PICAZO, J.; CASTILLO, A. y VALDERRAMA, M. (1997)
"Statistical criteria applicable to quality standards for wastewater reuse"
In: "Beneficial Reuse of Water and Solids". Ed. Water Environmental Federation. Pub. nº CP3702, (13): 85-92

# STATISTICAL CRITERIA APPLICABLE TO QUALITY STANDARDS FOR WASTEWATER REUSE

Jesús Picazo
Consejería de Salud de la Junta de Andalucía
Avda. del Sur, 13. 18071 Granada
SPAIN

Antonio Castillo
C.S.I.C. e Instituto del Agua (Univ. Granada)

Mariano Valderrama
Dpto. Estadística e Investigación Operativa (Univ. Granada)

ABSTRACT

This article discusses the suitability of different statistical criteria used in quality standards for wastewater reuse. A comparative analysis of standards reveals the heterogeneity of the various statistical tools employed: the most important are the geometric and arithmetic means (measures of central position) and the percentiles. The WHO and the French supervisory body favour the use of measures of central position, while other regulatory bodies apply percentiles.

The use of percentiles, with a limit for extreme values, is proposed as the most realistic and appropriate criterion, from a basically statistical standpoint and with regard to the process of sewage-treatment engineering. By means of the chosen percentile and limit value, this criterion enables the regulation of the accuracy requirement for various uses. It also leads to a quality control feedback system, as it requires the analysis of a greater quantity of samples due to the existence of values beyond the limits. The use of the arithmetic mean excessively penalizes the customary system control deficiencies in treatment plants, particularly those serving smaller populations. The geometric mean is not intended for possible situations where the parameter analyzed presents a zero value.

Key words: Municipal wastewater; reuse; quality standards; statistical criteria

RESUMEN

En el presente artículo se discute la idoneidad de los diferentes criterios estadísticos empleados en estándares de calidad para reutilización de aguas residuales. El análisis comparativo de directrices y normativas pone de manifiesto la heterogeneidad de los estadísticos utilizados; los más importantes son las medias geométrica y aritmética (medidas de posición central), y los percentiles. La OMS y la reglamentación francesa son partidarias del empleo de medidas de posición central, mientras que otras reglamentaciones aplican percentiles.

Desde una perpectiva fundamentalmente estadística y de ingeniería de procesos de depuración, se propone como criterio más realista e idóneo el empleo de percentiles, con limite para valores por exceso. Este criterio permite, mediante el percentil y el valor límite escogido, modular bien el nivel de exigencia para diferentes usos; además favorece el autocontrol de calidad, al obligar al análisis de un razonable mayor número de muestras por la existencia de valores fuera de límite. La utilización de la media aritmética grava en exceso las habituales deficiencias de control de las estaciones depuradoras, sobre todo de pequeñas colectividades. De forma complementaria, la media geométrica no está prevista para posibles situaciones de valor cero del parámetro analizado.

Palabras clave: Aguas residuales urbanas; reutilización; estándares de calidad; criterios estadísticos

# STATISTICAL CRITERIA APPLICABLE TO QUALITY STANDARDS FOR WASTEWATER REUSE

Jesús Picazo
Consejería de Salud de la Junta de Andalucía
Avda. del Sur, 13. 18071 Granada
SPAIN

Antonio Castillo
C.S.I.C. e Instituto del Agua (Univ. Granada)

Mariano Valderrama
Dpto. Estadística e Investigación Operativa (Univ. Granada)

## INTRODUCTION

Most countries have established quality standards for municipal wastewater reuse, with the objective of reducing the corresponding sanitary risks. One of the most frequently used quality parameters is that provided by the content of coliform bacteria (total and/or fecal).

The limits established for coliform bacteria, and their corresponding numerical expression, however, vary greatly from one country to another (table 1). This is due to the variety of epidemiological, social and cultural factors to be found, in addition to the different evaluations of the potential sanitary risk arising from reuse practices.

It is common to find quality standards that establish a certain maximum value which, in principle, must not be exceeded at any time; this is the case of the guidelines proposed by France (Conseil Superieur d`Hygiene Publique de France, 1991), Saudi Arabia, Kuwait, South Africa or Japan (EPA, 1992). Another group of guidelines establishes the maximum value for a particular measure that is representative of the numerical series obtained (expresed normally for the number of coliforms/100 ml). Among the most commonly used descriptive measures are the arithmetic mean, used by the United States (EPA, 1992) and Australia (National Health and Medical Research Council and Australian Water Resources Council, 1987), where it is used in combination with the establishing of an absolute maximum value that must not be exceeded in any reading. The geometric mean is the criterion adopted by the WHO at the Engelberg Conference (World Bank and WHO, 1985; OMS, 1989 and 1990). The other frequently used descriptive measure is that of the percentiles, used by the WHO until 1984 (OMS, 1973, WHO, 1984) and by countries such as Israel (EPA, 1992).

This article will comment upon certain features of the above descriptive measures, with respect to their suitability for establishing quality standards for reuse practices.

## NUMERICAL DESCRIPTIVE MEASURES

These descriptive measures, which are commonly used to establish the maximum number of coliforms permitted by the guidelines for municipal wastewater reuse, can be classified into central, or average, measures of position (arithmetic and geometric means) and non-central position measures (percentiles); in every case, these are values that are characteristic of a numerical series.

The arithmetic mean is a numerical indicator that is very frequently used in various fields; as is well known, it is derived from the sum of the observations in the sample divided by the number of observations. It is, therefore, a "central" number within the series of numerical values under consideration.

The main advantage of this indicator is its rapid calculation and ease of interpretation. However, as a descriptive measure, it has the disadvantage of being disproportionately affected by the existence of extreme values (outliers).

The geometric mean is often used in microbiology, and is also useful for the phenomena of exponential behaviour patterns (Martín Andrés and Luna del Castillo, 1989), or in those cases where the variable presents accumulative variations. The expression to calculate it is:

$$G = \sqrt[n]{x_1 \, x_2 \ldots x_n}$$

In practice, the calculation is performed by taking logarithms of the values of the variable and then estimating the arithmetic mean of these new values; the antilogarithm of this provides us with the geometric mean.

The geometric mean has the advantage of being relatively unaffected by extreme values. It has, though, the disadvantage that its calculation and interpretation is by no means simple. Moreover, in the case of observations with a zero value, the geometric mean also assumes this value and does not reflect the remaining observations; in this case, it cannot be calculated by logarithms, as the expression log0 ($\log 0 = -\infty$) is not defined.

The percentiles ($p_1$, $p_2$,... and $p_{99}$) correspond to those values which divide the sample into 100 intervals, with n/100 observations in each one. Thus the percentile $p_i$ is defined as that value, whether belonging to the sample or not, that has to the left of it (below its value) i% of the values of the ordered numerical series. As with the geometric mean, this descriptive measure is very little affected by the existence of extreme values. Like the arithmetic mean, percentiles are very simple to interpret. Nevertheless, they are not easy to calculate, though in the practical use of quality standards, the percentile value is fixed, and so does not have to be calculated.

## DISCUSSION

There is normally a very wide range of values for coliforms (total or fecal) in reclaimed municipal wastewater, even within the effluent from the same treatment plant. In practice, it is frequent to find variations from $10^1$ to $10^6$ fcu/100 ml. This variability is due to various factors, among which are the concentration in the input water and the variable performance, in time, of the treatment plants.

To guarantee an acceptable degree of variability, depending on the quality demanded of the effluent for the reuse practice in question, there are, as commented on above, various concentration limits, related or not to different numerical descriptive measures.

The adoption of a limit value that must not be exceeded under any circumstances (absolute value) is a strict and, in general, somewhat unrealistic criterion, which requires the functioning of treatment installations and processes to be both excellent and very reliable. It seems, therefore, more reasonable to relate such limit values to a numerical indicator (arithmetic mean, geometric mean or percentile) that is characteristic of the numerical series obtained.

As mentioned above, the arithmetic mean is exaggerated by the existence of an abnormally high value for coliform bacteria, within the series of values obtained. In fact, this frequently occurs, and thus it is difficult to comply with moderate limits for the quantity of colis measured when there exist spikes in the reclaimed wastewater effluents.

The geometric mean is much less affected by such abnormally high values. The choice of this as a measure to establish a limit for fecal coliform content is in agreement with the most recent criterion published by the WHO of not setting unjustifiably restrictive quality standards (WHO, 1984, OMS, 1990). However, as this criterion is included within a legal norm, it becomes necessary to note that coliform values equal to zero will not be taken into consideration, as this would lead to the geometric mean also having a value of zero and being insensitive, from then on, to the remaining data in the series, irrespective of the values obtained. It should be understood that it would be possible, as a deliberate and exceptional action, to completely eliminate coliforms from treatment plants by means of super-chlorination.

The adoption of limit values for a particular percentile of the series (e.g., 90% of the samples <10,000 col/100 ml) would avoid the problem of occasional extreme values. However, and in contrast to the geometric mean, a greater degree of reliability is required in treatment processes, because when the number of values is greater than the limit imposed (10% of the measures, in the above example), the limit established for the use in question would not be complied with. Percentiles offer, on the other hand, the advantage that under adverse conditions (when several data values lie outside the range) the situation can be remedied by increasing the number of analyses made. This would be, furthermore, a way of encouraging autoregulation after periods of incorrect functioning of the plant.

## CONCLUSIONS

Taking into account the normal distribution of the values for coliforms in treated wastewater, percentiles were decided to be the most appropriate numerical indicator of the maximum permissible value in quality standards for reuse practices. This criterion provides a good degree of control, by means of the chosen percentile and the value attributed to it, of the quality level demanded for different reuse practices. At the same time, it allows continuous regulation by the user or manager of the wastewater: the measurements effected mean that, at all times, the quality of the wastewater, and whether or not it lies within the margins established by the norm, are known. Thus the uncertainty is avoided that would otherwise be produced by the calculation of the arithmetic or geometric mean, which are only known when the complete series of measurements has been obtained. The adoption of absolute maximum limits, as a complementary form to the percentile, is a possibility that could be employed for applications of greater potential sanitary risk.

The characteristics of the geometric mean are fairly similar to those of the percentiles (especially when absolute maximum limits are applied), and the use of this measure is also to be recommended. However, the zero values within a series (which could easily be obtained by super-chlorination) would have to be discounted, in order to prevent a possible subterfuge intended to produce a geometric mean with a zero value.

The arithmetic mean is not suitable for the objectives sought by quality standards, as it is excessively influenced by high values (spikes), following which it is almost impossible to return to the set limits. Similarly, absolute maximum values are also inappropriate, as complying with these normally demands the exceptionally effective functioning of treatment plants, which is very difficult and highly expensive to achieve, above all in non-separative treatment systems (where rainfall and wastewater are combined).

## REFERENCES

Conseil Superieur d'Hygiène Publique de France, (1991) *Recommandations sanitares concernant l'utilisation, aprés épuration, des eaux résiduaries urbaines pour l'irrigation des cultures et des espaces verts*. Ministère Charge de la Santé. République Française. 40 p.

EPA, (1992) *Guidelines for Water Reuse*. U.S. Environmental Protection Agency. U.S. for International Development. 247 p. Washington, D.C.

Martín Andrés, A. y Luna del Castillo, J. de D., (1989) *Bioestadística para las Ciencias de la Salud*. 2ª ed. Norma S.A. Madrid. 614 p.

National Health and Medical Research Council and Australian Water Resources Council, (1987) *Guidelines for use of reclaimed water in Australia*. Australian Government Publishing Service. Canberra. 12 p.

OMS, (1973) *Aprovechamiento de efluentes: métodos y medidas de protección sanitaria en el tratamiento de aguas servidas*. Informe de una reunión de expertos de la OMS. Serie de Informes Técnicos, nº 517.

OMS, (1989) *Directrices sanitarias sobre el uso de aguas residuales en agricultura y acuicultura*. Informe de un Grupo Científico de la OMS. Serie Informes Técnicos nº 778.

OMS, (1990) *Directrices para el uso sin riesgos de aguas residuales y excretas en agricultura y acuicultura*. Medidas de protección de la salud pública. Mara, D. y Cairnscross, S. Ginebra. 214 p.

World Bank and WHO, (1985) *The Engelberg report. Health aspects of wastewater and excreta use in agriculture and aquaculture*. International Refeerence Centre for Waste Disposal (IRCWD), Dubendorf.

WHO, (1984) *Health aspects of reuses of treated wastewater for irrigation* (background document). World Health Organization. Intercountry seminar on wastewater reuse. Manama. 51 p.


| **REFERENCES** | **PERCENTILE** | **ARITHMETIC MEAN** | **GEOMETRIC MEAN** | **ABSOLUTE VALUE** |
|---|---|---|---|---|
| OMS, 1973 | $P_{80}$<100 FC | | | |
| WHO, 1984 | | | < 1.000 FC | |
| WB-WHO, 1985 | | | < 1.000 FC<br>< 10.000 FC | |
| CSHPF, 1991 | | | | < 200 FC<br>< 1000 FC |
| EPA, 1992 | | 0 to 200 FC | | < 14 FC<br>< 800 FC |
| SAUDI ARABIA | | | | < 2,2 TC |
| ISRAEL | $P_{80}$ < 12 TC<br>$P_{50}$ < 2,2 TC | | | |
| AUSTRALIA | | < 1.000 FC | | < 10 FC |
| JAPAN | | | | < 50 TC<br>< 1.000 TC |
| KUWAIT | | | | < 100 CT<br>< 10.000 CT |
| SOUTH AFRICA | | | | 0 TC<br>< 1.000 TC |
| BALEARIC IS. | | | < 1.000 FC | |
| ANDALUSIA | $P_{90}$ < 200 to 10.000 FC | | | |

*Table 1. Different limits for total or fecal coliform levels (nº / 100 ml) established for the irrigation of unprocessed crops*