Understanding the relationships between sanitation and health in Nicaragua and Honduras, through data mining tools

Ginevra Marina Lazerini  
*University Research Institute for Sustainability Science And Technology, Universitat Politècnica de Catalunya – Barcelona Tech*

Josep Nualart  
*University Research Institute for Sustainability Science And Technology, Universitat Politècnica de Catalunya – Barcelona Tech*

Sergio Ruiz-Cayuela  
*University Research Institute for Sustainability Science And Technology, Universitat Politècnica de Catalunya – Barcelona Tech*

Maialen Urbina  
*University Research Institute for Sustainability Science And Technology, Universitat Politècnica de Catalunya – Barcelona Tech*

Miquel Sànchez Marrè  
*Universitat Politècnica de Catalunya-BarcelonaTech*

Follow this and additional works at: [https://scholarsarchive.byu.edu/iemssconference](https://scholarsarchive.byu.edu/iemssconference)

Part of the Civil Engineering Commons, Data Storage Systems Commons, Environmental Engineering Commons, Hydraulic Engineering Commons, and the Other Civil and Environmental Engineering Commons

Lazerini, Ginevra Marina; Nualart, Josep; Ruiz-Cayuela, Sergio; Urbina, Maialen; Marrè, Miquel Sànchez; and Gibert, Karina, "Understanding the relationships between sanitation and health in Nicaragua and Honduras, through data mining tools" (2016). *International Congress on Environmental Modelling and Software*. 39.  
[https://scholarsarchive.byu.edu/iemssconference/2016/Stream-C/39](https://scholarsarchive.byu.edu/iemssconference/2016/Stream-C/39)

This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
Understanding the relationships between sanitation and health in Nicaragua and Honduras, through data mining tools

Ginevra Marina Lazerini\textsuperscript{(a)}, Josep Nualart\textsuperscript{(a)}, Sergio Ruiz-Cayuela\textsuperscript{(a)}, Maialen Urbina\textsuperscript{(a)}, Miquel Sànchez Marrè\textsuperscript{(b,d)}, Karina Gibert\textsuperscript{(c,d)}, karina.gibert@upc.edu
\textsuperscript{(a)} University Research Institute for Sustainability Science And Technology, Universitat Politècnica de Catalunya – Barcelona Tech, Campus Nord, Ed TG, Jordi Girona, 31, 08034, Barcelona.
\textsuperscript{(b)} Computational Sciences Department, Universitat Politècnica de Catalunya-BarcelonaTech, Campus Nord, Ed. Omega. Jordi Girona 1-3, 08034 Barcelona, Catalonia
\textsuperscript{(c)} Statistics and Operations Research Department, Universitat Politecnica de Catalunya – Barcelona Tech, Campus Nord, Ed C5. Jordi Girona 1-3, 08034 Barcelona, Spain
\textsuperscript{(d)} Knowledge Engineering and Machine Learning Group, Universitat Politecnica de Catalunya – Barcelona Tech, Campus Nord, Ed Omega. Jordi Girona 1-3, 08034 Barcelona, Spain

Abstract: The aim of this work is to analyze water and sanitation supply data from Nicaragua and Honduras by using different data mining tools. The data has been provided by SIASAR (Rural Water and Sanitation Information System), which is a water and sanitation management and information platform created through the joint effort of different Central American Governments and the World Bank. In the study data from a survey performed in all the rural communities in Nicaragua and in a sample of the rural communities in Honduras from 2012 to 2015 is analyzed. Database contains 10206 communities described by 23 numerical variables and 16 qualitative related to geography, demography, health, sanitation and access to water. A data mining approach is used to evaluate water and sanitation supply in rural communities in both countries, to later compare the results with the different political, social and environmental conditions. Global relationships between access to water, health implications, and policies is extracted for each of the countries. Whereas rural communities evaluated in Honduras are a part of the existing in all the country, in Nicaragua all the rural communities in the country were evaluated. Clustering using Ward’s method and Gower’s distance (compatible with both numerical and qualitative variables) shows three profiles of communities of increasing level of sanitation, health and water access supply: a) 3580 communities, mainly from Honduras, characterized by a systematic use of latrines and safe water, systematic hand washing, good level of hygiene and relatively high accessibility to water by gravity systems and to sanitation by hydraulic discharge. b) 2993 with little open air defecation, and little presence of rubbish and puddles; relatively high use of wells and pumps as a water source, and of latrines without cement floor for sanitation. c) 3633 communities mainly from Nicaragua, characterized by a high level of open air defecation, a significant presence of puddles and rubbish, high use of unsafe water, an unhealthy atmosphere and low level of hygiene; high use of wells and pumps as a water source, and of latrines without cement floor for sanitation. The SIASAR classification of communities (from A to D according to decreasing performance) is strongly associated with the obtained profiles ($\chi^2$p-value<0.05)

Keywords: preprocessing, data mining, postprocessing, sanitation infrastructures