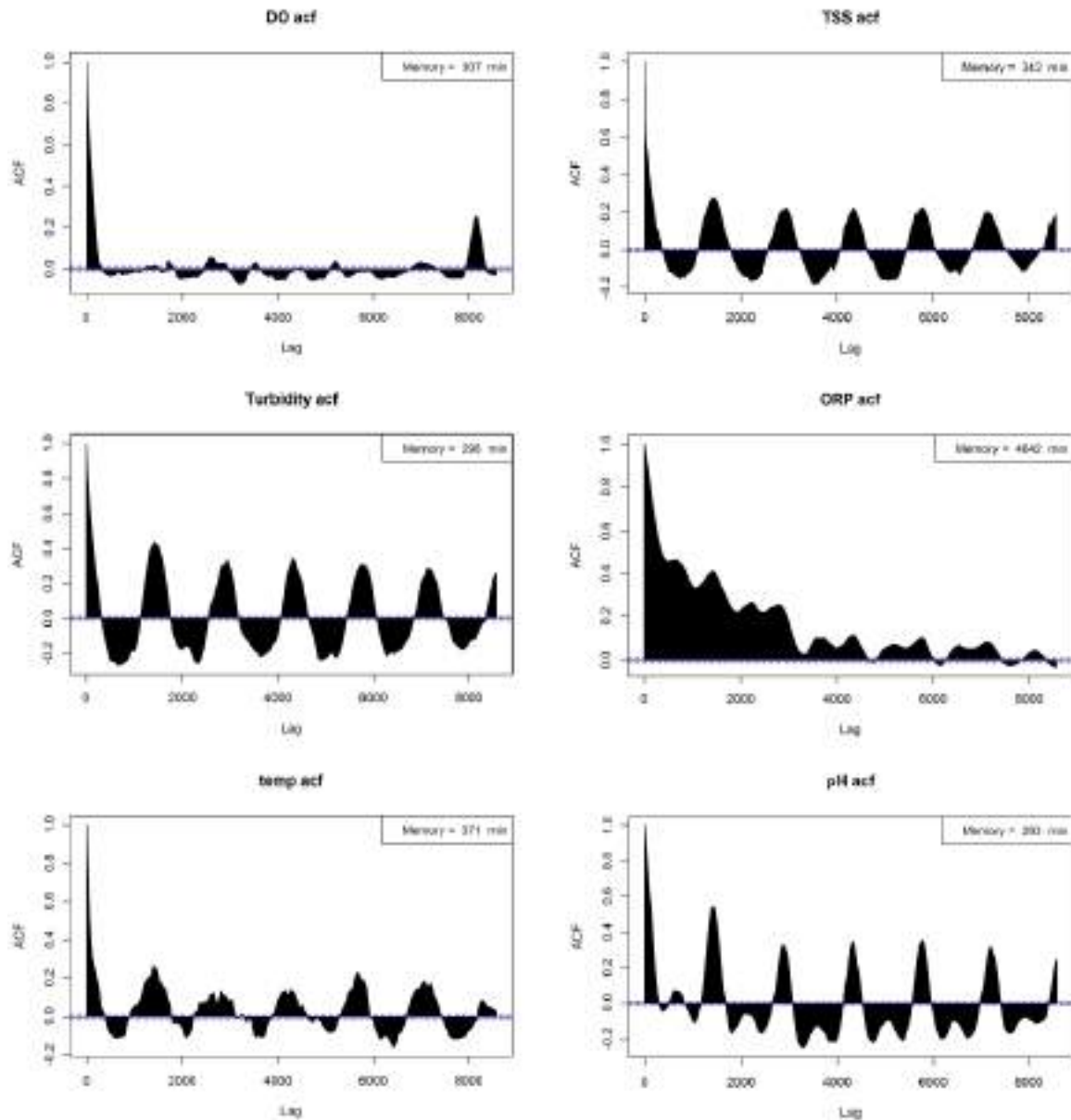


**Figure 4.21 Autocorrelation functions for DO, Total Suspended Solids, Turbidity, ORP, Temperature and pH; measured by uni-parametric (considering the complete measurement period)**

From the exposed results, recommended lengths of the mobile window for applying the Median Filter over the probability of dry weather signal, for each water quality parameters are: (i) 307 min for DO, (ii) 342 min for TSS, (iii) 296 min for Turbidity, (iv) 4643 min for ORP, (v) 371 min for Temperature and (vi) 293 min for pH.

Additional conclusions can be formulated from the results regarding the ACF of water quality parameters measured by means of uni-parametric sensors. The Turbidity and TSS showed a more periodical behavior than the other parameters. This leads to conclude that daily patterns for TSS and Turbidity expose a lower variability than parameters such as ORP and DO. On the other hand, the

ORP parameter showed to be a process with a long decay of the ACF. This fact means that ORP seems to be a process that is less sensitive to slight perturbations than the other water quality measurements (regarding uni-parametric technologies). However, selecting a time window with a length of about 78 hours will bring up additional limitations; no rainfall event will be detected in the output because no time window of this length registers an average probability of being part of dry weather lower than 0.05. Thus, for this water quality parameter, a reduction of time window's length was proposed. These results will be exposed later. By applying the Step Detection algorithm over the probability of being dry weather signal (filtered by means of the Median Filter, with the mobile windows' length proposed above), the result obtained for each water quality parameter's measurements can be listed is given by Figure 4.22.

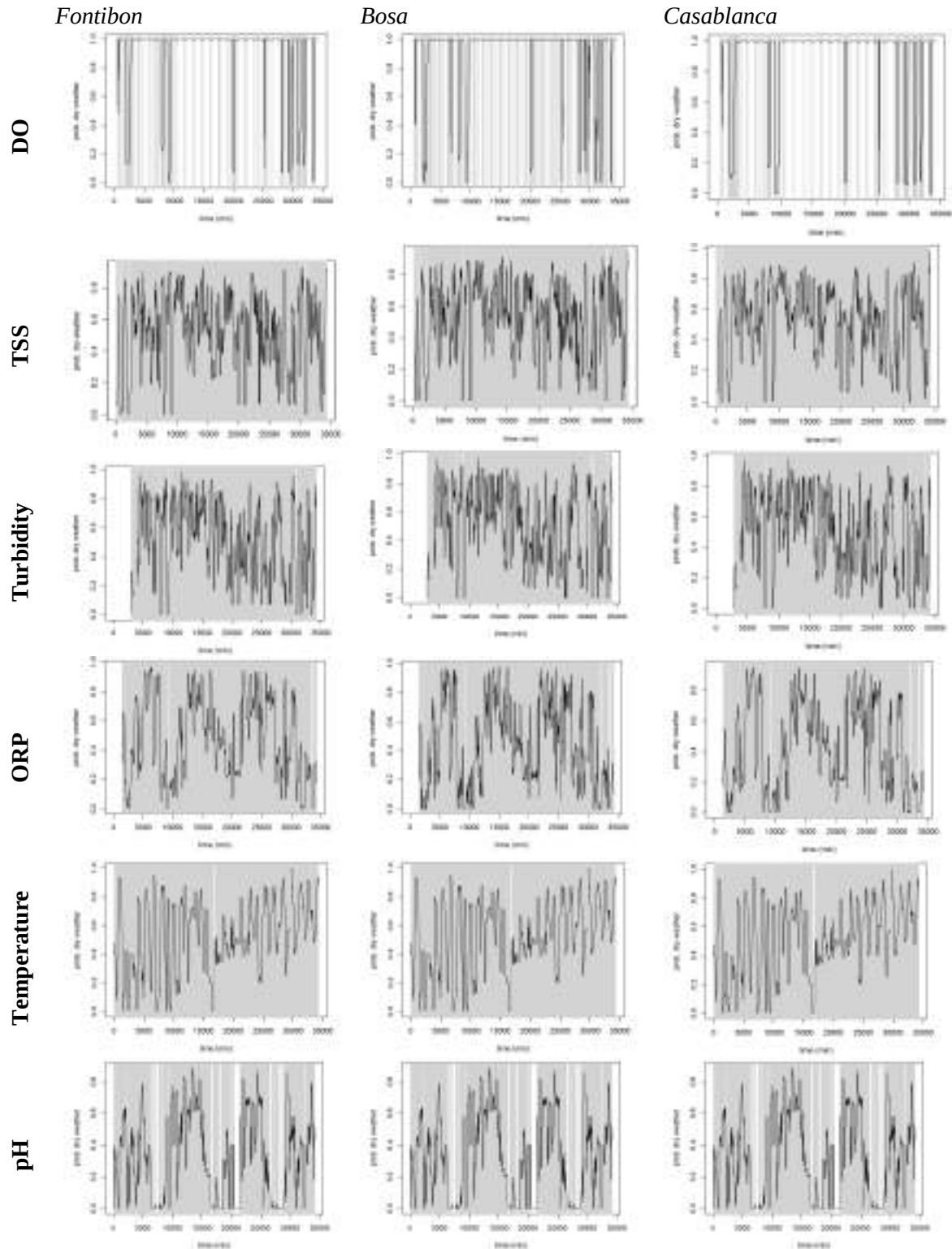


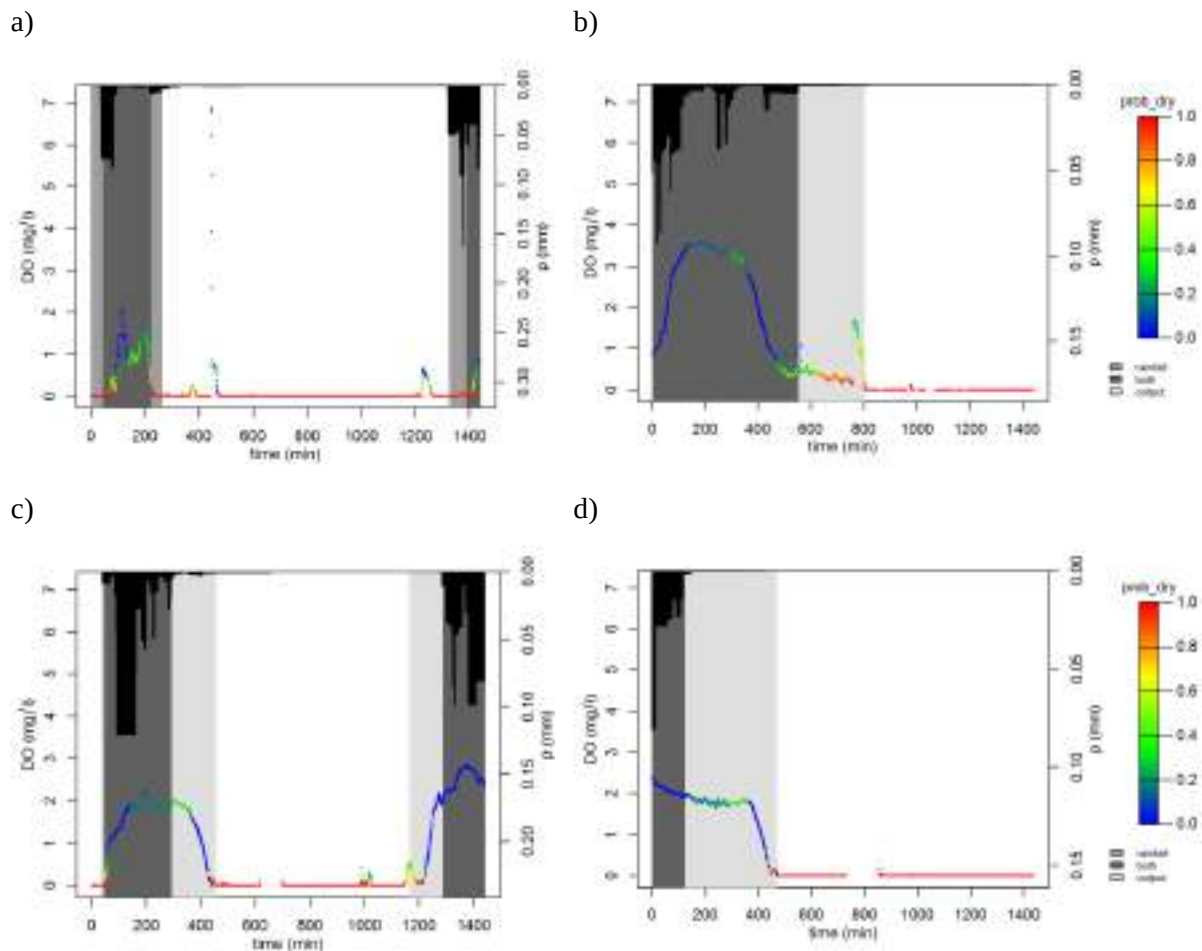
**Figure 4.22 Probability of dry weather signal filtered by a Median Filter for DO, TSS, Turbidity, ORP, Temperature and pH signals**

Regarding the results exposed in Figure 4.22, rainfall events were identified over the probability of dry weather signal, whenever the mean value between couples of points where abrupt changes

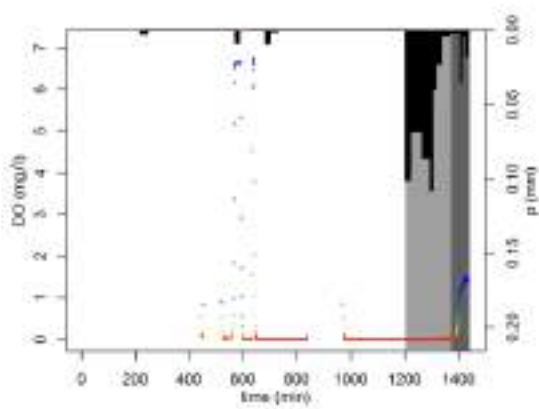
occurred (Step Detection technique) was lower than 0.05 (probability of being part of typical dry weather behavior under  $<0.05$ ).

For the case of DO a different approach was proposed: considering that most of the time the values of DO during dry weather are 0.0 mg/l (one-tailed distribution of the PDF, asymptotically to 0.0 mg/l), whenever the probability of being part of a rainfall event was lower than 0.95, it was considered that the output signal (DO) was ranging outside of its typical dry weather behavior. By means of the proposed methodology, the results presented in Appendix 8.10 were obtained for DO from rainfall registers at the *Fontibon* rain gauge station.

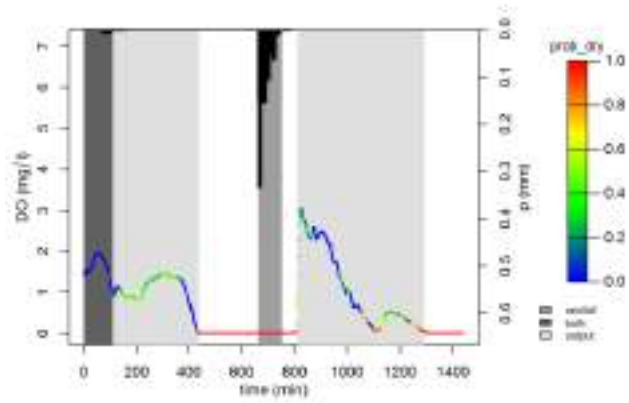
A total of six rainfall events (Appendix 8.10: a-b, c-d, e-f, g, h) were detected. Hence, it can also be stated that the temporal delimitation of rainfall and output seem to be consistent. Regarding the implementation of the methodology for the *Bosa* rain gauge station, the results are reported in Figure 4.23.



e)



f)



g)

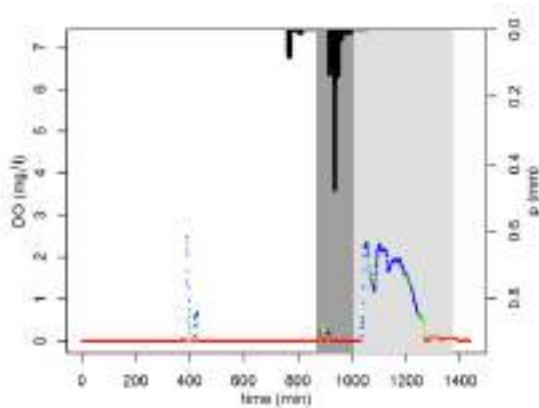


**Figure 4.23 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding DO (uni-parametric) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Bosa* rain gauge station.**

Referring to analyzes undertaken by means of the rainfall registered at the *Bosa* rain gauge station, a total of seven rainfall events (Figure 4.23: a, a-b, c, c-d, e-f, f, g) were detected. Accordingly with results obtained by the use of rainfall registered at the *Fontibon* rain gauge station, the temporal delimitation of rainfall pulses and time windows in which the output signal is ranging outside of its typical dry weather behavior seem to be consistent. For the DO measurements and the rainfall registered at the *Casablanca* rain gauge station, the results are given in Appendix 8.11.

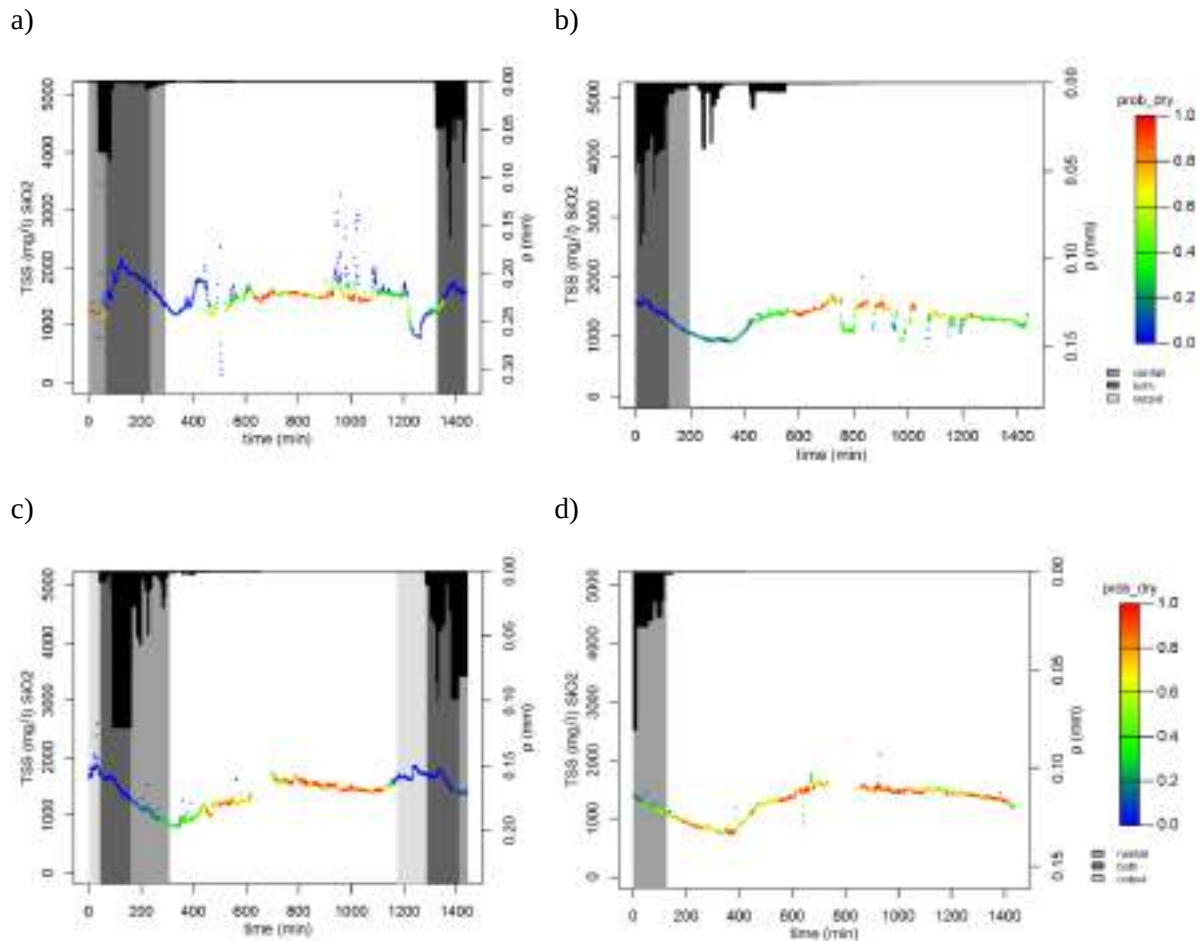
A total of seven rainfall events (Appendix 8.11: a, a, a-b, c-d, d, d, e) were reported by the use of the proposed methodology. For this case, the rainfall events’ temporal delimitation seems not to be as accurate as for the case of the *Fontibon* and *Bosa* rain gauge stations.

Subsequently, the Total Suspended Solids were analyzed by the proposed methodology. Results first are shown for rainfall registered at the *Fontibon* rain gauge station (see Appendix 8.12).

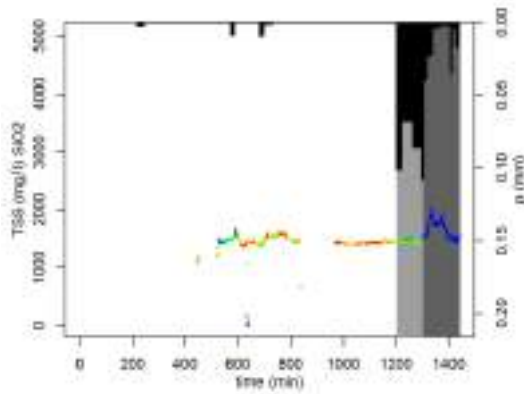
It can be stated that a total of five rainfall events (Appendix 8.12: a-b, c-d, d, e-f, g) were detected. Slight peaks, in which the probability of being part of typical dry weather behavior was under 0.05,

were observed. Thus, some of those peaks were not identified as rainfall events because of the following reasons: (i) the time window for which the output signal turned blue ( $< 0.05$ ) is smaller than the window used for median filtering the probability of dry weather signal, (ii) no rainfall pulse close was reported (greater than the time of response of the system, see details in section 4.1.3).

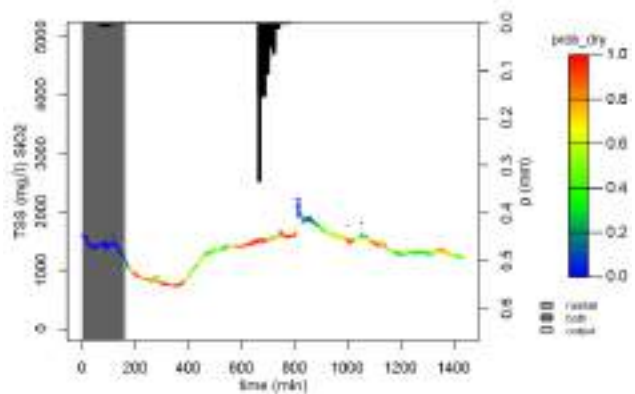
Regarding the TSS measurements undertaken by means of the uni-parametric sensor, the results obtained from the execution of the proposed methodology with rainfall registered at the *Bosa* rain gauge station are presented in Figure 4.24.



e)



f)



g)

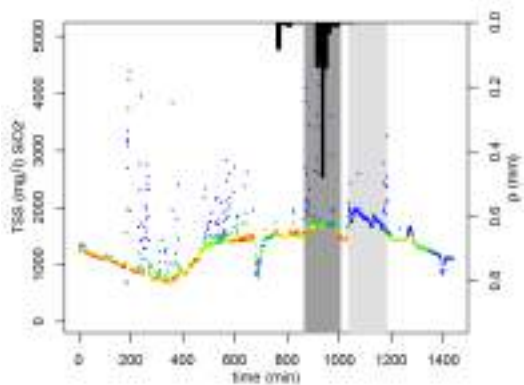


**Figure 4.24 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding TSS (uni-parametric) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Bosa* rain gauge station**

Comparatively, an additional event could be identified (Figure 4.24 a, a-b, c, c-d, e-f, g) by means of the rainfall registered at the *Bosa*, rather than with rainfall registered by the *Fontibon* rain gauge station. It can also be stated that the proposed methodology seems to be less sensitive to outlier presence, at least, for the studied data (important amount of outlier in Figure 4.24 g).

Results regarding the TSS (uni-parametric) and the rainfall registered at the Casablanca rain gauge station are shown in Appendix 8.13.

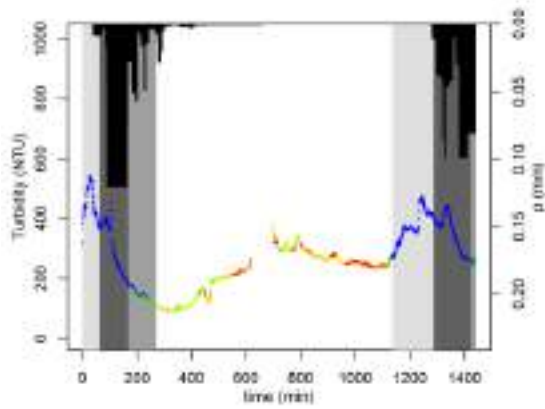
A total of two rainfall (Appendix 8.13: a, b-c) events could be identified by means of the rainfall reported at the *Casablanca* rain gauge station. Accordingly, it can be concluded that rainfall registered near the zone of the *Casablanca* rain gauge station seems not affecting the TSS measurements at the effluent of the studied catchment in a some way as the rainfall registered in the *Fontibon* or *Bosa* rain gauge stations.

The application of the proposed methods with the rainfall registered at the *Fontibon* rain gauge station, for Turbidity measured by means of a uni-parametric sensor at the effluent of *Gibraltar*’s urban catchment, brought up results presented in Appendix 8.14.

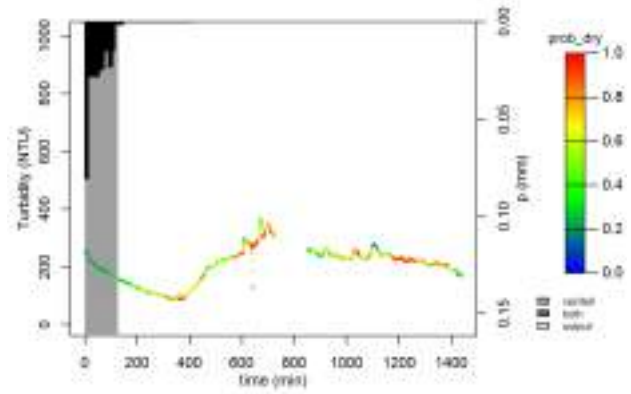
A total of five rainfall events (Appendix 8.14: a-b, b, c-d, d, e) (including the rainfall pulse and the output's time windows in which the signal ranged outside of its typical dry weather behavior) were identified. It can be observed that Turbidity peaks during a rainfall event seem to be more prominent than for uni-parametric TSS. Although Turbidity measurements seemed to report a lower number of outlier data than the TSS measurements, the proposed methods applied to Turbidity time series showed also to be insensitive to outlier or spurious data.

Regarding uni-parametric Turbidity measurements, specifically the rainfall registered at the *Bosa* rain gauge station. Results are given in Figure 4.25.

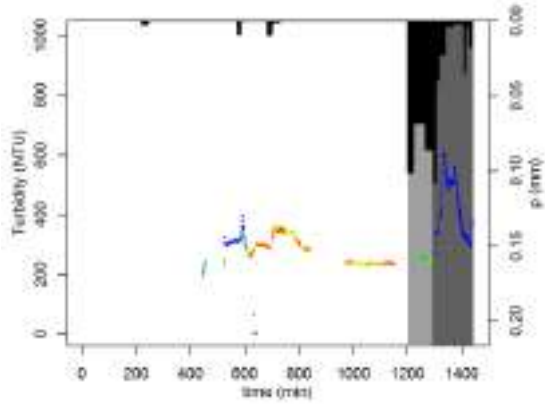
a)



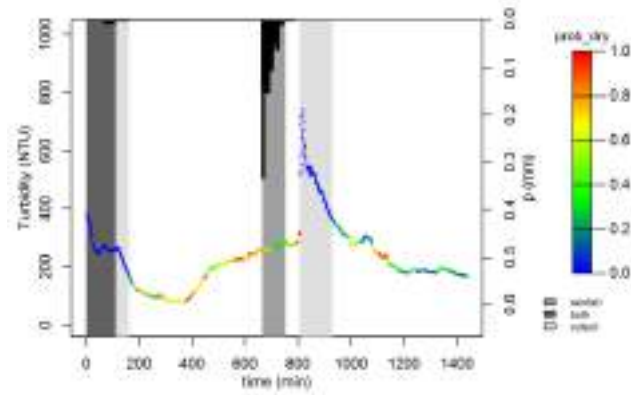
b)



c)



d)



e)

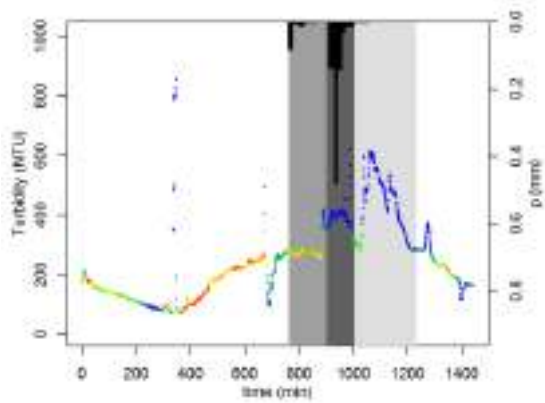


**Figure 4.25** Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding Turbidity (uni-parametric) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Bosa* rain gauge station.

A total of five rainfall events (Figure 4.25: a, a-b, c-d, d, e) were identified. Considering the rainfall registered at the *Casablanca* rain gauge station, just one rainfall event could be identified. In this sense, relationships between different events could not be assessed. Regarding the Turbidity

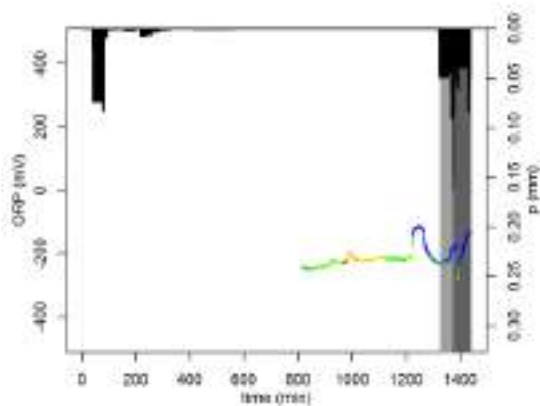
measurements, the rainfall events identified for the *Fontibon* and *Bosa* rain gauge stations occurred at the same moments during the measurement campaign.

For the case of ORP, as by its ACF the length of the mobile time window for applying a Median Filter over the probability of dry weather signal (colors) was assessed to be 4643 min, rainfall events with lower durations would hardly be detectable. Accordingly, a reduction of the length of the mobile time window in the Median Filter was proposed. By means of a trial-error procedure, the temporal delimitation seemed to be appropriate for a length of about 80 min. With this, the results of Appendix 8.15 were obtained by the use of rainfall registered at the *Fontibon* rain gauge station.

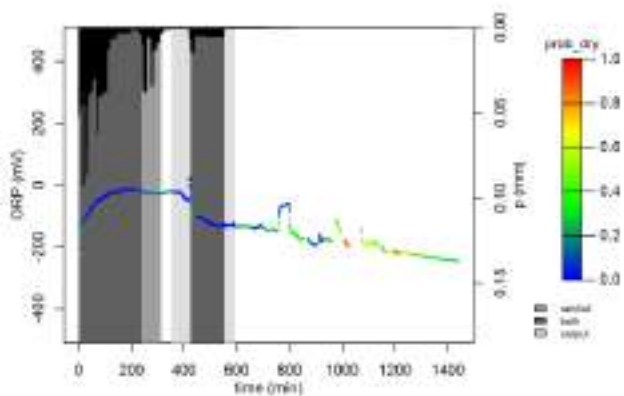
From these results, a total of five rainfall events (Appendix 8.15: a-b, b, c, d, e) could be detected. It can be appreciated that influence over the output signal (ORP) seems to be longer than for the other water quality parameters measured by the uni-parametric sensors. This appreciation is in agreement with the ACF functions reported in Figure 4.21 in which the ORP seemed to have a more highlighted inertia than the other parameters (measured by uni-parametric sensors). However, the process' memory of ORP reported by the ACF analysis seems to be an overestimation of the length of the mobile time window to be used for the proposed methodology; reason to reduce it (in about 80 min) until an appropriate delimitation of rainfall events is achieved (by means of a trial-error procedure).

For the case of rainfall registered at the *Bosa* rain gauges station, the ORP output signal at the effluent of the studied catchment allowed to identify the rainfall events reported in Figure 4.26.

a)



b)



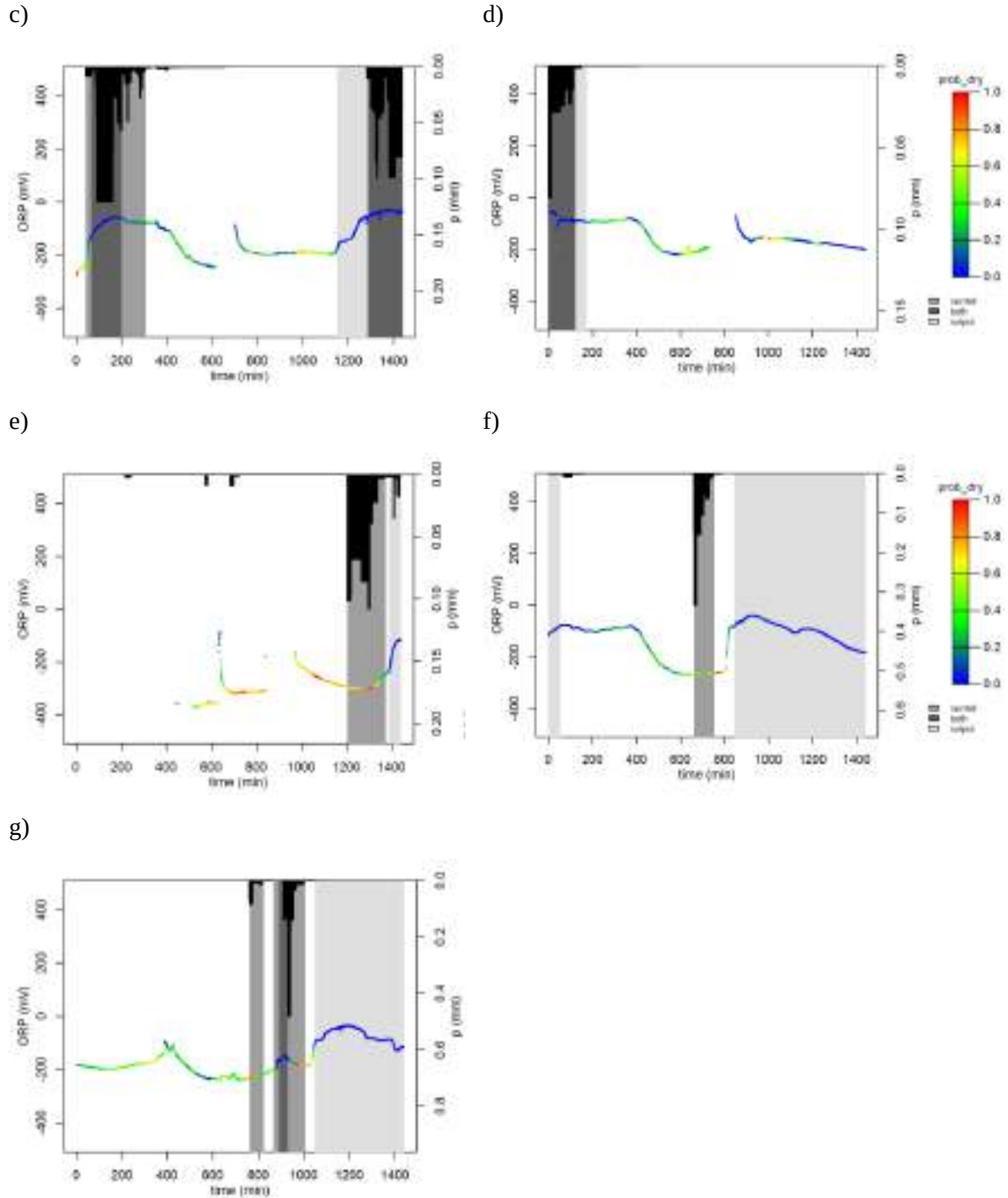


**Figure 4.26** Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding ORP (uni-parametric) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Bosa* rain gauge station.

A total of eight rainfall events (Figure 4.26: a-b, b, c, c-d, e-f, f, g, g) could be identified. Hence, the time of response (time that passes from the registration of the rainfall pulse to the identification of an “influence” time window over the output) for ORP seems to be higher than for other water

quality parameters measured by the uni-parametric sensors. This conclusion will be supported with further analysis (see Figure 4.37).

For the ORP measurements, the last result, regarding the rainfall registered at the *Casablanca* rain gauge station, is shown in Appendix 8.16. A total of four rainfall events (Appendix 8.16 a, b-c, c, d) could be identified. However, for the first rainfall event, several “NA” values were reported. From the observed delimitations, it can be stated that a generalizable relation between the length of the rainfall pulses and the length of the influence over the output signal, at least for this case, seems not to be present.

The Temperature measured by the uni-parametric sensor was analyzed by means of the proposed methodology and the rainfall registers at the *Fontibon*, *Bosa* and *Casablanca* rain gauge stations. For the analyzed cases, rainfall events could not be clearly established by rainfall registered for any of the analyzed rain gauge stations. For illustrative purposes, the unique identified event (Figure 4.27: b-c), for the *Bosa* rain gauge station is shown in Figure 4.27.

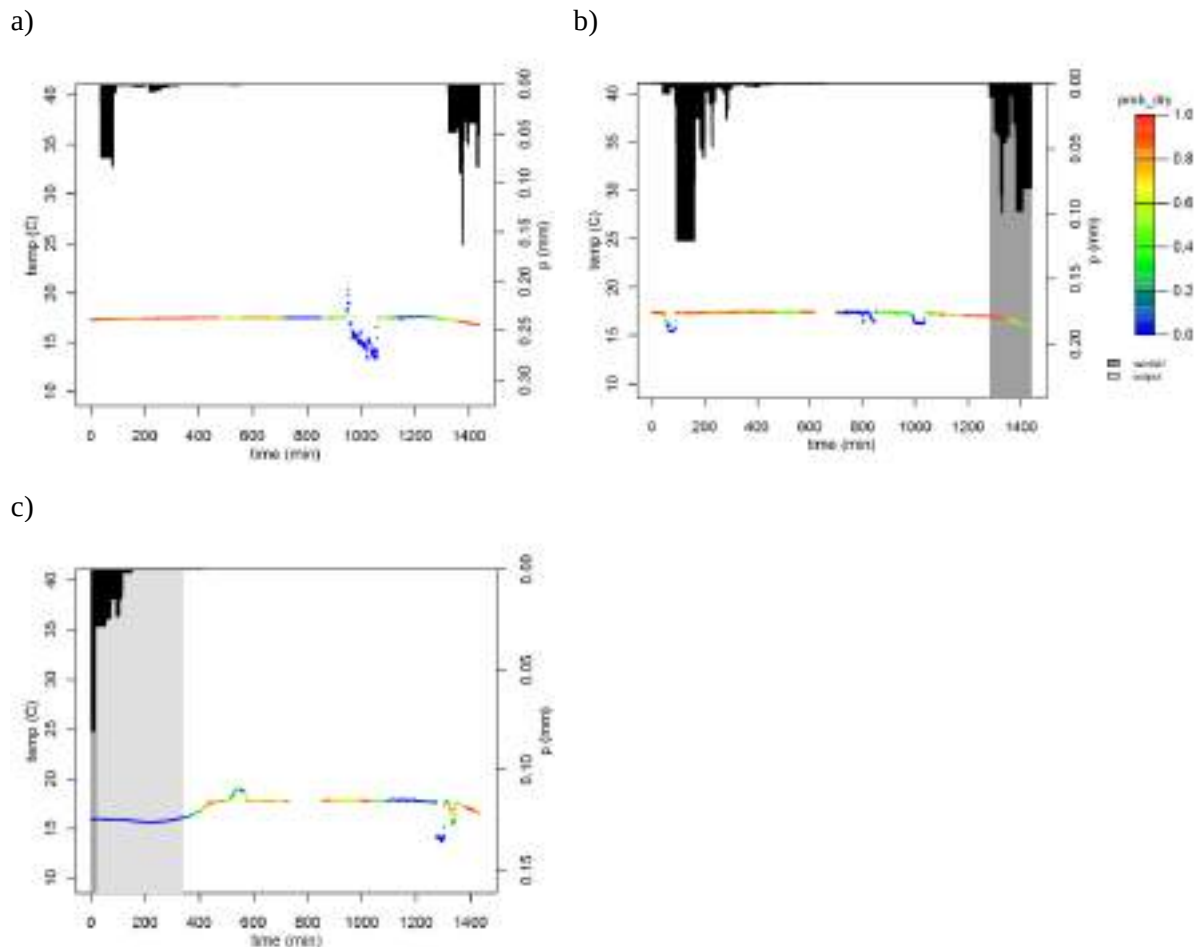


**Figure 4.27** Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding Temperature (uni-parametric) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Bosa* rain gauge station.

From the above results, it can be concluded that the Temperature seems to have a weak relation with rainfall, for the studied urban catchment. This statement is supported by the results obtained

from the PCA (see section 4.1.1.1) in which Temperature seemed not to have an influence over the differences between the variability of dry and rainy days.

At last measurement undertaken by means of a uni-parametric sensor, the pH was analyzed. The results were evaluated for the rainfall registered at the *Fontibon*, *Bosa* and *Casablanca* rain gauge station. As for temperature, the pH seemed to have a weak relation with rainfall, and was independently of the rainfall's spatial variability (given by registrations at the *Fontibon*, *Bosa* or *Casablanca* rain gauge stations). For illustrative purposes, the results obtained by applying the proposed methodology with rainfall registered at the *Bosa* rain gauge station are exposed in Figure 4.28.

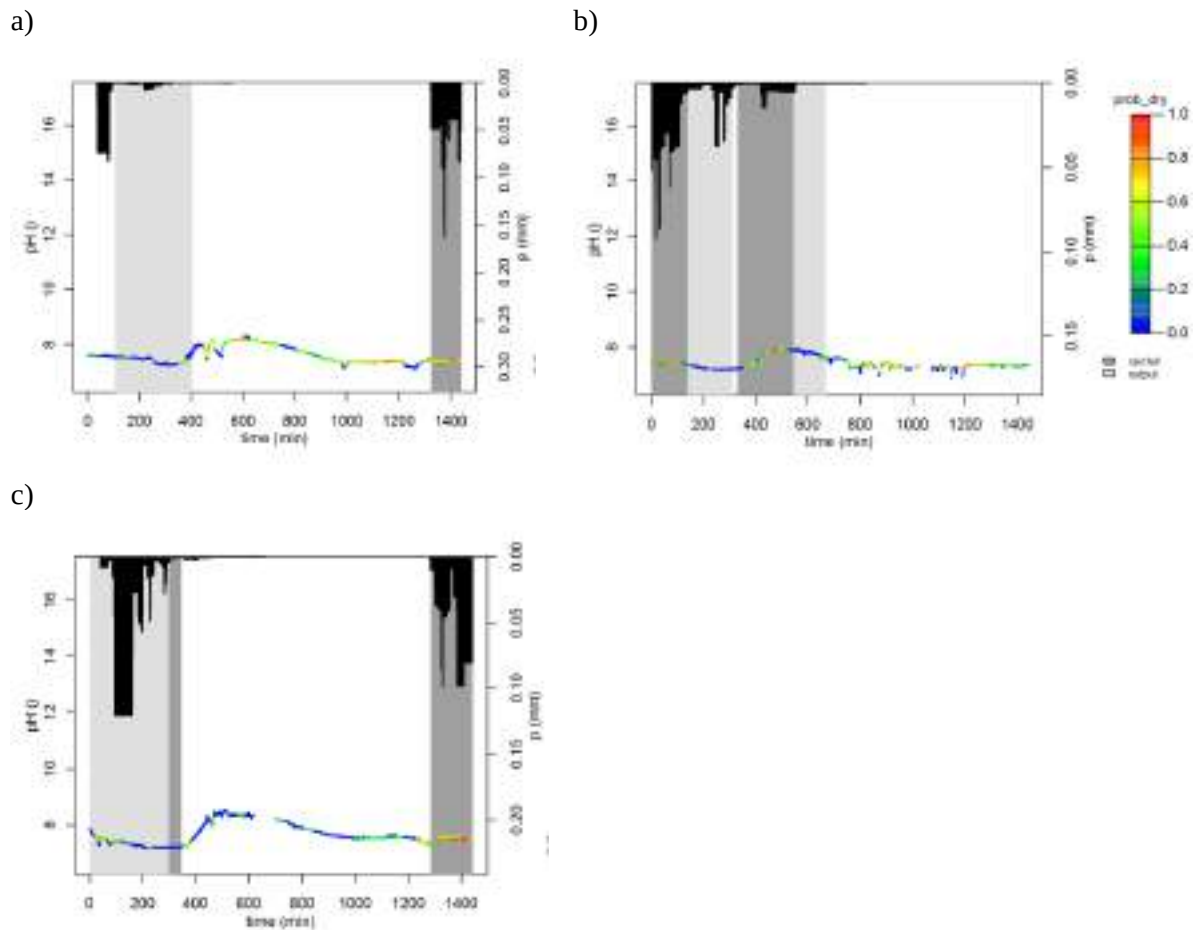


**Figure 4.28** Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding pH (uni-parametric) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Bosa* rain gauge station.

It can be appreciated that groups of blue points (low probability of dry weather values) seem to be scattered over the output signal, without exposing complete time windows ranging outside of the typical dry weather behavior. Accordingly, it can be stated that the pH, as the Temperature, seems not to be influenced by rainfall as the rest of the measured water quality parameters (DO, TSS, Turbidity and ORP). This result is also in concordance with the PCA calculations (see section 4.1.1.1). Therefore, whenever it is desired to assess the effect of the rainfall over water quality by

real-time operation, an ion sensitive sensor at the effluent of the studied urban catchment might be insufficient (as an alternative infra-red or UV-VIS spectrometry can be recommended).

Regarding the water quality parameters' time series obtained from measurements undertaken by the use of uni-parametric sensors, it was observed that the proposed methodology allows to identify rainfall events in terms of their cause (rainfall pulses) and their effect (output's time windows of influence), for several water quality parameters (Dissolved Oxygen, Total Suspended Solids, Turbidity and ORP), measured by means of uni-parametric sensors.

The ACF provided insights about the process' memory, for each water quality parameter measured by uni-parametric sensors. From these results, it can be stated that the ORP seems give time series that are heavily influenced by their past observations, fact that suggests a longer influence windows over the output signal. This statement was preliminarily verified by the results of rainfall events' identification, by the use of rainfall registers from different rain gauge stations (*Fontibon, Bosa and Casablanca*).

The PCA results (from section 4.1.1.1) were consistent with the results for the water quality parameters that seemed to be influenced by rainfall, detectable for the described calculations in this section: pH and Temperature seem to be the water quality parameters with a lower influence from rainfall, since not even a consistent detection of rainfall events could be accomplished.

The maximum number of rainfall events that could be identified (about eight events) by means of the proposed method was for ORP time series. The rainfall registered at the *Bosa* rain gauge station enabled to identify the highest number of rainfall events (from five to eight), for several water quality parameters (DO, TSS, Turbidity and ORP). Rainfall events' identification for the Turbidity and TSS (uni-parametric) were very similar, a fact that suggests a relation between TSS and Turbidity behavior, in terms of influence by rainfall, independently of the studied rain gauge station (*Fontibon, Bosa and Casablanca*). A relation between TSS and Turbidity has already been established in different studies (*e.g. Mishendo et al., 1999*).

#### 4.1.3.2.3 UV-VIS spectrometry

For UV-VIS spectrometry sensor two data sources were studied: (i) the spectra reported for each time-step and (ii) equivalent concentrations of TSS and COD obtained from the equipment's global calibration. Regarding the Spectral analysis, the wavelengths 205 nm and 745 nm were evaluated, due to their representativeness in terms of the differences of the variability between dry and rainy days (see details section 4.1.1.2.1).

##### 4.1.3.2.3.1 Spectral analysis

For the implementation of the proposed methodology, the first part refers to the application of a Median Filter over the probability of being a dry weather output signal, registered at the effluent of the studied catchment. For the execution of this filtering technique, an appropriate length of the mobile window which will be used for the filter can be estimated by means of the ACF (as a pointer of the memory of the occurring physical process). The ACF calculation for wavelengths 205 nm and 745 nm is exposed in Figure 4.29.

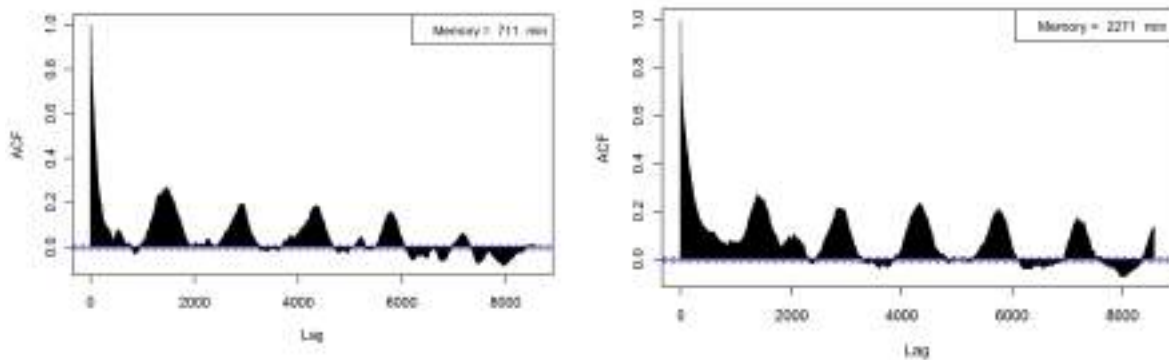


**Figure 4.29 Autocorrelation function for wavelength 205 nm (left) and 745 nm (right); measured by UV-VIS spectrometry (considering the complete measurement period)**

In agreement with the postulates formulated for the proposed methodology, it can be concluded from these results that an appropriate length for the mobile time window used for filtering the output signal (absorbance values registered in wavelength 205 nm and wavelength 745 nm) is about 711 min for the wavelength 205 nm and 2271 min for the wavelength 745 nm, respectively. However, the proposed length for the mobile window is too wide for the desired purpose (detailing time windows in which the probability of dry weather signal is lower than 0.05, in average). Therefore, by tuning the length of the mobile window proposed for the Median Filter (by a trial-error procedure), it was found that a window of about 80 min seems appropriate for detailing rainfall events over this output signal (wavelength 745 nm, measured by UV-VIS spectrometry, at the effluent of the catchment).

Regarding the calculations reported above, the Median Filter was then applied to the probability of being a dry weather signal (obtained from dry weather probabilistic characterization of a given time series), in order to expose groups of data with low probability values. Furthermore, the Step Detection Algorithm was applied for identifying, for a couple of points in which the probability of being a dry signal changed abruptly, time windows with an average of the signal (probability of being part of dry weather behavior) under 0.05. The results for wavelength 205 nm and wavelength 745 nm are presented in Figure 4.30.

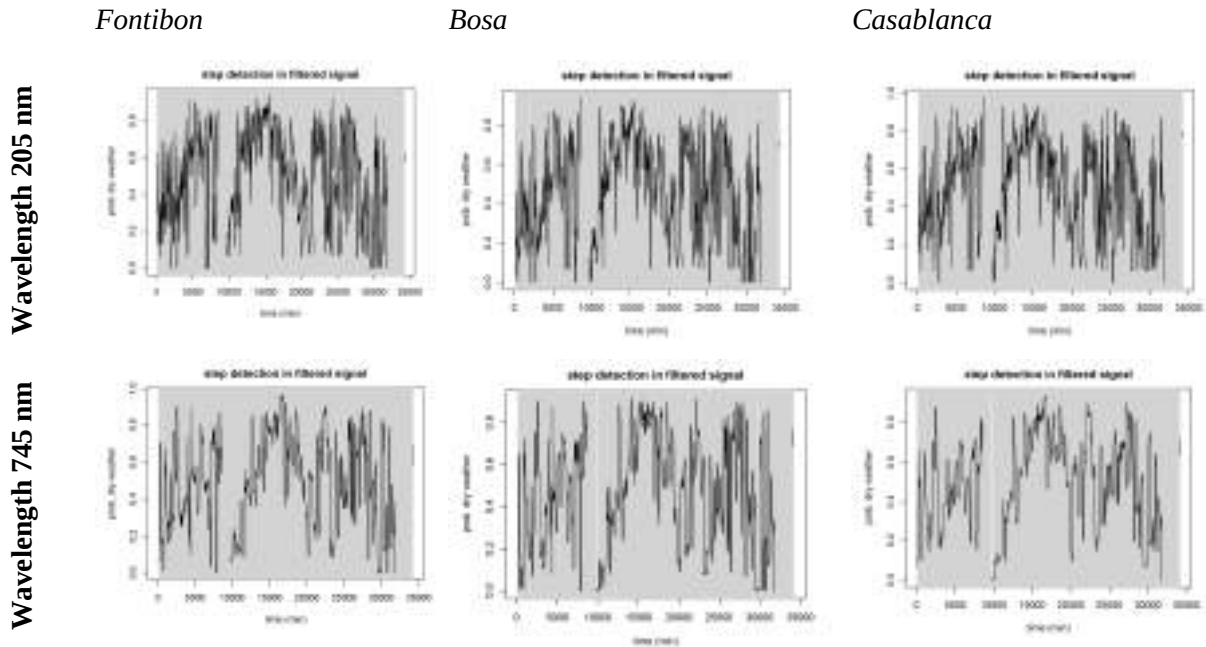


**Figure 4.30 Probability of dry weather signal filtered by a Median Filter, for wavelength 205 nm and wavelength 745 nm signals**

The proposed results show how the “probability of dry weather signal” obtained from the dry weather probabilistic characterization was filtered by the Median Filter (also supported by results from the ACF). Hence, the time windows in which the filtered signal was on average under 0.05, exposing a constant behavior (no abrupt changes), were detected.

Regarding the results obtained for wavelength 205 nm, a rainfall influence was not appreciated as clearly as for other cases (*e.g.* DO by uni-parametric sensor). In this sense, the results obtained by applying the proposed methods just for the rainfall registers at the *Bosa* rain gauge station (rainfall pulses) are reported in Figure 4.31.

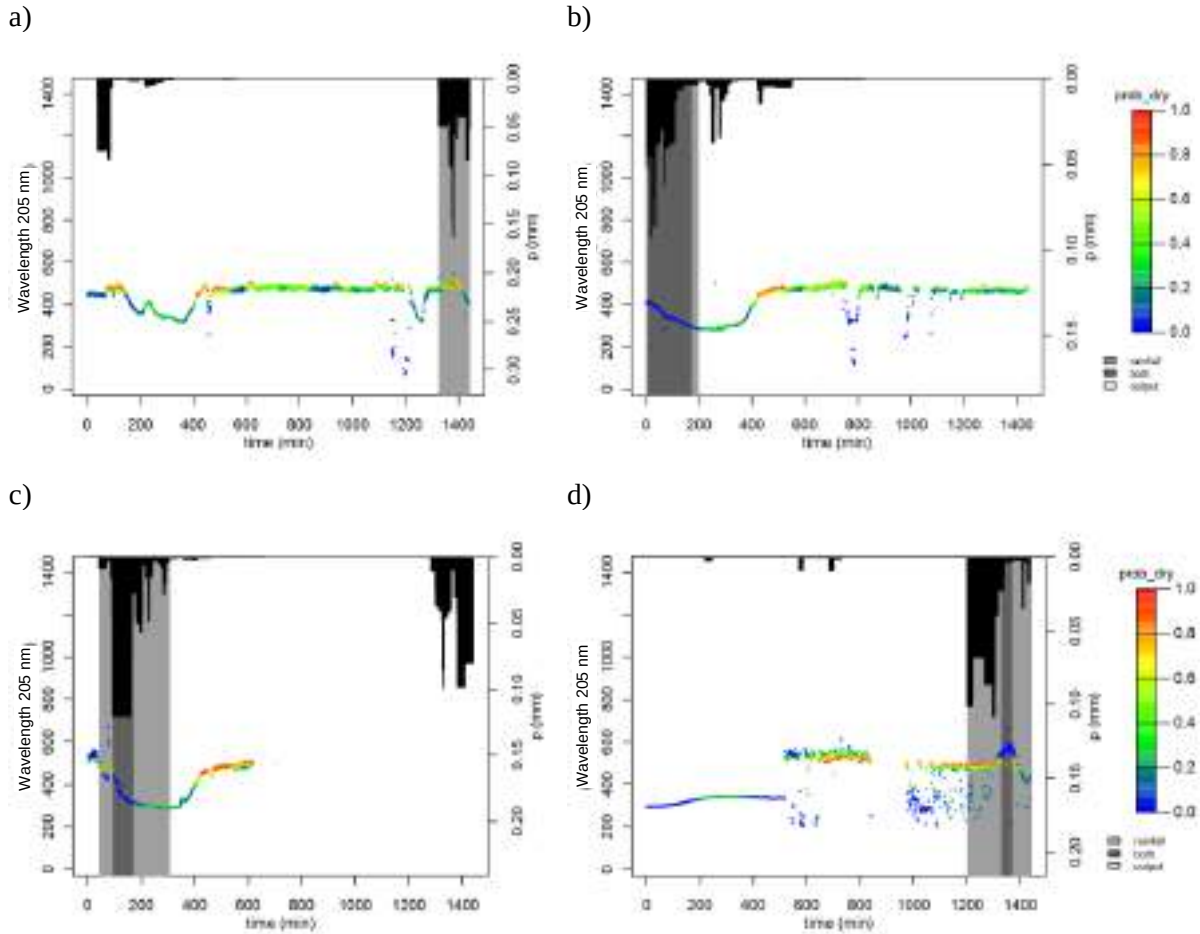


**Figure 4.31** Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding wavelength 205 nm (UV-VIS spectrometry) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Bosa* rain gauge station.

For three situations (Figure 4.31: a-b, c, d) a relation between a rainfall’s pulses and time windows of influence over the output could be established. Hence, rainfall pulses without a consequence over the output, in terms of lowering the probability of dry weather values, were detected. For the calculations undertaken on the absorbance values at the wavelength 745 nm, the results presented in Appendix 8.17 were obtained, concerning the *Fontibon* rain gauge station.

A total of two rainfall events could be identified (Appendix 8.17: a-b, c-d) by means of the proposed methods. In this sense, although the rainfall pulses seemed to be related with time windows in which the probability of being part of dry weather was frequently low; just three complete rainfall events could be reported. Additionally, the proposed method seems to be sensitive to the presence of outliers, for the studied output signal (Wavelength 745 nm).

For the rainfall registered at the *Bosa* rain gauge station, the following results were obtained:

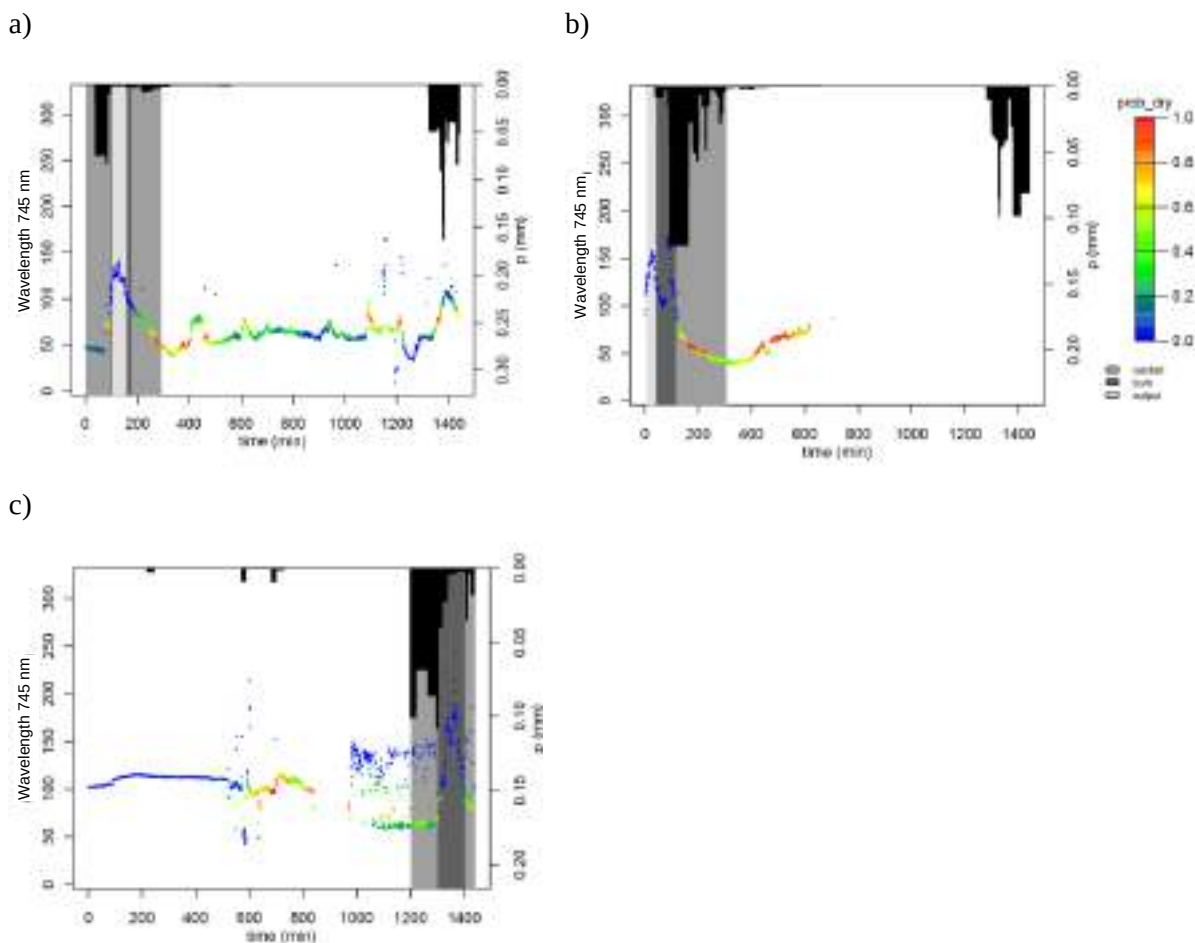


**Figure 4.32** Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding wavelength 745 nm (UV-VIS spectrometry) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Bosa* rain gauge station.

For this specific case, it was noticed that three rainfall events could be detected (Figure 4.32: a, b, c) by means of the proposed methods. However, windows ranging outside of the typical dry weather behavior, without a temporal closeness to rainfall pulses could be appreciated. Hence, rainfall pulses without a noticeable consequence (in terms of time windows ranging outside of the typical dry weather behavior) over the output were reported.

For the last analyzed scenario regarding the absorbance measurements registered by the spectra, the rainfall pulses registered in the *Casablanca* rain gauge station and their possible influence over the output were assessed (see Appendix 8.18).

Non conclusive results about the possible influence of rainfall registered at the *Casablanca* rain gauge station over the absorbance values registered in wavelength 745 nm, could be established.

#### 4.1.3.2.3.2 Equivalent water quality parameters

Equivalent COD and TSS concentrations could be reported by means of the equipment's global calibration. Additional analyses were undertaken with the purpose of exploring this data. By evaluating the ACF of the equivalent COD and TSS concentrations' time series, measured by the mentioned sensor, the results of Figure 4.33 were obtained.

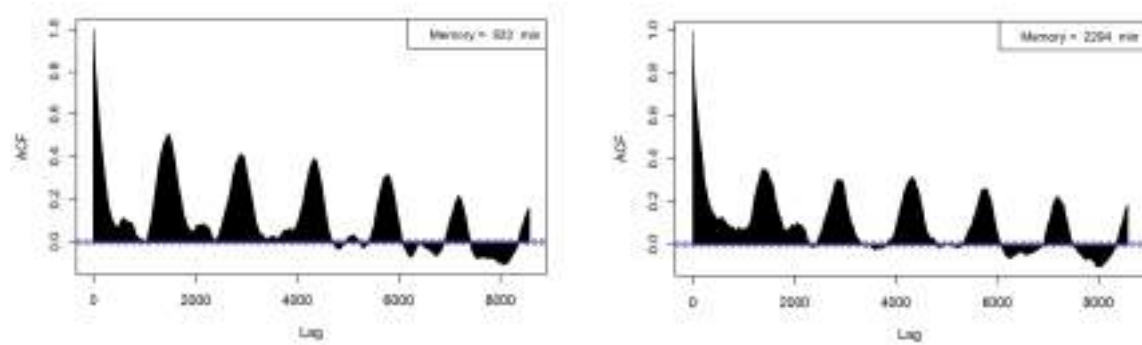


**Figure 4.33 Autocorrelation function for COD (left) and TSS (right); measured by UV-VIS spectrometry (considering the complete measurement period)**

From the ACF it can be noticed that the TSS is in agreement with the wavelength 745 nm's behavior. This fact suggests that the TSS values are correlated with the absorbance values registered at this wavelength. Hence, both ACF exhibit a periodical behavior. In this sense, the studied output signals seemed to have a daily periodical behavior at the effluent of the catchment. Thus, it can be stated that an appropriate length for the mobile time windows used in a Median Filter are the following: (i) for COD, the length might be around 922 min and (ii) for TSS, the length might be about 2294 min. It can be noticed that the memory of the process estimated for TSS measured by means of the uni-parametric and UV-VIS spectrometry sensors are noticeably different (342 min for the uni-parametric and 2294 min for the UV-VIS spectrometry, see section 4.1.3.2.2). This fact can lead to conclude that measurement undertaken with the uni-parametric sensor seems to have a stronger inertia than the measurements executed by the UV-VIS spectrometry sensor. In other words, the capacity of capturing temporal variability seems to be higher for the uni-parametric than for the UV-VIS spectrometry sensor.

Considering the very large length of the mobile window for applying the Median Filter proposed by means of the ACF function for the TSS time series, a tuning of the mentioned length was proposed with the purpose of finding smaller time windows ranging outside of the typical dry weather behavior (about 75 min seemed to be appropriate for TSS values measured with UV-VIS spectrometry).

Therefore, the probability of being part of dry weather, for each time-step of the TSS and COD time series, was calculated. The resulting signal (probability of dry weather signal) was filtered by means of a Median Filter, using the lengths of the mobile windows proposed above (ACF as an estimate of the memory of the physical process, Figure 4.33). The results are plotted in Figure 4.34.

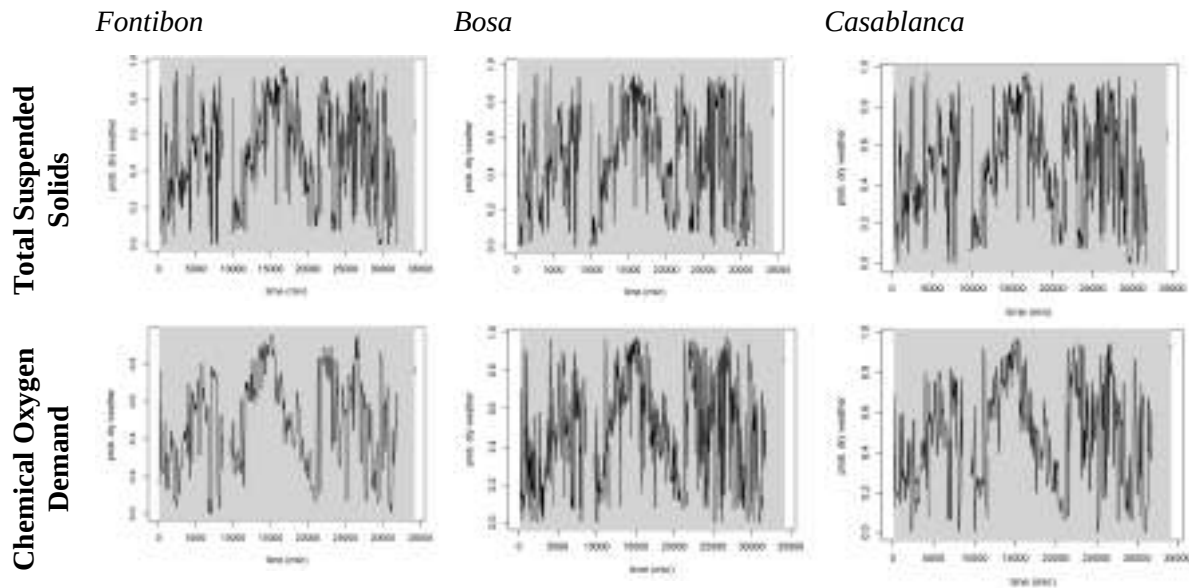


**Figure 4.34 Probability of dry weather signal filtered by a Median Filter, for TSS and COD signals**

Periods in which the probability of dry weather filtered signal exposed a low value (under 0.05) and no abrupt changes were detected (by means of the Step Detection technique), were classified as time windows ranging outside the typical dry weather behavior. Analyzing the TSS time series, the results were first reported for rainfall registered at the *Fontibon* rain gauge station.

From results in Appendix 8.19, it can be stated that a total of three rainfall events (Appendix 8.19: a, a-b, c-d) could be established by means of the proposed methods. Hence, the relation between time windows ranging outside the typical dry weather behavior and the rainfall pulses could not be assessed, due to the limited number of events identified.

Analyzing the rainfall registered at the *Bosa* rain gauge station, the results of Figure 4.35 were reported.

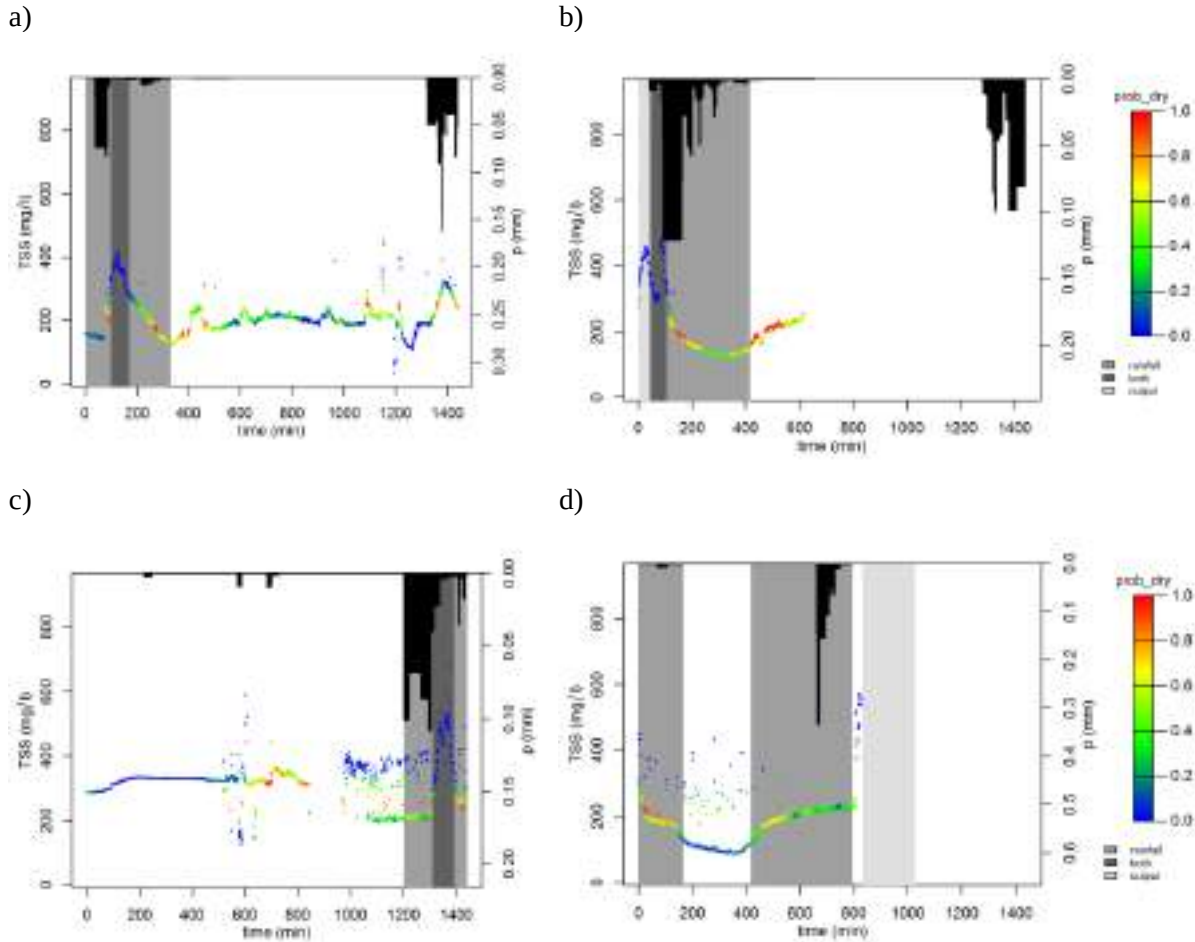


**Figure 4.35** Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding TSS (UV-VIS spectrometry) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Bosa* rain gauge station.

A relation between rainfall pulses and time windows ranging outside of the typical dry weather behavior could not be clearly established, because of the limited number of events identified (three events, Figure 4.35: a, b, c-d). For both cases (*Fontibon* and *Bosa*) the methodology exposes satisfactory results in terms of delimiting rainfall events (about three events), but exploring common characteristics from one event to another could not be found. No relevant results could be identified by means of the proposed methods for rainfall registered at the *Casablanca* rain gauge station.

Analyzing the results for the COD time series, the following results of Appendix 8.20 were reported for the *Fontibon* rain gauge station.

It can be stated that the proposed methodology does not report satisfactory results, in terms of rainfall events’ delimitation. The COD measurements by means of the UV-VIS spectrometry (global calibration) were also analyzed for the rainfall registered at the *Bosa* rain gauge station. The results are in Figure 4.36.

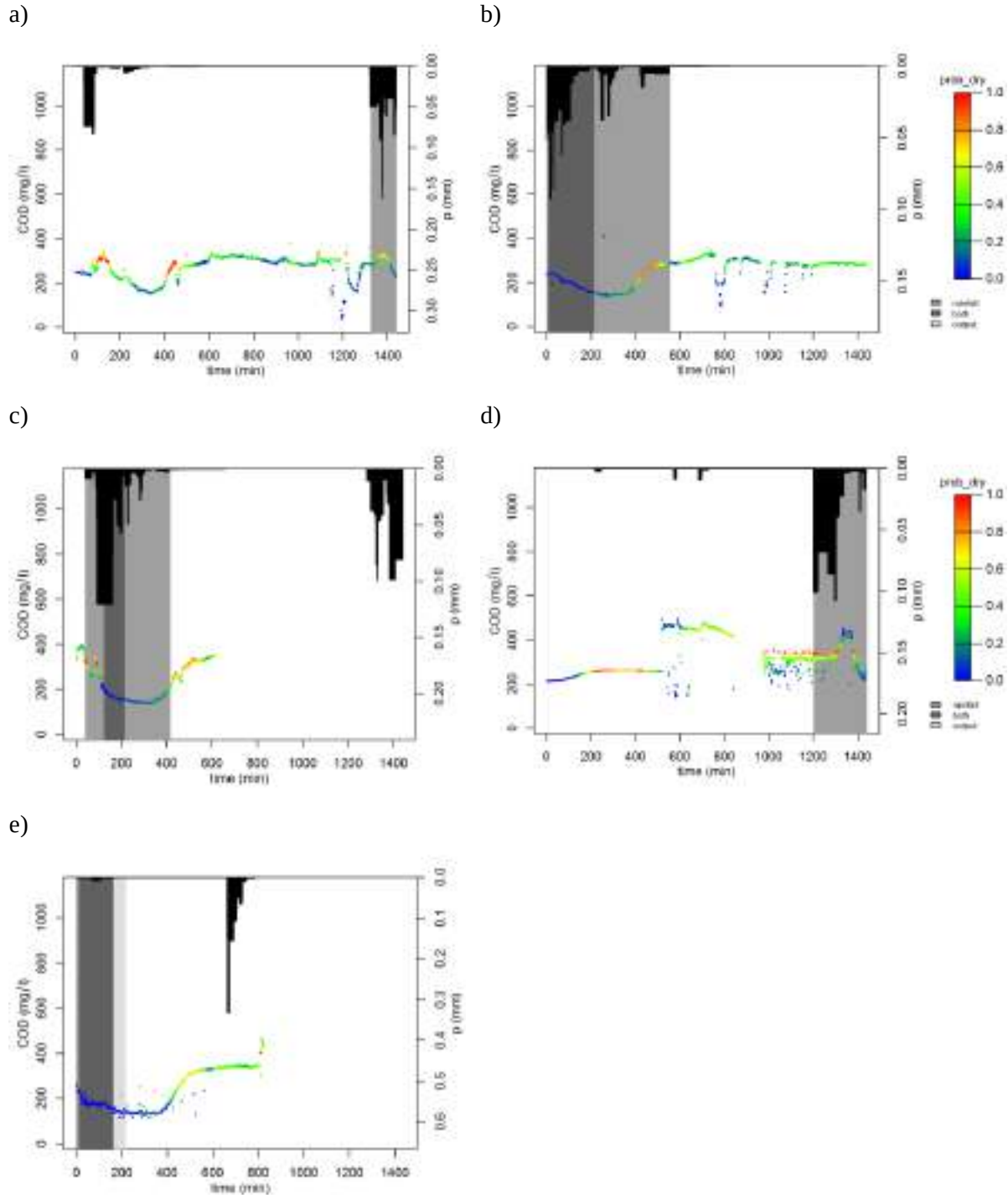


**Figure 4.36** Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding COD (UV-VIS spectrometry) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Bosa* rain gauge station.

Concerning the exposed temporal delimitations (in terms of rainfall pulses and time windows ranging outside of the typical dry weather behavior), it can be stated that the proposed methodology identified three rainfall events (Figure 4.36: a-b, c, d-e) (by rainfall registered at the *Bosa* rain gauge station) for the COD measurements. However, as only three rainfall events were delimited,

additional analysis exploring common behaviors among the events could not be executed. The analysis including rainfall registered at the *Casablanca* rain gauge station, as for the *Fontibon* rain gauge station, did not report satisfactory results, in terms of rainfall events' delimitation.

Relations between rainfall pulses and time windows ranging outside their typical dry weather behavior could not be explored for the case of UV-VIS spectrometry data. This fact can be explained by the limited number of rainfall events (lower or equal than three) that could be identified. One of the reasons that explain this situation is that many of the data were reported as "NA" (for the case of spectral analysis, as well as for the equivalent water quality measurements). In this sense, it is strongly recommended for incoming studies to implement the proposed methodology over longer time series, without many gaps.

By the identification of rainfall pulses which can be related with time windows ranging outside the typical dry weather behavior for the flow rate and the studied water quality parameters measured by means of the uni-parametric and UV-VIS spectrometry sensors, additional analysis can be executed. Recurrent relations between water quality parameters and rainfall characteristics can be explored (see details in section 4.2). However, a relevant analysis (principally for modeling and real time operation purposes) could be to estimate the response time of the output signal (water quality parameter) to a given rainfall pulse, registered at a specific studied rain gauge station (*Fontibon*, *Bosa* or *Casablanca*). Therefore, the "lag times" for each water quality parameter measured at the effluent of the catchment (by uni-parametric or UV-VIS spectrometry sensors), influenced by rainfall pulses registered at the *Fontibon*, *Bosa* or *Casablanca* rain gauge stations, were calculated. However, situations in which a time window ranging outside the typical dry weather behavior was registered to occur before a given rainfall pulse, led to negative lag times. This fact can be explained by the presence of rainfall at another point of the catchment before the registration of this rainfall at a given rain gauge station (rainfall's spatial variability). For those cases, lag times were not considered. The results obtained are summarized in Figure 4.37.

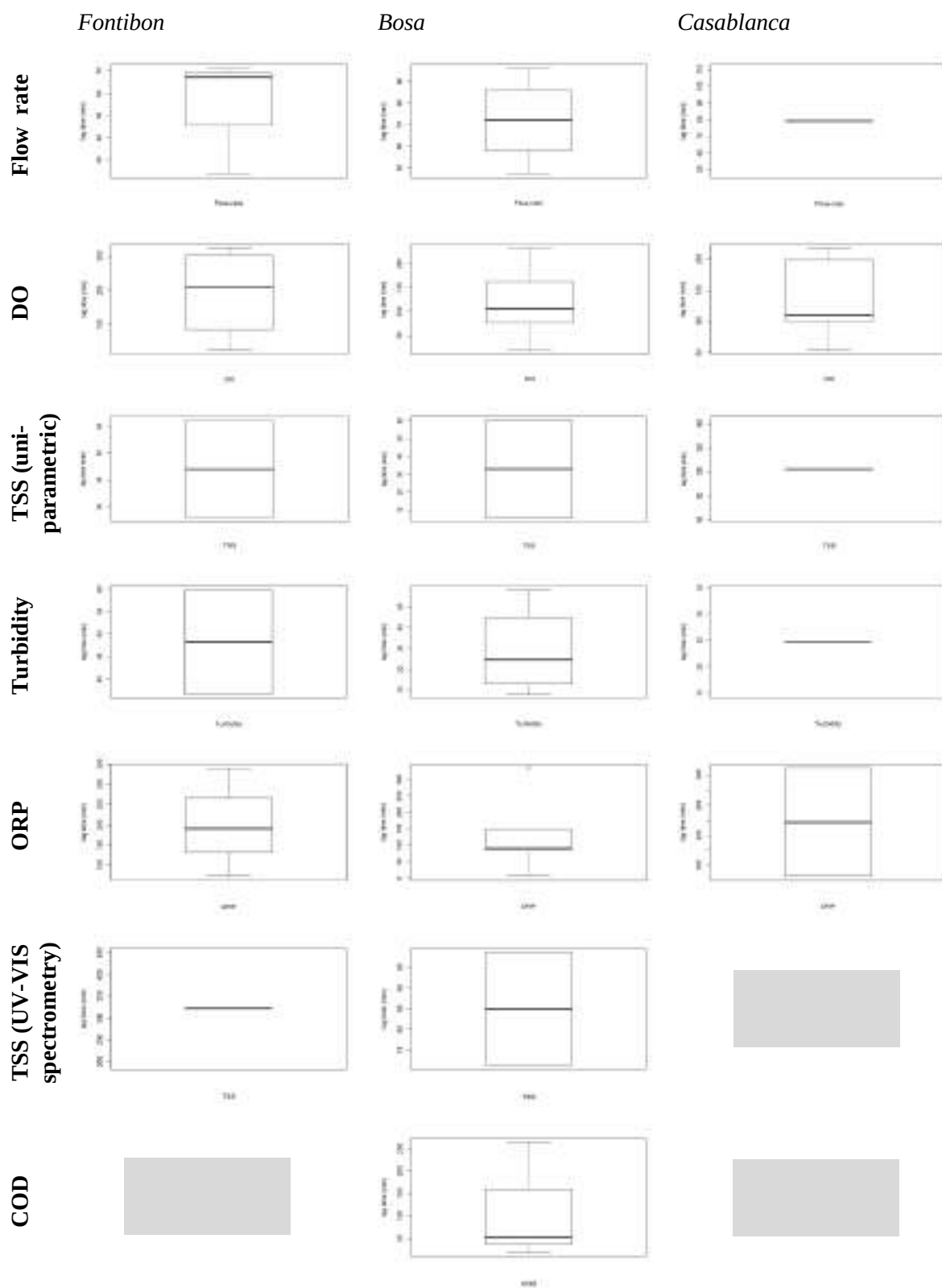


Figure 4.37 Lag times for measured water quality parameters at the effluent of the catchment, by rainfall registered at the *Fontibon*, *Bosa* and *Casablanca* rain gauge stations (grey boxes: no lag times reported)

From the exposed results it can be concluded that, for the analyzed water quality parameters, an important variability in the response time was observed for: (i) different water quality parameters, regarding the same rain gauge station and (ii) different rain gauge stations, for a given water quality parameter. Hence, a considerable variability was detected in the response time between a given rain gauge station and a specific water quality parameter, during different rainfall events. Regarding the flow rate measurements, the response time reported similar median values for the *Fontibon* and *Bosa* rain gauge stations. Nevertheless, for the case of the *Fontibon* rain gauge station, the variability was greater (Inter Quartile Distance). The unique response time identified from rainfall in the *Casablanca* rain gauge station was similar to the median values reported for the *Fontibon* and the *Bosa* rain gauge stations. For DO, TSS, Turbidity and ORP the response time was identified as greater for the *Fontibon* rain gauge station than for the *Bosa* rain gauge station, for about 75 % of the studied rainfall events. Hence, for water quality parameters in which more than one rainfall event could be identified by means of the rainfall registered at the *Casablanca* rain gauge station (e.g. DO and ORP), the response time was reported to be higher than for the case of the *Bosa* rain gauge station, with a high variability. For DO and ORP, the response time seemed to be greater than for parameters such as TSS and Turbidity. For Turbidity and TSS, the response times were reported more alike the hydrodynamic response time of the system (represented by flow rate), independently of the selected rain gauge station (*Bosa*, *Fontibon* or *Casablanca*). It can be stated that the hydrodynamic response time of the studied urban catchment seems to be shorter for the rainfall registered at the *Fontibon* than for the *Bosa* rain gauge station, regarding the flow rate results. On the other hand, most of the water quality parameters (DO, TSS, Turbidity and ORP) seemed to have faster responses for rainfall registered at the *Bosa*, rather than for the *Fontibon* rain gauge station.

Results of lag times for the spectral analysis and for the equivalent water quality parameters measured by UV-VIS spectrometry could not be assessed due to the presence of a high fraction of “NA” values (which also limited the detection of rainfall events). The response time for equivalent water quality parameters, could be identified just for the case of the *Bosa* rain gauge station. The TSS measured by UV-VIS spectrometry sensor exposed a similar response’s characterization than for TSS measurements undertaken by the uni-parametric sensor. In addition, it was detected that the COD had a response time higher than the TSS, for the studied rainfall events (identified by means of rainfall registered at the *Bosa* rain gauge station).

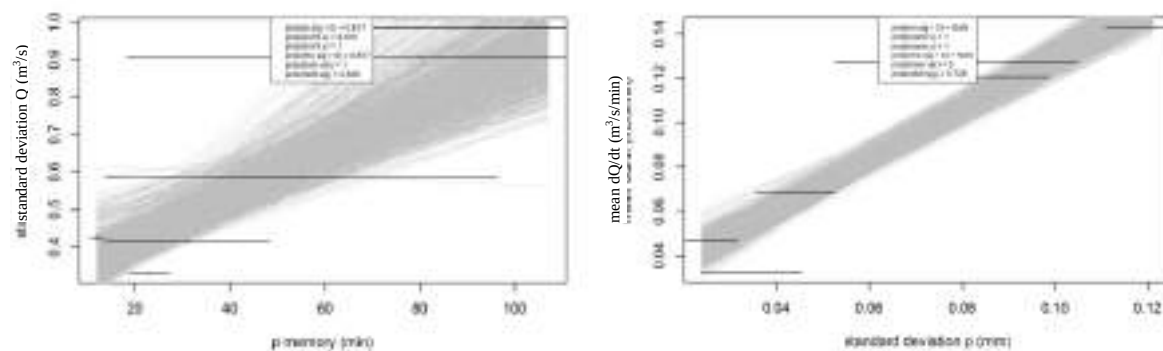
As is exposed in Table 3.1, the *Fontibon* rain gauge station is closer to the effluent than the *Bosa* rain gauge station. Considering that most of the analyzed water quality parameters (DO, Turbidity and ORP) reported response times higher for the *Bosa* than for the *Fontibon* rain gauge station, the distance of the rain gauge station to the effluent seems not to be the only fact that explains the differences on the time response. As a hypothesis, it can be stated that physical factors of the basin such as the pipes’ slopes and lengths can also have an influence in the travel of the constituents through a given urban catchment. This will be addressed again in section 4.3. Regarding the results of wavelength 205 nm and wavelength 745 nm, no positive lag time was reported.

## 4.2 Recurrent relations between rainfall pulses and water quality's time windows ranging outside of typical dry weather behavior

In section 4.1 results regarding rainfall event identification for water quality parameters registered by means of uni-parametric and UV-VIS spectrometry sensors were exposed. In this section, the proposed methods were applied to flow rate measurements undertaken by the use of an Echo Profile Flowmeter device. Thus, common characteristics between rainfall pulses and time window ranging outside typical dry weather behavior were explored, for the detected events. Assessment of common relations between rainfall and water quality is a function of the uncertainty reported for each rainfall or water quality characteristic, during a given rainfall event. In this sense, uncertainties related to rainfall and water quality characteristics and common relations assessed by the proposed association metrics (*e.g.* Correlation Test, Average Mutual Information, Linear Model and Partial Least Square Regression) were estimated by means of the Monte Carlo method, for 1000 random simulations (following a normal distribution). Common characteristics between rainfall and water quality parameters, for the studied rainfall events, were analyzed for rainfall registered at the *Fontibon*, *Bosa* and *Casablanca* rain gauge stations. The uncertainty on rainfall measurements was estimated based on the resolution of the pluviographs registered in the studied rain gauge stations, which was reported as 0.1 mm (see details in Section 3.2.3).

### 4.2.1 Echo Profile Flowmeter

Concerning the results obtained for flow rate measurements, the most relevant calculations will be shown in terms of identifiable repeated relations between data ranging outside the typical dry weather behavior and rainfall pulses, for the identified rainfall events. The uncertainty on the measurement of each flow rate value was estimated based on the resolution of the measurement device (1 % of the registered value, see details in section 3.2.1). This uncertainty value was then propagated over each statistic calculated for a given rainfall event (*e.g.* uncertainty in the standard deviation of the mean flow rate value) by use of Monte Carlo simulations. The probability of significance for each association metric (*e.g.* the probability that the Average Mutual Information is significant) was also assessed by means of this approach. The results obtained from rainfall registered at the *Fontibon* rain gauge station are listed in Figure 4.38.



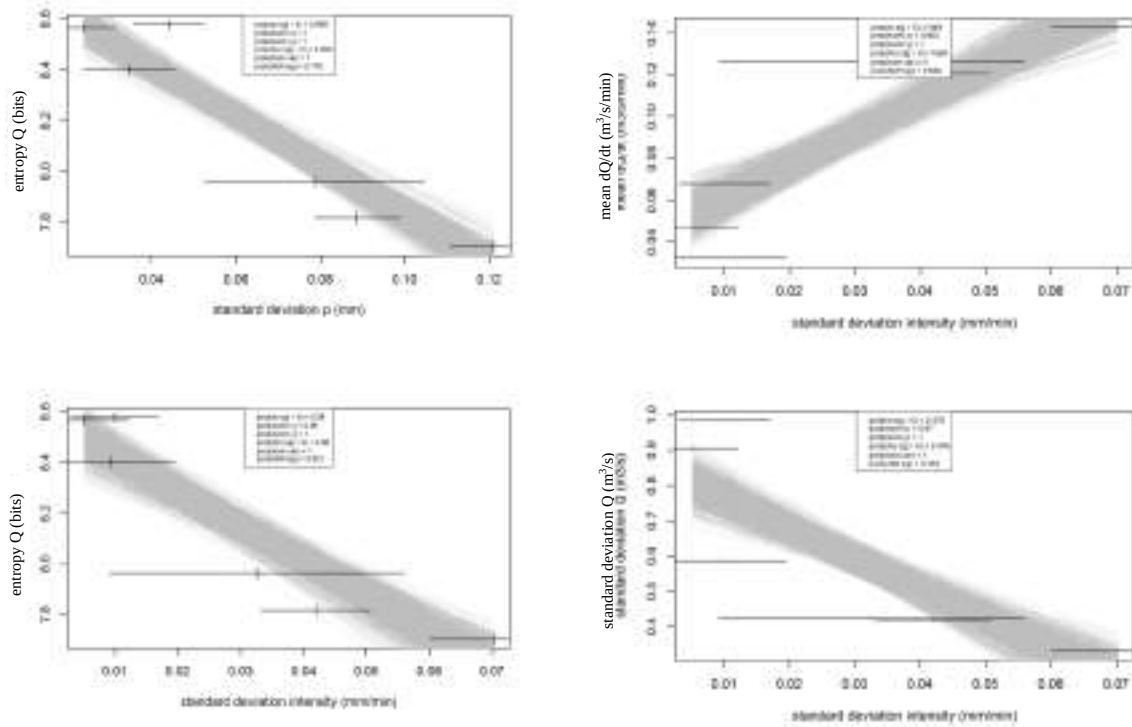


**Figure 4.38** Recurrent relations between flow rate and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the *Fontibon* rain gauge station

From the above results, several conclusions can be formulated. In terms of univariate analysis, it can be stated that the following rainfall characteristics (reported at the *Fontibon* rain gauge station) seemed to be related with the following water quality characteristics, estimable for a given rainfall event: (i) the standard deviation of the flow rate seems to be related with the decaying of the ACF of the rainfall, (ii) the standard deviation of the rainfall seems to be related with the mean flow rate gradient ( $dQ/dt$ ) and the entropy of the flow rate and (iii) the standard deviation of the rainfall's intensity seems to be related with the mean flow rate gradient ( $dQ/dt$ ), the entropy of the flow rate and the standard deviation of the flow rate.

Exploring the results obtained for the rainfall registered at the *Bosa* rain gauge station, the relations are highlighted in Figure 4.39.

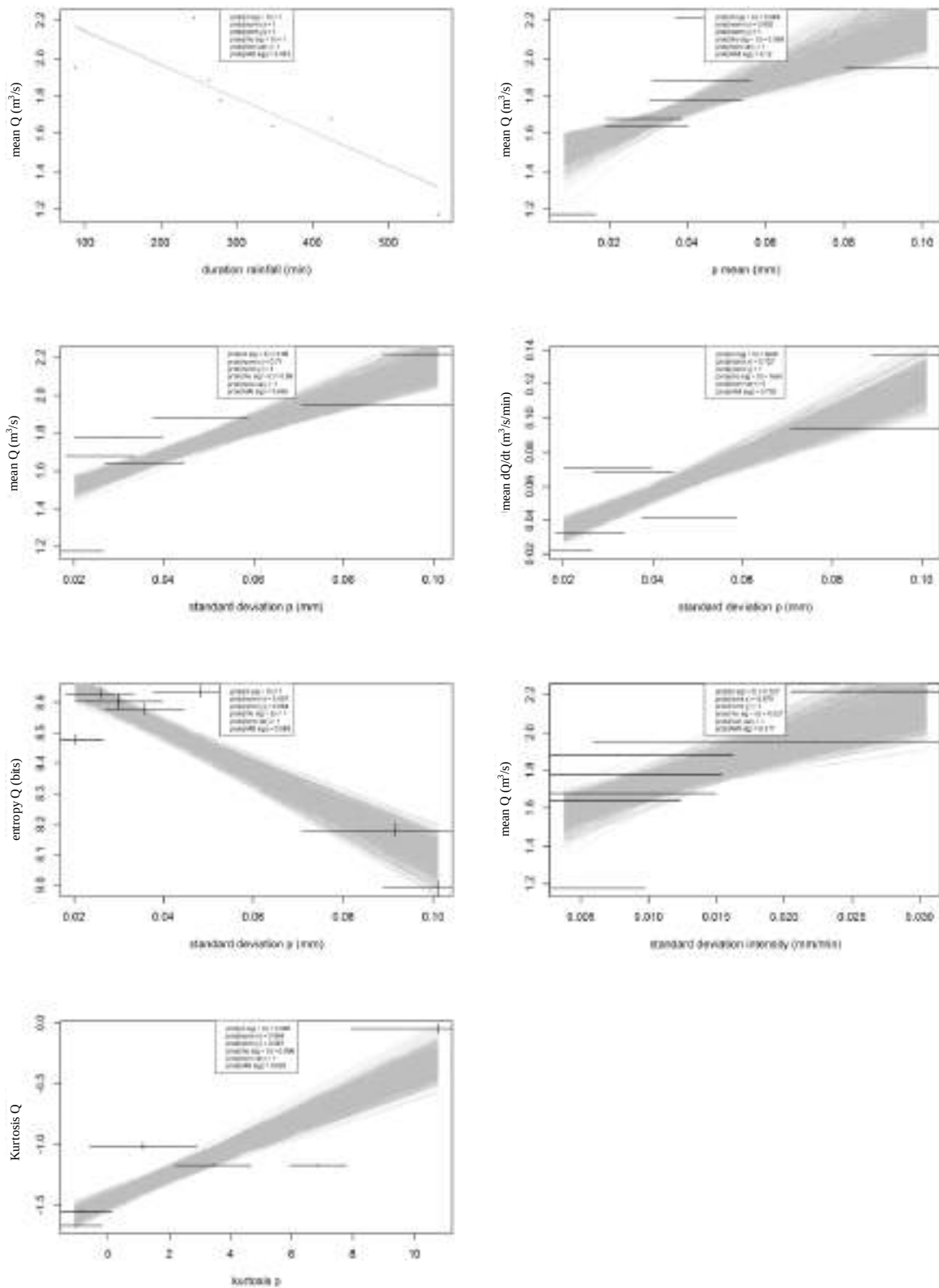


Figure 4.39 Recurrent relations between flow rate and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the Bosa rain gauge station

From the reported results, the following relations were identified for a rainfall event reported by means of rainfall registered at the *Bosa* rain gauge station: (i) the duration of the rainfall pulse seems to be related with the mean flow rate value, (ii) the mean rainfall seemed to be related with the mean flow rate value, (iii) the standard deviation of the rainfall seems to have a relation with the flow rate mean value, the mean flow rate gradient ( $dQ/dt$ ) and the flow rate entropy, (iv) the standard deviation of the rainfall intensity can have a relation with the mean flow rate value and (v) the kurtosis of the rainfall registrations seems to be related with the Kurtosis of the flow rate values.

Execution of a multivariate analysis, including different characteristics identifiable for rainfall pulses registered at a given rain gauge station as the input of the model, was undertaken. The model was evaluated with several characteristics from the output signal, for the identified rainfall events. The PLS analysis was undertaken for rainfall registered at the *Fontibon*, *Bosa* and *Casablanca* rain gauge stations. Representative results were obtained for the rainfall registered at the *Bosa* rain gauge station (see Figure 4.40).

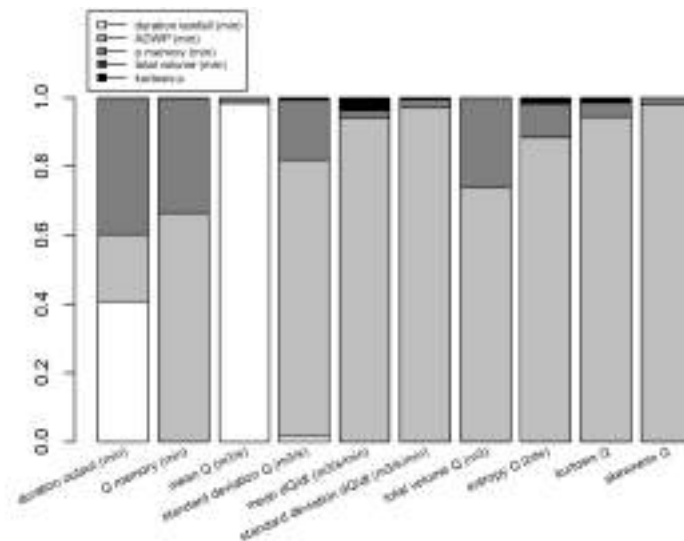


**Figure 4.40 Recurrent relations between flow rate and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the *Bosa* rain gauge station (PLS model)**

By considering uncertainty in the estimation of each variable (rainfall pulse or flow rate), the number of occasions in which each rainfall characteristic was the most important input variable were evaluated, in terms of explaining the output variable, by means of a PLS model (including Monte Carlo simulation). The duration of the influence over the output seems to be influenced by the ADWP, the duration of the rainfall and the memory of the precipitation (assessed by the ACF of the rainfall pulse), more than by the other studied rainfall variables. The memory of the flow rate values, during an event, seemed to be influenced mostly by the memory of the precipitation and the ADWP. The remaining studied variables, regarding the flow rate at the effluent of the catchment, were mostly explained by means of the ADWP, rather than by the other rainfall characteristics proposed.

The proposed conclusions obtained from the PLS regressions are even more hardly generalizable than the preliminary statements obtained from the univariate analysis. This statement can be supported by the fact that the PLS model does not include a formulation for evaluating the statistical

significance of the reported results. Therefore, although a given input variable of the model was identified as the most relevant variable, in terms of explaining a given output, this fact would not mean that a significant relation between the input and the output variables of the model can be established. Thus, PLS calculations bring up just relative insights about the hierarchy of the variables, in terms of explaining a given input.

The *Casablanca* rain gauge station calculations were not satisfactory due to the limited number of rainfall events detected by the proposed methodology (three rainfall events); as well as for the UV-VIS spectrometry estimations (in terms of the spectra and of the equivalent water quality parameters).

#### 4.2.2 Uni-parametric sensors

Online measurements of water quality parameters undertaken by means of the uni-parametric sensors were analyzed, in relation to the identified rainfall events, detected by using the rainfall registered at the *Fontibon*, *Bosa* and *Casablanca* rain gauge stations. With the purpose of establishing common characteristics among the different rainfall events, the proposed methods were executed. For a given rainfall event, the uncertainty in identifiable characteristics (*e.g.* mean, standard deviation, and entropy) of the rainfall pulse(s) linked to the event and the water quality data registered over a time window ranging outside the typical dry weather behavior, was assessed (Monte Carlo simulations). Furthermore, uncertainty in the association metrics established between rainfall and water quality characteristics during a rainfall event was evaluated (Monte Carlo simulations). The analyzed water quality parameters were: DO, TSS, Turbidity, ORP, Temperature and pH. The uncertainty of an isolated measurement was established based on the measurement resolution of the sensor: (i)  $\mu(\text{DO})=0.005$  mg/l, (ii)  $\mu(\text{TSS})=0.5$  mg/l, (iii)  $\mu(\text{Turbidity})=0.5$  NTU, (iv)  $\mu(\text{ORP})=0.5$  mV, (v)  $\mu(\text{Temperature})=0.005^{\circ}\text{C}$  and (vi)  $\mu(\text{pH})=0.005$  pH units (see details section 3.2).

Regarding the DO, the association metrics were implemented for the events identified by means of the rainfall registered at the *Fontibon*, *Bosa* and *Casablanca* rain gauge stations. The results obtained from rainfall registered at the *Fontibon* rain gauge station are exposed in Figure 4.41.

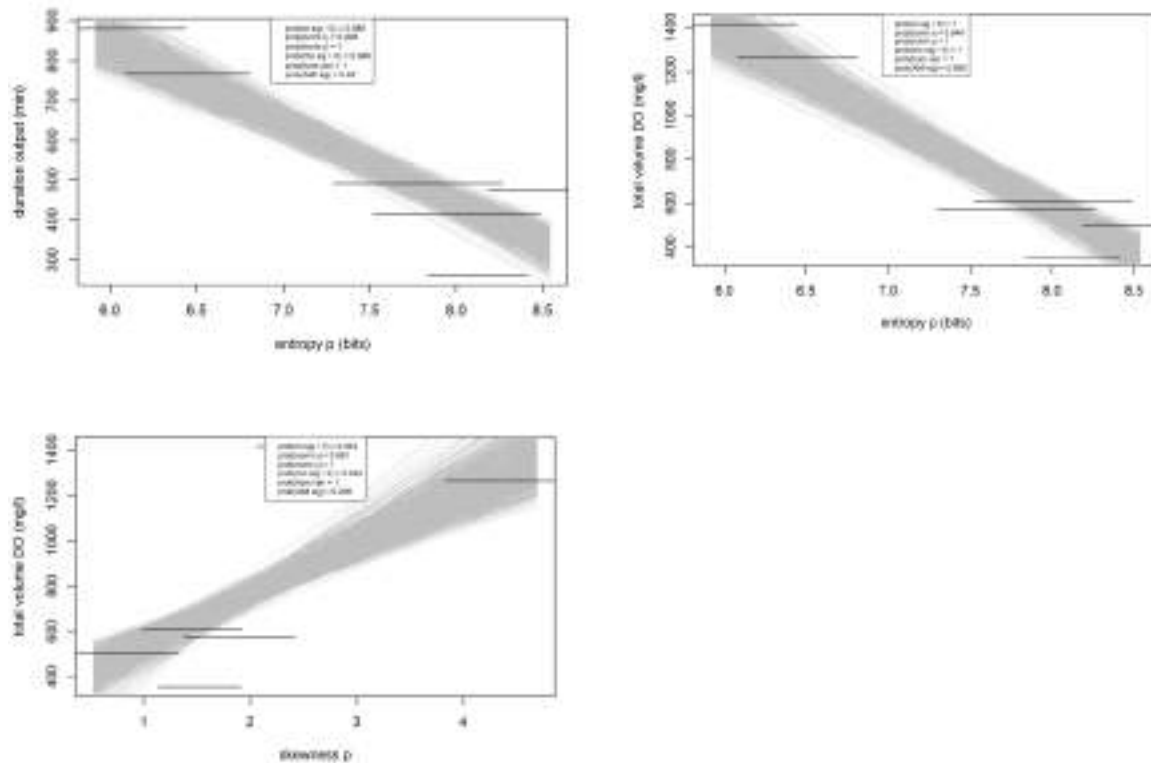


**Figure 4.41** Recurrent relations between the DO and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the *Fontibon* rain gauge station

From the above results, it can be stated that the duration of the output seems to be related with the entropy of the rainfall. Hence, the total DO seems to be related with the entropy of the rainfall, as well as with the skewness of the rainfall. In terms of a multivariate analysis, the results in Figure 4.42 were obtained from the PLS model, including different DO characteristics, for a given rainfall event:

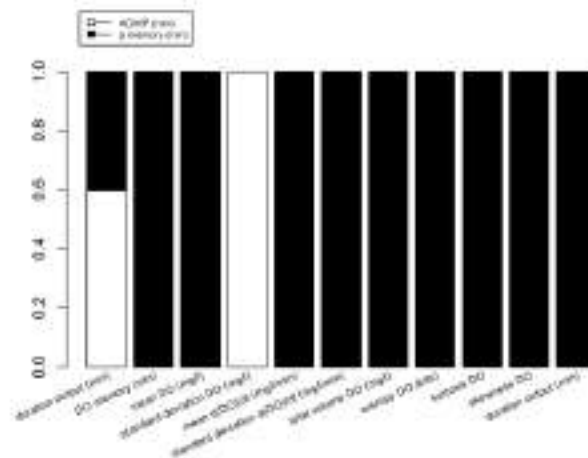


**Figure 4.42** Recurrent relations between DO and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the *Fontibon* rain gauge station (PLS model).

From the above results, it can be stated that, regarding a PLS approach, most of the output variables seem to be related with the memory of the rainfall pulse (estimated from the ACF, over particular rainfall pulses), rather than with the other studied rainfall variables (*e.g.* mean rainfall height, standard deviation of the rainfall height, gradients, entropy of rainfall intensity). However, the standard deviation of the DO during a given rainfall event seemed to be related with the ADWP, rather than with the rest of the rainfall's studied variables. The duration of the output seems to be more influenced by the ADWP and the memory of the rainfall pulse, rather than with the rest of the studied rainfall's variables.

Exploring the relations between the characteristics of the DO and the rainfall pulses registered by the *Bosa* rain gauge station, the relations of Figure 4.43 were perceived.

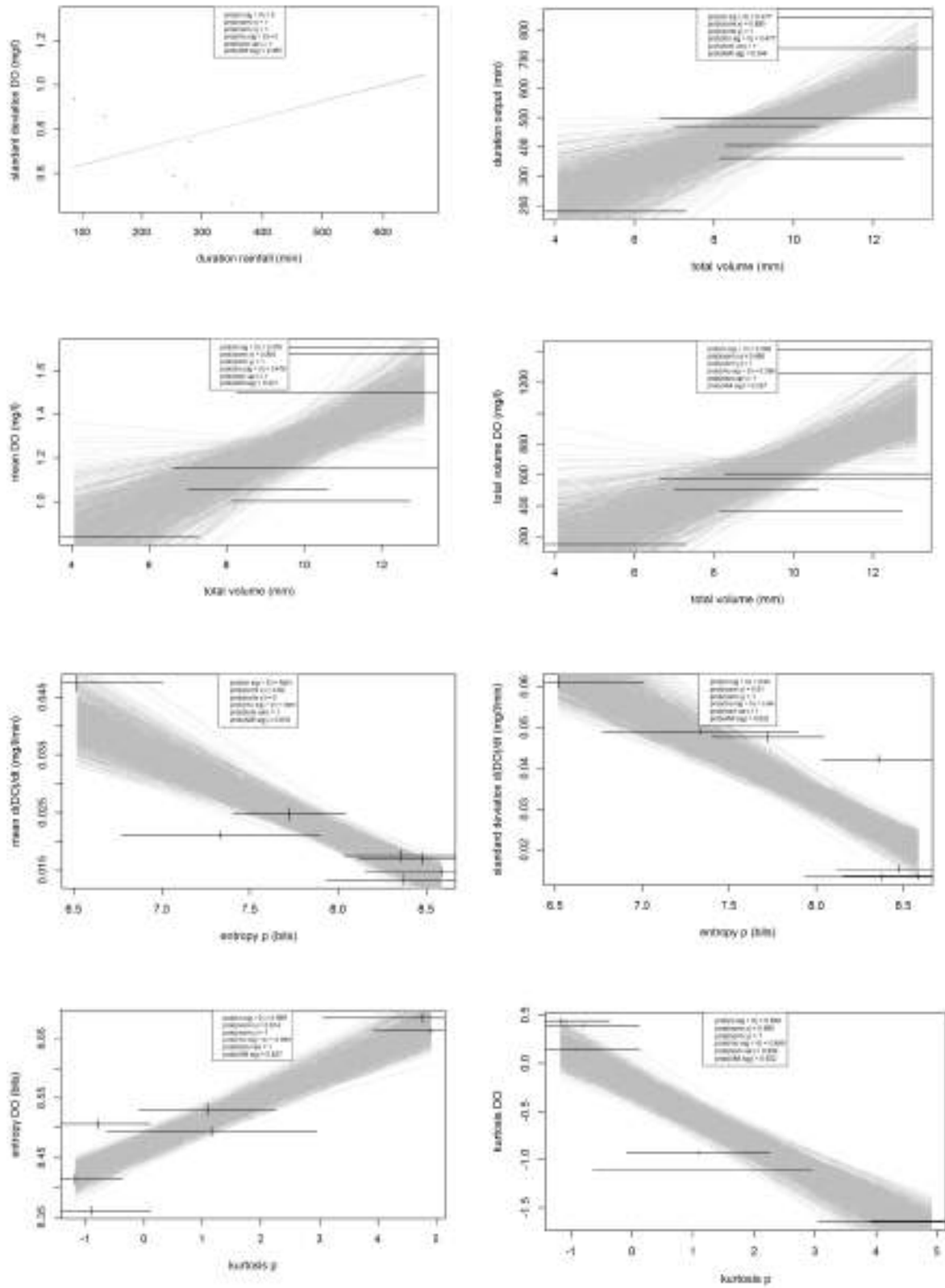


Figure 4.43 Recurrent relations between the DO and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the Bosa rain gauge station

From the above evaluations, it can be stated that the total volume of precipitation, registered at the *Bosa* rain gauge station, seems to be more related with the duration of the influence over the DO and with the total and the mean DO values, during a rainfall event. The entropy of the rainfall seems to be related with the mean and the standard deviation of the gradients ( $d(DO)/dt$ ), during a rainfall event registered at the *Bosa* rain gauge station. The Kurtosis of the rainfall seems to be related with the Kurtosis and the Entropy at the output of the catchment (DO signal), for a given rainfall event, registered at the *Bosa* rain gauge station. The last relation explored, reporting significant results in terms of the evaluated association metrics, was the duration of the rainfall against the standard deviation of the DO.

With the purpose of undertaking a multivariate analysis, a PLS model was proposed to identify which of the studied rainfall characteristics (reported from rainfall registered at the *Bosa* rain gauge station) seems to be more influent over the different studied variables, regarding the DO measurements, during a rainfall event. The results are exposed in Figure 4.44.

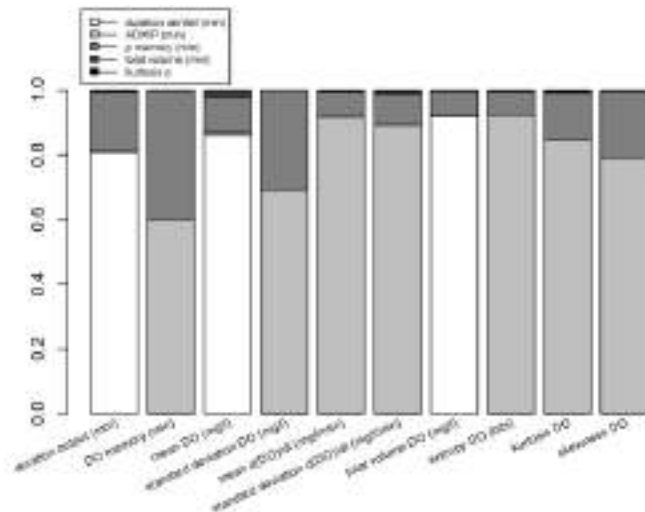


**Figure 4.44** Recurrent relations between DO and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the *Bosa* rain gauge station (PLS model).

In relative terms, considering a multivariate analysis, it can be concluded that most of the DO characteristics during a rainfall event (*e.g.* duration output, DO memory, mean DO and standard deviation DO), seem to be explained mostly by the ADWP and the duration of the rainfall. The DO memory (calculated from the ACF of the DO ranging outside the typical dry weather behavior) and the standard deviation of the DO, seemed to be more influenced, besides the ADWP, by the memory of the rainfall (calculated from the ACF of the identified rainfall pulses), rather than by the other rainfall variables calculable during a rainfall event.

The results obtained for rainfall registered at the *Casablanca* rain gauge station were reported in Figure 4.45.

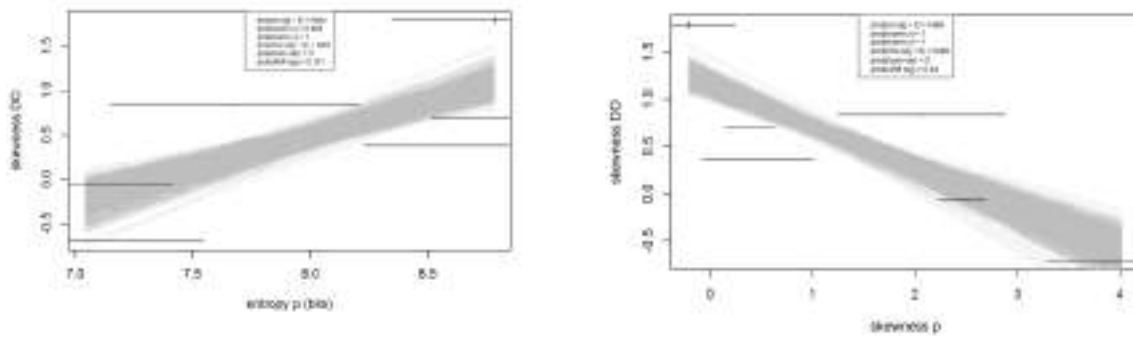


**Figure 4.45** Recurrent relations between the DO and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the *Casablanca* rain gauge station

From the above results, it can be concluded that recurrent relations in rainfall events could not be established as well as for the *Fontibon* and *Bosa* rain gauge stations. This seems to be related with the rainfall events' identification (in terms of rainfall pulses and time windows ranging outside the typical dry weather behavior), considering that for the *Bosa* and *Fontibon* rain gauge stations, the rainfall events' identification seemed to be more accurate than for the *Casablanca* case. The PLS analysis regarding rainfall registered at the *Casablanca* rain gauge station is reported in Figure 4.46.

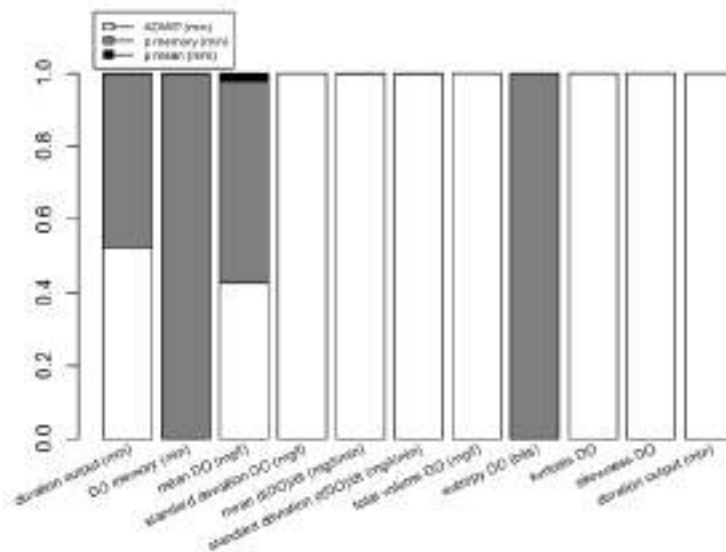


**Figure 4.46** Recurrent relations between DO and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the *Casablanca* rain gauge station (PLS model).

It can be appreciated in Figure 4.46 that the PLS results exposed a low variability in terms of the most relevant rainfall variables used to explain each of the DO variables, during a rainfall event, at the effluent of the studied catchment. In agreement with the exposed results, most of the water quality variables can be explained by means of the ADWP and the memory of the rainfall (estimated by the use of the ACF of the identified rainfall pulses), rather than by the other analyzed rainfall variables (e.g. mean intensity, standard deviation precipitation and entropy of the rainfall).

Regarding the TSS measurements undertaken by means of the uni-parametric sensors, recurrent relations between rainfall and water quality characteristics, among different rainfall events

(registered by the *Fontibon*, *Bosa* and *Casablanca* rain gauge station) were studied. The results for the rainfall registered at the *Fontibon* rain gauge stations, focusing in a univariate analysis are summarized in Figure 4.47.

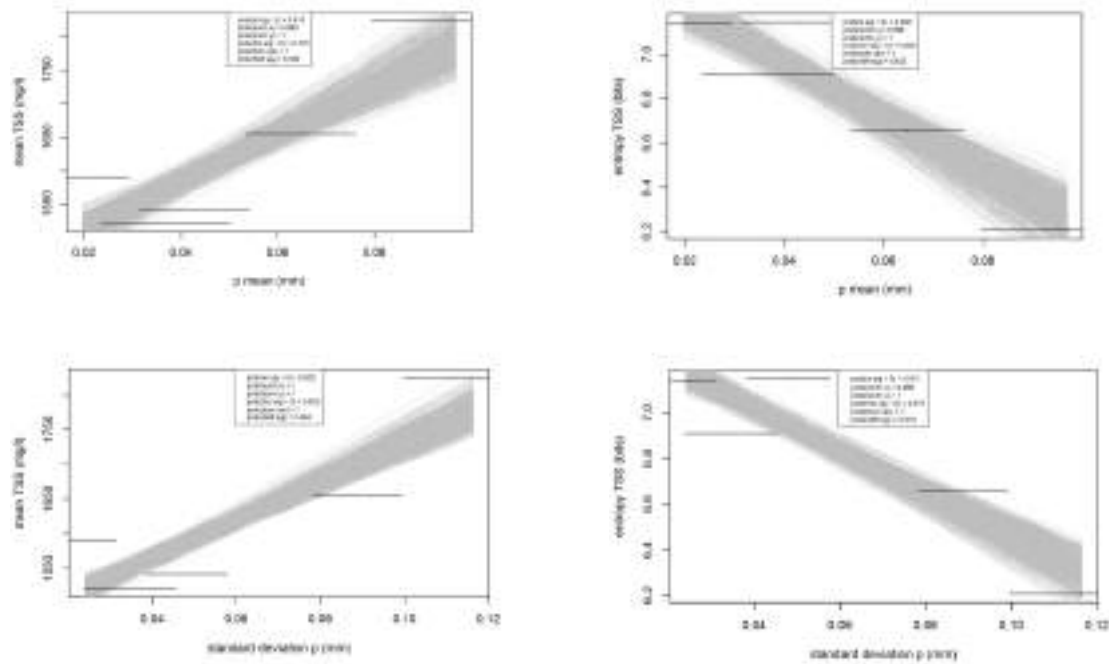


**Figure 4.47** Recurrent relations between the Total Suspended Solids and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the *Fontibon* rain gauge station

From these results, the following statements can be formulated: (i) the mean rainfall during an event seems to be related with the entropy and the mean TSS value and (ii) the standard deviation of the rainfall can be related with the entropy and the mean TSS value. In terms of a multivariate analysis, the PLS model did not perform appropriately, for the analyzed data.

By the use of the rainfall reported at the *Bosa* rain gauge station, the results of Figure 4.48 were obtained.





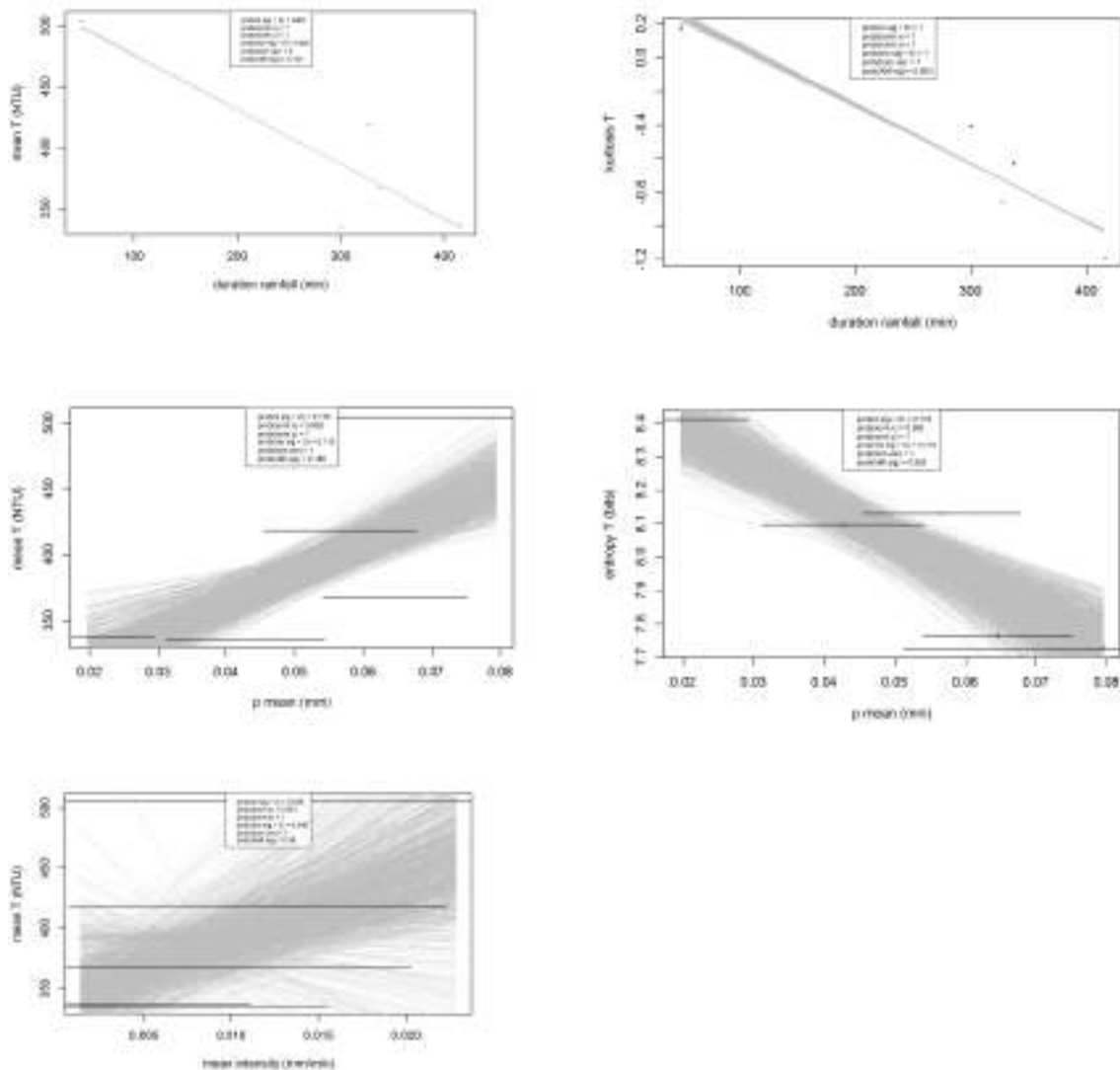
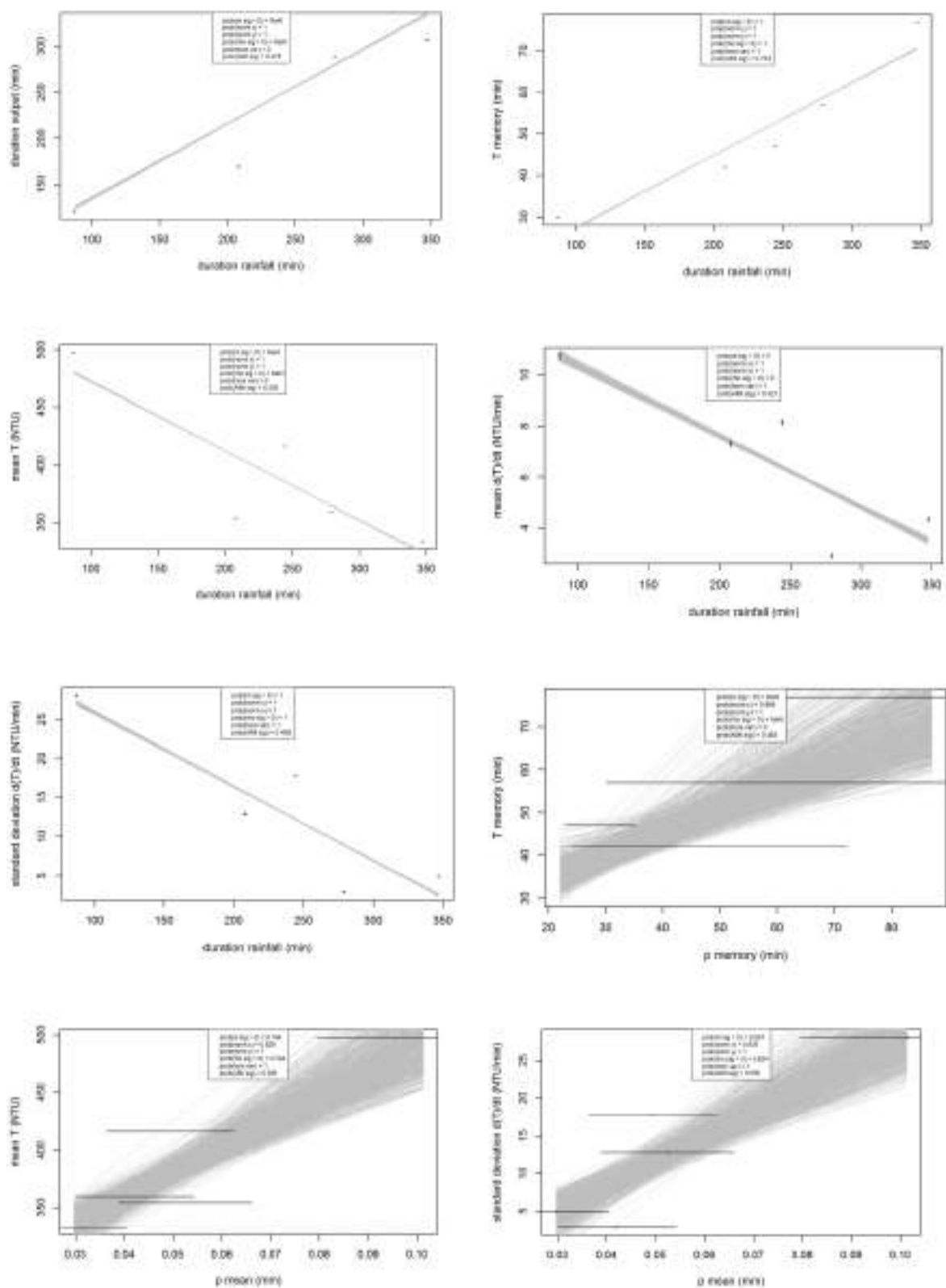


**Figure 4.50** Recurrent relations between the Turbidity and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the *Fontibon* rain gauge station

From the reported results, it can be stated that scarce significant relations were found between Turbidity and rainfall characteristics at the *Fontibon* rain gauge station. This fact can be explained by the limited number of rainfall events which were registered. Preliminary conclusions are: (i) mean precipitation seems to be related with the mean and the entropy of the Turbidity and (ii) the mean rainfall intensity seems to be related with the mean Turbidity, during a given rainfall event, by means of the explored association metrics (Pearson's Correlation Test, Linear Model Test and Average Mutual Information).

Hence, the results reported for rainfall events identified by the rainfall registered at the *Bosa* rain gauge station are presented in Figure 4.51.



More reliable relations between rainfall and Turbidity, than for the case of the *Fontibon* rain gauge station could be identified by means of the univariate association metrics proposed (Pearson's Correlation Test, Linear Model Test and Average Mutual Information). The rainfall's duration seems to be a characteristic of the rainfall pulse relevant for explaining: (i) the duration of the influence over the Turbidity time series, (ii) the memory of the Turbidity values influenced by rainfall (measured by ACF of the studied pollutograph), (iii) the mean of the Turbidity values, (iv) the mean of the Turbidity gradients ( $d(T)/dt$ ) and (v) the standard deviation of the Turbidity gradients ( $d(T)/dt$ ), for a given rainfall event. In addition, the mean rainfall during an event seems to be related with the mean Turbidity, as well as with the standard deviation of the Turbidity gradients ( $d(T)/dt$ ), measured at the effluent of the catchment. Hence, the memory of the physical process during the evolution of the pollutograph seems to be related with the memory of the rainfall pulse(s) which caused it (both memories estimated by means of the ACF). Addressing a multivariate approach, by the use of a PLS model and the rainfall registered at the *Bosa* rain gauge station, non-satisfactory results were obtained. Likewise, for the *Casablanca* rain gauge station (including univariate and multivariate analysis), results were not representative.

As for pH, Temperature and UV-VIS spectrometry measurements (spectra and equivalent water quality parameters) the proposed methods could not identify rainfall events appropriately, or the number of identified rainfall events was not enough for undertaking statistical analysis.

The last water quality parameter to be analyzed was the ORP (measured with uni-parametric sensor). Thus, by using the rainfall events identified for the *Fontibon* rain gauge station, the results were reported in Figure 4.52.

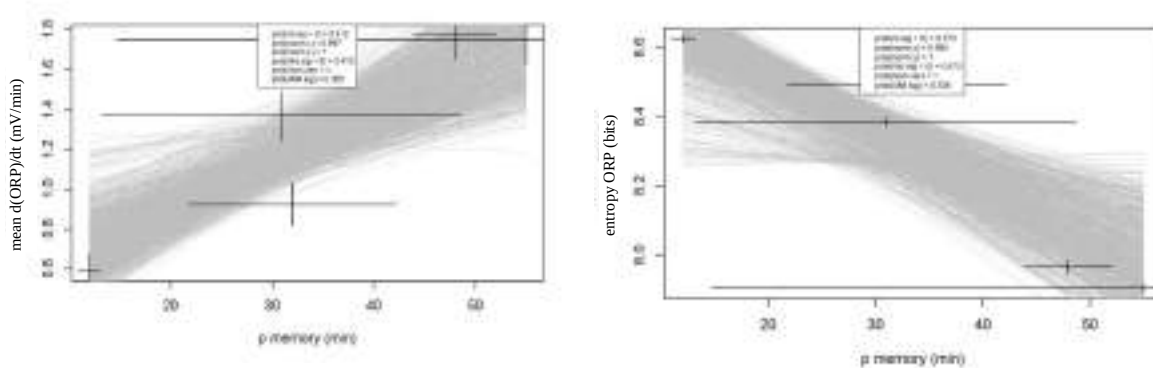


**Figure 4.52** Recurrent relations between the ORP and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the *Fontibon* rain gauge station

In agreement with the exposed results, a relation of ORP's variables with rainfall characteristics could not be established with a significance level higher than 30 %, through the studied association metrics. In this sense, although a sufficient number of rainfall events could be detected by means of the proposed methods, recurrent relations could not be established by univariate or multivariate approaches that were applied. Results obtained from rainfall registered at the *Bosa* rain gauge station are presented in Figure 4.53.

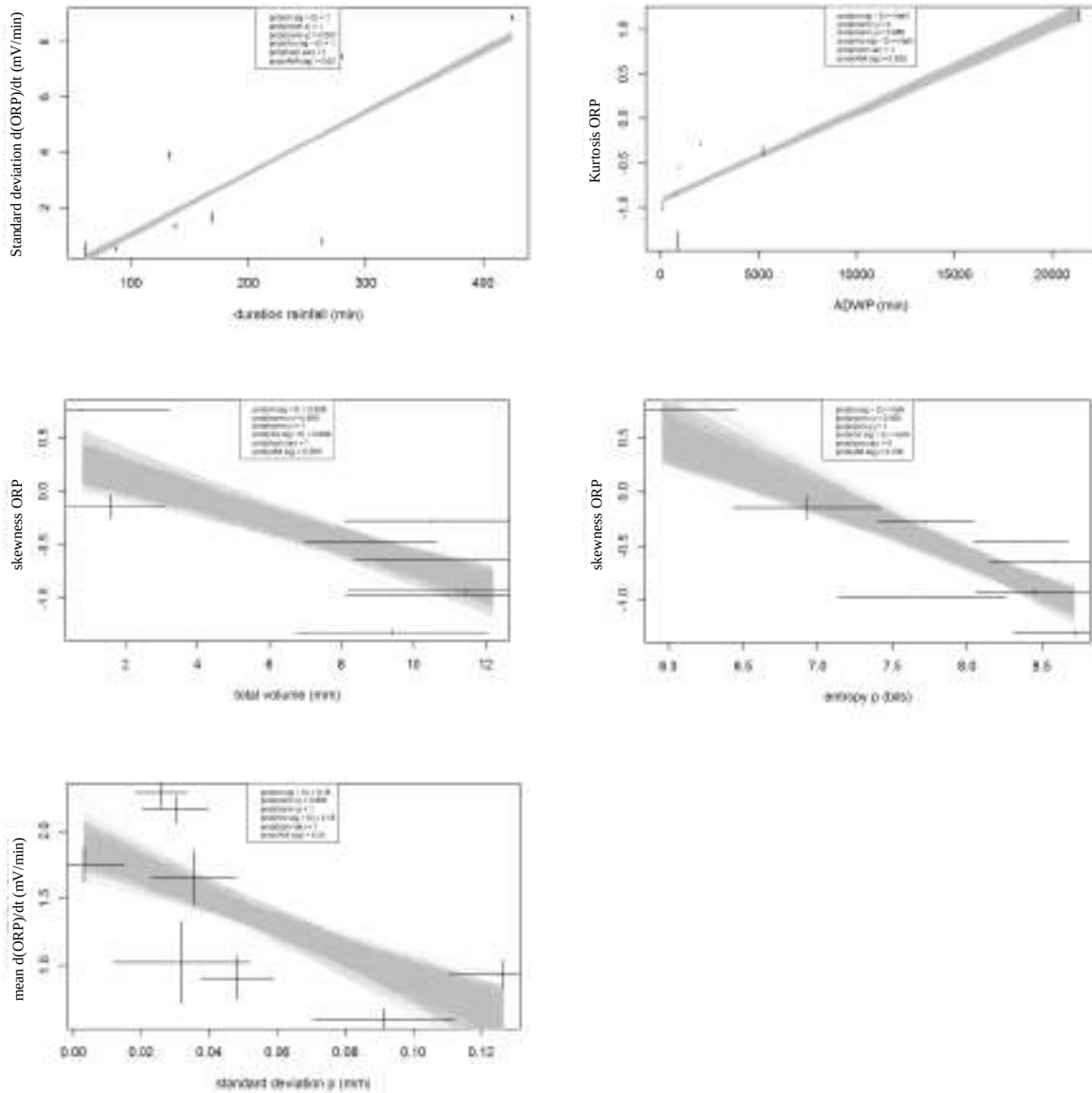


**Figure 4.53** Recurrent relations between the ORP and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the *Bosa* rain gauge station

It can be stated that the duration of the rainfall seemed to be related with the standard deviation of the ORP gradients ( $d(ORP)/dt$ ). ADWP seems to be related with the kurtosis of the probability density function (PDF) of ORP, during a rainfall event. The total rainfall volume and the rainfall entropy can be related with the skewness of the ORP's PDF, reported for a given rainfall event. The

standard deviation of the rainfall seems to have an influence over the mean ORP gradient ( $d(ORP)/dt$ ), in the course of a rainfall event, at the studied urban catchment.

For a multivariate analysis (executed with a PLS model), by the use of rainfall registered at the *Bosa* rain gauge station, the results of Figure 4.54 were obtained.

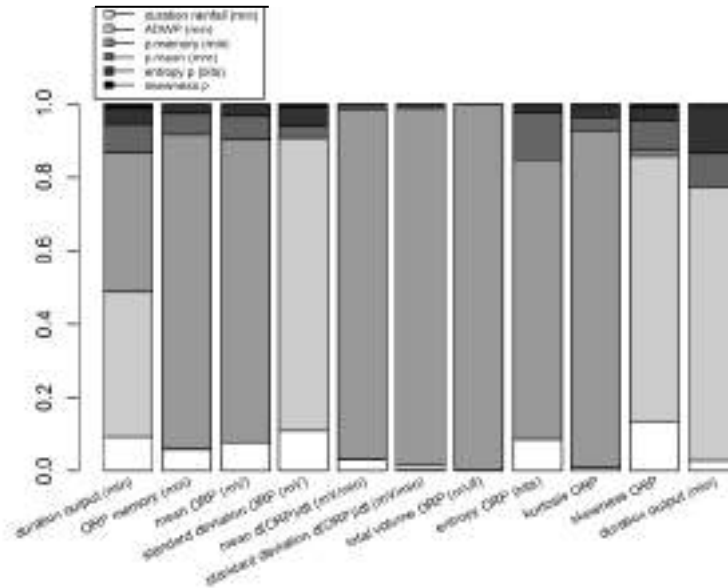


**Figure 4.54 Recurrent relations between ORP and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the *Bosa* rain gauge station (PLS model)**

It can be observed that the most important rainfall characteristics which explain the analyzed ORP's variables are the memory of the rainfall (estimated by the ACF of rainfall pulse, at a given event) and the ADWP, rather than the rest of the rainfall characteristics considered. Hence, the variables which are mostly influenced by rainfall memory are: ORP memory, mean ORP, mean  $d(ORP)/dt$ , standard deviation  $d(ORP)/dt$ , total volume ORP, entropy ORP and kurtosis ORP. For the ADWP, the most related variables with this rainfall characteristic seemed to be: the standard deviation of ORP, the skewness of ORP and the duration of the influence over the output. However, these results are just relative comparisons between the studied variables. In order to assess more general relations, more rainfall events and significance testing are strongly recommended. For the case of rainfall registered at the *Casablanca* rain gauge station, no significant relation between rainfall and ORP's characteristics during an event, with a probability over 30 %, was detected. However the multivariate analysis could be assessed, bringing up the relative importance of the variables for each rainfall event, without testing significance of the relations.

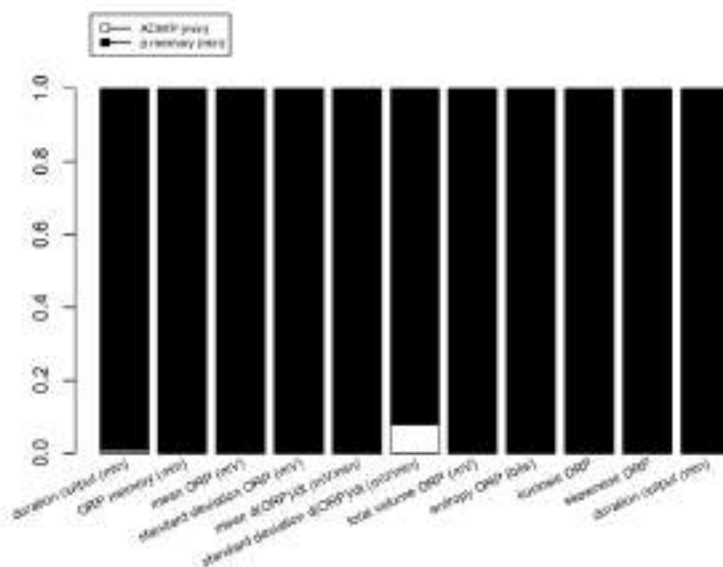


**Figure 4.55 Recurrent relations between ORP and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the *Casablanca* rain gauge station (PLS model)**

From the above results, it can be observed that the most important variable reported, over the other studied rainfall characteristics, was the memory of the rainfall. However, these results are just relative insights. In order to have more general conclusions, the previous statements should be supported with significance tests, given by the proposed univariate analyses (which were not conclusive for the ORP case, regarding the *Casablanca* rain gauge station). With the purpose of further clarification, multivariate analysis should also have significance testing (multivariate linear and nonlinear models), but considering the sample's size (about seven rainfall events), the number of parameters will be greater than the observations for the model, limiting their application. In this sense, a reasonable alternative is the PLS model (designed for a short number of observations, against the number of input variables); one should, however, keep in mind that the PLS model does not include significance testing. This fact makes the PLS model useful only for undertaking relative comparisons between the input variables, in terms of representativeness over the output of the model (water quality variables, in this case).

From the results referring to recurrent relations between rainfall characteristics and water quality's variables, it can be concluded that the rain gauge station which exposed the more clear relation in terms of: (i) amount of variables correlated, (ii) probabilities of significant relations, independently of the association metric implemented (Pearson's Correlation Test, Linear Model test or Average Mutual Information) and of the analyzed water quality parameter (Flow rate, DO, TSS, Turbidity or ORP), was the *Bosa* rain gauge station. However, the *Fontibon* rain gauge station exposed very clear relations regarding water quality parameters (specially for Turbidity and TSS). This leads to conclude that although the *Fontibon* showed significant relations over water quality, the rainfall registered at the *Bosa* rain gauge station seems to be more representative in terms of water quality measurements at the effluent of the studied urban catchment than the *Fontibon* rain gauge station (independently of the water quality parameter to be analyzed). For the *Casablanca* rain gauge station, the number of rainfall events detected by means of the proposed methods seems to be lower than for the other studied rain gauge stations. Hence, rainfall pulses registered by this rain gauge stations without a relation with the water quality's time windows ranging outside their typical dry

weather behavior were frequently noticed. Therefore, recurrent relations between rainfall characteristics and water quality could hardly be found as significant. This fact may lead to conclude that the influence of rainfall registered at the *Casablanca* rain gauge station over the effluent of the catchment (in terms of quantity and quality), seemed to be lower than for the case of the *Fontibon* or *Bosa* rain gauge stations. In addition, the identified recurrent relations can be useful for modeling and real time operation purposes. By the identification of representative statistical characteristics of a given rainfall pulse and of its corresponding pollutograph, modeling strategies can be built up more efficiently; modelers can evaluate the most appropriate relations that will lead to select a probabilistic or deterministic approach for representing the reality based on mathematical relations. Hence, real time operation is a very strong perspective for the present study.

Prediction techniques can also be proposed by means of the mentioned approach. Whenever a given rainfall pulse is shorter than the time of response of a water quality parameter (see section 4.1), by the calculation of a summarized behavior of the pollutograph based on the recurrent relations found, operative actions can be undertaken in the WWTP. These linkages between water quality and rainfall seem to be pertinent for decision-making processes, in an urban drainage system. For the case study, with perspectives of real time operation at the effluent of *Gibraltar*'s urban catchment, the installation of an automatic online system for rainfall reporting at the *Bosa* rain gauge station (more than for *Fontibon* or *Casablanca*) is strongly recommended.

### 4.3 Association of recurrent relations with physical characteristics of the urban catchment

For the case study (Bogotá, Colombia), the principal urban catchments are divided in sub-catchments, which are called UGAs (*Unidad de Gestión de Alcantarillados*). The UGAs that constitute the studied urban catchment (*Gibraltar*) have physical characteristics that are described in section 3.1.1. It can be noticed that, for this study, all characteristics mentioned in Table 2.1 are assumed to be constant over the whole UGA's surface (*e.g.* mean diameter, total pipes' length). Therefore, no difference in terms of physical characteristics can be attributed to one point or another, whenever they belong to the same UGA. In this sense, it will be stated that for all spatial points within a given UGA, the representativeness of rainfall influence will be given by the nearest rain gauge station to the centroid of the UGA.

With the purpose of applying the proposed methods, the first step is to label each rain gauge station, in agreement with the recurrent relations obtained. As discussed in section 4.2, the order in which the rain gauge station seemed to have an influence from higher to lower, independently of the studied water quality parameter, can be listed as follows: (i) *Bosa*, with a high influence, (ii) *Fontibon*, with a medium influence and (iii) *Casablanca*, with a low influence. Therefore, the following label is proposed: (i) *Bosa\_high\_infl*, (ii), *Fontibon\_med\_infl* and (iii) *Casablanca\_low\_infl*.

Regarding the case of study, no UGA had the *Casablanca* rain gauge station closer to its centroid than the other stations (*Fontibon* or *Bosa*). Accordingly, assigning a label to each UGA was limited to: *Bosa\_high\_infl* (UGAs 467, 501 and 470) and *Fontibon\_med\_infl* (UGAs 461, 462, 465 and 464).

A PCA was undertaken with the purpose of detecting explanatory variables of the differences between the variability of UGAs with a “High” (blue) or “Medium” (green) influence over the water quality at the effluent of the catchment in Figure 4.56. The data are represented by means of three principal components (comp. 1: 47 %, comp. 2: 77 % and comp. 3: 88 % of the cumulative variance).

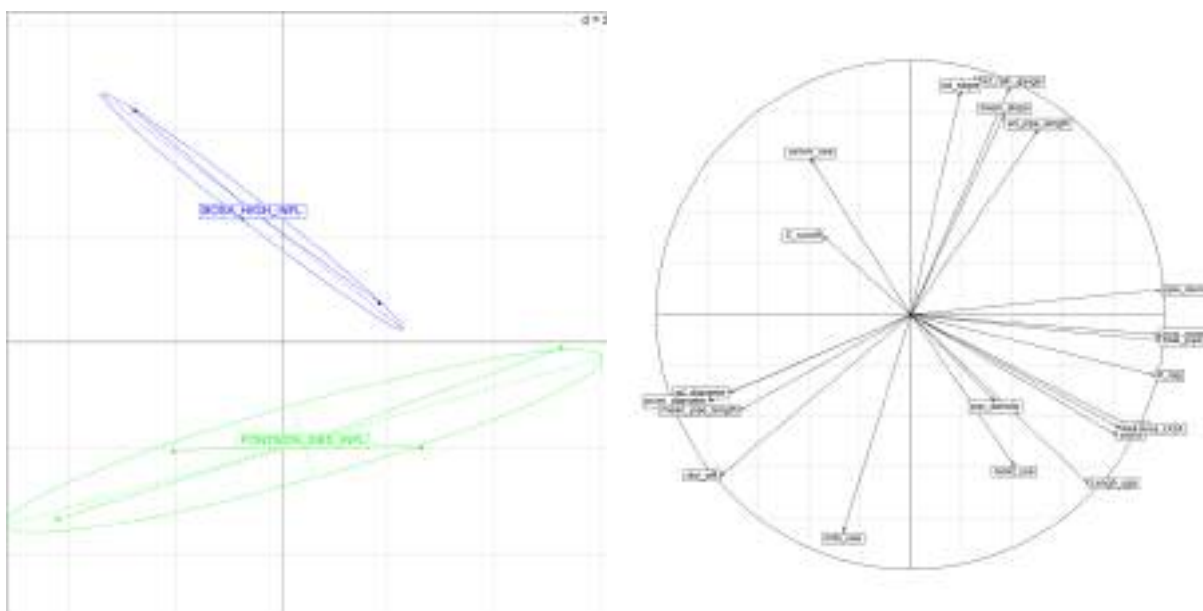


**Figure 4.56** PCA analysis regarding UGA’s characteristics and its influence from rainfall, given by their nearest rain gauge station (component 1: x-axis, component 2: y-axis)

From the above results, it can be concluded that the most relevant variables (regarding physical characteristics of the study zone) that explain the difference between UGAs with a “high” or “low” influence (in terms of quantity and quality) over the effluent of the studied catchment, can be listed the following in order of importance: (i) the distance to the rain gauge station which is mostly influencing the UGA, (ii) the variability of the pipes’ slope in the UGA, (iii) the industrial land use of the UGA, (iv) the mean pipes’ slope in the UGA and (v) the variability of the pipes’ length in the UGA. However, these are preliminary observations, and more conclusive statements can only be proposed based on further statistical tests and a more detailed spatial analysis of the physical characteristics of each point, rather than average values over the UGA’s surface. Therefore, these recommendations are strongly suggested for incoming studies.

Although *Bosa* showed a higher amount of recurrent relations between rainfall characteristics and water quality variables, representative of a given rainfall event than the *Fontibon* rain gauge station, the location of the mentioned rain gauge stations seems to be contradictory with this statement (*Fontibon* is closer to the effluent than the *Bosa* rain gauge station, see details in section 3.1.1). Hence, the additional identified physical characteristics that have a contribution over the influence of rainfall should bring up insights of the possible behavior of the water (in terms of quantity and quality) inside the studied urban catchment. For example, as the variability in the slopes and in the pipes’ length for an UGA seems to set a difference between a high or low influence from rainfall in terms of water quality, it can be stated that due to variability of velocities in the sewer pipe, settling and resuspension of sediments is encouraged. This can affect the response time for parameters such

as Turbidity or TSS, making them higher (statement in agreement with the results exposed in section 4.1.3). These conclusions were assessed for a range assignment given by the recurrent relations identified in section 4.2. However, they are just relative and preliminary insights. Further studies and analysis should be undertaken in order to have more generalizable and definite statements.

#### 4.4 Future Work

In addition to the results reported in the present study, further ideas were explored with the purpose of formulating perspectives for future work and incoming studies. The main two topics that were addressed, as an exploratory work, can be listed as follows: (i) implementation of the methodology for pollutant's loads and grouped analysis between water quality parameters (rather than individual calculations for each water quality parameter) and (ii) pollutograph's modeling (deterministic or semi-determinist approaches?)

Preliminary results of applying the methodology to the TSS loads (measured by Echo Profile Flowmeter and uni-parametric TSS concentration) will be exposed in the following lines. Calculations of the probabilistic characterization of dry weather behavior are presented in Figure 4.57.

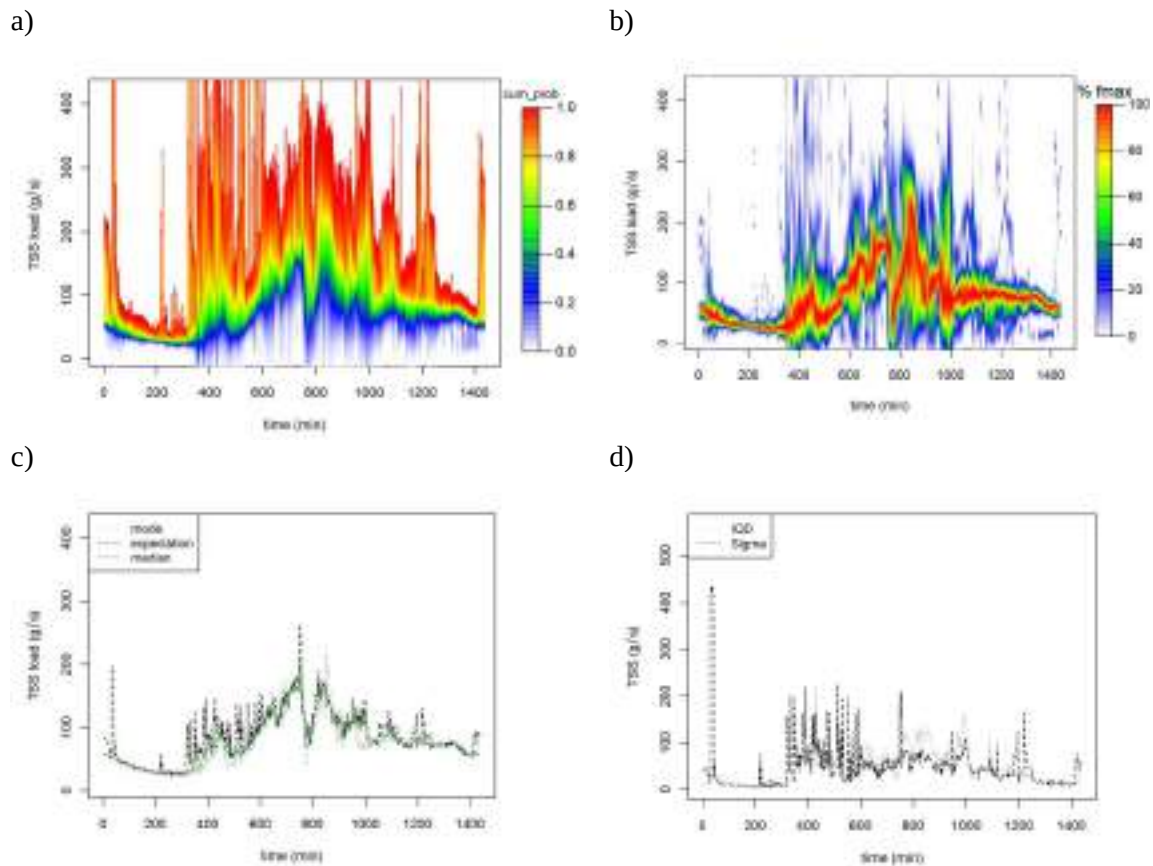


**Figure 4.57 Dry weather probabilistic characterization for TSS loads, using rainfall registrations at the Bosa rain gauge stations; a) Probability density function (PDF) for each time-step; b) Cumulative probability function (CPF) for each time-step; c) Mode, Median and Expectation for each time-step; d) Interquartile distance and standard deviation for each time-step**

It can be stated that a dry weather probabilistic characterization could also be assessed for the TSS loads at the effluent of the studied catchment, by the rainfall registered at the *Bosa* rain gauge station. TSS loads seemed to have a similar behavior as TSS concentrations measured by the uni-parametric sensor, in terms of the Mode, Expectation and Median values (see section 4.1.3.2.2). On the other hand, the variability curve (Figure 4.57 d) seems to be more alike to the variability of the flow rate curve (Figure 4.6 j).

With the purpose of implementing a Median Filter, the ACF function of the TSS loads was computed, in order to estimate the length of the mobile window to be used in the proposed filter.

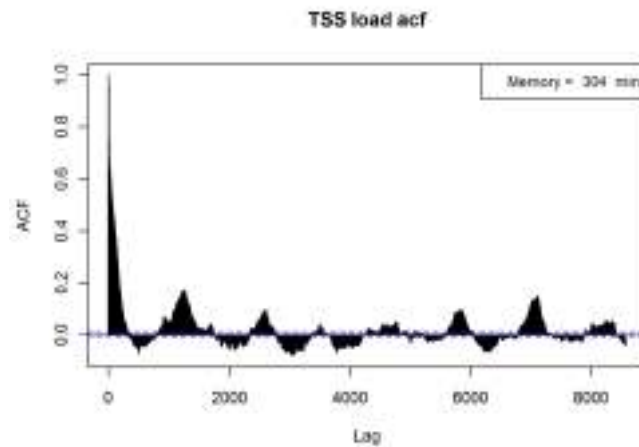


**Figure 4.58 Autocorrelation function for TSS loads**

From the results in Figure 4.58, it can be concluded that the ACF for TSS loads seems to be more alike to the ACF of Flow rate, rather than to TSS. The periodical behavior of the TSS ACF (see section 4.1.3.2.2) is lost as it is transformed to load (multiplied by the Flow rate value, registered for the same time-step). Hence, the memory of the process seems to be more alike to the flow rate (370 min for flow rate to 304 min for TSS load). By application of the Step Detection algorithm, the results of Figure 4.59 were obtained.

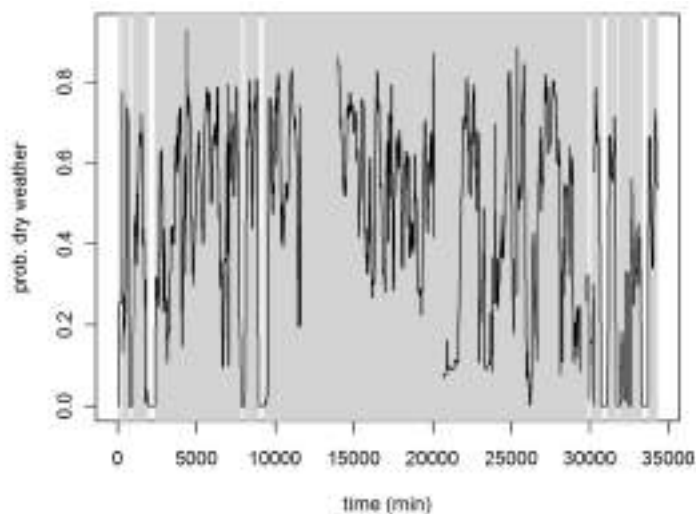
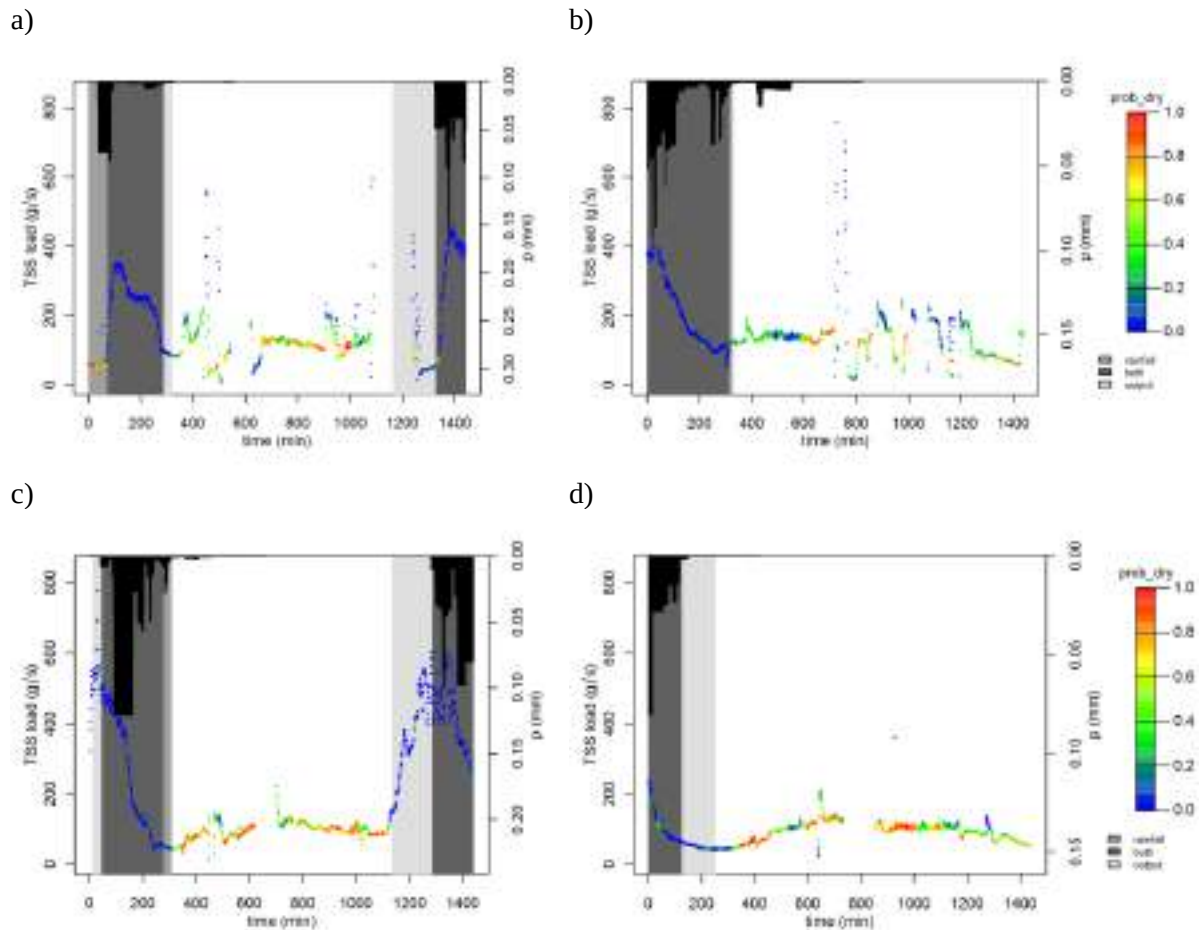


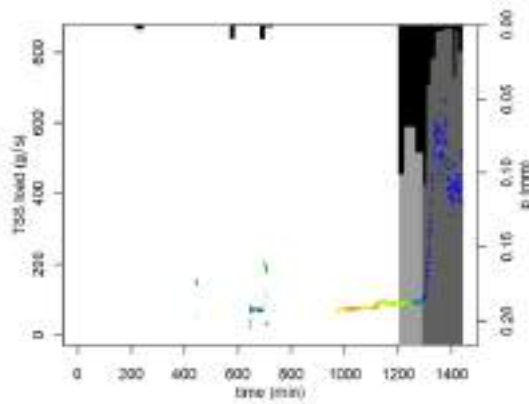
**Figure 4.59 Probability of dry weather signal filtered by a Median Filter, for TSS loads signal**

The “probability of dry weather” signal could be computed by means of the dry weather probabilistic characterization (Figure 4.57). Thus, the signal could be filtered by a Median Filter with 304 min’s mobile time window. By application of the Step Detection algorithm, time windows in which the mean of the “probability of being dry weather” signal was under 0.05, without reporting abrupt changes, were identified.

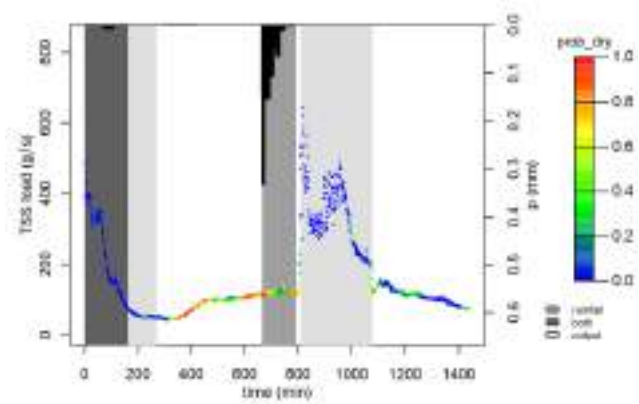
Including the rainfall pulses registered at the *Bosa* rain gauge station, the results for the proposed methods, in terms of identifying rainfall pulses and time window ranging outside their typical dry weather behavior were obtained (Figure 4.60).



e)



f)



g)

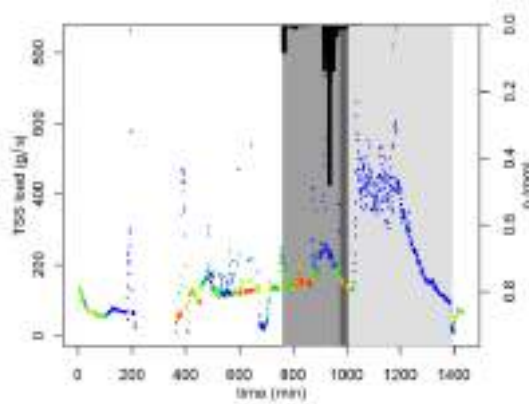


**Figure 4.60** Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding TSS loads (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Bosa* rain gauge station

By the proposed methods, a total of seven rainfall events (Figure 4.60: a, a-b, c, c-d, e-f, f, g) could be assessed. Hence, the identification of time windows ranging outside the typical dry weather behavior seems to be consequent with the results of rainfall pulses at the *Bosa* rain gauge station. These results can contribute with new insights for incoming studies, in terms of exploring the relation of rainfall with loads and also with combinations between water quality parameters time series (e.g. TSS with DO?), for the same time-step or for a lagged condition (time series convolution).

Regarding recurrent relations between rainfall registered at the *Bosa* rain gauge station and TSS loads’ characteristics, the results of Figure 4.61 were obtained.

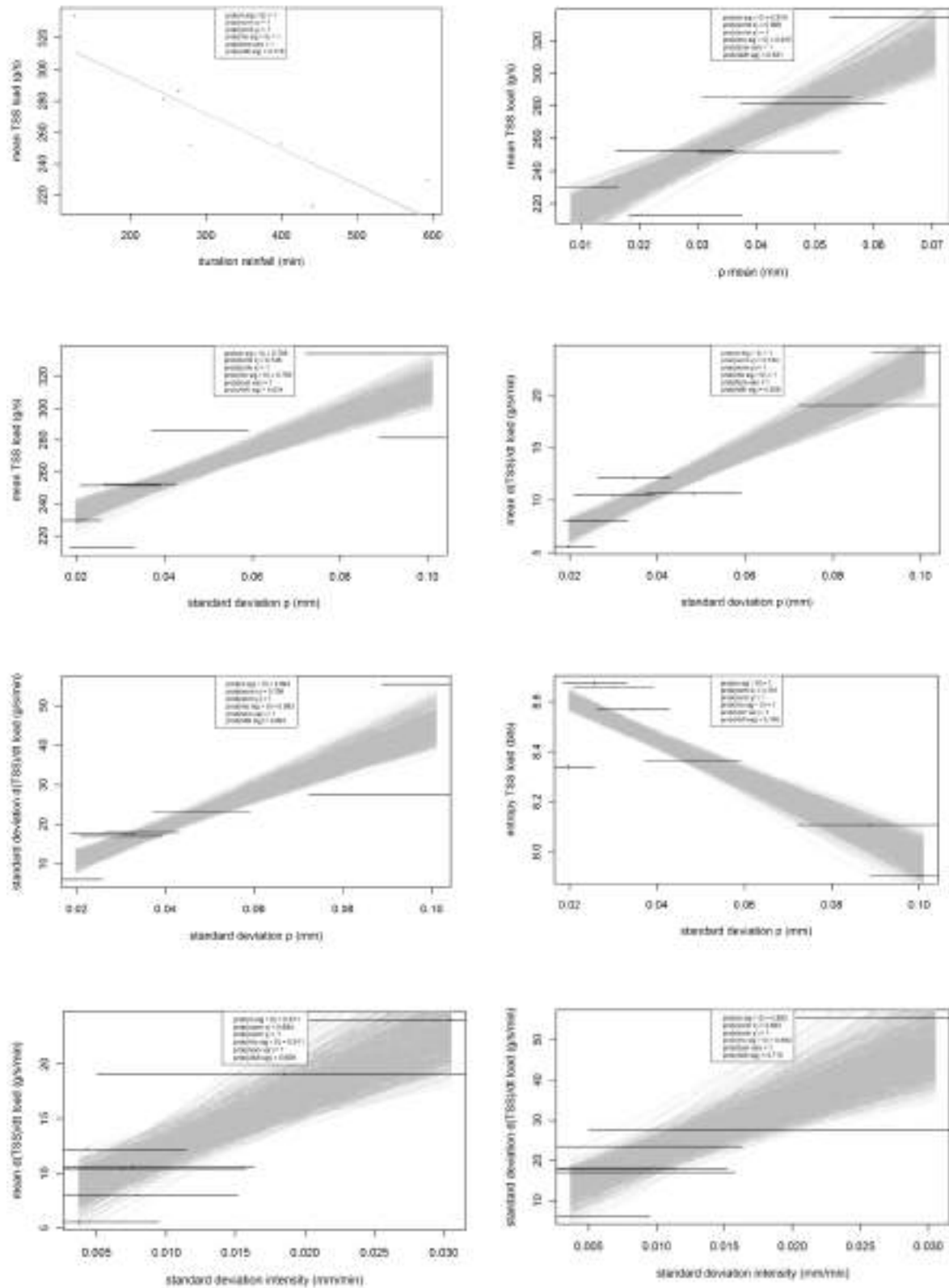


Figure 4.61 Recurrent relations between the TSS loads and rainfall statistics calculated for the identified rainfall events, by means of rainfall registered at the Bosa rain gauge station

From the above results, it can be stated that the TSS load seems to be influenced commonly by the standard deviation of the rainfall height and the standard deviation of the rainfall intensity. The mean TSS load during a rainfall event, seems to be more influenced by the mean rainfall height and by the duration of the rainfall pulse. The TSS loads and the TSS loads' gradients ( $d(TSS)/dt$ ) during a rainfall events seem to be influenced not only in their mean value, but also in their standard deviation and their entropy.

Preliminary associations can also be established for the modeling of pollutographs. Considering that several characteristics of the rainfall and the TSS loads during a given rainfall event are related, a simple model is proposed with the purpose of exposing the pertinence of establishing which water quality parameters (*e.g.* TSS loads) are indeed related with rainfall. In addition, by an appropriate temporal delimitation of the rainfall influence over water quality, data used for the model will not contain noisy values from the dry weather behavior. A SISO model is commonly used as a deterministic model of the rain-runoff process (Camacho *et al.*, 2010). The difference equation can be written as follows:

$$y(t) = b * x(t - \delta) - a * y(t - 1) \quad \text{Eq. 4.1}$$

where:

a,b- parameters of the model [M/L<sup>3</sup>];

$\delta$ - advective lag time [T];

y(t)- TSS loads at the output of the catchment [M/T];

t- time coordinate [T];

x(t)- rainfall [L<sup>3</sup>/T];

With the purpose of being consistent with units, the rainfall should be multiplied by the density of the TSS particles. However, regarding the calibration procedure, the model's parameters will absorb the dimensional inconsistency. Considering the TSS loads values and the rainfall registered during the identified rainfall events (one event for calibration and another for validation), the model was calibrated by minimizing the following goal function:

$$MSE = \sum_{j=1}^{\text{duration of the influence}} (TSS_{obsj} - TSS_{simj})^2 \quad \text{Eq. 4.2}$$

where  $j$  is the sub-index related to each time-step (one minute time-step). From the calibration procedure, a value of  $a = -0.992$  and  $b = 680.45$  exposed an appropriate fit (Figure 4.62).

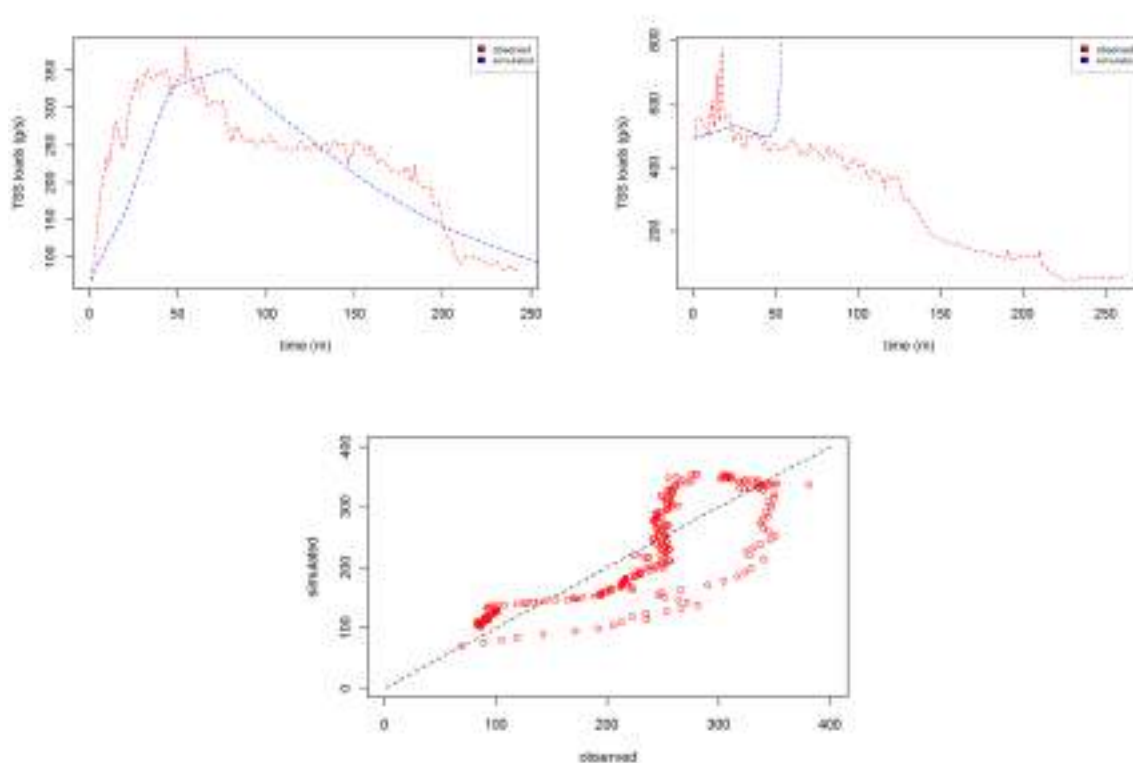


**Figure 4.62 Modeling results for TSS loads regarding the a) calibration rainfall event, b) validation rainfall event, c)observed-simulated values (calibration rainfall event)**

From this preliminary assessment, it can be observed that the calibration event exposed an appropriate adjustment (Figure 4.62). However, mathematical models based on equations such as Eq. 4.2 seem to have problems for reproducing the TSS loads' variability measured by means of high temporal resolution devices. The pollutographs produced by the model seem to be smoother than the observed values. Therefore, mathematical models capable to reproduce a more detailed behavior (deterministic or stochastic) of the pollutographs produced during rainfall events seem to be a future challenge for urban drainage modeling.

## 5 CONCLUSIONS

The present study aimed to establish the influence of rainfall on the water quality at the effluent of an urban catchment by the use of temporal high resolution measurements. The water quantity and quality parameters at the effluent of an urban catchment (*Gibraltar*: Bogotá, Colombia) were measured by means of infra-red, ion sensitive, UV-VIS spectrometry (water quality) and ultrasonic (water quantity: flow rate) sensors. Rainfall pulses were registered at three rain gauge stations (*Fontibon*, *Bosa* and *Casablanca*) within or in proximities of the studied urban catchment. A methodology for identifying periods in which water quality and quantity seemed to be ranging outside their typical dry weather behavior was designed, by the use of signal processing and time series techniques (Median Filter, Step Detection and Autocorrelation functions). Therefore, recurrent relations between rainfall event characteristics (in terms of rainfall pulses and measurements at the effluent) were evaluated. A preliminary association between the identified recurrent relations and the physical characteristics of the urban catchment was also proposed.

The proposed methodology seems to provide an accurate rainfall's influence delimitation over various water quality parameters measured at the effluent of the studied urban catchment. The rainfall pulses registered at a rain gauge station close to the studied urban catchment (principally for the *Bosa* rain gauge station) were consistent with the time windows in which the water quality signal was identified to be ranging outside its typical dry weather behavior. Therefore, the proposed technique can be implemented for real-time operation and modeling purposes, regarding the influence of rainfall over water quality.

For the case study, the water quality and quantity parameters which explain the differences between the variability of dry and wet weather data measured by means of the infra-red, ion sensitive and ultrasonic sensors were (ordered from the most to the least explicative): (i) the DO, (ii) the Flow Rate, (iii) the Oxidation Reduction Potential (ORP) (iv) the Turbidity and (v) the Total Suspended Solids (TSS). Therefore, it can be stated that parameters not directly related with solid particles (*e.g.* DO, flow rate and ORP) seem to be more related with rainfall, rather than parameters that describe particulate pollution (TSS and Turbidity). This fact may lead to conclude that resuspension and settling phenomena in the wastewater (including stormwater) effluent of the studied urban catchment exhibits a complex behavior, with variations that can be more difficult to relate to dry or wet weather conditions than for water characteristics non-related exclusively with solid particles (*e.g.* DO, flow rate and ORP). In addition, the pH and the Temperature seemed not to be altered from dry to rainy weather. Therefore, whenever it is desired to assess the effect of the rainfall over water quality by real-time operation, an ion sensitive sensor at the effluent of the studied urban catchment might be insufficient (as an alternative infra-red or UV-VIS spectrometry can be recommended).

For the case of measurements undertaken by UV-VIS spectrometry, the absorbance values registered by the boundary wavelengths of the UV-VIS range (200 nm to 750 nm) seem to better explain the differences between variability of dry and wet weather days' data. Thus, the representativeness of the mentioned differences seems to be reduced as the wavelength's value gets closer to the center of the UV-VIS range (for about 400 nm). In terms of the COD and TSS values

reported by the global calibration, both water quality parameters seemed also be influenced by rainfall: the greater TSS values for rainy days data and the greater COD values for dry weather registers. Considering the fact that by the use of an infra-red sensor the rainfall influence on water quality could be assessed, the implementation of more sophisticated technologies (UV-VIS spectrometry) can be proposed in view of additional purposes (*e.g.* detailing into water quality contamination). However, these analyses might be supported with longer time series data and are only attributable to the studied urban catchment.

Regarding a proposed dry weather probabilistic characterization, and specifically the measurements undertaken by means of infra-red, ion sensitive and Ultrasonic sensors, the DO exposed a value of 0.0 mg/l for most of the dry weather data. Hence, the Probability Density Function (PDF) was assessed as one-tailed during all time-steps of a typical dry weather day, contrasting to the PDFs obtained for the ORP, the Turbidity, the TSS and the flow rate, which were described as two-tailed. Therefore, *a priori* operation rules can be formulated regarding the DO's time series (*e.g.* whenever the DO is greater than 0.0 mg/l, rainfall is influencing the effluent), as its behavior related to dry/wet weather was less difficult to understand than for the case of the remaining analyzed water quality parameters. Turbidity, TSS and pH showed to have a peak value at midday during dry weather conditions. On the other hand, two peaks were reported for the case of the ORP. In addition, no highlighted peak(s) could be appreciable regarding the DO, the flow rate and the Temperature's dry weather characterization. The typical dry weather behavior of the TSS and the Turbidity seems to be alike (in terms of a high peak at midday), although the Turbidity's peak seems to be sharper than the one reported for the TSS's curve; the TSS curve can be smoother. The flow rate was the parameter which exposed most variability during dry weather, while the DO was the parameter with the lowest variability. The dry weather characterization undertook can be considered to provide a preliminary insight of the dry weather behavior of pollutants in the studied urban catchment. This characterization can be an input for WWTP design, warning systems, wastewater treatment technologies' implementation, or outliers' detection of online data.

Concerning the UV-VIS spectrometry data, the wavelength of 745 nm exposed a similar behavior to the TSS values reported by the global calibration, including a lower variability. Regarding the COD, the probabilistic curve for representing the dry weather behavior seemed to be alike to that of wavelength 745 nm, also including similarities with wavelength 205 nm (smoothness of the curve). Therefore, the COD seems to have components from both the UV range (200 nm to 380 nm) and from the VIS range (380 nm to 750 nm), in terms of its dry weather behavior. This fact may lead to conclude that a certain amount of organic matter is also contained in particulate matter. The variability of the TSS measured by the infra-red sensor was greater than for the UV-VIS spectrometry measurements. Therefore, models of TSS at the effluent of the studied catchment might consider this type of sensor used for measuring the input data. The models calibrated by the use of the infra-red sensor's data could include a component regarding additional variability (*e.g.* volatility or noisy process) complementary to models calibrated by the use of UV-VIS spectrometry measurements.

The memory of the physical process was estimated by the Autocorrelation function (ACF) of the analyzed water quality parameters' time series. Regarding the flow rate, the DO, the TSS (measured by infra-red sensors), the Turbidity, the pH, the Temperature, the COD and the wavelength 205 nm,

the memory of the process was assessed in a range from about 200 to 400 min. On the other hand, the ORP, the wavelength 745 nm and the TSS (measured by UV-VIS spectrometry) reported larger values as the memory of the process (more than 2000 min). Regarding the TSS, the ACF showed that measurements undertaken by UV-VIS spectrometry contain more inertia than those for the case of infra-red sensors. The ACF of the water quality time series can bring up additional benefits: *e.g.* response time estimations of different components regarding a given urban drainage. These approximations can contribute with new insights for operation and real-time guidelines as coupled elements: operation strategies can be formulated based on the evaluation of the response time of the system vs. the time of execution by the operators. Furthermore, the integrated and dynamical urban drainage modeling can be supported by this kind of linkages.

The method for event's temporal delimitation proposed in the present thesis seems to be insensitive to outlier presence by using the length of the Median Filter's mobile window to assess the memory of the process. Therefore, the approach for identifying rainfall events at the effluent of the catchment showed an appropriate temporal consistency with rainfall registrations (especially for the registers at the *Bosa* rain gauge station). The proposed methodology seems to expose an appropriate performance for data with outliers. This fact can be an advantage in terms of real-time implementation, considering that a pre-processing treatment of the signal seems not to be strictly necessary (however, it is recommended).

Regarding the response time of the catchment to a given rainfall event, the following facts were observed: (i) the flow rate had similar median response's times for the *Fontibon* and *Bosa* rain gauge stations, with a higher variability for *Fontibon*; (ii) the times for DO, TSS, Turbidity and ORP were identified as greater for *Fontibon* than for *Bosa*, in about 75 % of the studied rainfall events; (iii) the responses of DO and ORP seemed to be larger than for TSS and Turbidity, which were reported more alike to the hydrodynamic time of response of the system (represented by flow rate), independently of the rain gauge station (*Bosa*, *Fontibon* or *Casablanca*); (iv) the hydrodynamic response time seems to be shorter for the rainfall registered at the *Fontibon* than for *Bosa*; (v) DO, TSS, Turbidity and ORP seemed to have faster responses for rainfall registered at the *Bosa*, rather than for *Fontibon*; (vi) the COD reported higher lags than the TSS, for rainfall registered at the *Bosa* rain gauge station. Therefore, it can be concluded that for real time operation purposes, ORP or UV-VIS spectrometry's response time to rainfall might not be considered representative of the complete effluent: prior to a noticeable change of ORP, other water quality parameters might be already influenced by rainfall (*e.g.* flow rate). Regarding modeling purposes, the advective lag of the system can be selected depending on the water quality parameter to be modeled (for the case of ORP, the response time of the system might be greater than for flow rate or TSS).

Statistical characteristics (*e.g.* the mean, standard deviation and entropy) of water quality measurements taken during the identified time windows ranging outside the typical dry weather behavior and rainfall pulses related to each identified time window were calculated (including their uncertainty assessed from a Monte Carlo simulation). Therefore, with the purpose of detecting recurrent and significant relations between characteristics of rainfall and water quality, different linear and non-linear association techniques (*e.g.* Pearson's correlation test, AMI) coupled to Monte Carlo simulations are proposed. The results obtained when applying this methodology to the case

study reported the following statistically significant relations : (i) Turbidity seems to be related with the duration of the rainfall pulses and with the mean rainfall height, (ii) DO seems to have an association with the total rainfall volume and with the rainfall pulse's entropy, (iii) flow rate was found to be related with mean rainfall height, standard deviation of rainfall heights and the mean rainfall intensity, (iv) TSS behavior was found to be linked with characteristics of rainfall such as the duration, the mean rainfall height, the standard deviation of the rainfall heights and the memory of the process (evaluated by the ACF of each specific rainfall pulse), (v) ORP seemed to be connected with the characteristics of rainfall events such as the duration, the ADWP, the total rainfall volume, the entropy and the standard deviation of the rainfall pulse(s). Regarding the studied rain gauge stations, it was found that rainfall registered at the *Bosa* rain gauge station seems to be more correlated with characteristics of the effluent in terms of quality and quantity, rather than the *Fontibon* and *Casablanca* rain gauge stations. With the purpose of taking advantage of the established relations between rainfall and water quality in a real-time operation system, rainfall and water quality characteristics should be computed. Hence, simpler statistics (*e.g.* the mean, standard deviation and the total volumes) seem to be more advantageous than not so easily computable characteristics in real-time operation (*e.g.* the memory of the process and the entropy).

Linkages with physical characteristics of the basin as preliminary conclusion for the study case, are that the principal sources of variability between the spatial influence of a given rain gauge station (in terms of the amount of recurrent relations with water quality and quantity at the effluent) with different sub-basins of the urban catchment are: (i) the distance to the rain gauge station which is mostly influencing a specific sub-basin, (ii) the variability of the pipes' slope at a given sub-basin, (iii) the industrial land use regarding a sub-basin, (iv) the mean pipes' slope for a given sub-basin and (v) the variability of pipes' length at a sub-basin. Thus, the heterogeneity of a given urban catchment can be considered for designing its instrumentation network (in terms of rainfall and water quality). The places in which physical characteristics (*e.g.* sewer network topology, land uses) exhibit the greater variability should be monitored with more detail.

## 6 RECOMMENDATIONS

The analyzed time series obtained from online measurements at the effluent of the studied urban catchment were undertaken for a period of about 25 days. Therefore, the implementation of the proposed methodology for delimitating rainfall events, should be studied for longer online measurement periods. Furthermore, time series with a lower amount of “NA” data can also bring up additional benefits in terms of identifying a greater number of rainfall events (more than eight, as for the present study). Hence, measurement of additional water quality parameters (e.g. Conductivity, Nitrogen or Total Organic Carbon) can be an option to be explored with the purpose of characterizing the influence of rainfall over a more detailed description of the wastewater. Long time series (e.g. 10 years) with not so many missing data can be an appropriate environment for applying non-linear and more robust methods (e.g. non-linear prediction, chaotic analysis, fractal and multi-fractal spectral analysis); including additional explorations regarding long-term behaviors such as trends, periodicities or persistence.

It can also be recommended to implement the proposed methods not only at the effluent, but also at other points inside a larger urban catchment; this will lead to include in the analysis additional variability provided by elements such as land use, rainfall spatial heterogeneity, traffic analyses, geotechnical considerations and social dynamics. This type of study zones may represent new challenges regarding the assessment of the variability over water quality produced exclusively by means of the rainfall phenomena.

Concerning the designed methodology, it can be recommended for future studied to test additional non-linear filters (e.g. detectors or ranklets) and Step Detection Algorithms (Kolmogorov-Smirnov or Bartlett) rather than the ones proposed for the present study (Median Filter and Mann-Whitney). A previous treatment of the water quality signal by removing outliers (e.g. Hampel Filter) can also be recommended, in order to test the benefits of implementing pre-processing treatment regarding the method's performance.

Most of the measurement sensors implemented for the present study require a local calibration of the equipment (UV-VIS spectrometry and infra-red). This practice leads to report more accurate values than with the global calibration. For future studies, a local calibration of the analyzed wastewater is recommended, with the purpose of obtaining more precise data of water quality. These calibrations will bring up accuracy in the relations between the rainfall and the water quality, useful for decision-making purposes, especially in the operation of a WWTP at the effluent of the catchment.

Additional metrics of the memory of the process (e.g. auto average mutual information function) can also be implemented for future studies. Considering the importance of the response time of a system to external perturbations, it can be strongly recommended for modeling and real time operation to consider the variability of the response time; especially keeping in mind that the hydrodynamic response time seemed not to be equivalent to the response time of the constituents of the wastewater and stormwater.

The evaluation of rainfall influence on, not only the individual water quality parameters, but also on the associated total loads (multiplying the TSS concentration by flow rate value, at the same time-step) is strongly recommended, based on the promising results obtained from a preliminary assessment of the TSS loads. Considering that the total loads can be understood as a particular case of the convolution between the flow rate and a water quality parameter time series, studying the rainfall influence over convolved time series (by also including lags) of different water quality parameters (*e.g.* COD and TSS) seems to be another interesting option to be explored.

The modeling of pollutographs produced during a rainfall event seems to be an additional challenge regarding urban drainage scientific problems; the variability observed by means of high temporal resolution measurement devices is a fact that mathematical models seem to underestimate. Hence, exploring mathematical models based on deterministic, but also stochastic components can be strongly recommended for incoming studies (with the purpose of assessing the variability and the noisy behavior of the pollutographs).

Additional recommendations pointing to real time and modeling purposes can be formulated. Considering that the most relevant rainfall registers were the data at the *Bosa* rain gauge station in terms of the influence over water quality (for water quality parameters in which a relation with rainfall could be established), additional instrumentation over this specific zone in proximities of the urban catchment can be proposed. Therefore, the online rainfall data, in addition with an appropriate dry weather characterization, will lead to establish a real time operation system at a WWTP located at the effluent of the studied urban catchment.

## 7 REFERENCES

- A.H.E.P.E. - Asociación de Historia de la Estadística y de la Probabilidad en España** (2006). *Historia de la Estadística y de la Probabilidad*. 1 ed., Delta Publicaciones Universitarias, Madrid, España. ISBN-8496477258. pp. 117-129.
- Alvarado, J. A.** (2008). *Fundamentos de inferencia estadística*. Pontificia Universidad Javeriana, Bogotá, Colombia. pp. 243-244.
- Arce G.R.** (2005). *Nonlinear Signal Processing: A Statistical Approach*, Wiley:New Jersey, USA. pp. 81.
- Athayde D.N., Shelly P.E., Driscoll E.D., Gaboury D., and Boyd. G.** (1983). *Results of the Nationwide Urban Runoff Program: Volume I - Final Report*. Water Planning Div., U.S. Environ. Prot. Agency. Washington, DC .pp. 12-18.
- Aumond, M., Joannis, C.,** (2008). *Processing sewage turbidity and conductivity recorded in sewage for assessing sanitary water and infiltration/inflow discharges*. In: Proceedings of the 11th International Conference on Urban Drainage, Edinburgh, Scotland, UK, 31 Aug.- 5 Sept., 10 p
- Ayyub, B., Mccuen, R.** (2003). *Probability, Statistics, and Reliability for Engineers and Scientists*. 2nd ed, Chapman and Hall / CRC Press LLC, Boca Raton, USA. ISBN 1-58488-286-7
- Barraud, S., Gibert, J., Winiarski, T., Bertrand-Krajewski, J.-L.,** (2002). Implementation of a monitoring system to measure impact of stormwater runoff infiltration. *Water Science and Technology* 45 (3), 203-210.
- Bertrand-Krajewski, J.-L., Barraud, S., Gibert, J., Malard, F., Winiarski, T., Delolme, C.,** (2007). *The OTHU case study: integrated monitoring of stormwater in Lyon, France (Chapter 23)*. In: Fletcher, T., Deletic, A. (Eds.), *Data Requirements for Integrated Urban Water Management*. Urban Water Series - UNESCO IHP. Taylor and Francis, London (UK), ISBN 9780415453455, pp. 303-314.
- Brombach, H., Weiss, G., Fuchs, S.** (2005). A new database on urban runoff pollution: comparison of separate and combined sewer systems. *Water Science & Technology*, 51 (02), 119-128.
- Camacho, L., Díaz-Granados, M., Giraldo, E., Saénz, J., Herrera, M.** (2002). *Instrumentación y análisis ambiental de una cuenca urbana en Bogotá: Implementación de modelos simplificados lluvia-escorrentía*. XV Nacional de Hidráulica e Hidrología.
- Cammann, K., Schroeder, A.** (1979). *Working with Ion-Selective Electrodes*. ISBN 0-387-09320-6. Springer Verlag, New York, USA. 225p.

**CEM, Centro Español de Metrología**, (2008). *Vocabulario internacional de metrología – Conceptos fundamentales y generales, y términos asociados (VIM)*, 3ª edición, Madrid, España, 88p.

**Charniak, E.** (1993). *Statistical language learning*. Massachusetts Institute of Technology, USA. P 137.

**Chib, S.** (1998). *Estimation and comparison of multiple change-point models*. Journal of Econometrics, 86(2), 221-241. ISSN 0304-4076.

**Chocat, B.** (1997). *Encyclopédie de l'Hydrologie Urbaine et de l'Assainissement*. Paris, France: Tec et Doc, 1997.

**Díaz Granados, M., Camacho, L.A.** (2003). Modelación hidrológica y ambiental de una Subcuenca de la ciudad de Bogotá. *Seminario Internacional La Hidroinformática en la Gestión Integrada de los Recursos Hídricos, Agua 2003*, Cartagena/Colombia 1-3 October 2003. Pp. 328-335.

**Dray, S., Dufour, A.B.** (2007): The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*. 22(4): 1-20.

**Duncan, H.P.** (1999). Urban Stormwater Quality: A Statistical Overview. Cooperative Research Centre for Catchment Hydrology. 80 p.

**Ellis, J.B., Revitt, D.M., Scholes, L., Sieker, H., Helm, B., Winger, J., Picouet, C. and Soutter, M.** (2006). *Review of the adaptability and sensitivity of current stormwater control technologies to extreme environmental and socio-economic drivers*. Section B of Deliverable 2.1.1. EU 6th Framework SWITCH Project. Pp. 10-34.

**Estupiñán, H.A., Rodríguez, E.A., Camacho, L.A.** (2009). *Evaluación de la contaminación por escorrentía urbana y su modelación mediante la simulación de los procesos de acumulación y lavado*. Seminario Internacional: Un Nuevo Paradigma en el Manejo Integrado del Agua en Áreas Urbanas, Agua 2009, Cali/Colombia, 9-13 November 2009, 12 p.

**Fearnhead, P., Liu, Z.** (2007). *On-line inference for multiple change point problems*. Journal of the Royal Statistical Society Series B, 69(4), 589-605. ISSN 1369-7412.

**Fletcher, T.D., Deletic, A.** (2007). *Data requirements for integrated urban water management*, Urban Water Series - UNESCO IHP. Taylor and Francis, London (UK), ISBN 9780415453455, 337 p.

**Gnecco, I., Berretta, C., Lanza, L.G., La Barbera, P.** (2005). Storm water pollution in the urban environment of Genoa, Italy. *Atmospheric Research*, 77(1-4), pp.60-73.

**Gordon, J.** (2012). cpm: Sequential Parametric and Nonparametric Change Detection.

- Gruber, G., Winkler, S., Pressl, A.,** (2005). Continuous monitoring in sewer networks an approach for quantification of pollution loads from CSOs into surface water bodies. *Water Science and Technology* 52 (12), 215-223
- Grüning, H., Orth, M.,** (2002). Investigations of the dynamic behaviour of the composition of combined sewage using online analyzers. *Water Science and Technology* 45 (4e5), 77-83.
- Harkat, M.F., Djelel, S., Doghmane, N.** (2007). Sensor Fault Detection, Isolation and Reconstruction Using Nonlinear Principal Component Analysis[J] *International Journal of Automation and Computing* , 2007,V4(2): 149-155.
- Hausser, J., Strimmer, K.** (2012). entropy: Entropy and Mutual Information Estimation. R package version 1.1.7. <http://CRAN.R-project.org/package=entropy>
- Hernández, O.** (1998). *Temas de análisis estadístico multivariado* (1. Ed). Editorial de la Universidad de Costa Rica. San José, Costa Rica. 170p.
- HMV Ingenieros** (2007). *Predimensionamiento planta de tratamiento de aguas residuales canoas. Producto 2: Metodología de muestreo*. Informe Inédito. EAAB. Informe de consultoría Contrato: 1-02-26100-806-2006. 68 p.
- Hochedlinger, M., Kainz, H., Rauch, W.** (2006). Assessment of CSO loads based on UV/VIS-spectroscopy by means of different regression methods. *Water Science and Technology* 54 (6-7), 239-246.
- Jacobsen, J.L., Madsen, H.** (1996). Grey box modelling of oxygen levels in a small stream. *Environmetrics*, 7: 109–121
- Lacour, C.,** (2009). *Apport de la mesure en continu pour la gestion de la qualité des effluents de temps de pluie en re'seau d'assainissement [Interest of continuous monitoring for the control of water quality during wet weather in sewer systems]*. PhD thesis, Université Paris Est, Marne-la-Valle'e, France, 278 p.p appendix.
- Lacour, C., Joannis, C., Chebbo, G.,** (2009). Assessment of annual pollutant loads in combined sewers from continuous turbidity measurements: sensitivity to calibration data. *Water Research* 43 (8), 2179-2190.
- Lai, T.L.** (1995). *Sequential Changepoint Detection in Quality Control and Dynamical Systems*. Journal of the Royal Statistical Society. Series B - Methodological, 57(4), 613-658. ISSN 00359246. URL <http://www.jstor.org/stable/2345934>
- Lee, J.H., Bang, K.W., Ketchum, L.H., Choe, J.S., Yu, M.J.** (2002). First flush analysis of urban storm runoff. *The Science of the total environment*, 293 (1-3), pp.163-75.
- Mason, L., Gunst F., Hess, L.** (2003). *Statistical Design and Analysis of Experiments, with Applications to Engineering and Science*. 2nd ed. Wiley-Interscience, New Jersey. 729 p.

- Métadier M., Bertrand-Krajewski J.-L.** (2012). Pollutographs, concentrations, loads and intra-event mass distributions of pollutants in urban wet weather discharges calculated from long term on line turbidity measurements. *Water Research* 46 (20), 6836-56.
- Mevik, B.H., Wehrens, R.** (2007). *The pls package: principal component and partial least squares regression in R*. J. Stat.Softw. 18: 1–28.
- Mevik, B.H., Wehrens, R., Liland, K.H.** (2011). pls: *Partial Least Squares and Principal Component regression*. R package version 2.3-0. <http://CRAN.R-project.org/package=pls>
- Mishendo, M.I., Hovenier, J.W., Travis, L.D.** (1999). *Light Scattering by Nonspherical Particles: Theory, Measurements, and Applications*, Academic Press Inc 1999, 690 p.
- Montenegro García, A.** (2010). *Análisis de Series de Tiempo*. Bogotá: Editorial Pontificia Universidad Javeriana, Bogotá, Colombia. 396 p.
- NIVUS Gmb H** (2012). Online measurement technology, NIVUS *on-line*, Germany.
- Pettitt, A.N.** (1980). A Simple Cumulative Sum Type Statistic for the Change-Point Problem With Zero-One Observations. *Biometrika*, 67(1), 79-84. ISSN 0006-3444.
- Pitt, R., Bannerman, R., Clark, S., Williamson, D.** (2004). “Source of pollutants in urban areas (Part 1)—Older monitoring projects.” *Effective modeling of urban water systems*, Monograph 13, CHI, Guelph, Ontario, Canada.
- Priestley M.B.** (1981). *Spectral Analysis and Time Series*. Academic Press, London, U.K
- R Development Core Team** (2012). R: A Language and Environment for Statistical Computing; R Foundation for Statistical Computing: Vienna, Austria. <http://www.R-project.org>
- Roberts, S.W.,** (1959). \Control chart tests based on geometric moving averages." *Technometrics*, 42(1), 97{101. ISSN 0040-1706. doi:<http://dx.doi.org/10.2307/1271439>.
- Rodríguez, J.P., Díaz-Granados, M.A., Penagos, J.C., McIntyre, N., Achleitner, S., Maksimović, Č.** (2011). A methodology to assess dry weather flow pattern uncertainty for integrated urban drainage modelling. *Proceedings of the 12th International Conference on Urban Drainage*, Porto Alegre, Brazil, 11-16 September 2011, 9 p.
- Rosenzweig, R., Moon, H.S., Smith, J.A., Baeck, M.L., Jaffe, P.R.** (2008). Variation in the instream dissolved inorganic nitrogen response between and within rainstorm events in an urban watershed. *Journal of Environmental Science and Health Part A* (2008) 43, 1223-1233.
- s::can.** (2007). Manual s::can spectrometer probe y ana::pro. Vienna: Liquid Monitoring Networks.
- Saget, A.** (1994). *Data base of stormwater quality: distribution of discharged pollutant loads and sizes of interception constructions*. Ph.D. thesis, Ecole Nationale des Ponts et Chaussées,

Paris, France, p. 333 (in French).

**Schilperoort, R.P.S.** (2011). *Monitoring as a tool for the assessment of wastewater quality dynamics*. PhD thesis, TU Delft, The Netherlands, June 2011, 320 p.

**Schölkopf, B., Smola, A.J.** (2002). *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. Cambridge, MA: MIT Press.

**Schreiber, T., Schmitz, A.** (2000) Surrogate time series. *Physica D*, 142, 346–382.

**Steuer, R., Kurths, J., Daub, C. O., Weise, J., Selbig, J.** (2002). The Mutual Information: Detecting and Evaluating Dependencies between Variables. *Bioinformatics*, 18(Suppl. 2), S231–S240.

**Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., Farmer, J.D.** (1992) Testing for nonlinearity in time series: the method of surrogate data. *Physica D*, 58, 77.

**Tiefenthaler, L.L., Schiff, K.C.** (2003). *Effects of rainfall intensity and duration on first flush of stormwater pollutants*. In the 2001-2002 Southern California Coastal Water Research Project (SCCWRP), pp.209-215.

**Veldkamp, R., Henckens, G., Langeveld, J.G., Clemens, F.,** (2002). *Field data on time and space scales of transport processes in sewer systems*. In: Proceedings of the 9th International Conference on Urban Drainage, Portland, OR, USA, 8-13 September, 15 p.

**Walpole, E., Myers H., Sharon, Myers, L., Ye, K.** (2006). *Probability & Statistics for Engineers & Scientists*. 8th ed. Prentice Hall.

**Warith, M.A., Kennedy, K., Reitsma, R.** (1998). Use of sanitary sewers as wastewater pre-treatment systems. *Waste Management*, 18(4), 235-247.

**Warnes, G., Bolker, B., Bonebakker, L., Gentleman, R., Liaw W., Lumley T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., Venables, B.** (2012). gplots: Various R programming tools for plotting data. R package version 2.11.0. <http://CRAN.R-project.org/package=gplots>

**Winkler, S., Saracevic, E., Bertrand-Krajewski, J.L., Torres, A.** (2008). Benefits, limitations and uncertainty of in situ spectrometry. *Water science and Technology*, 57(10), 1651.

**Wissenschaftlich Technische Werkstätten Gmb H** (2010). Online measurement technology, WTW on-line, Germany.

**WMO** (1996). *Guide to Meteorological Instruments and Methods of Observation*. 6th ed. WMO Rep. 8, World Meteorological Organization, Geneva, Switzerland. ISBN 978-92-63-10008-5.

**Wuertz, D., Chauss, P., King, R., Gu, Chong., Gross, J., Scott, D., Lumley, T., Zeileis, A., Aas, K.** (2012). *fBasics: Rmetrics - Markets and Basic Statistics*. R package version 2160.81.

<http://CRAN.R-project.org/package=fBasics>

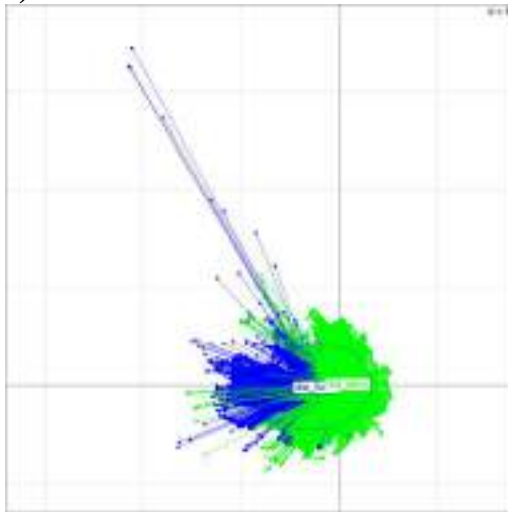
**Zeileis, A. & Grothendieck, G.** (2005). zoo: S3 Infrastructure for Regular and Irregular Time Series. *Journal of Statistical Software*, 14(6), 1-27. URL <http://www.jstatsoft.org/v14/i06/>

## 8 APPENDICES

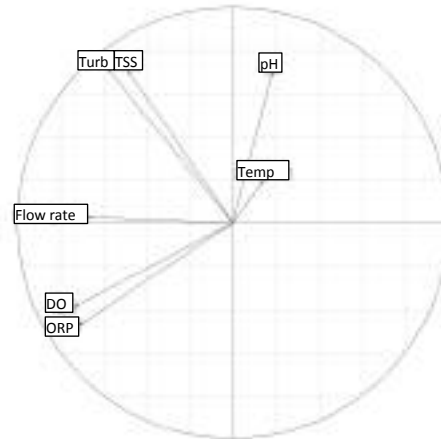
Appendix 8.1 Data base of the measurements undertaken at the effluent of the studied catchment

Appendix 8.2 Implementation of the proposed methods in R-project

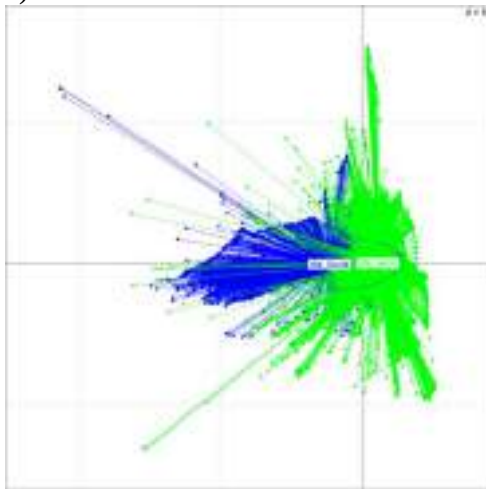
a)



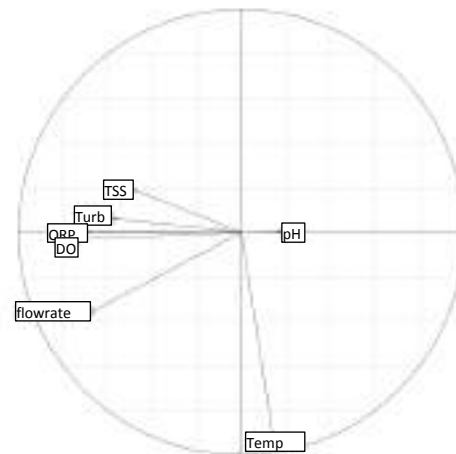
b)



c)

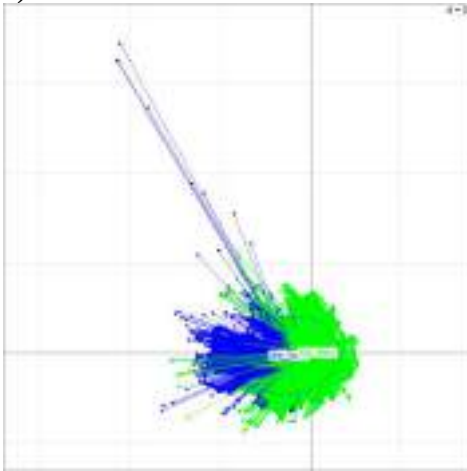


d)

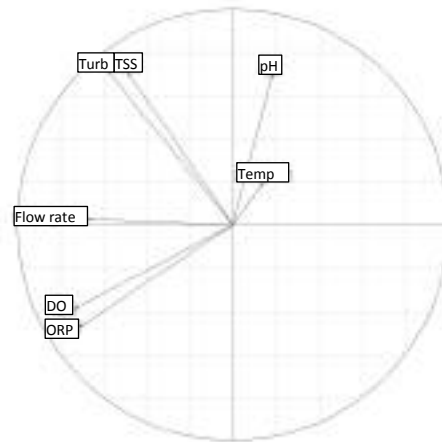


Appendix 8.3 PCA analysis for the *Fontibon* rain gauge station (rainy days' data blue group, dry days' data green group); a), b) separation given by principal components 1 (x-axis) and 2 (y-axis); b), c) separation given by principal components 1 (x-axis) and 3 (y-axis).

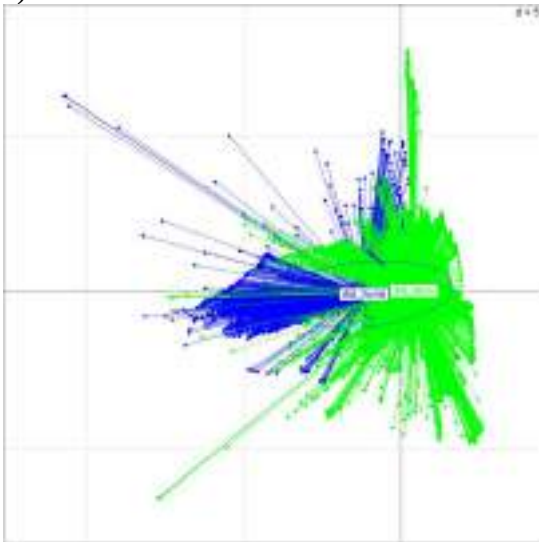
a)



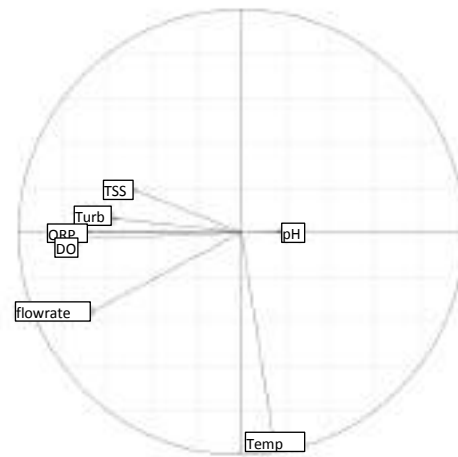
b)



c)

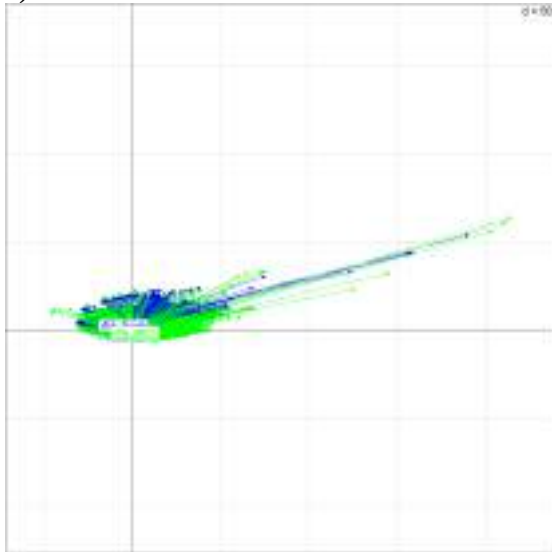


d)

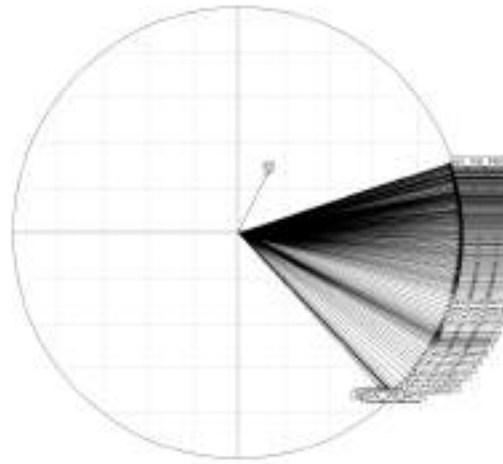


**Appendix 8.4 PCA analysis for the *Casablanca* rain gauge station (rainy days' data blue group, dry days' data green group); a), b) separation given by principal components 1 (x-axis) and 2 (y-axis); b), c) separation given by principal components 1 (x-axis) and 3 (y-axis).**

a)

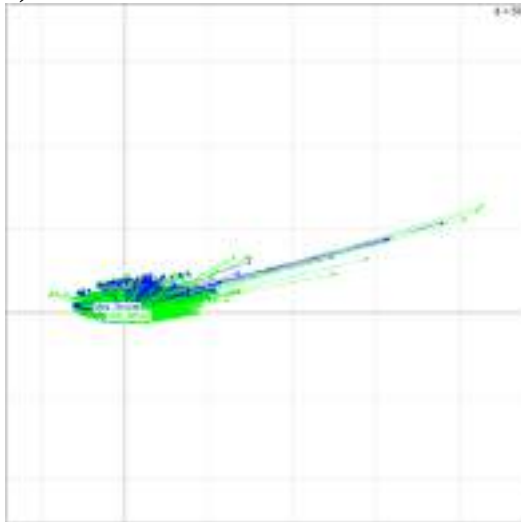


b)

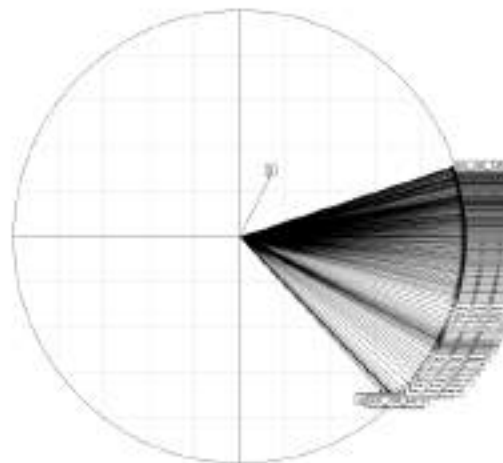


**Appendix 8.5 PCA analysis for the *Fontibon* rain gauge station (rainy days' data blue group, dry days' data green group); a), b) separation given by principal components 1 (x- axis) and 2 (y- axis).**

a)

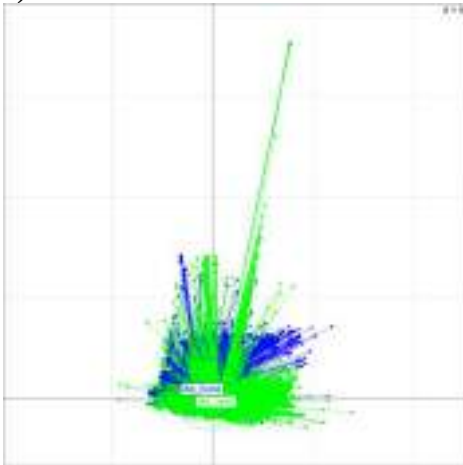


b)

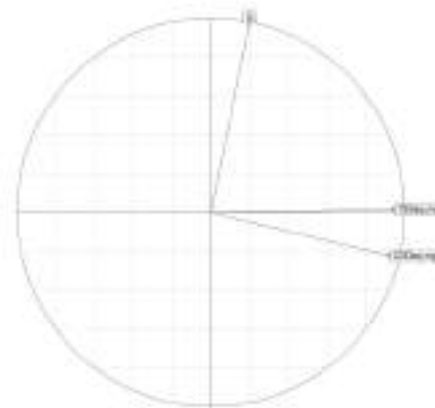


**Appendix 8.6 PCA analysis for the *Casablanca* rain gauge station (rainy days' data blue group, dry days' data green group); a), b) separation given by principal components 1 (x-axis) and 2(y-axis).**

a)

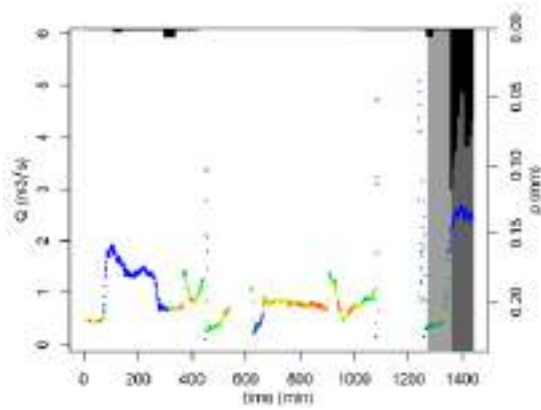


b)

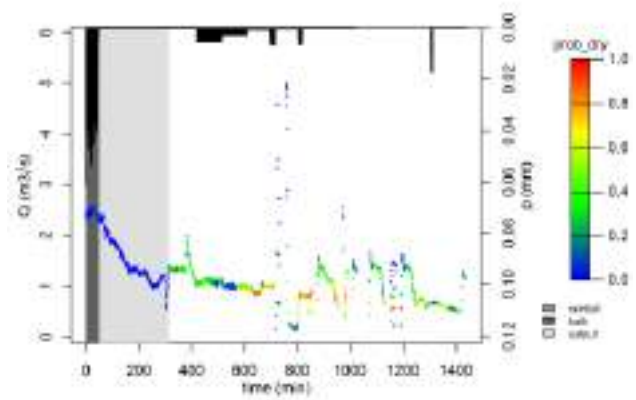


**Appendix 8.7. PCA analysis for the *Fontibon* rain gauge station (rainy days' data blue group, dry days' data green group); a), b) separation given by principal components 1 (x-axis) and 2 (y-axis).**

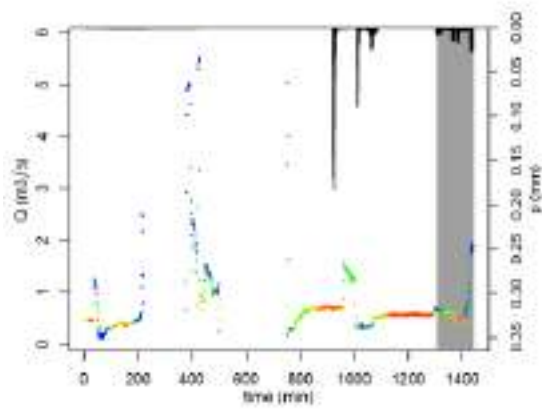
a)



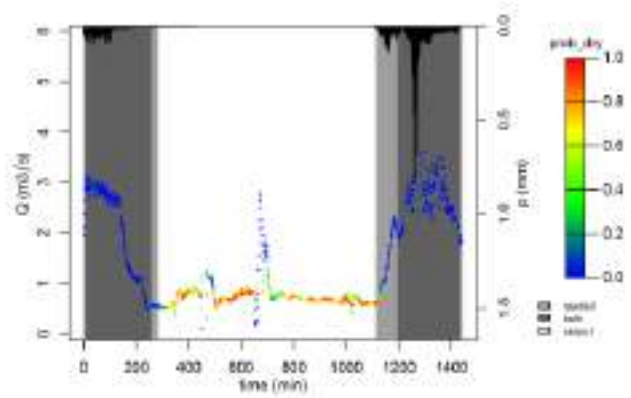
b)



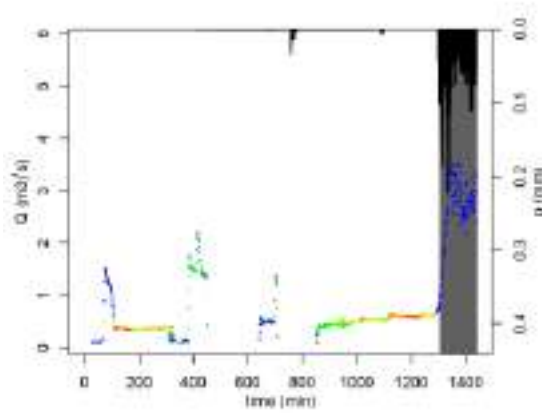
c)



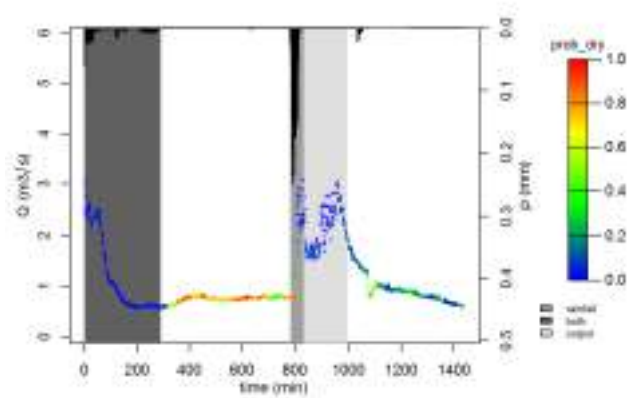
d)



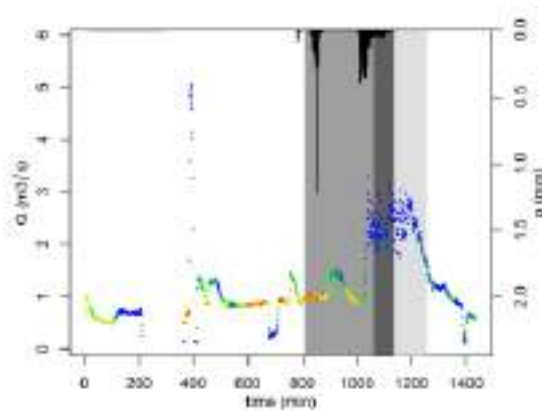
e)



f)

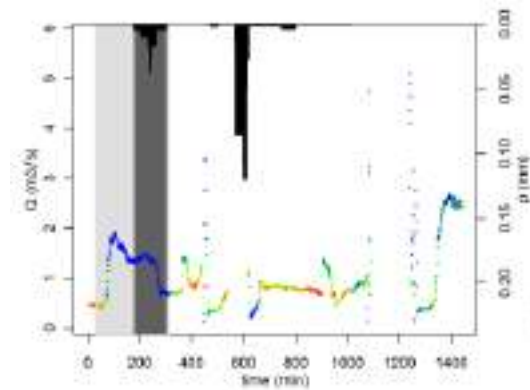


g)

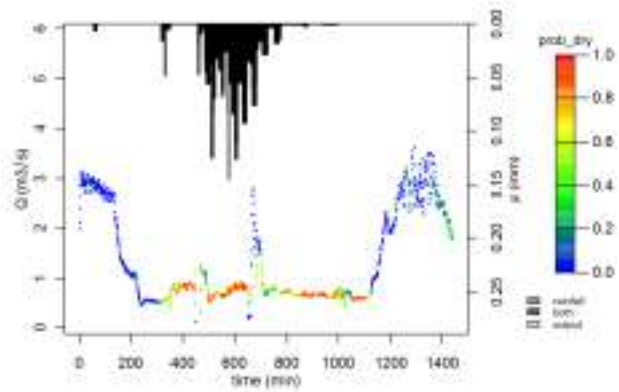


**Appendix 8.8 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding Flow rate (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Fontibon* rain gauge station**

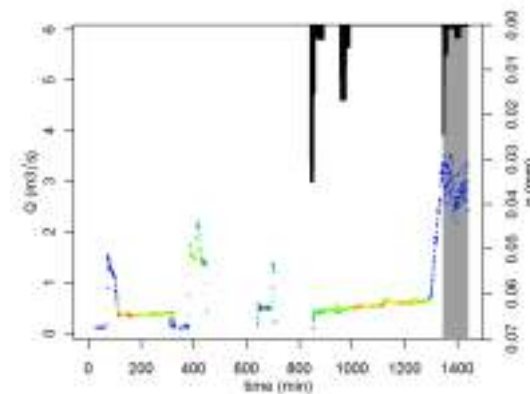
a)



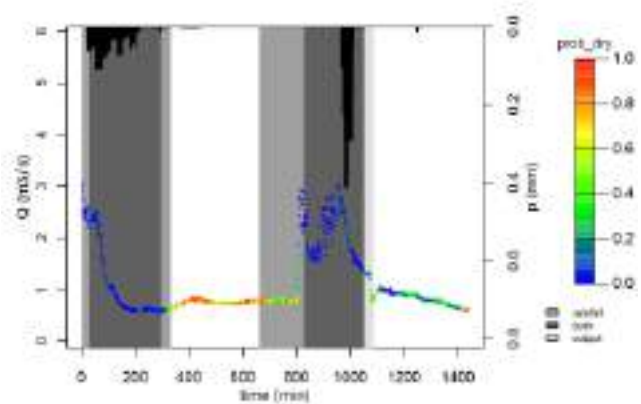
b)



c)

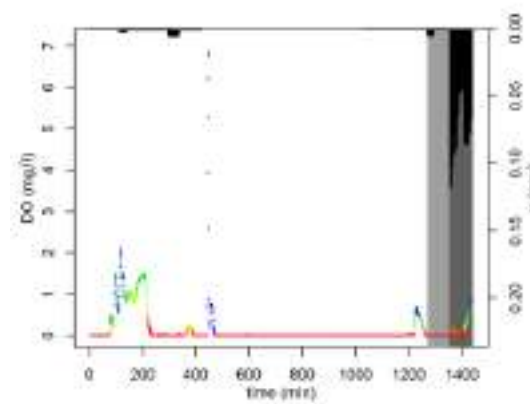


d)

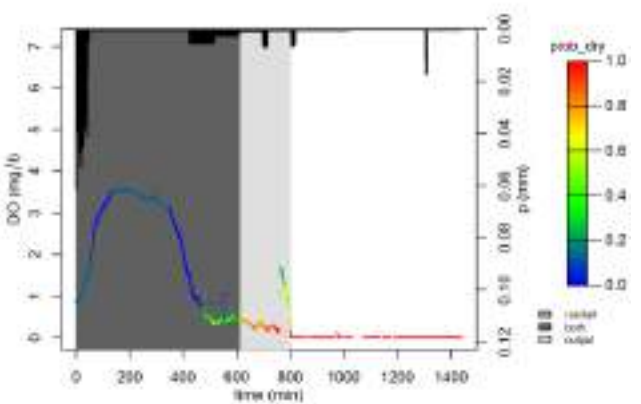


**Appendix 8.9 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding Flow rate (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Casablanca* rain gauge station.**

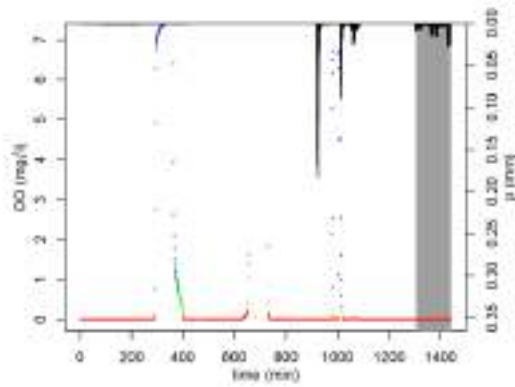
a)



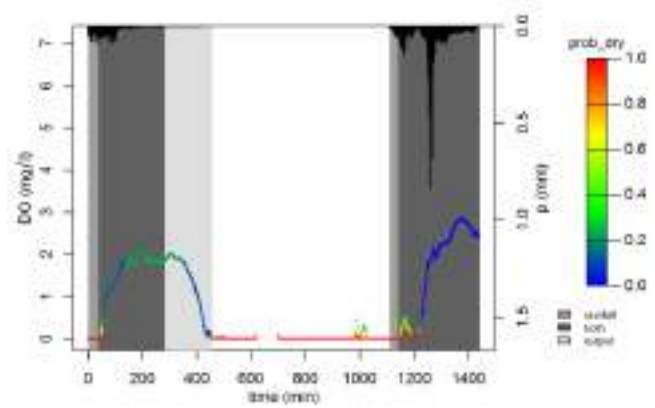
b)



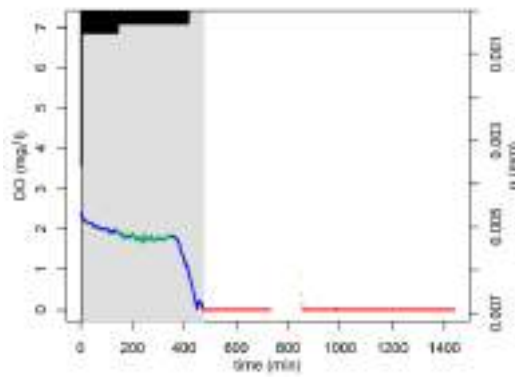
c)



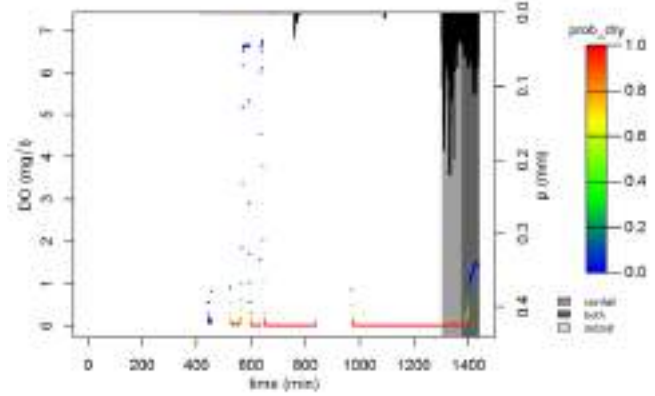
d)



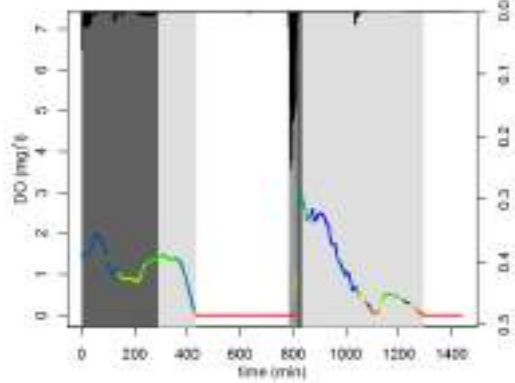
e)



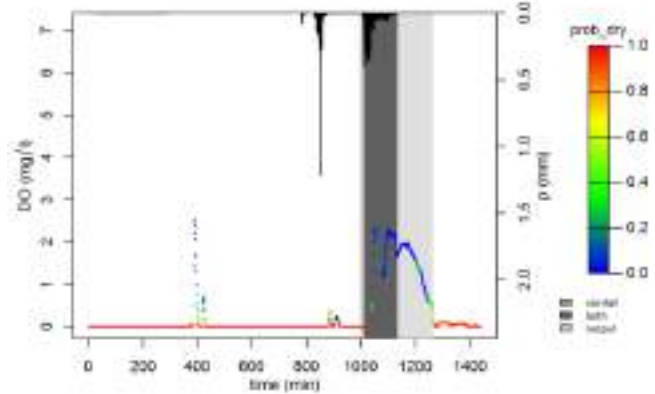
f)



g)

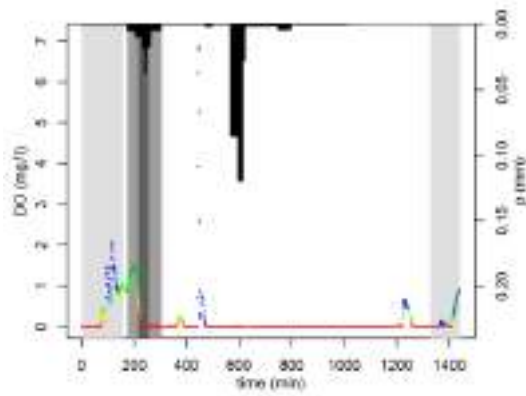


h)

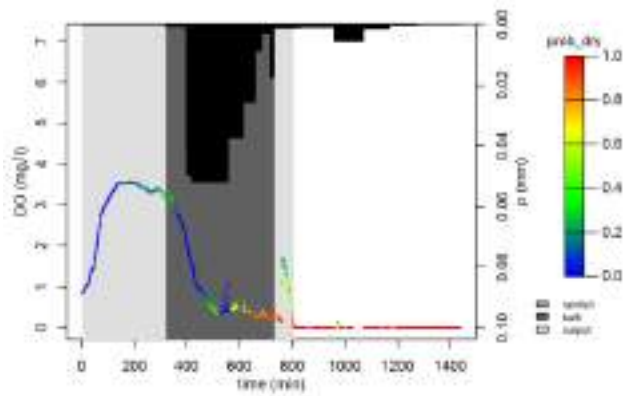


**Appendix 8.10 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding DO (uni-parametric) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Fontibon* rain gauge station**

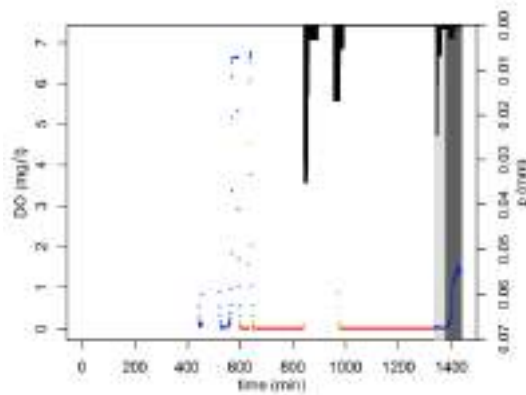
a)



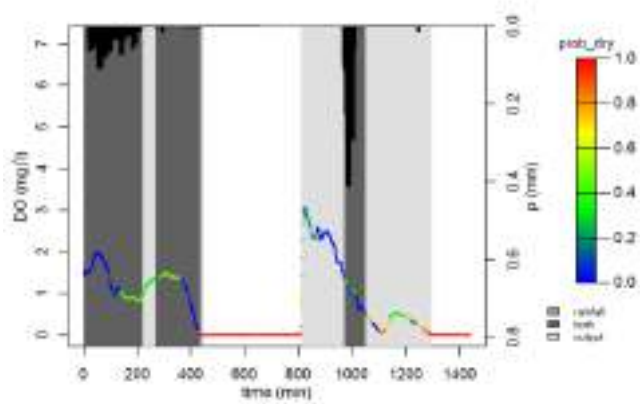
b)



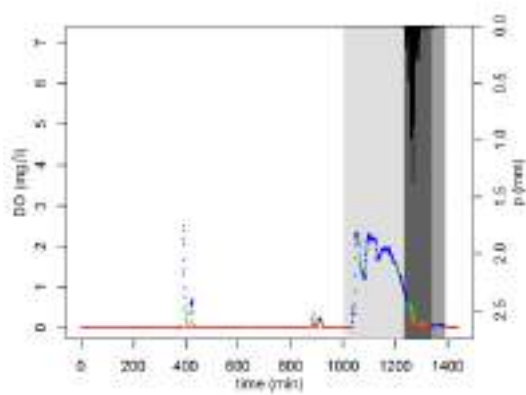
c)



d)

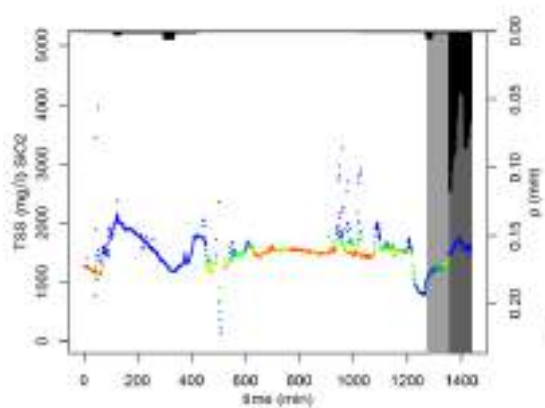


e)

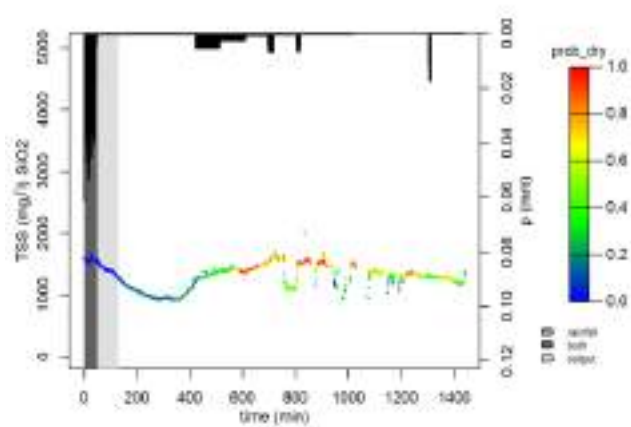


**Appendix 8.11 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding DO (uni-parametric) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Casablanca* rain gauge station.**

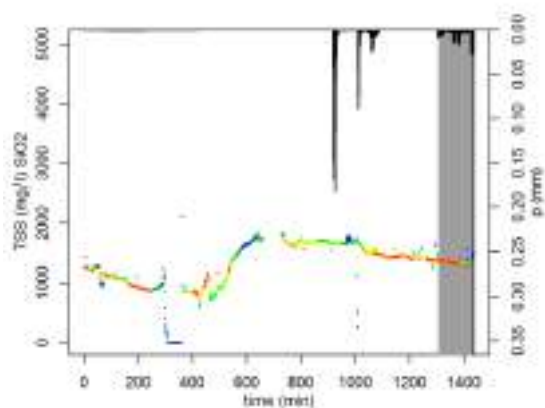
a)



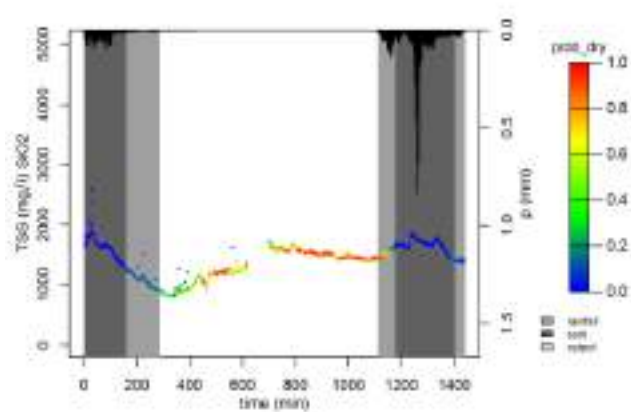
b)



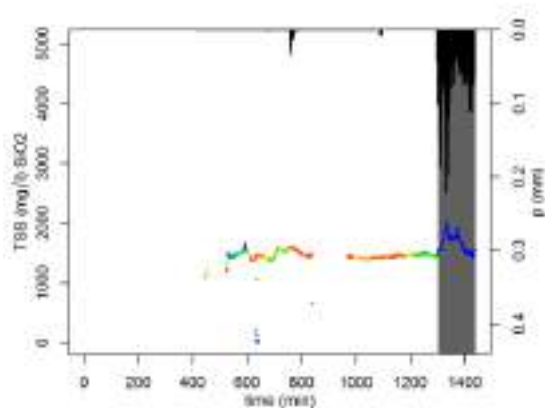
c)



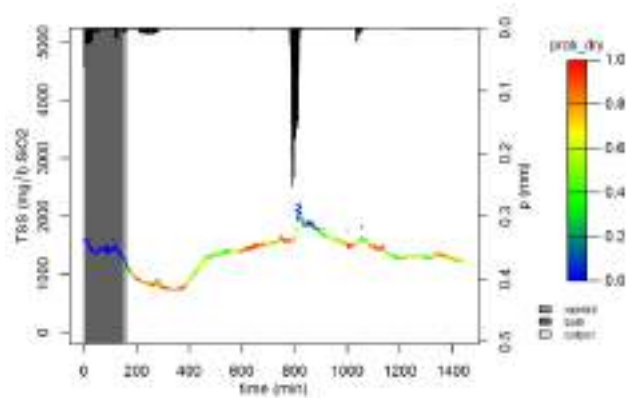
d)



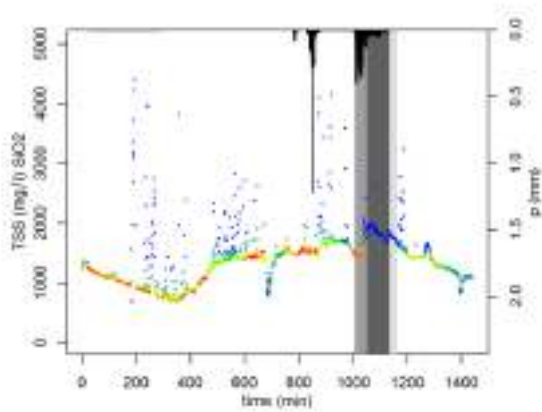
e)



f)

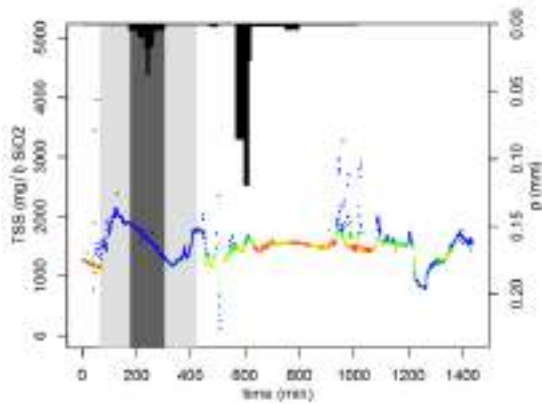


g)

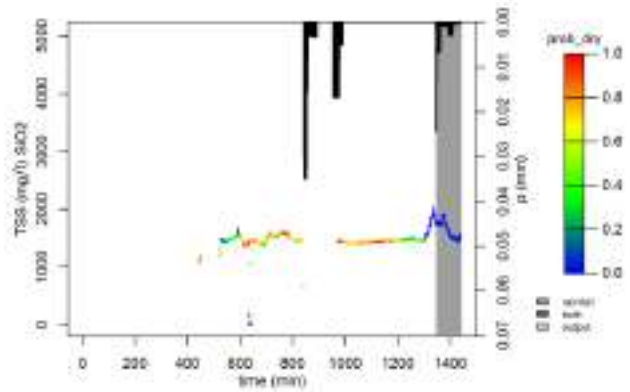


**Appendix 8.12 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding TSS (uni-parametric) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Fontibon* rain gauge station**

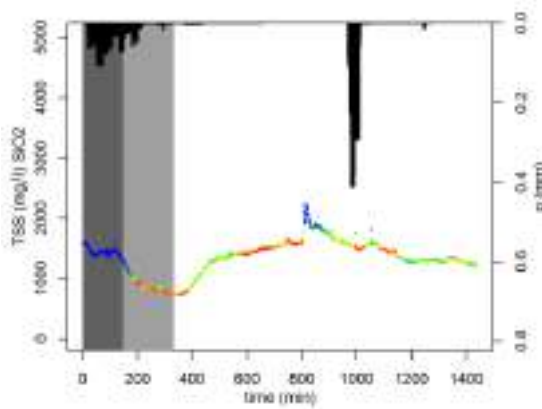
a)



b)

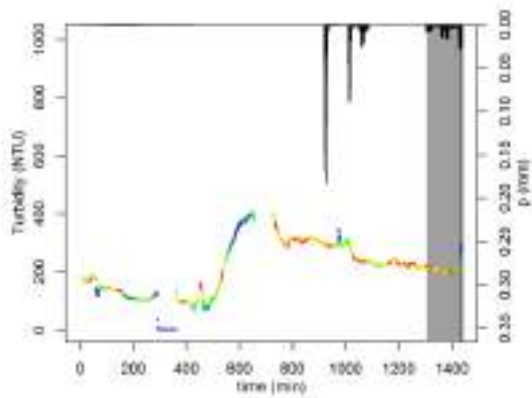


c)

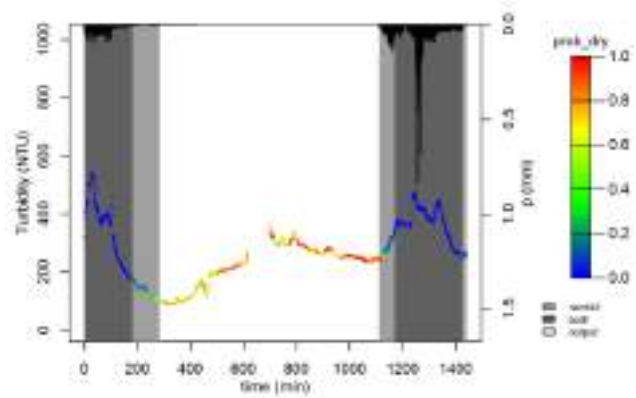


**Appendix 8.13 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding TSS (uni-parametric) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Casablanca* rain gauge station.**

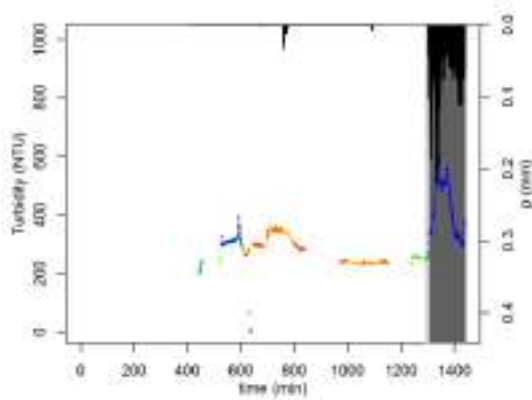
a)



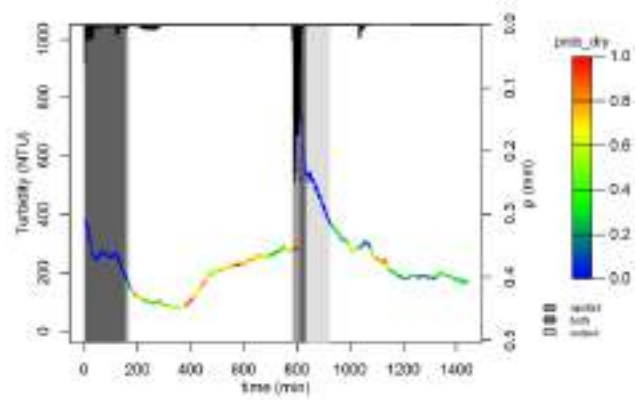
b)



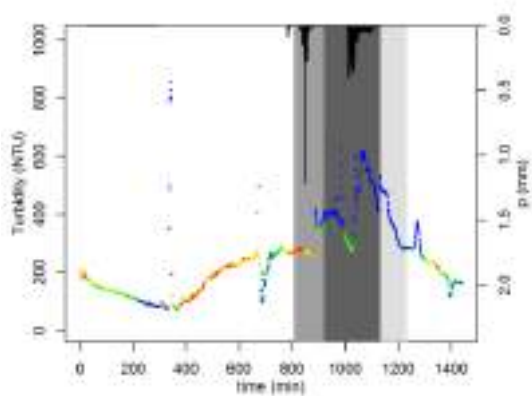
c)



d)

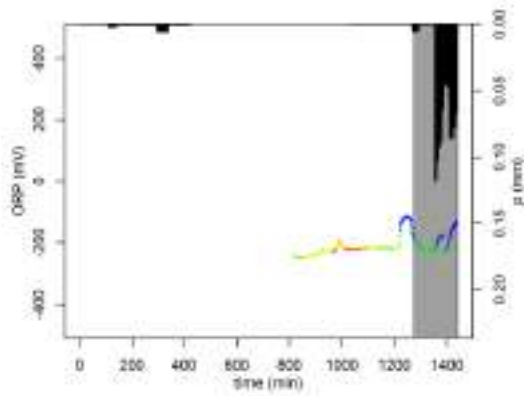


e)

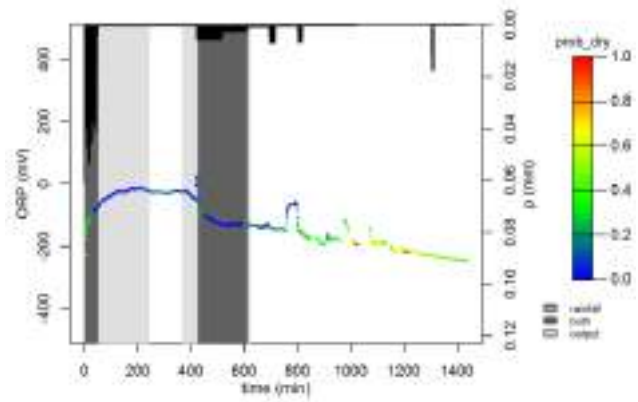


**Appendix 8.14 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding Turbidity (uni-parametric) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Fontibon* rain gauge station.**

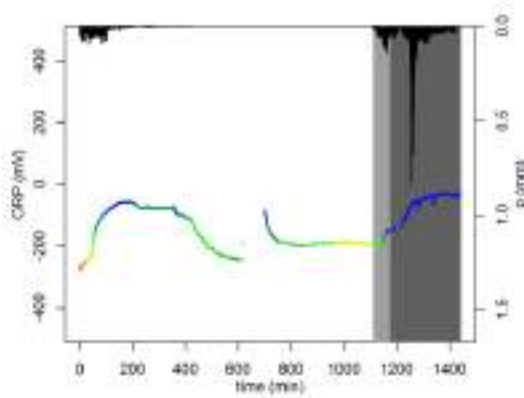
a)



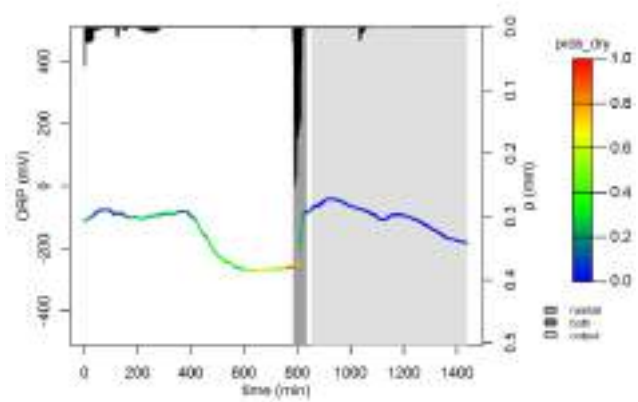
b)



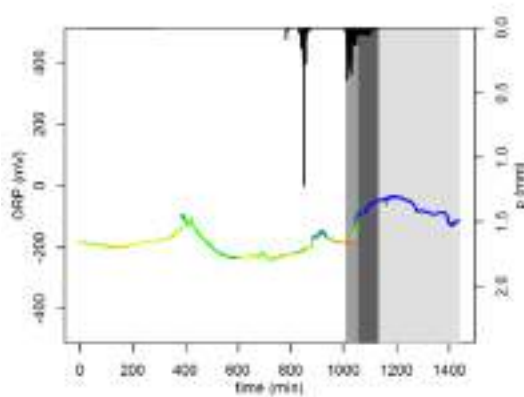
c)



d)

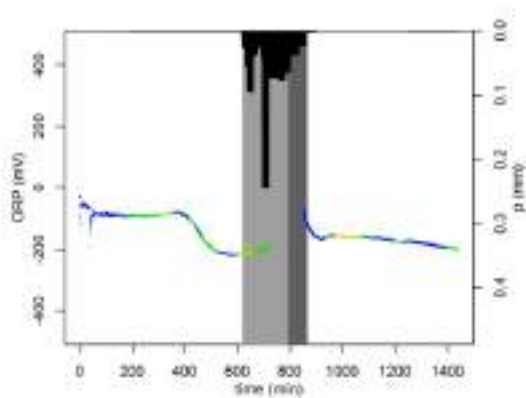


e)

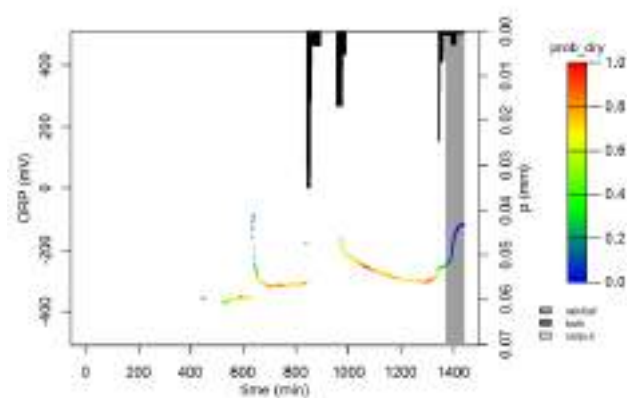


**Appendix 8.15 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding ORP (uni-parametric) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Fontibon* rain gauge station.**

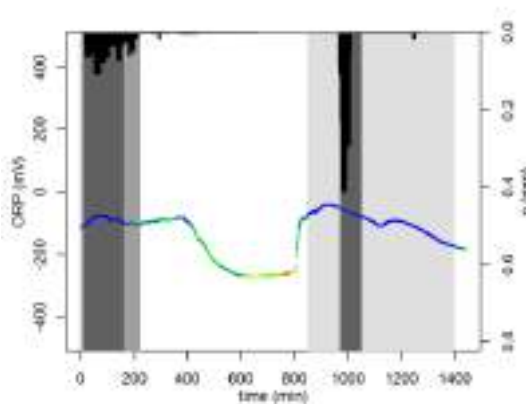
a)



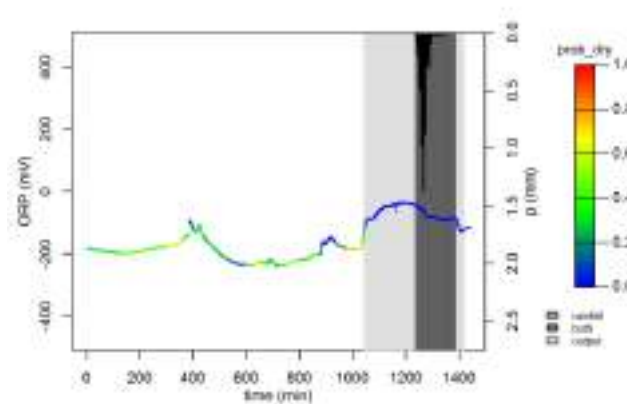
b)



c)

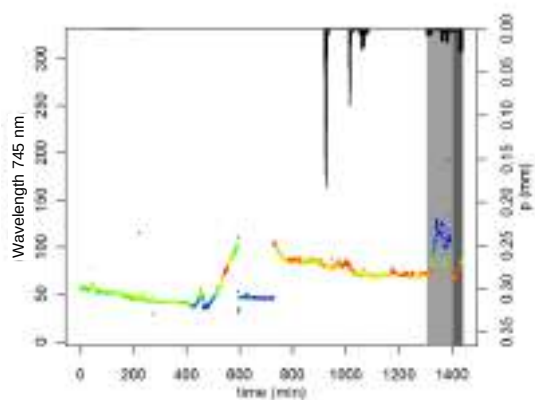


d)

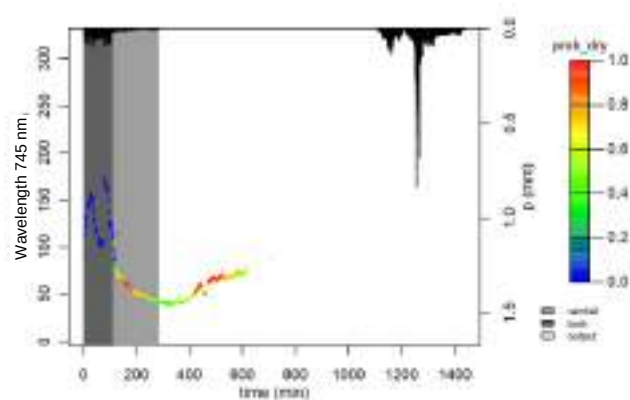


**Appendix 8.16 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding ORP (uni-parametric) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Casablanca* rain gauge station.**

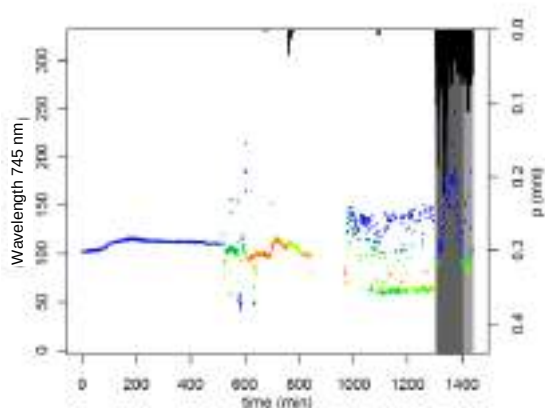
a)



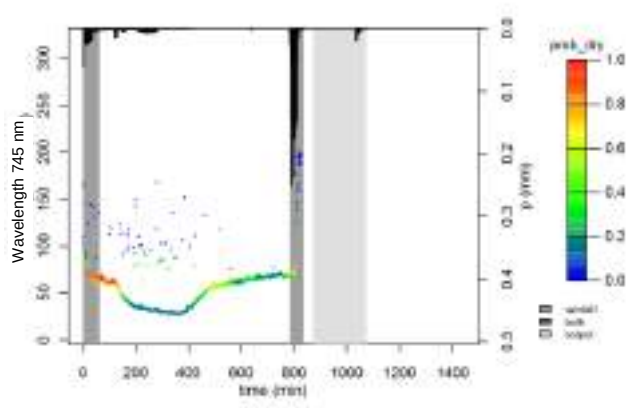
b)



c)

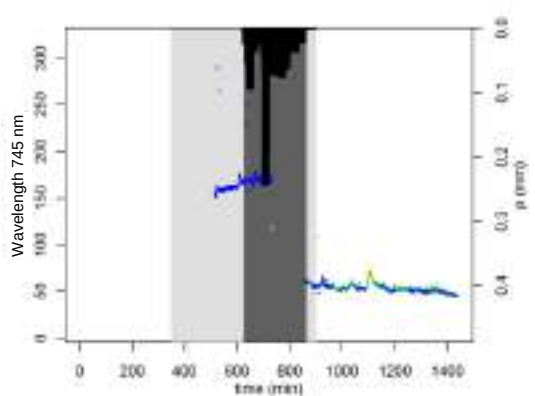


d)

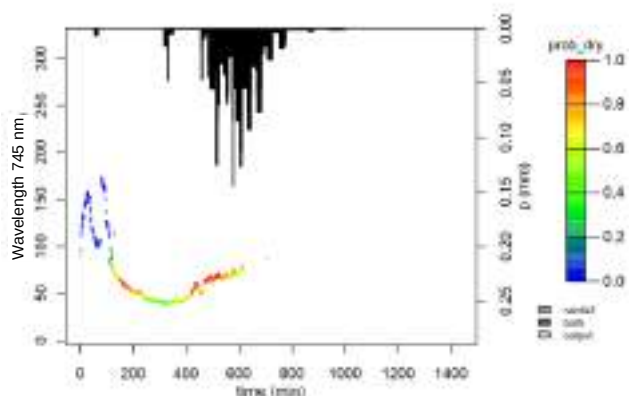


**Appendix 8.17 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding wavelength 745 nm (UV-VIS spectrometry) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Fontibon* rain gauge station.**

a)

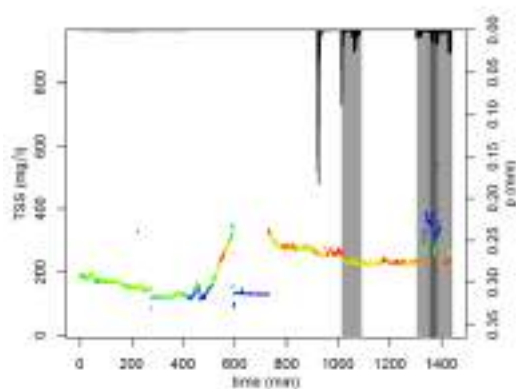


b)

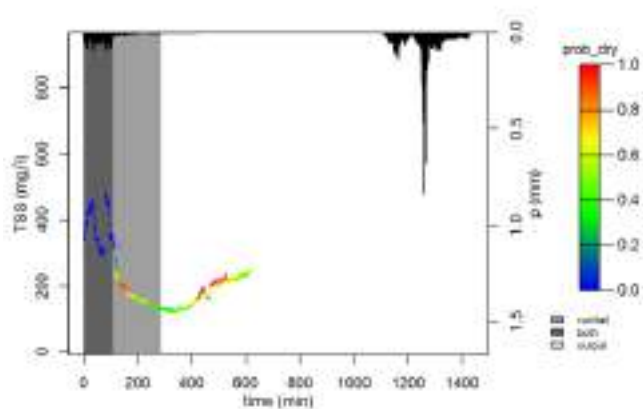


**Appendix 8.18 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding wavelength 745 nm (UV-VIS spectrometry) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Casablanca* rain gauge station.**

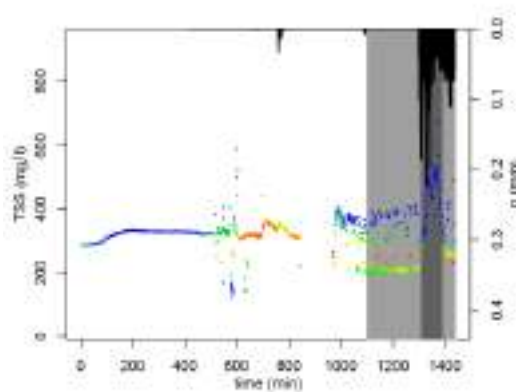
a)



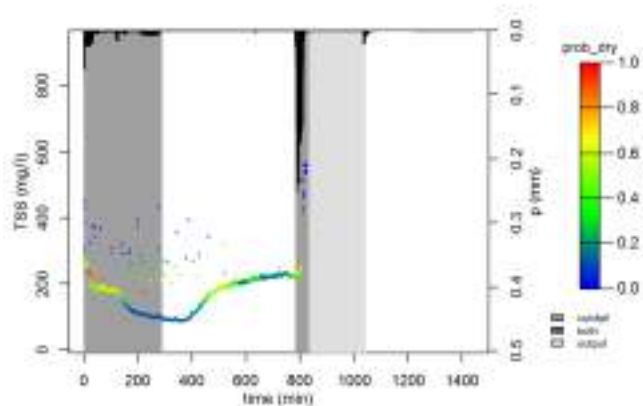
b)



c)

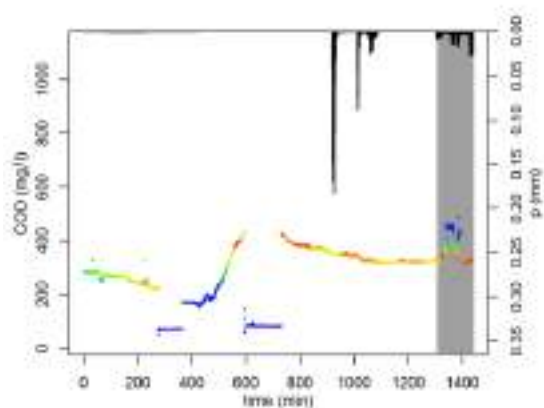


d)

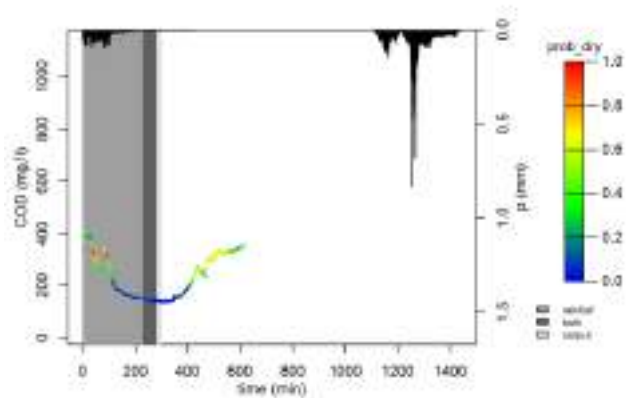


**Appendix 8.19 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding TSS (UV-VIS spectrometry) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Fontibon* rain gauge station.**

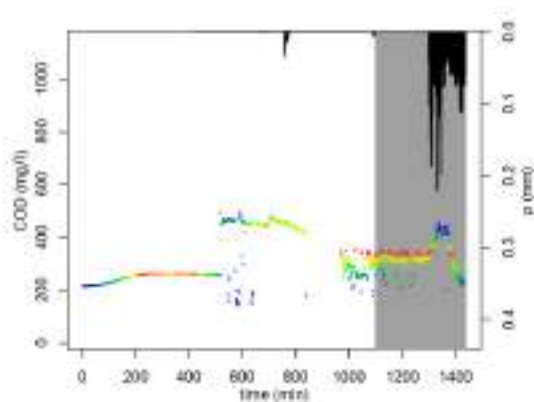
a)



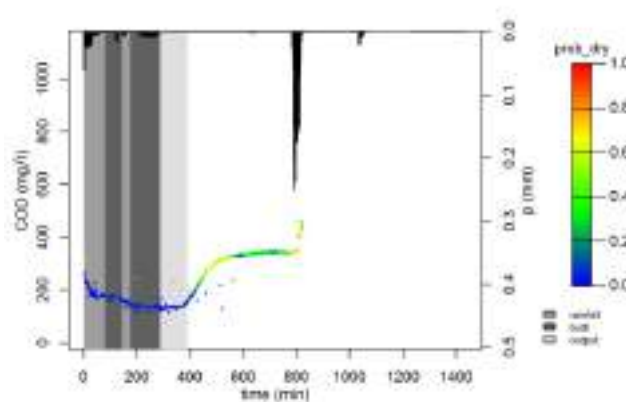
b)



c)



d)



**Appendix 8.20 Rainfall pulses (“rainfall”), time windows raining outside of typical dry weather behavior regarding COD (UV-VIS spectrometry) (“output”) and the overlap between the rainfall pulse and the influenced output (“both”), for rainfall registered at the *Fontibon* rain gauge station.**