Correlation Analysis between Illiteracy and Flooding Disasters in Cundinamarca, Colombia

Lozano-Rivas, W. A., Moyano Fonseca, D. and González Acero, A. M.

***Abstract* — This study summarizes the results obtained from a correlational analysis of reported disasters in 40 municipalities of the River Bogota's basin in Colombia (South America). Environmental, social, and economic variables were selected, for which the strength of association was defined with the use of a multiple linear regression model. The ability to predict the occurrence of flood disasters depends mostly on the Population Density and Total Illiteracy Rate, with a probability of error of P<0.001 (0.1%) and correlation coefficient R=0.738. This study concludes that for the basin of River Bogota there could be a significant mathematical correlation between the total illiteracy rate and the number of reported flood disasters.**

***Index Term* — floods; illiteracy; multiple correlation; natural disaster; poverty.**

I. Introduction

The Department of Cundinamarca of the Andean region, where the Bogotá River basin is located, is one of the 32 territorial divisions of Colombia. Cundinamarca has several climates, which accounts for the diversity in land use and the large variety of agricultural produce. The number and magnitude floods and landslides affecting the Department of Cundinamarca in the harsh rainy seasons of 2010 and 2011 confirmed that, in spite of the institutional efforts of both departmental and national authorities to reduce the number of disasters, they are still occurring frequently. Between 2010 and 2011 in Cundinamarca alone, some 458 emergency situations due to floods, landslides, gales or hailstorms were reported [1].

The Bogotá River in consists of 19 sub-basins and covers a surface area of 589,143 hectares, which corresponds to 32% of the department's total surface as shown in Fig. 1. The basin spans the source of the river in the municipality of Villa Pinzón in the northeast of the department (located 3300 meters above sea level) to the southwestern municipality of Girardot, where the river meets the Magdalena River (at 280 meters above sea level). Along its 308 km path, the Bogota River is affected by agricultural and industrial activities and it receives the outflow of non-treated sewage water of some 9 million inhabitants, including the 7 million population of the country's capital city Bogota D.C. [2].

Economically, Cundinamarca plays a highly significant role in the country, given that its Gross Domestic Product (GDP) is 5.23%, which places it as the fifth regional economy after Bogotá D.C., and the departments of Antioquia, Valle and Santander. If one includes Bogotá D.C., which is located in Cundinamarca but holds the status of "special district", the department as a whole accounts for more than 30% of the national Gross Domestic Product [3]. In spite of this economic power, 2004 estimates indicate that 54% of its population is poor and more than 18% extremely poor (indigent) [4].

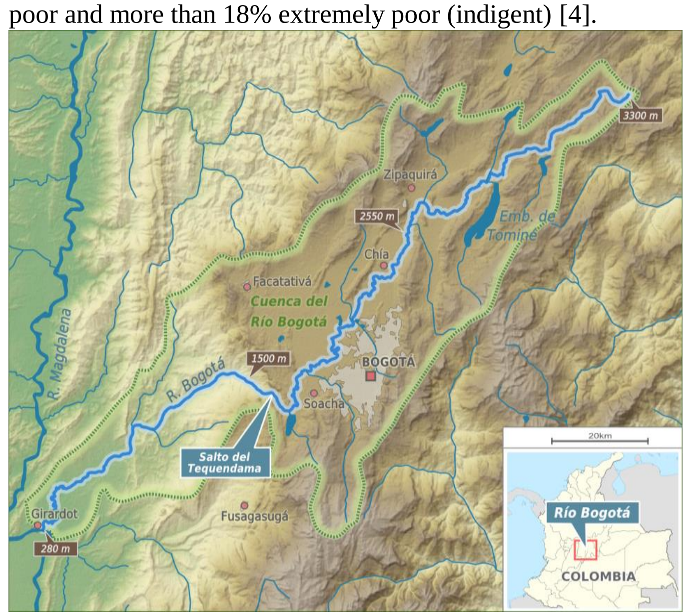

Fig. 1. Bogota River basin
(VMAP0. SRTM3 USGS (2004), Shuttle Radar Topography Mission, 3 Arc Second, Filled Finished, University of Maryland (Global Land Cover Facility, www.landcover.org))

In addition, the main economic activity in the Department of Cundinamarca is agriculture, and most municipalities have low rates of industrial and technical development. Although poverty is far from being defined [5], it may be described as "a low standard of living because of insufficient resources to

This study was financed by the Piloto of Colombia University within the framework of the project "Assessment of Environmental Services associated with the Water Ecosystems of Bogotá D.C. and their impact on Human Wellbeing". It was carried out by the Environmental and Sustainability Research Group.

William A. Lozano-Rivas, MSc, PhD; is a Full Professor in the School of Environmental Sciences of the Piloto of Colombia University. 110231, Bogotá D.C., Colombia. Phone: (+57)(1)3322976; (+57)3105850427 (e-mail: wlozanorivas@gmail.com).
David Moyano Fonseca is an Urban Environmental Management Graduate in the School of Environmental Sciences of the Piloto of Colombia University. Bogotá D.C., Colombia (e-mail: david-moyano@unipiloto.edu.co).
Ana M. González Acero is a Specialist in Management and Adjunct Professor in the School of Environmental Sciences of the Piloto of Colombia University. Bogotá D.C., Colombia (e-mail: ana-gonzalez1@unipiloto.edu.co).

avoid the deprivation," a definition which suggests a correlation between resources, infrastructure and services [6]. Resource scarcity in turn leads marginal communities to occupy risky areas, where they often have no access to healthcare, drinking water or basic sanitation, which then makes them even more vulnerable to natural disaster [5,7].

Furthermore, countries and regions that depend heavily on the primary sector are more likely to experience the effects of natural disasters, particularly those of hydro-meteorological origin [8,9]. Eighty percent of the poor in Latin America, 60% of the poor in Asia, and 50% of the poor in Africa live in low-productivity marginal zones [10]. It has also been proven that disasters are a source of and a contributing factor to poverty [11], inasmuch as poverty can be defined as a decrease in the ability to respond to adverse situations [12].

A. Disasters in Colombia, perception and reality

A common perception of natural disaster is that it is an event (or set of events) caused by natural perturbations and that a given community becomes a victim of nature's destructive power. In Colombia, as in many other parts of the world, many believe that such disasters are a form of divine punishment. This conviction generates a sense of impotence and conformity in the population, making preventive interventions more difficult [13]. One of the most common beliefs is that a disaster is produced by powerful natural or supernatural forces acting against human beings [14].

Most disasters in Colombia are however actually caused by deficient land planning, a known factor of vulnerability. Poverty and vulnerability are directly proportional and complementary variables, associated with social exclusion and spatial segregation, which are frequently the product of a system characterized by errors, individual interests and informal urban planning and management [15].

B. Hazard, Disasters and Vulnerability

The concepts of risk, hazard and vulnerability have been ascribed many definitions and meanings, due to the fact that they have been approached by many different disciplines. In the 1940s, the risk/hazard paradigm was suggested by Gilbert F. White. Indeed in the mid-twentieth century, there was considerable research and policy interest in understanding human occupancy of hazard zones, societal adjustments for reducing the impact and the acceptance of the communities in question, when it came to living with the risks associated with hazardous conditions [16].

A natural hazard is a phenomenon that may cause or contribute loss of life, injury or other detrimental health impacts, property damage, loss of livelihoods and services and may also lead to social and economic disruption and/or environmental damage [17]. A disaster meanwhile is defined as a serious disruption to the functioning of a community or a society which involves widespread human, material, economic and/or environmental losses and impacts, which exceeds the ability of the affected community or society to cope using its own resources [17]. The impacts caused by a given disaster are determined not only by the nature of the hazard but also the level of vulnerability in the community. Vulnerability is defined as the characteristics and circumstances of a community, system or asset that make it susceptible to the damaging effects of a hazard [17].

Furthermore, vulnerability has three components: sensitivity to hazards; the ability to respond to disasters; and the ability to cope with the immediate aftermath of disasters. These three components are exacerbated in conditions of poverty, which, along with the lack of infrastructure and accelerated urban growth, increase the vulnerability of human populations [18,19].

II. Material and Methods

The article establishes a correlation for the results obtained from an evaluation of those disasters reported in 40 municipalities located in the basin of the Bogota River in the Department of Cundinamarca in 2010 and 2011. The data for each of the used variables were obtained from the Sistema de Información Geográfica para la Planeación y el Ordenamiento Territorial (SIGOT-Geographic Information System for Territorial Planning and Organization). The statistical analysis was performed using SigmaStat (version 3.5.) statistical software (*Systat Software, Inc.*) with a level of significance of 95% ($\alpha = 0.05$) and a P value < 0.05.

III. Selecting variables for analysis

The data were collected within the framework of the project "Assessment of Environmental Services associated with the Water Ecosystems of Bogotá D.C and their impact on Human Wellbeing" carried out by the Urban Environmental Research Group of the Piloto of Colombia University.

Three (3) components were taken into account in the selection of independent variables:

1. Environmental: existence of strategic ecosystems and their status; areas used for productive purposes and degraded areas.
2. Social: number of inhabitants per municipality; educational attainment of the population and use of resources.
3. Economic: quality of life of the population and the availability of resources in the municipal administration.

The selected variables which represent these components and which information was available, are shown in Table I.

TABLE I
INDEPENDENT VARIABLES ANALYZED

Acronym	*Description*	*Definition*
FAP	% of Forest Area	*(area of forest use/ total municipal area)*100*
CAP	% of Cultivated Areas	*(cultivated area in the municipality/ total municipal area)*100*
PAP	% of Pasture Area	*(pasture area / total municipal area)*100*
PD	Population Density	*Total number of inhabitants/ municipal area in square kilometers*
TIR	Total Illiteracy Rate	*Total population aged 15 and over who cannot read nor write / total population aged 15 and over*
CAR	Compliance in Allocation of Resources for Water and Basic Sanitation	*Pursuant to articles 356 and 357 of the National Constitution of Colombia, these resources are transferred by the nation, through the General Participation System (SGP, its acronym in Spanish), to territorial entities, departments, districts and municipalities for the financing of the local provision of services such as health and education*
QLI	Quality of Life Index	*Measures and characteristics of the living conditions of the poor, including variables associated with housing, public services and the number of people in a household.*
FD	Financial Dependency on the funds transferred by the nation through the SGP	*(received transfers from SGP / total revenue)*100*
GOR	Generation of Own Resources through Taxation	*(tax revenue/ total revenue)*100*

IV. CALCULATION

In order to define the strength of association for each of the nine (9) independent variables analyzed for the 40 municipalities in the Department of Cundinamarca (Colombia) with the number of flood disasters reported (dependent variable), the Multiple Linear Regression Method was applied in several stages. For each of these stages the model was purged, eliminating those variables with a very low correlation coefficient "R" and keeping those with a relatively high "t statistic." Likewise, the variables with multicollinearity (redundancy), determined by the high values of the Variance Inflation Factor (VIF), were purged.

V. RESULTS AND DISCUSSION

A. First Multiple Linear Correlation Model

Table II shows the results of the first linear correlation analysis, carried out for the Number of Reported Flood Disasters (DISASTERS) using the nine (9) independent variables.

TABLE II
RESULTS FOR FIRST MULTIPLE LINEAR CORRELATION ANALYSIS

Variable	**t statistic**	**P value**	**Variance Inflation Factor (VIF)**	**Incremental or Type I sum of squares (SSIncr)**	**Marginal or Type III sum of squares (SSMarg)**
FAP	1.02	0.320	4.03	20.51	6.84
CAP	-0.27	0.790	1.70	14.82	0.48
PAP	0.81	0.430	4.16	29.08	4.27
PD	3.56	0.001	3.46	162.88	82.89
TIR	2.82	0.008	2.26	61.71	52.04
CAR	0.26	0.798	1.37	1.14	0.44
QLI	0.87	0.391	6.27	2.96	4.96
FD	0.61	0.548	8.18	2.96	2.42
GOR	0.27	0.790	8.91	0.47	0.47

DISASTERS = -19.967 + (0.0833 * *FAP*) - (0.0223 * *CAP*) + (0.0532 * *PAP*) + (0.00659 * *PD*) + (0.452 * *TIR*) + (0.00876 * *CAR*) + (0.137 * *QLI*) + (0.0392 * *FD*) + (0.0172 * *GOR*)
R = 0.776
F = 5.036
P <0.001
PRESS = 443.08

B. Discussion of the First Multiple Linear Correlation Model

The correlation coefficient, "R" (0.776) partially explains a positive correlation between all the independent variables and the dependent variable. These values are supported by the analysis of variance "F" statistic, which, although not high (F=5.036), stands along a low P value (P<0.001) thud confirming the likelihood of a correlation between the analyzed variables.

The higher "t statistic" values are indicative of the independent variables that have the most influence on the dependent variable, which are namely *Population Density "PD"* (t=3.56) and *Total Illiteracy Rate "TIR"* (t=2.82). These also have a low P value, are likely to be right.

There are some redundant independent variables (where multicollinearity is present), a fact evidenced by the large Variance Inflation Factor "VIF" values: *Generation of Own Resources "GOR"* (VIF=8.91), *Financial Dependency "FD"* (VIF=8.18), *Quality of Life Index "QLI"* (VIF=6.27), *% of Pasture Area "PAP"* (VIF=4.16) and *% of Forest Area "FAP"* (VIF=4.03). These variables could therefore be excluded from the regression equation.

The comparison between the values of SSIncr (Incremental or Type I sum of squares) and SSMarg (Marginal or Type III sum of squares) for each variable identifies those which contribute the most, both incrementally (Incr) and independently (Marg), to the prediction of the dependent variable (in order of priority: *PD, TIR, FAP, QLI, PAP, FD, CAP, GOR, CAR*). As was expected, the variables with the higher "t statistic" values also had high values for these indicators. This could suggest the inconvenience of eliminating the variables *FAP* and *QLI* although they present a high Variance Inflation Factor (VIF).

Taking into account the analysis, the *FD* and *GOR* variables (which show the higher VIF values) and the *CAP* and *CAR* variables (which have very low "t statistic" values but and very high "P value") were eliminated for the second adjustment of the model.

C. Second Multiple Linear Correlation Model

Table III shows the results, after eliminating both the redundant variables and those with a low "t statistic," which account for the ability of the independent variable to predict the behavior of the dependent variable.

TABLE III

RESULTS FOR SECOND MULTIPLE LINEAR CORRELATION ANALYSIS

Variable	**t statistic**	**P value**	**Variance Inflation Factor (VIF)**	**Incremental or Type I sum of squares (SSIncr)**	**Marginal or Type III sum of squares (SSMarg)**
FAP	1.93	0.062	2.21	20.51	21.93
PAP	1.43	0.161	2.76	16.53	12.06
PD	5.34	<0.001	2.06	190.26	167.80
TIR	3.00	0.005	2.14	60.94	52.90
QLI	0.83	0.411	2.94	4.08	4.08

*DISASTERS = -14.804 + (0.110 * FAP) + (0.0729 * PAP) + (0.00725 * PD) + (0.443 * TIR) + (0.0848 * QLI)*

R = 0.771

F = 9.936

P <0.001

PRESS = 332.87

D. Discussion of the Second Multiple Linear Correlation Model

It is evident that the "PRESS statistic (Predicted Residual Error Sum of Squares)" is lower here than in the first model. This indicates that eliminating the redundant variables *FD* and *GOR* and those with a low "t statistic" (*CAP* and *CAR*) allows for a model with a higher predictive capacity.

E. Final Multiple Linear Correlation Model

The same procedure of model purging was applied and two variables among those chosen were prioritized due to their ability to explain the behavior of the dependent variable "*Number of Reported Flood Disasters" (DISASTERS)*, namely *Population Density (PD)* and *Total Illiteracy Rate (TIR)*. Table IV shows the results for each of these variables.

TABLE IV

RESULTS FOR FINAL MULTIPLE LINEAR CORRELATION ANALYSIS

Variable	**t statistic**	**P value**	**Variance Inflation Factor (VIF)**	**Incremental or Type I sum of squares (SSIncr)**	**Marginal or Type III sum of squares (SSMarg)**
PD	6.65	<0.001	1.24	203.43	267.52
TIR	3.28	0.002	1.24	64.99	64.99

*DISASTERS = -1.611 + (0.00711 * PD) + (0.374 * TIR)*

R = 0.738

F = 22.172

P <0.001

PRESS = 282.59

F. Discussion of the Final Multiple Linear Correlation Model

At a 95% (alpha=0.05) Confidence Level the following is observed:

1. Among the variables chosen for the study, the ability to predict the occurrence of disasters is given chiefly by the *PD* and *TIR* variables, with a probability of error of P <0.001 (0.1%).
2. The other variables assessed show a probability of a Type I error of 5%, hence why they were excluded.
3. Although a high level of adjustment was not achieved (R=0.738), the dependent variable (*Number of Reported Flood Disasters (DISASTERS)*) can be partially explained by the *Population Density (PD)* and *Total Illiteracy Rate (TIR)* (F=22.172).

The relationship between *Population Density (PD)* and the number of reported disasters can be explained by the fact that a disaster is a phenomenon directly related to the number of inhabitants in the area where it occurs.

The *Total Illiteracy Rate (TIR)* can be related to poverty levels in less developed countries. Likewise, a relationship between poverty and flood disasters has been established as a consequence of a marginal situation. Nonetheless, no mathematical relationship supporting this observation had previously been established for Colombia.

As for the relationship between education, productivity and poverty, the findings published here could lend support to the notion that Colombian communities with lower levels of education are also less capable of coping with flood disasters due to their reduced organization and response capacity [20], as evidenced by Boksh Moral [5] for the city of Rajshahi in Bangladesh, who demonstrated the correlation coefficient of 0.83 between flood frequency and the level of educational attainment.

The vulnerability of poor communities when a flood disaster occurs could be reduced by improving educational attainment whilst increasing the level of access to and participation in governmental risk management programs for these sectors of the population.

VI. CONCLUSIONS

This study permits the authors to state that in the

Department of Cundinamarca (Colombia) there is a significant correlation between the *total illiteracy rate (TIR)* and the *number of reported flood disasters (DISASTERS)*. Such findings do not only indicate a relationship between poverty and flood risk vulnerability, which has been established as a considerable anecdotal [21] but rather they also suggest a direct correlation between limited educational attainment and the ability to respond to the effects of a natural hazard.

Through this study, a significant issue has been determined, which shows many different variables and interrelations which deserve to be evaluated not only from a mathematical or engineering perspective but also for their sociological, psychological, cultural and even behavioral aspects. Finally, the findings here could be corroborated for other regions in Colombia as well as for poor regions around the world.

ACKNOWLEDGMENT

The authors would like to thank the Directorate of Research of the Piloto of Colombia University for providing the funding of the project “Assessment of Environmental Services associated with the Water Ecosystems of Bogotá D.C. and their impact on Human Wellbeing” of which the research project discussed in his article formed a part.

REFERENCES

[1] Sistema Nacional para la Prevención y Atención de Desastres - SINPAD, «Consolidado de Emergencias en Colombia,» [Online]. Available: http://www.sigpad.gov.co/sigpad/emergencias_detalle.aspx?idn=41.
[2] W. A. Lozano-Rivas, Elementos para la Recuperación de Ríos Urbanos. Caso de estudio: Río Torca (Tramo 1), Bogotá D.C., Bogotá D.C.: Universidad Antonio Nariño, Fondo Editorial: Secretaría Distrital de Medio Ambiente, 2011.
[3] Departamento Administrativo Nacional de Estadística - DANE, «Producto Interno Bruto - Cuentas,» 2 Mayo 2012. [Online]. Available: http://www.dane.gov.co/.
[4] Secretaría de Planeación de Cundinamarca, «Pobreza e Indigencia por Departamentos,» Secretaría de Planeación de Cundinamarca, Bogotá D.C., 2004.
[5] J. Boksh Moral, «Model of correlation between Perceived Local Environment and Urban Poverty: An Evidence from Rajshahi City, Bangladesh,» *Akademika (Mei-Ogos),* pp. 117-125, 2010.
[6] S. Ringen, The Possibility of Politics, New York: Oxford University Press, 1987.
[7] A. Rahman, Environment and Poverty: Key Linkages for Global Sustainable Development, Dhaka: UPL, 1998.
[8] C. Benson, «The economic impacts of natural disasters in Fiji. Working Paper 98,» Overseas Development Institute, London, 1997.
[9] C. Benson y E. Clay, «Understanding the economic and financial impacts of natural disasters,» World Bank, Washington D.C., 2004.
[10] World Bank, «World Bank Development Report: Attacking Poverty 1:170,» World Bank, Washington D.C., 2000.
[11] IPCC - Intergovernmental Panel of Climate Change, «Fourth Assessment Report. AR4,» Geneva, Switzerland, 2007.
[12] A. Sen, The Standard of Living, Cambridge: Cambridge University Press, 1987.
[13] C. Cárdenas, Antecedentes y origen del sistema nacional para la prevención y atención de desastres en Colombia; una primera aproximación al tema. Proyecto UNDRO mitigación de riesgos en Colombia fase II. Dirección nacional para la prevención y atención de desastres, Bogotá D.C.: Dirección Nacional para la Prevención y Atención de Desastres, 1993.
[14] G. Romero y A. Maskrey, Cómo Entender los Desastres Naturales. Los Desastres no son Naturales, Lima: Red de Estudios Sociales en Prevención de Desastres en América Latina. LA RED, 1993.
[15] R. D'Ercole, S. Hardy, P. Metzger y J. Robert, «Vulnerabilidades urbanas en los países andinos,» *Boletín del Instituto Francés de Estudios Andinos,* nº 38, pp. 401-410, 2009.
[16] S. L. Cutter, C. T. Emrich, J. J. Webb y D. Morath, «Social Vulnerability to Climate Variability Hazards: A Review of the Literature. Final Report to Oxfam America,» Hazards and Vulnerability Research Institute. Departmentof Geography. University of South Carolina, Columbia, 2009.
[17] UNISDR - United Nations International Strategy for Disaster Reduction Secretariat, Terminology on disaster risk reduction, Bangkok, 2009.
[18] P. Freeman, Infraestructure, natural disasters and poverty, Laxenburg, Austria: International Institute for Applied Systems Analysis, 1999.
[19] M. Bahry, *The relationship between household livelihood profile and fire and flood-related vulnerability,* 2007.
[20] W. A. Lozano-Rivas y T. Bolaños Silva, Elementos para la comprensión de los servicios ambientales en la gestión integral del agua: una mirada desde Bogotá, Bogotá D.C.: Universidad Piloto de Colombia, 2012.
[21] P. Narsey Lal, R. Singh y P. Holland, Relationship between natural disasters and poverty: a Fiji case study. SOPAC Miscellaneous Report 678, United Nations International Strategy for Disaster Reduction. Secretariat's 2009 Global Assesment Report on Disaster Reduction, 2009.