Developing a wastewater treatment monitoring tool

C. Capilla, S. Barceló & J. M. Prats
Departamento de Estadística e Investigación Operativa, Universidad Politécnica de Valencia, Spain.

Abstract

In this paper we discuss and illustrate with an example the application of statistical methods to develop a monitoring tool in a wastewater treatment process. The objective is to control and monitoring the system performance using the database information of process parameters and quality characteristics, which are recorded in form of time series. This type of database is traditionally used to summarize the short or long term variability of effluent quality parameters such as suspended solids or the five-day biological oxygen demand BOD$_5$. The summaries are then used to check whether the environmental quality standards are fulfilled. The temporal dependence between the quality variables however can be exploited using statistical models, not only to keep a check on the environmental standards but also to forecast critical quality characteristics, whose analytical determination implies a temporal delay in the measurement process, e.g. BOD$_5$, and therefore to predict in advance process upsets and out-of-control situations. The application of time series models is considered in this work and the performance of statistical monitoring tools is studied by simulation of some special disturbances which may occur in the process.

1 Introduction

Wastewater treatment is a complex and dynamic process that usually operates under varying conditions and is regularly subject to checking of compliance with environmental standards. Information of these processes comes in the form of multiple time series. Traditional approaches to analyze the effluent quality and its potential environmental impact, do not model the characteristic serial correlation and the dynamic relationships among them adequately. Most published reports only contain data summaries such as histograms and statistics as the average and dispersion parameters (Berthouex, Hunter and Pallensen [1]). They frequently fail to understand the dynamic mechanisms that govern
the process performance and the interactions between process parameters and effluent characteristics, which are critical to fulfil the environmental security standards. In this paper we describe a statistical approach which integrates times series modelling and monitoring charts. Figure 1 shows a schematic drawing of the components of this approach. The procedure can be useful both to continuously analyze whether the treatment process is under control and then to check whether the effluent meets the quality standards. The former provides a stable and then predictable process performance and the later allows to establish a system to check of effluent compliance with the environmental standards. We illustrate the procedure with an example of a secondary treatment plant, characterized by measurements of suspended solids and biochemical oxygen demands.

The outline of the paper is the following: in Section 2 the process dynamics is described using a linear transfer function model. Section 3 discusses some statistical monitoring charts to be applied to magnitudes derived from the time series model in order to detect outlying situations which may be associated to process upsets. The performance of the monitoring tools is then evaluated by Monte Carlo simulation of some particular process changes that may subsequently lead to unpredictable or unacceptable effluent quality. Finally, Section 4 contains a summary and some concluding remarks.
2 Using a Transfer Function model

The serial correlation in the observations from the process can be studied using statistical univariate and multivariate dynamic models. We consider the approach in Box, Jenkins and Reinsel [2] and Berthouex and Box [3], that model the process dynamics with autoregressive-integrated moving average models (ARIMA) or linear transfer functions to analyze dynamic relationships between variables. Both types of models allows to obtain predictions in advance and then compare expected performance with observed results to early detect process anomalies or unacceptable levels.

Figure 2: Run-chart of a) BOD$_5$, SS and b) FR in a treatment process realization

Figure 2 represents a realization of the wastewater treatment process we use as an example to apply the times series models mentioned above. The variables charted are input flow rate FR (m$^3$), suspended solids SS (mg/l) and BOD$_5$ (mg/l). A logarithmic transformation has been applied to both the output BOD$_5$ and the exogenous variables. Experience and many other works with data from wastewater treatment plants (e.g.[1] and [3]) indicate that errors in the measurements use to be proportional to the variables magnitude, rather than be constant at all levels, and the logged data have approximately constant variance. On the other hand, the data are monthly averages and the results are
then useful to evaluate the performance of the process in the long-term, which
is relevant when the pollutants are bioaccumulative or cause chronic toxicity.
The delay in the BOD₅ measurements, which necessarily are not available until
almost a week after the sampling moment, implies that there is not any impor-
tant loss of information from the analysis with the monthly data.

Using univariate ARIMA models to the parameters of interest in the example
provides for each series separately understanding of their variability and predic-
tions in advance. The interactions among the variables can be analyzed however
by means of a transfer function model. This model is also useful for prediction
of BOD₅ using as exogenous variables the FR and SS, whose measurements
have no delay. A previous study (Barceló and Capilla [4]) with the same data in
Fig.2, compares the BOD₅ forecasting accuracy of its univariate ARIMA model
with that provided by the transfer function approach, and indicates that the
latter produces more accurate results in terms of the variance of the one-step
ahead forecast or mean square error (MSE). Therefore, we discuss and study in
this paper the integration of a statistical monitoring procedure when the system
dynamics is described using the latter type of model. In particular, the transfer
function model fitted to the transformed data in Figure 2 is

\[
BOD_{5t} = (\omega_1 B - \omega_2 B^2)SS_t + (\psi_0 - \psi_1 B)FR_t + (1 - \theta_1 B - \theta_2 B^2)a_t
\]  

(1)

where residual perturbations \(a_t\) are white noise, i.i.d. \(N(0,\sigma_a)\), and the output
and input variables are expressed in deviation units from their mean values.
According to 1, the BOD₅ deviation from its mean value changes when the
current and previous flow rate is modified, with a steady-state gain in terms
of the parameters of the model equal to \((\psi_0 - \psi_1)\). In relation to the other
exogenous variable, we can see that BOD₅ is related to the suspended solids
at the two previous periods \(t - 1\) and \(t - 2\), and the steady-state gain is in this
case \((\omega_1 - \omega_2)\).

The model describes the dynamic relationship between the variables under
normal process operation, i.e. when there are common causes of variability.
It allows to obtain the forecasts of BOD₅ to anticipate the delayed laboratory
results and then detect any process upset. An observed value greater or smaller
than the prediction may indicate that there is a special cause of variability
which might have produced a one-shot effect on the measurement, a temporary
change or a level shift in the quality variable. In next section we discuss some
of these possible anomalies and statistical monitoring schemes, which allow the
comparison of observed and predicted values when the common cause system is
described by means of 1, in an integrated system as described in Figure 1.

3 Statistical monitoring procedures

The differences between the forecasts and observations may indicate out-of-
control situations and unacceptable process performance. The comparison can
be done through statistical analysis of the prediction error, which under the
common system of variability is white noise \(a_t\). We consider in this section
several process disturbances that may be consequence of real problems in the process and which modify the structure of the prediction errors. The statistical procedures we consider to detect this modification are based on quality control charts (Montgomery [5]) applied to the BOD₅ prediction errors. In particular we study the performance of the Shewhart control chart and the more sophisticated CUSUM chart, both for individuals and with the design parameters adjusted to have approximately the same in control percentage of false alarms. These control charts, the run-charts and the ARIMA individual plots of the output and input variables, provide a monitoring system which may perform several roles: exploratory tools of the variability, early detection of problems and its identification when process meta-information is available.

3.1 Out of control situations

Time series data as the one we use in the wastewater treatment example, may be subject to unexpected disturbances that are often regarded as outliers. These outliers can be the consequences of unusual weather conditions, process overloading, changes and malfunction of equipment or operations. In this paper we consider several outlying situations occurring at an unexpected time t₀, which are described using the process model 1 as follows.

A one-shot effect or additive outlier:

\[ BOD₅_t = \omega(B)SS_t + \psi(B)FR_t + \theta(B)a_t + \delta_1 I_t(t₀) \]  

where \( I_t(t₀) = 1 \) for \( t = t₀ \), and \( I_t(t₀) = 0 \) otherwise (pulse function).

Temporary level change with dynamic dampening effect:

\[ BOD₅_t = \omega(B)SS_t + \psi(B)FR_t + \theta(B)a_t + \frac{\delta_1}{1 - \delta_2 B} I_t(t₀) \]  

where \( I_t(t₀) \) is a pulse function.

Level shift

Step level shift

\[ BOD₅_{t₀} = \omega(B)SS_{t₀} + \psi(B)FR_{t₀} + \theta(B)a_{t₀} + \delta_1 I_{t}(t₀) \]  

where \( I_{t}(t₀) = 1 \) for \( t \geq t₀ \), and \( I_{t}(t₀) = 0 \) otherwise (step function).

Linear trend in level shift

\[ BOD₅_{t₀} = \omega(B)SS_{t₀} + \psi(B)FR_{t₀} + \theta(B)a_{t₀} + \frac{\delta_1}{1 - B} I_{t}(t₀) \]  

where \( I_{t}(t₀) \) is a step function.

In Figure 3 we plot the application of the two control charts, Shewhart and CUSUM for individuals, to the prediction errors \( a_t \) of a process realization with a simple one-time effect of magnitude \( \delta_1 = 0.8 \) introduced at \( t₀ = 50 \) (see equation 2). In both cases the problem is clearly detected.
3.2 Performance of the monitoring procedures

The performance of the two statistical charts when applied to the BOD$_5$ prediction errors, i.e. observation minus prediction, is evaluated for early detection of the outlying situations described above by a simulation study with 100 runs. The outlying situations explained are introduced in each run after a burn-in period of 50 observations. In each realization the MSE of the BOD$_5$ and the run length, or number of observations until the chart detects the problem, are registered. The magnitudes of the perturbations depend on the parameters $\delta_1$ and $\delta_2$, as stated in equations 2 to 5. We have considered in every case the values $\delta_1 = 0.06, 0.2$ and $0.8$ and $\delta_2 = 0.06$ and $0.08$. Figure 4 shows the average value of MSE results in each simulation as a measure of the resulting dispersion in the output associated to the corresponding disturbance. The MSE under normal operation is $\sigma^2_{\text{ref}} = 0.08$. We observe that the MSE in the Figure 4 charts are clearly greater than this value, in spite of implementing a statistical chart that detects the problem a few observations after it is introduced. The dispersion in the effluent quality, which may also be due to unacceptable performance, is very affected by the perturbations studied, and this effect would be even greater if no statistical approach where used to detect the anomaly. It can be seen that the MSE increment is particularly greater when the effect is a linear trend level.
shift. In all cases the CUSUM chart is more efficient in detecting smaller perturbations than the Shewhart chart, and the resulting MSE error is lower. This is also associated to a smaller average run length as Figure 5 shows (it is reported there the average number of observations since the problem is introduced in the simulation until the chart detects it).

![Graph](image)

**Figure 4:** MSE simulation results in the outlying situations

In Figure 5, we observe that in the range of deviations studied, the CUSUM and Shewhart average run length are very similar, with the exception of the smallest $\delta_1$ value with temporary change ($\delta_2=0.08$, and the two level shift cases. In the situation of step level shift effect, the CUSUM performs worse than the Shewhart with intermediate $\delta_1$ values. When the anomaly produces a linear trend level shift, the CUSUM is clearly more efficient than the Shewhart.
Figure 5: Average run length in the simulation results of the outlying situations

4 Concluding remarks

In this paper we have considered the example of a wastewater treatment process. The purpose has been to discuss the implementation of an integrated system that uses a transfer function model to predict effluent characteristics which usually are known with delay, and which applies a statistical monitoring tool to the prediction errors. This allows early detection of process upsets and unacceptable performance according to environmental standards. The identification of the cause of the outlying situation can be done, once detected by the monitoring tool, using process meta information and with the help of run-charts of the effluent observations and flow rate. The integrated system has been evaluated by
simulation of several effects which could be associated to real process problems. The results are satisfactory in terms of the time of detection (average run length) and indicate that the monitoring tools are important to avoid further increment in output MSE or variability.

Acknowledgements

We acknowledge the process engineer M. Marcos for providing the wastewater treatment data. The authors wish to thank Universidad Politécnica de Valencia for partially funding this work with the research project 20010948.

References


