Geostatistics – a tool applied to the distribution of *Legionella pneumophila* in a hospital water system

Pasqualina Laganà¹, Umberto Moscato², Andrea Poscia², Daniele Ignazio La Milia², Stefania Boccia², Emanuela Avventuroso², Santi Delia¹

¹ Department of Biomedical and Dental Sciences and Morphofunctional Imaging. Section of Biotechnologies and Medicine Preventive. University of Messina, Italy
² Department of Public Health – Università Cattolica del Sacro Cuore – Rome, Italy

Abstract

**Introduction.** Legionnaires’ disease is normally acquired by inhalation of legionellae from a contaminated environmental source. Water systems of large buildings, such as hospitals, are often contaminated with legionellae and therefore represent a potential hazard for the hospital population. The aim of this study was to evaluate the potential contamination of *Legionella pneumophila* (LP) in a large hospital in Italy through georeferential statistical analysis to assess the possible sources of dispersion and, consequently, the risk of exposure for both health care staff and patients.

**Materials and Method.** LP serogroups 1 and 2–14 distribution was considered in the wards housed on two consecutive floors of the hospital building. On the basis of information provided by 53 bacteriological analysis, a ‘random’ grid of points was chosen and spatial geostatistics or FAIk Kriging was applied and compared with the results of classical statistical analysis.

**Results.** Over 50% of the examined samples were positive for *Legionella pneumophila*. LP 1 was isolated in 69% of samples from the ground floor and in 60% of sample from the first floor; LP 2–14 in 36% of sample from the ground floor and 24% from the first. The iso-estimation maps show clearly the most contaminated pipe and the difference in the diffusion of the different *L. pneumophila* serogroups.

**Conclusion.** Experimental work has demonstrated that geostatistical methods applied to the microbiological analysis of water matrices allows a better modeling of the phenomenon under study, a greater potential for risk management and a greater choice of methods of prevention and environmental recovery to be put in place with respect to the classical statistical analysis.

**Key words**

*Legionella pneumophila*; hospital; water systems; Geostatistics; Kriging

INTRODUCTION

Legionnaires’ disease is normally acquired by inhalation or aspiration of legionellae from a contaminated environmental source. Water systems of large buildings, such as hospitals, are often contaminated with legionellae and therefore represent a potential hazard to patients and health care staff [1, 2, 3, 4, 5, 6]. Several papers have shown a clear association between the presence of legionellae in hot water systems and the occurrence of legionellosis [7, 8, 9, 10, 11]. For risk evaluation of nosocomial *Legionella* infection, surveillance of hospital water systems is needed with traditional culture techniques or molecular typing methods [12, 13, 14, 15, 16].

In hospital as well as other health care facilities, waterborne diseases may originate from bacterial colonization of water pipes, cooling towers, spa pools, taps, showers, and water supplies [17, 18, 19, 20, 21, 22]. Several reports have concerned the epidemiological surveillance of pathogenic bacteria in Italian hospitals [23, 24, 25, 26, 27, 28]. Istituto Superiore di Sanità (ISS, National Institute of Health), reported that during 2011 in Italy nosocomial cases of legionellosis were 65 (6.4%) of total cases reported, of which 33 (50.8%) were of certain nosocomial origin, and 32 (49.2%) possibly of nosocomial origin [29].

Water systems represent suitable environments for the growth and multiplication of *Legionella* spp., Gram-negative bacteria which survive at different pHs and temperatures [30, 31, 32, 33, 34, 35]. This is encouraged by water temperatures below 50°C, areas of stagnation and sludge formation, the presence of amoebae and other bacteria [36], and the materials used in the piping.

In hospitals, the problem of Legionellosis should be tackled and managed with the methodology of Risk Analysis, the logic of which suggests that when it is not possible to eliminate the hazard, it is necessary to rely on risk control. For *Legionella* this results into control of three basic parameters: (1) the amount of the germ in the water, through analytical monitoring; (2) the presence of virulence factors in isolated strains; (3) the receptivity in the guests.

Being able to estimate the distribution of *Legionella* in a healthcare facility allows extension of the remediation areas also not monitored but adjacent to premises controlled by elimination, or reducing, thus, the risk of infection in patients, health workers and in visitors.

This study evaluates the distribution of *Legionella pneumophila* serogroup 1 and serogroups 2–14 in some
hospital wards to assess the risk of exposure for both health care staff and patients. Also, it was evaluated the dispersion in the air in relation to the critical points identified, and hypotheses formulated for the determination of possible sources of dispersion according to the water supply facilities.

Geostatistics was applied in the study. It is a discipline that deals with evaluating the spatial autocorrelation of the data, attempting to verify whether observations carried out on neighbouring points actually have a lower variability compared to observations made between distant points. The objective, therefore, is to evaluate the effect of the position of the measuring point on the variability of the data observed. Apart from evaluating the spatial variability, geostatistical features of spatial interpolation techniques that can provide estimates on the value of a variable in a position where the measurement was not made on the basis of the detected data of neighbouring points.

**MATERIALS AND METHOD**

**Sampling.** The water distribution system of the ground and first floors of pavilion E of the 'G. Martino' University Hospital in Messina, Italy, was examined for *Legionella pneumophila* 1 and *L. pneumophila* 2–14. Incoming cold groundwater is provided by the municipality and disinfected with chlorine dioxide, the water reaches the hospital by means of one pipeline, which leads to a centralized tank where the water is stored. The water does not undergo further chlorination after it is collected from the Messina town pipeline. From the centralized tank, the cold water is distributed to each building by electric-motor pumps which send it through a pipeline that runs across the basements of all the buildings. Under each building there is a boiler which produces heated water (average temperature of approximately 13–48 °C) that rises again to supply the wards located at each floor.

Samples of the heated water were collected after flushing for 5–10 min, and sterilization of the water tap, at the start of daily activities from taps using 1 L-sterile glass bottles. Care was taken to take samples both from floors and wards located both on the left and right sides inside each building to obtain a representative sampling of the hygienic-sanitary conditions.

**Isolation and identification of *Legionella* spp.** To recover *Legionella* spp. from water samples the standard procedures reported in the Italian Guidelines for the prevention and control of legionellosis (Gazzetta Ufficiale della Repubblica Italiana, No.103, 5 May, 2000) were used. 1 L water samples were concentrated to 10 mL through 0.2 μm porosity membrane filters, and incubated at 50 °C for 30 min in a thermostatic bath. Concentrated and unconcentrated samples were spread on duplicate plates of Buffered Charcoal-Yeast Extract (BCYE) Agar Base Medium (Oxoid, Ltd., Milan, Italy), incubated for 10 days at 36–37 °C in a moist chamber with 2.5% CO₂. The suspected colonies were isolated and confirmed as *Legionella* spp. after screening their inability to grow on a culture medium without cysteine. *Legionella* spp. counts were reported in colony forming units/litre (CFU/L) according to the number of colonies per plate and to the dilutions performed on the original sample. The isolates were further identified as *L. pneumophila* serogroup 1, *L. pneumophila* serogroups 2–14, or *Legionella* spp using the microagglutination Legionella Latex Test Kit (Oxoid).

**Statistics.** The data obtained were analyzed by means of classical statistics by Stata/IC 12.1 for Mac. Data distribution was evaluated according to Shapiro-Wilk test. Differences in distribution of *Legionella* (by floors, or according to the serogroups) have been evaluated by rank-sum test for independent samples, while the correlation between time of isolation and CFU was tested by the rho Spearman coefficient. A P value 0.05 was considered statistically significant.

**Geostatistics.** Sampling points were chosen with a ‘random’ system according to elaborations of geostatistical models. Geostatistics offers a wide variety of useful methods to model the processes of pollutants diffusion as a result of random (aleatory) functions [37]. In fact, the geostatistical methods analyze the behaviour of microorganisms (analysis variables), their spatial correlations, and draw the spatial structure of their variability in order to estimate the current values that are not directly measured (biological, physical or chemical) in different matrices (air, water or soil), or in order to predict the expected future levels of pollutants in the matrices themselves [38, 39].

The directional variograms are processed according to the results of the sampling technique called ‘Croix de Sondage’ and the variogram models are proposed as spatial structures of the variability in particle sampled areas [40]. Placing the sampling points at regular distances so as to form a cross – the croix de sondage – which allowed development of the spacing between sampling points, defined as neighborhood areas, in which the phenomenon could be described as not yet ‘independent’.

A ‘random’ grid of points was chosen, then, according to the distance between points thus calculated (spatial geostatistics or FAIk Kriging) of variogram models previously processed, so that the points were included in a spacing between 4 and 7 m with an over estimate of 10 meters. This choice was made to ensure the most accurate and, at the same time, prudential estimation [41].

The parameters of this model consist in the distance beyond which the measured values should be considered to be independent from the observed phenomenon (defined range), and the variance between independent samples (defined sill). In addition, the spatial structure of variability that differs from the observed values, is defined geometrically anisotropic (otherwise assume isotropic values). Finally, the expected values of the sampling grid nodes were processed through the software Faipack (DOS version, Release 1.00 – X.93), an automatic software tool for performing linear geostatistical analysis of spatial data [38, 39] in relation to an interpolation bi/three-dimensional. This allows the building of a continuous function random on the plane z=f(x,y), based on the Z values of the microorganisms sampling performed, fixing the sampling time (t) as a function of the algorithm lim. →∞, in order that the data interpolation can be performed as a random function of the only inferential spatial relationship [41, 42, 43, 44, 45]. Maps of estimated values are thus generated which are useful for describing the areas of *Legionella* spread contamination in the hospital water pipe system.
RESULTS

Over 50% of the 53 examined samples were *Legionella* positive. *L. pneumophila* 1 were isolated in 69% of samples from the ground floor and in 60% of samples from the first floor; *L. pneumophila* 2–14 in 36% of samples from the ground floor and 24% from the first one (Tab. 1, 2).

**Table 1.** Proportion of *Legionella* 1 and 2–14 isolated from the samples drawn from the water distribution system of ground and first floor of the pavilions E

<table>
<thead>
<tr>
<th>Legionella</th>
<th>Proportion</th>
<th>Std. Err.</th>
<th>[CI 95%]</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. pneumophila</em> total</td>
<td>79.20%</td>
<td>5.60%</td>
<td>70.20–88.20%</td>
</tr>
<tr>
<td><em>L. pneumophila</em> 1</td>
<td>62.20%</td>
<td>6.70%</td>
<td>58.80–65.60%</td>
</tr>
<tr>
<td><em>L. pneumophila</em> 2–14</td>
<td>24.50%</td>
<td>9.50%</td>
<td>16.50–32.50%</td>
</tr>
</tbody>
</table>

Despite the high level of *L. pneumophila* 1 and 2–14 found in the water distribution system of the pavilion, no cases of Legionnaires’ disease had been reported among patients or among health care workers. Table 3 shows in detail the wards and sampling sites, and min-max of serogroups isolated.

The mean concentration of *Legionella* did not show significant differences between floors (p>0.05). Also, the different serogroups of *L. pneumophila* (1 vs. 2–14) did not show any significant difference in the distribution on the same floor, even if on the first floor there was a higher CFU of *L. pneumophila* 2–14 than on *L. pneumophila* 1.

Analysis of the trend of Legionella CFU in different years showed a significant increase in the concentrations of *Legionella* in the latest samples (rho = 0.31; p<0.01), with the maximum CFU in the samples from two years ago (200.000 CFU) due to *Legionella* 2–14 isolation (rho 0.25; p>0.05) (Fig. 1).

It is noteworthy that the main structure of the model of the parameters’ spatial variability is the same for all the measurements carried out, and that the proposed values of range and anisotropy show small differences in their variability, where the value expressed by the sill appears to be more dispersed around the central value of reference. However, the *Indicative Goodness of Fit* (IGF), which gives a measure of how well the model ‘adjusts’ to the directional variograms, lies almost in all models in the vicinity of its optimal value equivalent to 0.

Observe the difficulty to ‘filter’ the trends estimated and applying the theories of universal kriking, the use of the FAI-k has allowed, in this study, the generation of models of Covariance Generalized of the order K-type variable from 0 – 1 (non-stationary models) (Tab. 5), with jackknife chosen according to the following calculation algorithm:

\[
\gamma(h) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij}(h) - \frac{1}{n} \sum_{i=1}^{n} \gamma_{ii} - \frac{1}{n} \sum_{j=1}^{n} \gamma_{jj} + \frac{1}{n} \sum_{i=1}^{n} \gamma_{ii} \sum_{j=1}^{n} \gamma_{jj} \]

To facilitate understanding of the Geostatistical results obtained, Figures 2 and 3 are concordant estimation maps with their geostatistical parameters of linear kriking evaluation of sampling models.

**Table 3.** Floor, ward, sampling sites, serogroup isolates and percentage of *Legionella* isolated from the water distribution system

<table>
<thead>
<tr>
<th>Floor</th>
<th>Ward</th>
<th>Type of site sampling</th>
<th>Serogroup (C.F.U. min-max)</th>
<th>Pos</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Anesthesia and Intensive Care Unit;</td>
<td>laboratory; room nurses; kitchen; surgery; toilets</td>
<td><em>Legionella pneumophila</em> 1</td>
<td>69%</td>
</tr>
<tr>
<td></td>
<td>First Aid; Pharmacy; Radiology; Nuclear Medicine</td>
<td></td>
<td>(25–40000e+1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>Legionella pneumophila</em> 2–14</td>
<td>25%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(300–50000e+1)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Anesthesia and Intensive Care Unit;</td>
<td>haematological room; hospital room; laboratory; medical room; nurses room; operating room; kitchen; sterilization room; storeroom; surgery; toilets</td>
<td><em>Legionella pneumophila</em> 1</td>
<td>59.5%</td>
</tr>
<tr>
<td></td>
<td>Hematology; Transfusional Medicine</td>
<td></td>
<td>(100–100000e+1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>Legionella pneumophila</em> 2–14</td>
<td>24.3%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(100–100000e+1)</td>
<td></td>
</tr>
</tbody>
</table>
DISCUSSION

Although classical statistics is very helpful to organize and interpret data from the samples drawn from the water pipe distribution system, it would be appropriate to combine such statistics with other tools in order to better understand the way of colonization of *Legionella* and, therefore, to highlight the situations that require more thorough investigation. For this reason, combining the methods of classical statistical analysis with the techniques of georeferencing analysis may be of particular use to understand the source of the contamination, because it is possible to put together the information about the maintenance of the pipe system (i.e. restructuring interventions, changes in materials over time) and the results of routine monitoring activity.

In particular in this study, statistical descriptive classical analysis, while showing differentiated values of the *Legionella*’s concentration at different water sampling points, even in the same hospital wards and on the same floors, represents an alert index useful for intervention on a water plant, instead of cannot ‘estimate’ the probability of risk of the spread of the contamination in the environment and, ultimately, risk to the patient. It also allows a limited range of choices for the management and prevention of the risk, not allowing estimation of the possible real source of bacterial contamination in the water plant and, consequently, implementing the choices for more effective prevention.

In this sense, the plans analysis of departments and water sampling points, overlaid with maps of geostatistical iso-estimation, is used to assess the degree of risk at the point of sampling and in its surroundings, both risk assessments applied to the closed circles. In the presented study, the level of risk related to *Legionella pneumophila* 1 and 2–14 apparently seems low for the hospital departments assessed, and analysis of the maps would lead to the implementation, as well as the known preventive measures (de-calcification, sanitization with chlorine agents, etc.), a more accurate assessment of the state of the water system, and consider replacing some of its components (valves, end portion of the water pipe system, sanitation in recirculation water pipes in the interested

| Table 5. Models of *Legionella* spatial diffusion Generalized Covariance |
|---------------------------------|-----------------|----------------|-------------|-------------|-----------------|-----------|-----------------|-------------|
| Floor 0 | 1.625 | 1.0E+8 | 0.000 | -2.936E+06 | 0.000 | 0.000 | 0.944519 | 1.0E+8 | 0.000 | -2.936E+06 | 0.000 | 0.000 | 0.944519 |
| Floor 1 | 2.028 | 1.1E+8 | 0.000 | -6.083E+06 | 0.000 | 0.000 | 1.160316 | 1.1E+8 | 0.000 | -6.083E+06 | 0.000 | 0.000 | 1.160316 |
| Floor 0 | 1.750 | 1.9E+8 | 0.000 | 3.5705E+03 | 0.000 | 0.000 | 1.356938 |
| Floor 1 | 1.861 | 5.2E+8 | 0.000 | -1.342E+07 | 0.000 | 0.000 | 1.328014 |

Figure 2 (a). Map showing the estimation of distribution of LP 1 on the ground floor using geostatistical analysis. The various concentrations of C.F.U. isolated: each gradient of grey refers to different range. The sections in the circle show the Nuclear Medicine and Radiology wards, corresponding to the highest levels of *L. pneumophila* 1.

Figure 2 (b). Estimation of the distribution of LP 1 on the first floor. The sections in the circle shows the Haematology and Transfusional Medicine wards, corresponding to the highest levels of *L. pneumophila* 1. There is a remarkable correlation between the ground floor and the first floor.

Figure 3 (a). Estimation of the distribution of LP 2-14 on the ground floor using geostatistical analysis. Various concentrations of C.F.U. isolated: each gradient of grey refers to a different range. Sections in the white circle show the First Aid, Anaesthetics and Intensive Care wards, corresponding to the highest levels of *L. pneumophila* 2-14.

Figure 3 (b). Estimation of the distribution of LP 2-14 on the first floor. Sections in the black circle show the First Aid ward, corresponding to the highest levels of *L. pneumophila* 2-14. There is a remarkable correlation between the ground floor and first floor.
departments concerned, etc.), as well as increase the level of active epidemiological surveillance on patients and staff of the departments concerned.

The good performance of the IGF, the K-order always in a range between 0 – 1 (anisotropic Gaussian model) and the Jackknife around 1 for the majority of the analysis performed, indicate values of an accurate estimate of the model with respect to real sampling, with an excellent reliability interpolation of the neighbourhood areas on octants selected for the graphics processing maps of iso-estimate.

The advantages of the study are represented, therefore, by the demonstration that geostatistical methods applied to microbiological analysis of water matrices allows a better modeling of the phenomenon under study, a greater potential for risk management and a greater choice of methods of prevention and recovery of the environment to be put in place with respect to the classical statistical analysis alone, which, moreover, allows evaluation of the sampling point, but not to estimate the surrounding areas. An even smaller number of samples useful for determining the model applicability in all hospital scenarios, although the statistical analyses are reliable and indicate a high level of significance, the small number of sample units useful for determining the model applicability still represents a limit of the presented study. Moreover, the projections estimation for areas far from sampling points, based on the number of available samples, do not appear to be more than 96%, which renders the study an significant object to the limit value of ≤ 0.05, but no higher. The speculative hypotheses compared in this study are represented by the implementation of the samples, with analysis of all floors of the building in relation to the spatial distribution of the water, studying the correlation between water pipes and Legionella through an assessment of spreading, and, finally, not in the air of hospital environments but within the water system, in order to identify the sources of risk and the spread of Legionella ‘reservoirs’.

Short Geostatistical Glossary

Anisotropic/Isotropic: anisotropic is the property of being directionally dependent, as opposed to isotropy, which implies identical properties in all directions.

Geostatistics: a discipline introduced in the 1960s by Matheron in the geo-mining ambit, in order to allow a content estimation of particular minerals in soils from a few samples distributed in space. From the 1990s – 2000 onwards the geostatistics has been applied to environmental parameters, to assess the probability of spread in dynamic complex matrices of the pollutants or of the microorganisms.

Interpolation: a method of constructing new data points within the range of a discrete set of known data points.

Jackknife: a re-sampling technique especially useful for variance and bias estimation.

Kriging: a geostatistical technique that allows interpolation of the value of a magnitude considering the values that it assumes in the near points.

Variogram: a basic tool by which the linear geostatistical studies andcharacterizes the spatial variability.

REFERENCES


