Desalination and Water Treatment
www.deswater.com
doi: 10.1080/19443994.2015.1029533

57 (2016) 2669–2684
February

Taylor & Francis
Taylor & Francis Group

Shewhart-type control charts and functional data analysis for water quality analysis based on a global indicator

C. Iglesias[a,*], J. Sancho[b], J.I. Piñeiro[a], J. Martínez[b], J.J. Pastor[b], J. Taboada[a]

[a]*Department of Natural Resources and Environmental Engineering, University of Vigo, 36310 Vigo, Spain, Tel. +34 986811930; emails: carlaiglesias@uvigo.es (C. Iglesias), jpdiblasi@uvigo.es (J.I. Piñeiro), jtaboada@uvigo.es (J. Taboada)*
[b]*Centro Universitario de la Defensa, Academia General Militar, 50090 Zaragoza, Spain, Tel. +34 976739500; emails: jsanchov@unizar.es (J. Sancho), jmtorres@unizar.es (J. Martínez), jjpastor@unizar.es (J.J. Pastor)*

Received 20 October 2014; Accepted 10 March 2015

ABSTRACT

Water quality is a major concern, particularly when the water is used for human consumption. We study the variability of the Ebro River (Spain) water quality through a global quality index (GQI) using two methods: functional data analysis (FDA) and Shewhart-type control charts for statistical process control (SPC). The aim of this study is to identify abnormal values of this quality indicator. We used the data collected in 2008 at the El Bocal station, which is a strategic location. Temperature, ammonium content, nitrate content, conductivity, dissolved oxygen, pH, and turbidity were measured every 15 min. These physical–chemical parameters were used to calculate the GQI. The results obtained using SPC reflect the causes of specific variation in May, July and October. However, no functional outlier was detected when using FDA. According to our results, we conclude that Shewhart-type control charts could be used to search for and eliminate abnormal values in a water quality analysis based on global indicators. The FDA methodology is not appropriate for this case study because of the type of functions obtained from the available data.

Keywords: Water quality; Outlier; Water quality index; Water quality monitoring

1. Introduction

Water is likely the most valuable natural resource, and greater efforts are being made at all levels of our society to achieve an optimal management of both superficial and underground streams. Water quality is a main concern for both the environment and human health because it is fundamental to our nutrition.

In general terms, the following types of water are distinguished based on their use: public supply, domestic, irrigation, livestock, aquaculture, industrial, mining, and thermoelectric power [1]. Maintaining minimum water quality levels and achieving a proper control of streams and their uses are essential for the sustainable management of this natural resource in accordance with legal requirements. This research work focuses on domestic water quality, specifically on drinking and pre-drinking water.

*Corresponding author.

Presented at the International Congress on Water, Waste and Energy Management
16–18 July 2014, Porto, Portugal

In the recent past, EU legislation only regulated water quality, not its quantity. However, the findings of a ministerial seminar on water policy of the European Community held in Frankfurt in 1988 highlighted the need for Community legislation that would address the ecological quality of water.

Until the adoption of Directive 2000/60/EC [2], which aims to maintain and improve the aquatic environment of the Community, referring primarily to water quality, the two directives on the quality and methods of measurement, frequency of sampling, and analysis of surface water intended for the abstraction of drinking water that articulated the control of water quality were as follows:

(1) Directive 75/440/EEC [3], which defined the quality required of surface water intended for the abstraction of drinking water in the Member States (transposed into Spanish legislation by Ministerial Order 11/05/88).
(2) Directive 79/869/EEC [4], which defined the methods of measurement, frequency of sampling, and analysis of surface water intended for the abstraction of drinking water (transposed into Spanish law by OM 08/02/88).

In Spain, Royal Decree 140/2003 [5] defines the health criteria with which the water intended for human consumption must comply. It considers some basic parameters, including turbidity, conductivity, hydrogen ion concentration or pH, and ammonium content, which are also considered as indicator parameters according to Council Directive 98/83/EC [6]. The current EU Water Framework Directive [2] considers biological, hydromorphological, chemical, physicochemical, and general elements when classifying the ecological status of rivers. Those general elements include thermal conditions, oxygenation conditions, salinity, acidification status, and nutrient conditions.

Any attempt to define water quality must be based on a set of physicochemical characteristics that are commonly called quality indicators. Several classifications can be made from these indicators, but two of them are the most widely accepted. The United States Environment Protection Agency distinguishes between primary indicators (inorganic chemicals, organic chemicals, radioactive compounds, micro-organisms) and secondary indicators (color, turbidity, suspended solids). Conversely, the Organization for Economic Cooperation and Development proposes that indicators can be classified into two main groups: pressure indicators (which measure actions that generate change) and response indicators (which measure corrective actions arising from the change).

Of the possible potential unit indicators, the most commonly measured in the literature [7] are pH, water temperature, conductivity, dissolved oxygen, redox potential, turbidity, level, flow, total ammonia, nitrates, phosphates, absorbance, mercury, nitrites, total nitrogen, total phosphorus, chlorophyll, and phycocyanin. These indicators have been used in various ways to define global quality indexes (GQI). When selecting the appropriate unit indicators, anywhere from a minimum of 2 to an infinite number could be considered. The selection is made according to circumstances such as standards, time or location, and usually follows the criteria derived by experts.

Despite the importance of controlling water quality, there is no easy-to-construct index with wide application [8], although several general quality indexes are found in the literature (see [7]). These indexes include some that are used worldwide and have been validated in several studies, though they were originally developed for the specific conditions of a single region or country [9]. This is the case of the ICA indexes of the National Health Foundation of the United States (NSF) and Dinius' ICA [10]. Whereas the first focuses on water intended for human consumption, the second considers other uses, including agriculture, fisheries, industry, and recreation.

Abbasi [11] reviews the existing water quality indexes, including many others aside from those mentioned above. A multiplicative weighed index that ranges from 0 to 100 classifies rivers as very bad, bad, medium, good or excellent quality [12,13] has been applied to the study of many rivers [13–16]. This index uses physicochemical properties such as dissolved oxygen, biochemical oxygen demand, turbidity, total solids, nitrates, phosphates, pH, and temperature. A new subjective quality index WQI_{sub} is presented in Ref. [17] and incorporates a subjective constant into an equation that represents the general state of the water. Pesce and Wunderlin [13] present two other indexes: WQI_{obj} (which is an objective WQI) and WQI_{min} (which considers only dissolved oxygen, conductivity, and turbidity). Other authors have focused their efforts on defining a fuzzy water quality index using inference [18–22]. The Environmental Protection Index of Taiwan has adopted its own classification system and uses a river pollution index that is based on only four parameters: dissolved oxygen, biochemical oxygen demand, suspended solids, and ammonia nitrogen [23,24]. The index proposed by Lamontagne and Provencher [25] is widely recognized and was used by most of the Spanish

Hydrographic Confederations until the implementation of the Water Framework Directive.

The index used in this study is not presented as a solution for this issue but as a tool that incorporates the main parameters available in the control stations of the Ebro River. In this research work, we study the GQI proposed in Ref. [8], which is based on the four levels of water quality for human consumption set by Directive 75/440/EEC.

Directive 75/440/EEC subdivides the surface water intended for the abstraction of drinking water quality into four categories: A1, A2, A3, and +A3. These four categories correspond to different water qualities and are characterized by mandatory values and guide values. They are directly related to the type of treatment processes for the purification of water:

(1) Category A1: simple physical treatment and disinfection, e.g. rapid filtration and disinfection.
(2) Category A2: normal physical treatment, chemical treatment and disinfection, e.g. prechlorination, coagulation, flocculation, sedimentation, filtration, and disinfection (final chlorination).
(3) Category A3: intensive physical and chemical treatment, refining treatment, and disinfection, e.g. chlorination to "break point", coagulation, flocculation, sedimentation, filtration, refining (activated carbon), and disinfection (ozone, final chlorination).
(4) Category +A3: surface water whose physical, chemical, and microbiological indicators are below the mandatory limit values for type A3. This lower quality water could be used as an exception if suitable processes (including mixing) raised all of the characteristics of water quality to a level that is consistent with the quality standards of drinking water.

In Spain, inland water quality monitoring is performed by a total of nine hydrographic confederations. These organizations have several essential roles, including hydrographical planning, resource management, public hydraulic domain protection, and dam security plans. The Ebro Hydrographic Confederation [26] is in charge of managing surface water and groundwater in the Ebro basin, which is the largest by discharge volume in Spain. Among its tasks, the Confederation protects and controls the water quality of the river through the Automatic Hydrological Information System (AHIS).

AHIS comprises a number of stations distributed in different Spanish basins that measure several parameters to control the status of the watershed, alert in case of flood risk, and instantly make known the availability of water resources. Taking into account the several legal requirements involving water quality, we considered the following seven parameters in our study: temperature, ammonium content, nitrate content, conductivity, dissolved oxygen, pH, and turbidity.

The case study considers a strategic location of Ebro River, El Bocal, which is located after the confluence of the Aragón and Ebro Rivers and at the beginning of the Imperial Channel of Aragón (an important irrigation channel). Thus, the quality of the water measured at this point is directly related to that of the drinking water of the city of Zaragoza. The drinking water supply of Zaragoza city is a mixture of 50% among which is received from the reservoir of Yesa that comes from the mountains (the Pyrenees) and the other 50% comes from the Imperial Channel of Aragón.

We study the variability of water quality using a GQI based on European Directive 75/440/CEE [8]. Although Directive 2000/60/EC repealed the abovementioned European Directive, the purpose of the study is unaffected because the proposed methodology is valid for any index defined for water quality monitoring. In other words, the aim of this study is to develop a methodology that is suitable for detecting abnormal values of any water quality indicator. These abnormal values (or outliers) could be noisy data or indicators of abnormal behaviors in the controlled system, thus representing useful information that may lead to significant discoveries [27].

Hence, the variability of the Ebro River water quality is studied in this work through a GQI using two methods: functional data analysis (FDA) and Shewhart-type control charts for statistical process control (SPC).

We have successfully applied FDA to detect outliers in different environmental studies [27–31], occasionally demonstrating many and obvious advantages over SPC [32]. Many other works can be found in the literature in which FDA has been applied in the following ways with satisfactory results: to determine the influence of several biotic and abiotic factors on marine fish species [33], to characterize microbial communities [34], and to detect air pollution [30].

SPC has its origin in manufacturing processes and the need to develop a methodology for efficiently controlling product quality. It is a strategy for production and process optimization that has been commonly used for decades [35].

SPC is based on the concept of the variation of a process and its relationship with quality. According to Shewhart, understanding the variability of a monitored process allows for its correction and reduction [36,37]. The variability of a process can be natural (random) or non-natural (assignable to a cause). A process is said to be under statistical control when the only variability present is natural, unavoidable, and inherent to the process itself. Conversely, assignable causes should be detected and eliminated to improve the process [36,38]. FDA is used here to study any variation in the quality of the Ebro River and to determine the causes of such variation.

For details on the FDA and SPC methodologies, the reader referred to the works of Refs. [39] and [40], respectively.

2. Methodology

2.1. Study area and available data

The Ebro basin is located in the northeast of the Iberian Peninsula and is the largest Spanish watershed, with a total area of 85,362 km^2. It is drained by the Ebro River, which runs through the peninsula in a NW–SE direction and flows into the Mediterranean Sea. The Ebro River is the largest river in Spain (the second in the Peninsula after the Douro River) and the second longest river in the Peninsula, after the Tagus River. Though it is a large river, its flow is irregular; at the end of the summer, its drought is so intense that the flow can be a tenth of the medium annual flow of the river.

To control the state of this watershed, the Ebro basin has 41 control stations: 20 are in the upper part of the basin, and 21 are on the lower Ebro River. This research work focuses on the data collected at El Bocal station (Fig. 1), which is located on the upper Ebro River in the municipality of Fontellas (Navarra), 482 km from the river mouth and 79.5 km from the city of Zaragoza, one of the most important cities in Spain; Zaragoza's 682,000 inhabitants in 2013 make it the 5th most populous city.

The importance of this gage point is its strategic location; El Bocal is found just after the confluence of Aragón (the main tributary of the Ebro) and Ebro Rivers and at the beginning of the Imperial Channel of Aragón. This channel is one of the most important hydraulic works in Europe. It was built in the eighteenth century from Fontellas (Navarra) to Fuentes de Ebro (Zaragoza) with the aim of improving the irrigation of the old Imperial Irrigation Ditch of Aragón. The new channel conducted water of Ebro River to Zaragoza and established a transport service for passengers and goods. Currently, the Ebro Hydrographic Confederation is in charge of managing this irrigation and navigation channel.

The database used in this work contains measurements that were taken every 15 min from January to December 2008 of the following physical and chemical properties: temperature, ammonium content, nitrate content, conductivity, dissolved oxygen, pH, and turbidity. This time period includes the months when the Expo, a world's fair focused on "Water and Sustainable Development", occurred in Zaragoza.

The complete database is available on the Ebro Hydrographic Confederation's website [26] to any interested user by means of prior registration.

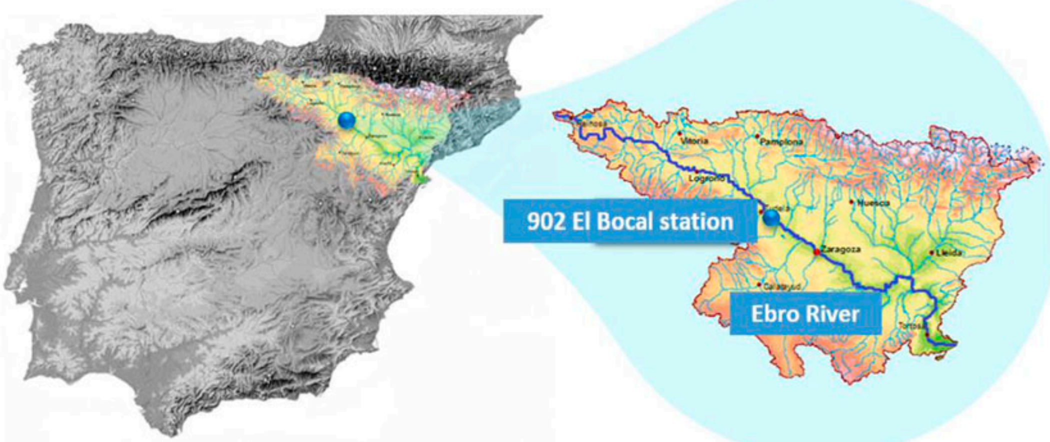

Fig. 1. Location of the Ebro basin and the automatic measurement station 902 El Bocal, the source of the data in this study.

2.2. GQI definition and analysis

The calculation procedure for the physical and chemical classification is the result of applying the criteria set out in Art. 5 of Directive 75/440/EEC.

The corresponding tables are used to calculate the compliance with the mandatory values. The indicator is in accordance with the appropriate quality group, i.e. A1, A2, or A3, when:

(1) In at least 95% of the samples, the indicator is consistent with the corresponding value in Table 1 of the Directive.
(2) For non-compliant samples (the remaining 5%):
   (i) The assessed value of the indicator should not exceed more than 50% of the limit set value (temperature is excluded from this limitation).
   (ii) There is no danger to public health resultant.
   (iii) Consecutive water samples taken at statistically suitable intervals do not differ from the values of the corresponding indicators.

Indicators that are identified as mandatory may be excluded from the calculation if the exception is justified by exceptional weather or geographical reasons.

To measure water quality with a global index, Beamonte et al. [8] defined a vector $v = (a, b, c, d)$ called the "stewardship quality vector". This vector shows the number of indicators that belong to a certain level of quality (A1, A2, A3, +A3) of a total of k indicators: a for A1 quality, b for A2, c for A3, and d for +A3.

Given two water samples M_1 and M_2 defined by $v_1 = (a_1, b_1, c_1, d_1)$ and $v_2 = (a_2, b_2, c_2, d_2)$, the quality of M_1 is better than that of M_2 if and only if one of these constraints is true:

(1) $d_1 < d_2$;
(2) $d_1 = d_2$ and $c_1 < c_2$;
(3) $d_1 = d_2$, $c_1 = c_2$ and $b_1 < b_2$.

Thus, the worst quality possible is given by $v = (0, 0, 0, k)$, and the best is represented by $v = (k, 0, 0, 0)$.

Furthermore, this stewardship quality vector can be used to define a GQI [8]:

$$\mathrm{GQI}(a,b,c,d) = \frac{1}{6}\left(s_1^3 + 3s_1^2 + 2s_1\right) + \frac{1}{2}\left(s_2^2 + s_2\right) + a + 1 \tag{1}$$

where $s_1 = a + b + c$ and $s_2 = a + b$. Unlike other indexes whose values are from 0 to 100, this index ranges from 1 to $\frac{(k+3)(k+2)(k+1)}{6}$.

The procedure for calculating the GQI is the following:

(1) The administrative quality (A1, A2, A3, +A3) of every measure is calculated, considering the maximum and minimum allowed values set by law for each physical and chemical parameter.
(2) Then, the administrative quality of each day is established with these criteria: if more than 95% of the measurements of the day are qualified as A1, then the day is A1. If not, the percentage of A2 measurements is checked; if it is greater than 95%, the day is qualified as A2. If not, the same criterion is applied to A3; if none of the previous conditions is true, the day is qualified as +A3. This is done for each parameter (temperature, ammonium content, nitrate content, conductivity, dissolved oxygen, pH, and turbidity).
(3) Once the daily quality for each property is known, the daily index GQI is calculated according to Eq. (1).

This GQI index was analyzed using two methods, FDA and Shewhart-type control charts for SPC.

2.2.1. Functional data analysis

FDA focuses on observations of a random continuous process that are taken at discrete points

Table 1
Faulty indexes per month in 2008

Month	Days	Faulty GQI values	Month	Days	Faulty GQI values
January	31	18	July	31	24
February	29	21	August	31	24
March	31	22	September	30	12
April	30	18	October	31	28
May	31	5	November	30	17
June	30	15	December	31	12

(functional data). In this methodology, a functional outlier is defined as a curve generated by a stochastic process whose distribution differs from that of the rest of the curves, which are identically distributed [29,30,34,41,42].

First, a smoothing process is applied with the aim of transforming the vector problem into a new functional problem. The initial annual data are divided into 12 subsets of the new functional data, representing the measurements of each variable per month.

Then, the functional depth concept is applied: in a Euclidian space, those points closer to the center have greater depth. In the case of functional analysis, functional depth is a measure of the centrality of a certain curve within a set of curves or functions. We use the H-modal depth concept to calculate this functional depth [42]:

$$\mathrm{MD_n}\ (x_\mathrm{i}, h) = \sum_{k=1}^{n} K\left(\frac{\|x_\mathrm{i} - x_\mathrm{k}\|}{h}\right) \tag{2}$$

where x_i and x_k are the curves of the managed dataset, h is the bandwidth parameter, and $K{:}R^+ \rightarrow R^+$ represents a kernel function. According to the results of the functional depth, outliers are detected: because outliers are abnormal values, their depth is considerably lower than that of the other curves, which represent normal, expectable values.

The detection of outliers is based on the Z-score:

$$z_\mathrm{i} = \frac{x_\mathrm{i} - x_\mathrm{r}}{\sigma_\mathrm{r}} \tag{3}$$

where x_i represents the mean value of the observations at point i, x_r is the mean value for all measures (used as a reference value), and σ_r represents the reference standard deviation. Possible outliers have a Z-score equal to or greater than two, whereas clear candidate outliers have a Z-score of three or greater [43].

For further information on FDA, the depth concept and the Z-score method, the reader referred to [27,30,39,42,43].

2.2.2. Statistical process control

The second approach to water quality analysis is Shewhart-type control charts for SPC. In a controlled process, there is always a certain variability due to common (or natural) causes or to special (or assignable) causes [36]. Though the first are inherent to the process, the latter are related to causes outside the process and can be eliminated if they are correctly identified. If there is only natural variability in a process, then it is said to be under statistical control.

SPC is a technique that allows for monitoring, analyzing, predicting, controlling, and improving the variability of a certain characteristic using control charts. Shewhart control charts [36] define an upper control limit (UCL) and a lower control limit (LCL), and a process is under statistical control if all points are between the UCL and LCL. The central line (CL) represents the mean value of the studied property. UCL and LCL would represent the tolerance of the process for a certain target value CL.

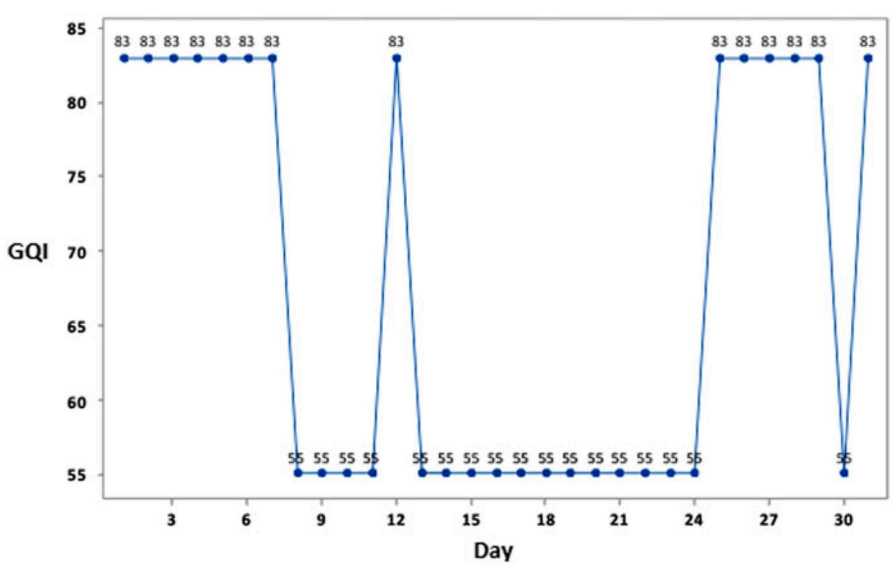

Fig. 2. Daily GQI in January 2008.

The studied quality characteristic is considered to follow a normal distribution $N(\mu_y, \sigma_y)$, and the control limits can be calculated as follows:

$$\begin{aligned} UCL &= \mu_y + k \cdot \sigma_y \\ LCL &= \mu_y - k \cdot \sigma_y \\ CL &= \mu_y \end{aligned} \tag{4}$$

where μ_y is the mean value, σ_y is the standard deviation, and k is a measure of the tolerance. Warning and control limits are usually defined; warning limits are set at a distance of $\pm 2\sigma$ from the CL, and control limits are set at a distance of $\pm 3\sigma$.

To use Shewhart control charts, the samples should be grouped according to a criterion related to the cause of the special variation that is under study if the cause of variation is suspected to be seasonal; for example, data must be grouped into rational seasonal subgroups. In the case study, the special causes of variation in the quality index pattern are expected to be monthly.

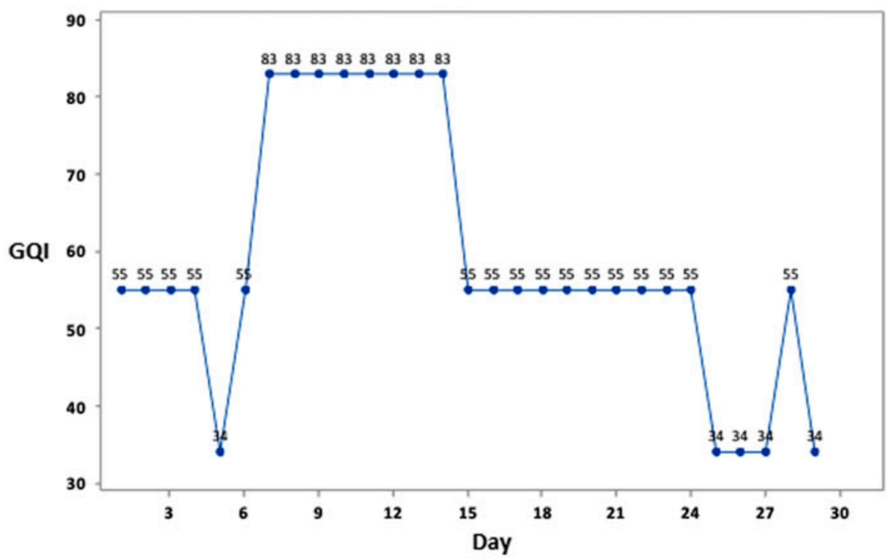

Fig. 3. Daily GQI in February 2008.

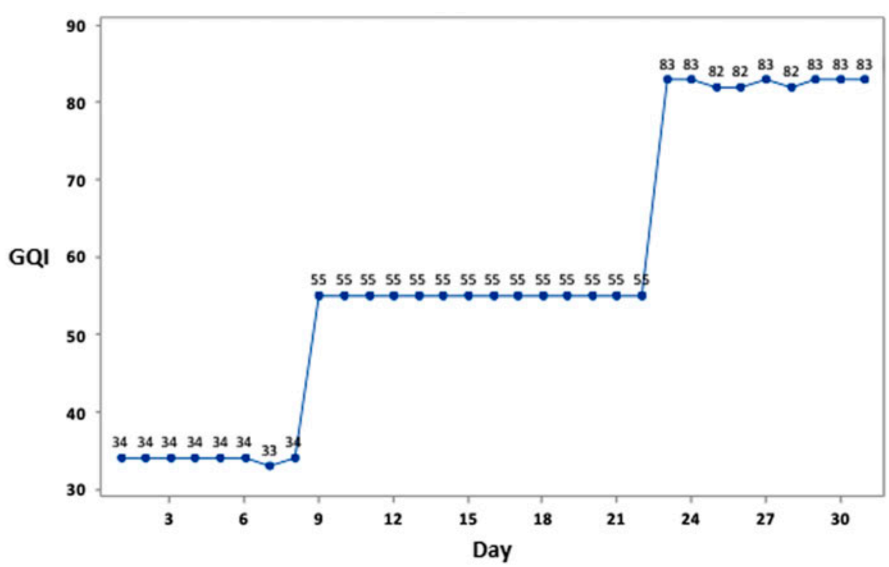

Fig. 4. Daily GQI in March 2008.

There are many types of control charts, and many rules can be used to interpret them and analyze the variability of the process, whether the variation is small or large. Control charts for attributes allow for controlling the qualitative characteristics of quality instead of measurable magnitudes [44], such as the proportion of defective units. As in the graphs of the variables, the control of attributes represents a statistical process vs. time or the number of the selected samples, where the CL signifies the expected value for the statistical control limits and defines the rejection region.

In general, there are two groups of control charts for attributes: one is formed by p and np-type control charts, which compares the attribute with a standard and classifies it as defective or not defective, and another includes c and u-type control graphics and is used for more complex products where the existence of a defect does not necessarily invalidate the product. These types of graphs allow for the classification of a product according to the number of defects it has.

Usually, the election between using control charts for variables or for attributes is simple. However, both

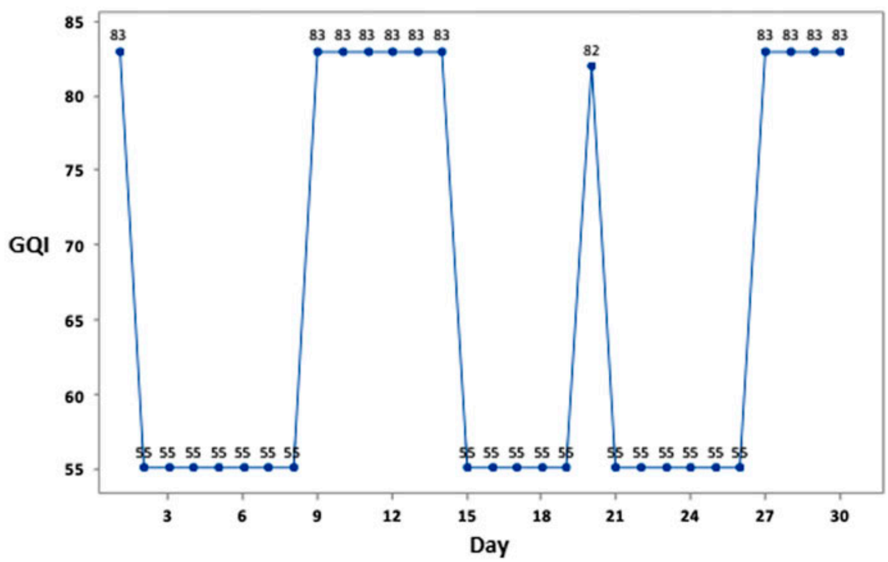

Fig. 5. Daily GQI in April 2008.

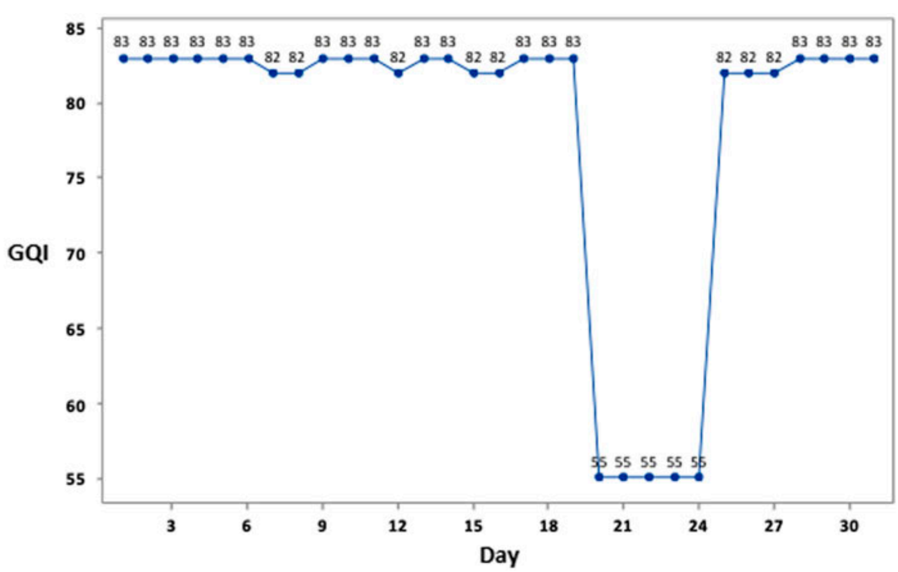

Fig. 6 Daily GQI in May 2008.

were analyzed in this case study: in addition to studying the variables, a control chart for attributes was used based on the proportion of defective elements (P-chart). Additional information regarding the variability of the GQI was then obtained.

The type of control chart for attributes, which monitors the fraction of nonconforming observations (in the case study, ICQ values under 60), has been widely used in industry because of its suitability for interpreting and reducing the sources of variability in manufacturing processes [45]. In fact, other fields such as medicine are also adopting this methodology for continuous quality control [37]. Furthermore, P-charts allow for a later capacity analysis based on the defects detected.

A summarized explanation of the SPC process and the stages that it includes is presented below [46]. Phase 1 is based on the study of historical data to determine whether it is under control, and phase 2 is focused on analyzing sequentially taken observations to detect changes in a process that is under statistical

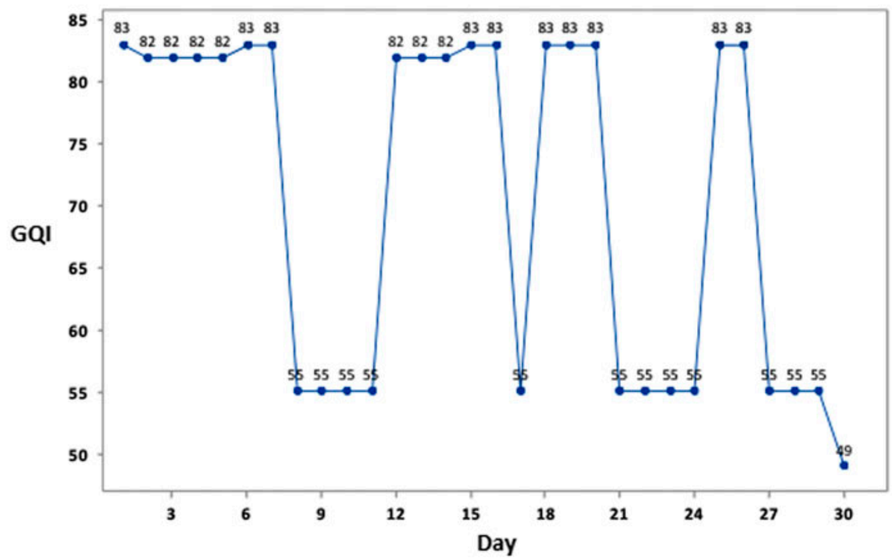

Fig. 7. Daily GQI in June 2008.

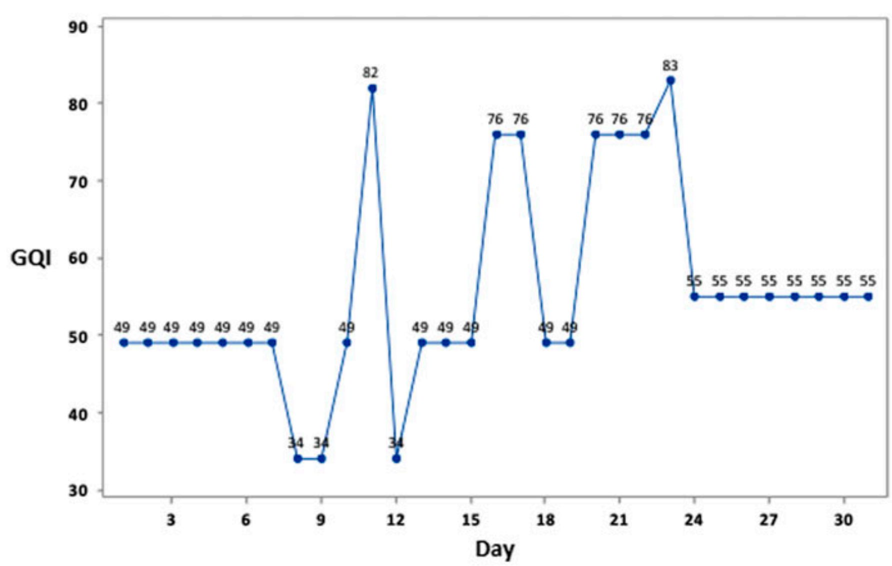

Fig. 8. Daily GQI in July 2008.

control. Hence, the aim of phase 1 is to determine the experimental control limits that determine whether the process is under statistical control and are used to control the incipient production. Out-of-control points are detected and studied to determine the assignable causes of variation. These points are then deleted from the sample, and the limits are recalculated; this process is repeated until all of the points forming the sample are under control.

Once the appropriate control limits are set, they are applied to control charts in phase 2 for the real-time monitoring of the process. In this stage, different rules are applied to detect deviations and patterns. Thus, we can say that phase 1 performs a retrospective analysis and that phase 2 performs a prospective analysis of the process.

According to the characteristics and objectives of the phases of SPC, this research work represents the

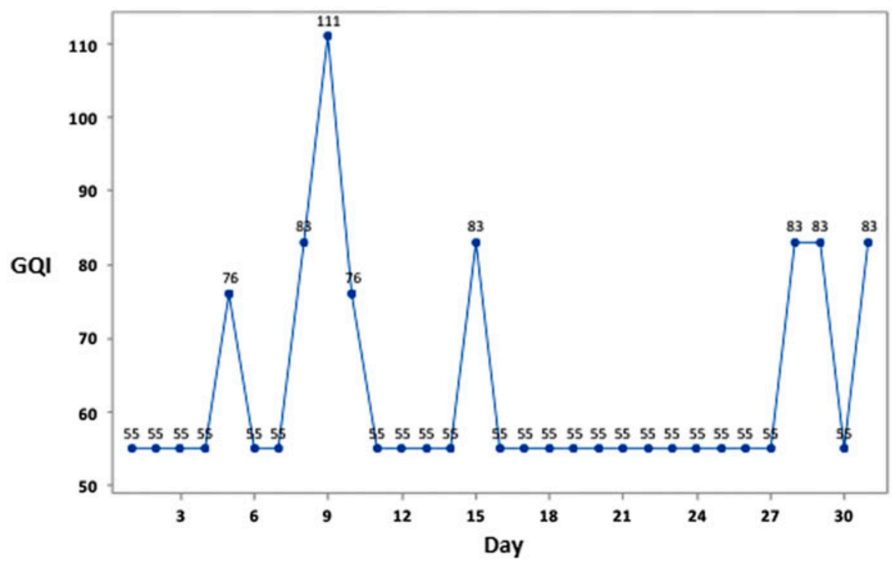

Fig. 9. Daily GQI in August 2008.

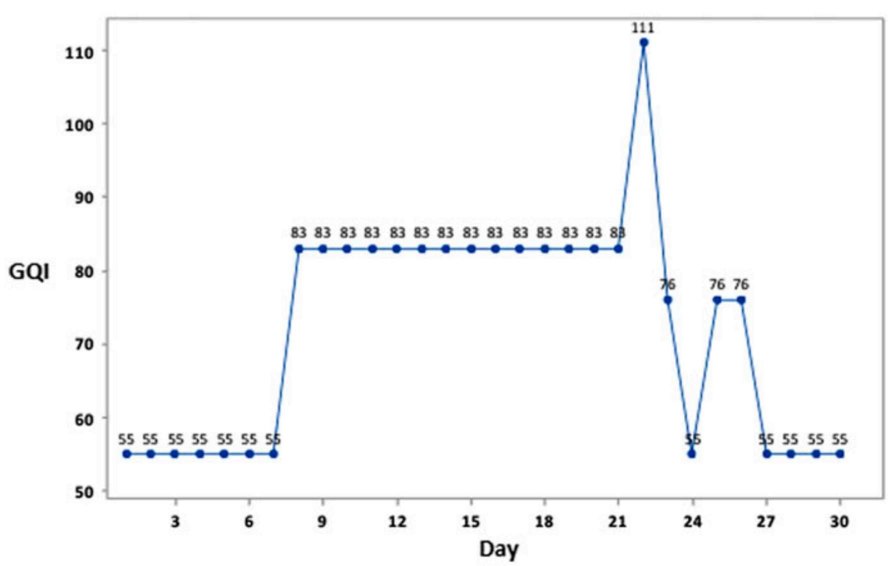

Fig. 10. Daily GQI in September 2008.

first stage of the process: data are analyzed retrospectively with the aim of identifying abnormal observations (outliers) and determining their assignable causes.

3. Results and discussion

Following the methodology described above, we calculated the corresponding GQI for each day of the year 2008. Seven parameters are considered, so $k = 7$ and GQI ranges from 1 to 120. The following figures show the daily GQI values (Figs. 2–13). It is a stepwise function that does not fit any of the following distributions: normal, exponential, lognormal, Weibull, and extreme value.

Normality and control charts are usually linked, assuming that the data fit a normal distribution. However, SPC can be applied to study non-normal data, regardless of whether we know that it fits some other distribution or whether its fit is simply unknown [46].

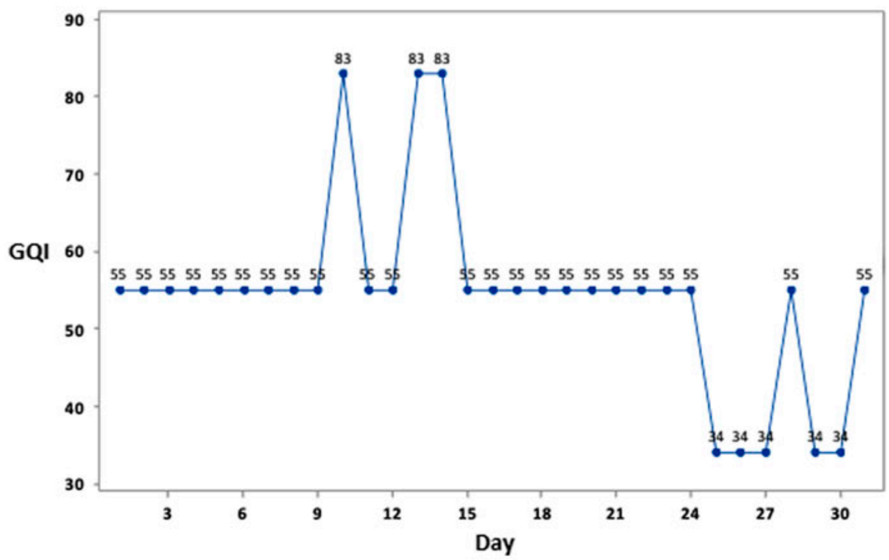

Fig. 11. Daily GQI in October 2008.

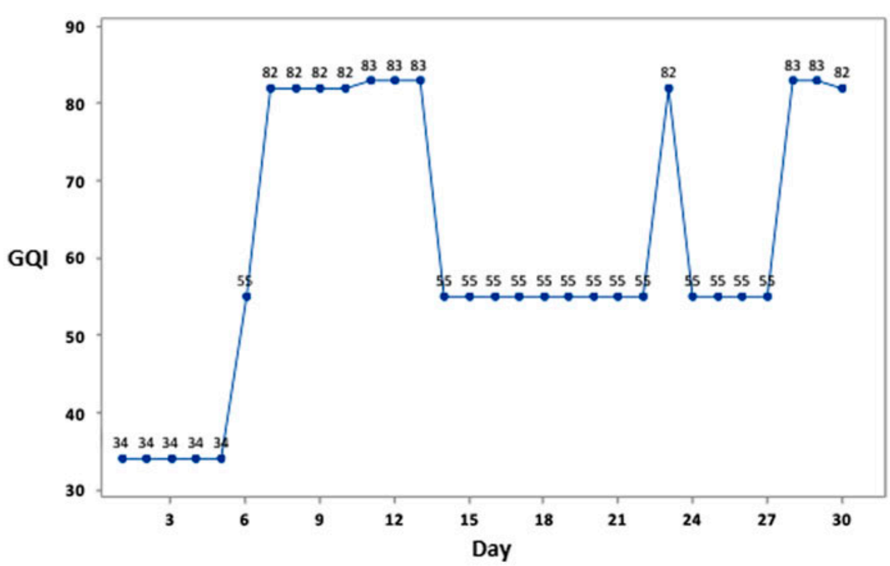

Fig. 12. Daily GQI in November 2008.

In this case, normal theory results could be used to obtain the probability limits for the control charts.

In the following sections, the variability of the GQI is studied using a control chart for variables, the Xbar-S (Fig. 14), and a control chart for attributes, the P-chart (Fig. 16).

The control charts obtained using the above-mentioned methodology are shown in Figs. 14 and 16. These are standard Shewhart control charts, where the UCL and LCL were calculated as three-sigma limits, with an in-control ARL (average run length) equal to 370.4 (a false alarm occurs every 370.4 observations).

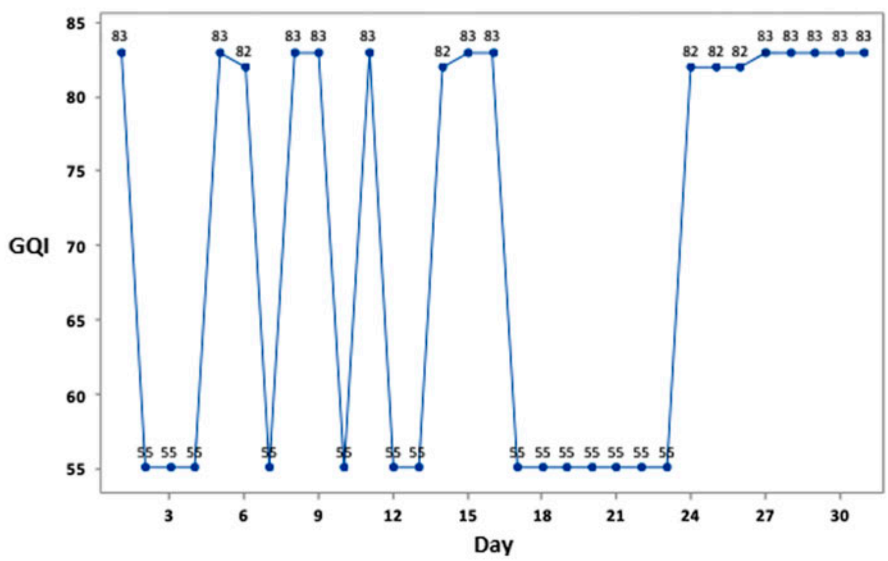

Fig. 13. Daily GQI in December 2008.

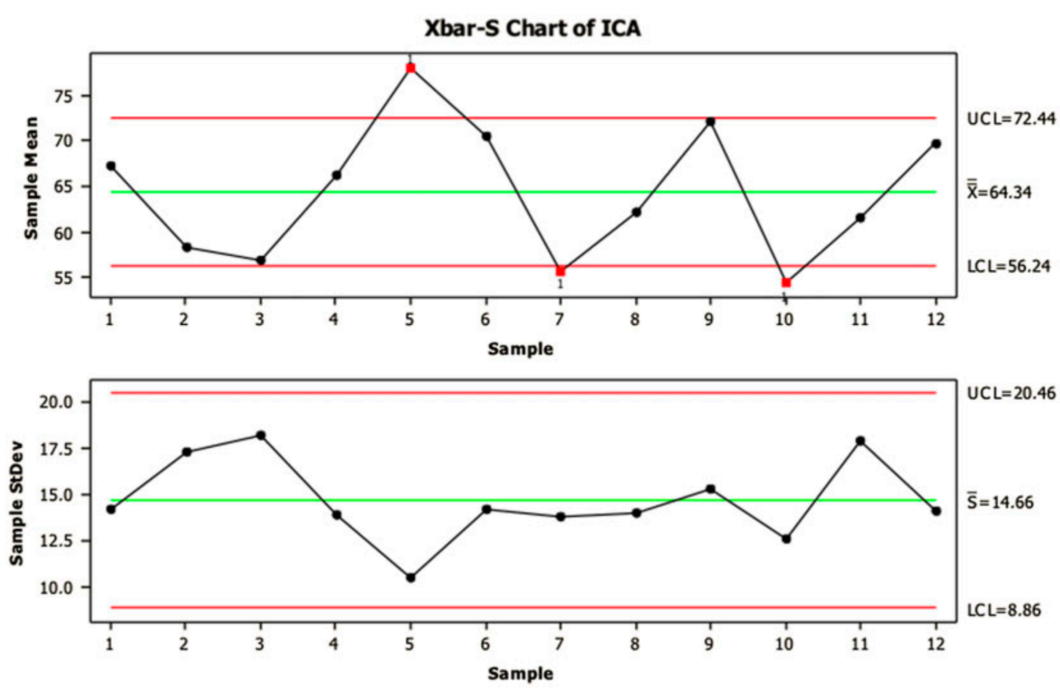

Fig. 14. Resulting Shewhart-type control chart for 2008's GQI. The upper graph shows the control chart of the mean values, and the lower graph shows the control chart of the standard deviations.

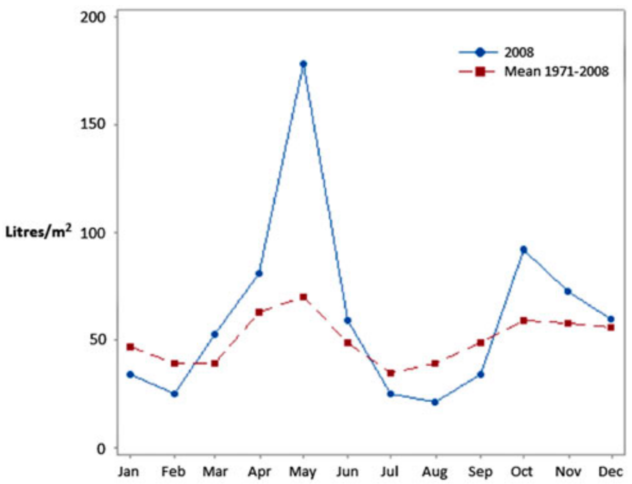

Fig. 15. Average monthly rainfall in the Ebro River Basin 1971–2008. Year 2008. Data: CHE.

The autocorrelation of data, which is likely to occur in a time series, has been studied using the partial autocorrelation function of GQI (for further details, the reader referred to Ref. [47]). No evidence of autocorrelation was found: according to the graphs obtained for each month, there is no autocorrelation for GQI values for intervals greater than one day.

The results obtained using Shewhart-type control charts (Fig. 14) reflect causes of specific variation that may be deleted if they are identified in May, July and October (represented by red dots). It should be noted that Fig. 14 contains both the control chart for the mean values and the control chart for the standard deviations. This simultaneous monitoring is commonly performed in SPC because it considers not only the variability of the process but also the variability of the different rational subgroups [46]. In other words, both the variability of the process over time and the instantaneous variability are considered.

When interpreting these outliers, it is important to note that 2008 was anomalous from the point of view of the current flow along the Ebro River for three main reasons:

(1) May 2008 was a month of heavy rains in the Ebro River Basin: 178 l/m^2, compared with the average of 70 l/m^2 for the period 1971–2007 (Fig. 15). This heavy precipitation during the spring of 2008 explains the high river flows and therefore the corresponding outlier in May.
(2) From June 14 to September 14, 2008, Zaragoza held the Expo. The location of the exhibition complex in the Ranillas Meander, which passes through Zaragoza, motivated the maintenance (by artificial means) of higher than normal flows during the summer months. This explains the abnormal value in July.
(3) Finally, the outlier in October is justified by the first rains of autumn (92 l/m^2 during the month of October 2008, compared with 59 l/m^2 for the period from 1971 to 2007) after a very dry summer (59, 25, 21, and 34 l/m^2 during the months of June, July, August, and September, compared with 49, 35, 39, and 49 l/m^2 on average during the period of 1971–2007) (Fig. 15).

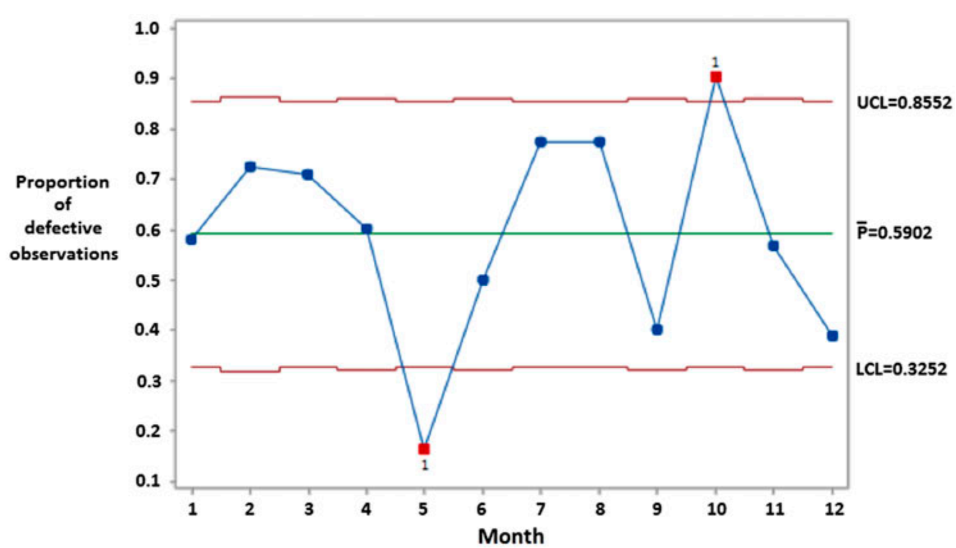

Fig. 16. P-Chart of the rational monthly subgroups of GQI in 2008.

In this case study, we constructed a P-chart with rational subgroups representing the 12 months of 2008 and a limit value of 60 for the GQI. According to this limit, any index lower than 60 is classified as faulty. The results of the application of this limit are shown in Table 1, and the corresponding control graph is shown in Fig. 16.

Two outliers can be observed in the P-chart, one in May and one in October. The reasons explained above regarding the outliers detected in the Xbar-S control diagram are equally applicable to the interpretation of the outliers detected in this P-chart.

For the FDA, in this case study, no functional outlier was detected. Here, the functional analysis methodology proved to be inadequate because the function is stepwise. The detection of outliers using this methodology provides good results when the analyzed functions are complex (see [27–30]), but its application to this GQI is nonsense because of the characteristics of the function itself.

4. Conclusions

The quality of the Ebro River in 2008 was evaluated using a GQI by means of two methods: SPC and FDA.

SPC usually has limited validity when it is used to study environmental parameters because of the non-normality of data distribution, the effect of autocorrelation in time series data, and the greater variability between rational subgroups than within the rational subgroups. However, the results of the case study suggest that SPC could be successfully used for the study of water quality based on global indicators. Abnormal values with assignable variability were found in the measurements taken in Station No. 902 of the Ebro River in 2008. After eliminating these outliers and recalculating the control limits, the control charts could be used for the real-time monitoring of the defined GQI.

The functional analysis methodology is a novel approach that allows treating data as a set of continuous measurements with functional outliers as outcomes and analyzing trends, regardless of abnormal data that may represent measurement errors (this method focuses on the analysis of functions rather than mean values and detects abnormal patterns rather than abnormal observations). Additionally, functional analysis presents another advantage in that a normal distribution of the measurements is not required. However, this methodology also has some drawbacks, such as the time series data not being sufficiently derivable. In the case study, despite varying the number of basis functions, the method is unable to fit the analyzed function, and no outlier is detected. Hence, in this case, FDA is not appropriate for the detection of outliers because of the type of functions obtained from the available data, which are stepwise.

Acknowledgements

The authors would like to thank the Ebro Hydrographic Confederation for collaborating in this study. This research work was supported by the Spanish Ministry of Education, Culture, and Sport through Grant FPU 12/02283, the Spanish Ministry of Economy and Competitiveness through Project ECO 2011-22650 and the University of Vigo through a predoctoral contract (2014).

References

[1] S.S. Hutson, N.L. Barber, J.F. Kenny, K.S. Linsey, D.S. Lumia, M.A. Maupin, Estimated Use of Water in the United States in 2000, 2004. Available from: <http://pubs.water.usgs.gov/cir1268>.

[2] European Parliament and the Council, Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy, 2000.

[3] Council of the European Union, Council Directive 75/440/EEC of 16 June 1975 Concerning the Quality Required of Surface Water Intended for the Abstraction of Drinking Water in the Member States, 1975.

[4] Council of the European Union, Council Directive 79/869/EEC of 9 October 1979 concerning the methods of measurement and frequencies of sampling and analysis of surface water intended for the abstraction of drinking water in the Member States, 1979.

[5] Ministerio de Sanidad y Consumo, Royal Decree 140/2003 of 7 February by which health criteria for the quality of water intended for human consumption are established, Spain, 2003.

[6] Council of the European Union, Council Directive 98/83/EC of 3 November 1998 on the quality of water intended for human consumption, 1998.

[7] N. Fernández, A. Ramírez, F. Solano, Physico-chemical water quality indices—A comparative review, Rev. BISTUA, 2(1) (2004) 19–30.

[8] E. Beamonte, J. Bermúdez, A. Casino, E. Veres, Un indicador global para la calidad del agua, in: 27 Congr. Nac. Estadística E Investig. Oper., Lleida, 2003.

[9] V. González Meléndez, O. Caicedo Quintero, N. Aguirre Ramirez, Application of water quality indices NSF, DINIUS and BMWP, Rev. Gestión Y Ambient. 16 (2013) 97–108.

[10] S.H. Dinius, Social accounting system for evaluating water resources, Water Resour. Res. 8 (1972) 1159–1177.

[11] T. Abbasi, S. Abbasi, Water Quality Indices, Elsevier, Amsterdam, 2012.

[12] R.A. Deininger, A water quality index for rivers, Water Int. 5 (1980) 16–21.
[13] S. Pesce, D.A. Wunderlin, Use of water quality indices to verify the impact of Córdoba City (Argentina) on Suquía River, Water Res. 34 (2000) 2915–2926, doi: 10.1016/S0043-1354(00)00036-1.
[14] A.B.M. Egborge, J. Benka-Coker, Water quality index: Application in the Warri River, Nigeria, Environ. Pollut. Ser. B, Chem. Phys. 12 (1986) 27–40, doi: 10.1016/0143-148X(86)90004-2.
[15] E. Sánchez, M.F. Colmenarejo, J. Vicente, A. Rubio, M.G. García, L. Travieso, R. Borja, Use of the water quality index and dissolved oxygen deficit as simple indicators of watersheds pollution, Ecol. Indic. 7 (2007) 315–328, doi: 10.1016/j.ecolind.2006.02.005.
[16] D. Sarala Thambavani, T.S.R. Uma Mageswari, Water quality indices as indicators for potable water, Desalin. Water Treat. 52 (2013) 4772–4782, doi: 10.1080/19443994.2013.834517.
[17] V. Conesa Fernandes-Vitora, Methodological Guide for Environmental Impact Evaluation, third ed., Mundi-Prensa, Madrid, 1997.
[18] N.B. Chang, H.W. Chen, S.K. Ning, Identification of river water quality using the Fuzzy Synthetic Evaluation approach, J. Environ. Manage. 63 (2001) 293–305, doi: 10.1006/jema.2001.0483.
[19] S. Dahiya, B. Singh, S. Gaur, V.K. Garg, H.S. Kushwaha, Analysis of groundwater quality using fuzzy synthetic evaluation, J. Hazard. Mater. 147 (2007) 938–946, doi: 10.1016/j.jhazmat.2007.01.119.
[20] F. Nasiri, I. Maqsood, G. Huang, N. Fuller, Water quality index: A fuzzy river-pollution decision support expert system, J. Water Resour. Plan. Manage. 133 (2007) 95–105, doi: 10.1061/(ASCE)0733-9496(2007)133:2(95).
[21] A. Lermontov, L. Yokoyama, M. Lermontov, M.A.S. Machado, River quality analysis using fuzzy water quality index: Ribeira do Iguape river watershed, Brazil, Ecol. Indic. 9 (2009) 1188–1197, doi: 10.1016/j.ecolind.2009.02.006.
[22] Y. Icaga, Fuzzy evaluation of water quality classification, Ecol. Indic. 7 (2007) 710–718, doi: 10.1016/j.ecolind.2006.08.002.
[23] S.M. Liou, S.L. Lo, S.H. Wang, A generalized water quality index for Taiwan, Environ. Monit. Assess. 96 (2004) 35–52, doi: 10.1023/B:EMAS.0000031715.83752.a1.
[24] C.-P. Yang, P.-C. Chiang, E.-E. Chang, T.-F. Lin, C.-M. Kao, Using a systematic approach to develop water quality management strategies in the Nankan River, Taiwan, Desalin. Water Treat. 51 (2013) 6808–6823, doi: 10.1080/19443994.2013.792526.
[25] M.P. Lamontagne, M. Provencher, Méthode de determination d´un indice d'appréciation de la qualité des eaux selon diferentes utilisations (Method for the determination of an index for the appreciation of water quality according to different uses), Ministère de Richesses Naturelles, Quebec, 1977.
[26] Ebro Hydrological Confederation, Website of the Ebro Hydrological Confederation, (n.d.). Available from: <http://www.chebro.es> (accessed June 12, 2014).
[27] J. Martínez, P.J. Garcia-Nieto, L. Alejano, A.N. Reyes, Detection of outliers in gas emissions from urban areas using functional data analysis, J. Hazard. Mater. 186 (2011) 144–149, doi: 10.1016/j.jhazmat.2010.10.091.
[28] J. Sancho, J. Martínez, J.J. Pastor, J. Taboada, J.I. Piñeiro, P.J. García-Nieto, New methodology to determine air quality in urban areas based on runs rules for functional data, Atmos. Environ. 83 (2014) 185–192, doi: 10.1016/j.atmosenv.2013.11.010.
[29] C. Díaz Muñiz, P.J. García Nieto, J.R. Alonso Fernández, J. Martínez Torres, J. Taboada, Detection of outliers in water quality monitoring samples using functional data analysis in San Esteban estuary (Northern Spain), Sci. Total Environ. 439 (2012) 54–61, doi: 10.1016/j.scitotenv.2012.08.083.
[30] J. Martínez, Á. Saavedra, P.J. García-Nieto, J.I. Piñeiro, C. Iglesias, J. Taboada, J. Sancho, J.J. Pastor, Air quality parameters outliers detection using functional data analysis in the Langreo urban area (Northern Spain), Appl. Math. Comput. 241 (2014) 1–10, doi: 10.1016/j.amc.2014.05.004.
[31] J.I. Piñeiro Di Blasi, J. Martínez Torres, P.J. García Nieto, J.R. Alonso Fernández, C. Díaz Muñiz, J. Taboada, Analysis and detection of outliers in water quality parameters from different automated monitoring stations in the Miño river basin (NW Spain), Ecol. Eng. 60 (2013) 60–66, doi: 10.1016/j.ecoleng.2013.07.054.
[32] J. Sancho, J.J. Pastor, J. Martínez, M.A. García, Evaluation of harmonic variability in electrical power systems through statistical control of quality and functional data analysis, Procedia Eng. 63 (2013) 295–302, doi: 10.1016/j.proeng.2013.08.224.
[33] K. Lekve, G. Ottersen, N.C. Stenseth, J. Gjøsæter, Length dynamics in juvenile coastal Skagerrak cod: Effects of biotic and abiotic processes, Ecology. 83 (2002) 1676–1688. Available from: <http://www.scopus.com/inward/record.url?eid=2-s2.0-0000109086&partnerID=tZOtx3y1>.
[34] J.B. Illian, J.I. Prosser, K.L. Baker, J.I. Rangel-Castro, Functional principal component data analysis: A new method for analysing microbial community fingerprints, J. Microbiol. Methods 79 (2009) 89–95, doi: 10.1016/j.mimet.2009.08.010.
[35] D. Bramwell, An introduction to statistical process control in research proteomics, J. Proteomics 95 (2013) 3–21, doi: 10.1016/j.jprot.2013.06.010.
[36] W.A. Shewhart, Economic Control of Quality of Manufactured Product, Van Nostrand Company, New York, NY, 1931.
[37] A. Duclos, N. Voirin, The p-control chart: A tool for care improvement, Int. J. Qual. Health Care 22 (2010) 402–407.
[38] J.F. MacGregor, T. Kourti, Statistical process control of multivariate processes, Control Eng. Pract. 3 (1995) 403–414.
[39] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, second ed., Springer-Verlag, New York, NY, 2005.
[40] D.C. Montgomery, G.C. Runger, Applied Statistics and Probability for Engineers, fifth ed., Wiley, New York, NY, 2010.

[41] M. Febrero, P. Galeano, W. González-Manteiga, Outlier detection in functional data by depth measures, with application to identify abnormal NO_x levels, Environmetrics 19 (2008) 331–345, doi: 10.1002/env.878.
[42] A. Cuevas, M. Febrero, R. Fraiman, On the use of the bootstrap for estimating functions with functional data, Comput. Stat. Data Anal. 51 (2006) 1063–1074, doi: 10.1016/j.csda.2005.10.012.
[43] ISO/IEC, Guide 43-1:1997. Proficiency testing by interlaboratory comparisons—Part 1: Development and operation of proficiency testing schemes, 1997.
[44] R. Acero, J. Pastor, J. Sancho, M, Ingeniería de la calidad (Quality engineering), Centro Universitario de la Defensa, Torralba, 2014.
[45] V. Amirzadeh, M. Mashinchi, A. Parchami, Construction of p-charts using degree of nonconformity, Inf. Sci. 179 (2009) 150–160, doi: 10.1016/j.ins.2008.09.010.
[46] D. Montgomery, Introduction to Statistical Quality Control, sixth ed., Wiley, New York, NY, 2009.
[47] J.D. Hamilton, Time Series Analysis, Princeton University Press, Princeton, NJ, 1994, doi: 10.2307/1270781.