Post-print version of 'Jato-Espino, D., Sillanpää, N., Andrés-Doménech, I., Rodriguez-Hernandez, J. (2018). "Flood Risk Assessment in Urban Catchments Using Multiple Regression Analysis". **J. Water Resour. Plann. Manage.**, 144(2): 04017085. DOI: 10.1061/(ASCE)WR.1943-5452.0000874'.

Flood risk assessment in urban catchments using multiple regression analysis

Daniel Jato-Espino[1]; Nora Sillanpää[2]; Ignacio Andrés-Doménech[3]; Jorge Rodriguez-Hernandez[4]

Abstract

Flood assessment in urban catchments is usually addressed through the combination of Geographic Information Systems (GIS) and stormwater models. However, the coupled use of these tools involves a level of detail in terms of hydrological modelling which can be beyond

---

[1] Postdoctoral Researcher, GITECO Research Group, Universidad de Cantabria, Av. de los Castros 44, 39005, Santander, Spain (corresponding author). E-mail: jatod@unican.es

[2] Postdoctoral Researcher, Department of Built Environment, Aalto University School of Engineering, P.O. Box 15200, 00076 Aalto, Finland. E-mail: nora.sillanpaa@aalto.fi

[3] Associate Professor, Instituto Universitario de Investigación de Ingeniería del Agua y del Medio Ambiente (IIAMA), Universitat Politècnica de València, Cno. de Vera s/n, 46022 Valencia, Spain. E-mail: igando@hma.upv.es

[4] Associate Professor, GITECO Research Group, Universidad de Cantabria, Av. de los Castros 44, 39005, Santander, Spain (corresponding author). E-mail: rodrighj@unican.es

the scope of overall flood management planning strategies. This research consists of the development of a methodology based on Multiple Regression Analysis (MRA) to assess flood risk in urban catchments according to their morphologic characteristics and the geometrical and topological arrangement of the drainage networks into which they flow. Stormwater models were replaced by a combination of Multiple Linear Regression (MLR), Multiple Non-Linear Regression (MNLR) and Multiple Binary Logistic Regression (MBLR), which enabled identifying influential parameters in the maximum runoff rates generated in urban catchments, modelling the magnitude of peak flows across them and estimating flood risk in the nodes of sewer networks, respectively. The results obtained through a real urban catchment located in Espoo (Finland), demonstrated the usefulness of the proposed methodology to provide an accurate replication of flood occurrence in urban catchments due to intense storm events favored by Climate Change, information that can be used to plan and design preventive drainage strategies.

Keywords

Flood Risk Management; Geographic Information System; Multiple Regression Analysis; Urban Hydrology

Introduction

The interactive effects of urbanization and Climate Change are one of the most important challenges with which human beings will have to deal as a collective in the future (Hoornweg et al. 2011). Their combination inflicts a particularly forceful impact on stormwater drainage in urban catchments. Growing urbanization contributes to increasing runoff volume and accelerating the time until peak flow occurs, whilst Climate Change is expected to result in an intensification of the hydrological cycle, which might lead to more violent rainfall events (Huntington 2006). In the end, this might result in localized floods along the sewer networks used to drain urban catchments, which are incapable of conveying such large amounts of water.

Urban floods have important consequences in physical, economic, environmental and social terms (Tingsanchali 2012). These effects have traditionally been divided into tangible and intangible, depending on whether they can be monetized or not (Smith and Ward 1998). Examples of the former include damage to property or loss of profits, whilst the latter stand for aspects like loss of life or negative impacts on the well-being and environment. The potential impacts of urban floods can also be classified into direct and indirect, according to their spatiotemporal scale. Hence, direct damage is related to any loss caused by the immediate interaction of floods with human beings, properties and the environment, whilst indirect effects concern those which are beyond the limits of the flooded area (Hammond et al. 2015). All these potential consequences are expected to become more severe and frequent due to the combination of Climate Change with high density of population and large impervious areas (Huong and Pathirana 2013).

However, despite the increasingly important threat posed by this phenomenon, flood management keeps being usually addressed in literature through stormwater models, which are often used in combination with Geographic Information Systems (GIS). Knebl et al. (2005) integrated them in a study to highlight the importance of the degree of imperviousness in urban catchments, which resulted in a decrease in infiltration capacity of the terrain and increased flood risk. Barco et al. (2008) coupled both components to prove that imperviousness and depression storage were the most influential factors in the generation of flow rates in a large urban catchment located in Southern California. Dongquan et al. (2009) combined stormwater models with GIS based on elevation-related data such as flow direction, raster-vector conversion or catchment division to automate the process of rainfall-runoff modelling. Guan et al. (2015) merged both tools to determine the increase in peak flow and runoff volume caused by urbanization in a catchment located in Espoo (Finland) and proposed different alternative drainage techniques to mitigate these effects. Eshtawi et al. (2016) coupled three existing models (SWAT, MODFLOW and MT3DMS) to quantify surface-groundwater interactions in increasingly developed areas, demonstrating the strength of using integrated hydrologic models in the sustainable urban water planning process. Jato-Espino et al. (2016a) presented a GIS-based stormwater modelling approach that demonstrated the potential of permeable pavements and green roofs to attenuate floods in urban catchments in comparison with conventional sewer networks. Beck et al. (2017) described a semi-distributed approach to estimate runoff reductions (TELR) to inform stormwater management decisions, including a series of algorithms for rainfall-runoff transformation and routing and specifications to implement Best Management Practices (BMPs). Hanington et al. (2017) developed and calibrated a fine-scaled quasi-2D hydro-dynamic model (IWRM-LXQ) for interprovincial water resource planning and management, arguing that their approach was especially suitable for assessing hydraulically complex systems at a provincial or district level.

The use of tools as those mentioned above is complex and entails a substantial time investment, aspects which are in conflict with the simplicity and promptness required by administrative and public entities to design flood management planning strategies (Ashley et al. 2007). Stormwater models are especially demanding in terms of characterization and simulation to be used by general public or non-modelling planners (Elliott and Trowsdale 2007), since they involve making decisions related to infiltration and routing processes and calibrating the parameters that influence the transformation of rainfall to runoff. In contrast, the GIS-related tasks required for creating the input data into these models to run stormwater simulations, which mainly concern catchment delineation and the determination of subcatchment imperviousness and average slope, are easy to compute using basic editing and statistical tools (Jato-Espino et al. 2016a). Furthermore, GIS are widely implemented systems for multiple purposes related to the management of all kinds of spatial data, so that using them for flood management does not involve an innovation with respect to common resources and practices.

As a result of these considerations, the objective of this research was to develop a methodology for flood risk assessment omitting the use of stormwater models. This was achieved through the integration of Multiple Linear Regression (MLR), Multiple Non-Linear Regression (MNLR) and Multiple Binary Logistic Regression (MBLR), which were combined to play the role of stormwater models in a simpler manner. The integrated application of different types of Multiple Regression Analysis (MRA) provided a cutting-edge approach to replicate peak flow rates and predict flooding probabilities in sewer-catchments and opens new lines of research and in the field of urban water planning and management. The usefulness of the proposed methodology was evaluated through a case study of an urban catchment located in Espoo (Finland), which provided the precipitation and flow data required to validate the results at the nodes forming its sewer network.

Methodology

Since the delineation of urban catchments using Geographic Information Systems (GIS) is a widely studied topic in literature (Knebl et al. 2005; Guan et al. 2015; Jato-Espino et al. 2016a), the methodology proposed in this research only focused on reproducing the role of stormwater models using Multiple Regression Analysis (MRA). The general purpose of MRA is to predict the value of a dependent variable or predictand based on the values of a series of independent explanatory variables or predictors.

In the context of this research, Multiple Linear Regression (MLR) was used as an exploratory analysis to identify relevant parameters to the occurrence of peak flow rates. Next, the identified parameters were incorporated into a Multiple Non-Linear Regression (MNLR) framework, in order to boost the prediction accuracy of the magnitude of runoff that might be generated in urban catchments and conveyed by sewer networks as a result of severe rainfall events. The magnitude of runoff was presented through the values of Maximum Lateral Inflow (MLI, l/s) and Maximum Total Inflow (MTI, l/s) produced in urban catchments. The former stands for the peak of apportionment of surface runoff from the subcatchment areas to each node in the sewer network, whilst the latter also includes the contribution from preceding nodes and conduits. Finally, Multiple Binary Logistic Regression (MBLR) models were created from the estimates of MLI and MTI to determine the probability of flooding in sewer networks, facilitating the adoption of measures for preventing urban flood events at strategic sites and maximizing their positive impact and effectiveness. Fig. 1 provides a graphical scheme of the proposed approach based on the sequential application of these three types of MRA.

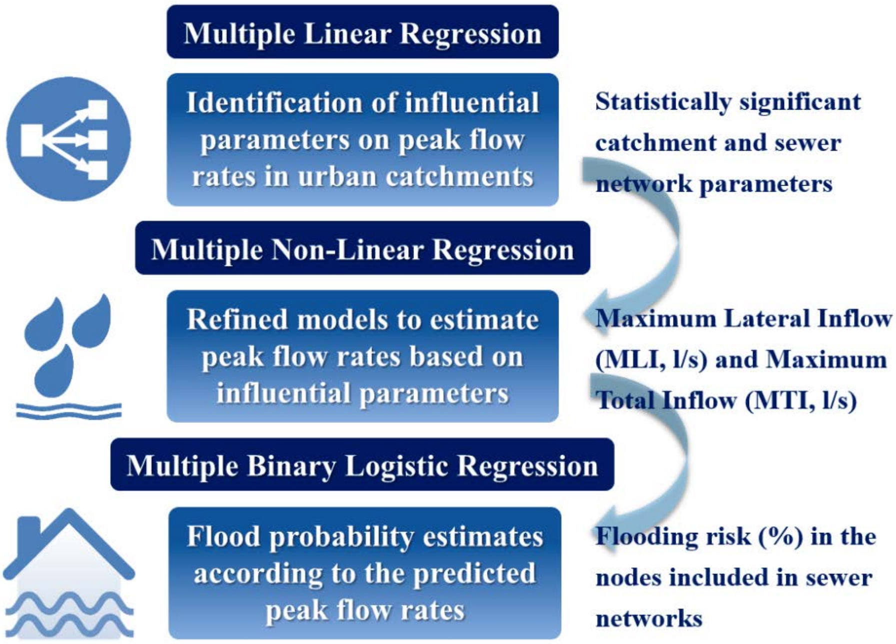

Fig. 1. Scheme of the proposed methodology based on Multiple Regression Analysis (MRA)

The development of these analyses stemmed from a set of catchment parameters considered for estimating MLI, whose combination with a list of predictors related to the two main elements forming sewer networks (nodes and conduits) enabled modelling both MTI and flooding probability. Table 1 links these parameters to the visual objects required to represent urban drainage systems: subcatchments, nodes and conduits.

Table 1. Parameters for estimating Flooding (%), Maximum Lateral Inflow (MLI, l/s) and Maximum Total Inflow (MTI, l/s) in urban catchments

Predictand	**Sub-predictand**	**Visual object**	**ID**	**Predictor**
Flooding (%)	MLI (l/s)	Catchment	$x_{1.1}$	Subcatchment area (ha)
			$x_{1.2}$	Degree of imperviousness in the subcatchment (%)
			$x_{1.3}$	Subcatchment width (m)
			$x_{1.4}$	Average slope in the subcatchment (%)
	MTI (l/s)	Sewer network	$x_{2.1}$	Node invert elevation (m)
			$x_{2.2}$	Preceding length of conduit (m)
			$x_{2.3}$	Cumulative preceding length of conduits (m)
			$x_{2.4}$	Subsequent length of conduit (m)
			$x_{2.5}$	Preceding diameter of conduit (m)
			$x_{2.6}$	Cumulative preceding diameter of conduits (m)
			$x_{2.7}$	Subsequent diameter of conduit (m)
			$x_{2.8}$	Preceding slope of conduit (%)
			$x_{2.9}$	Cumulative preceding slope of conduits (%)
			$x_{2.10}$	Subsequent slope of conduit (%)

These predictors were selected to result in a set of basic variables easy to acquire and/or compute, in order to facilitate the implementation of the proposed methodology worldwide. Therefore, the length and depth of conduits (pipes) and nodes (manholes), respectively, were the inputs required to parameterize the drainage network in the study area, whilst the subcatchments forming it were characterized according to their area, percentage of imperviousness, slope and width, which was estimated dividing their area by the average length of the flow paths from the furthest drainage points. All these catchment-related variables were determined using GIS-based editing and zonal statistics tools.

As pointed out by Yao et al. (2017), who highlighted the complexity of urban hydrology for water resources planning and management, the relationships between rainfall-runoff processes and spatial patterns is a key factor to manage flood risks in small catchments. In fact, the need for designing strategies for flood risk prevention based on proactive spatial planning has been identified as a crucial aspect to ensure the resilience of urban socio-ecological systems (Hegger et al. 2016).

Identification of relevant parameters for peak flow generation

The justification of which parameters related to the morphology of urban catchments and the geometry of sewer networks (Table 1) contributed more to explaining the values of MLI and MTI associated with different storm events was carried out using MLR. MLR consists of modelling the relationship between n predictors x_n and a predictand y through a linear expression as formulated in Eq. (1) (Aiken et al. 2003):

$$y = b_0 + \sum_{i=1}^{n} b_i \cdot x_i + \varepsilon \tag{1}$$

where b_i is the weight that indicates the importance of each predictor in the model. All the information that cannot be provided by the independent variables is completed by b_0 and ε, which are a constant and the residuals, respectively. A significance level of $\alpha = 0.05$ (Fisher 1925) was the threshold which determined whether a parameter was statistically significant for estimating MLI and MTI or not.

Therefore, this first step had the sole purpose of justifying the selection of parameters that contributed the most to reach high values of MLI and MTI from a physical point of view.

The use of MLR was deemed essential to ensure the hydrological and hydraulic validity of the proposed approach, since the clarity of these contributions might be distorted if evaluated through MNLR, due to the increased complexity involved by the inclusion of non-linear terms. This course of action was adopted based on the results of similar previous studies (Jato-Espino et al. 2017), which applied MNLR to refine the prediction accuracy of relevant parameters identifiable using MLR.

Modelling of Maximum Lateral Inflow and Maximum Total Inflow

Once the identification of catchment and sewer network parameters proving to be statistically significant for the generation of peak flow rates was accomplished, MNLR was used to combine them in the creation of non-linear equations for predicting both MLI and MTI with high accuracy. Unlike Eq. (1), the formula associated with MNLR also includes interactions and quadratic terms, as shown in Eq. (2):

$$y = b_0 + \sum_{i=1}^{n} b_i \cdot x_i + \sum_{i=1}^{n} b_{ii} \cdot x_i \cdot x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} \cdot x_i \cdot x_j + \varepsilon \qquad (2)$$

where b_{ii} and b_{ij} are the weights that indicate the importance of the quadratic and interaction terms in the model.

Since the objective of this paper was the design of a methodology to assess flood risk in urban catchments for water resources planning and management, the predicted R^2 coefficient was chosen to measure the goodness-of-fit of MNLR, in order to validate the proposed approach for modelling new cases. This coefficient consists of removing each observation from the dataset in the MNLR, predicting the regression equation and calculating how well the model estimates the omitted observation.

To further ensure the validity of the MNLR models built, their residuals were analyzed based upon the assumptions of normality, homoscedasticity and independence. The fulfillment of these assumptions was graphically verified using the following residual plots (Osbourne and Waters 2002): normal probability plot (Q-Q plot), standardized residuals vs. standardized predicted values plot and standardized residuals vs. observation order plot.

The equations derived from these MNLR were used to estimate the values of MLI and MTI in urban catchments straightforward through a series of predictors as those listed in Table 1. Such equations were further processed to incorporate rainfall intensity (I) into them, in order to allow predicting the magnitude of flooding for different precipitation scenarios accordingly. To this end, MLR was used again to determine the constant b_0 and weights b_i, b_{ii} and b_{ij} of the terms in the MNLR models as a function of I through Eq. (3). The estimates obtained using this expression were inputs in the calculation of the flooding probability of the nodes forming urban sewer networks through MBLR.

$$b_0 \vee b_i \vee b_{ii} \vee b_{ij} = b_0' + b_1' * I \tag{3}$$

Prediction of flooding probability

MBLR was used to integrate the knowledge generated through the application of the other types of MRA considered in the methodology (MLR and MNLR) to assess flood risk in urban catchments. MBLR is another variant of MLR characterized by the dichotomous nature of the predictand, which only has two possible outcomes. Hence, MBLR predicts the probability Pr that a certain characteristic is present in the predictand y from the values of the predictors x_i. The analytic expression for a MBLR model based on the logit link function is given in Eq. (4):

$$\Pr(y = 1 \mid x_i) = \frac{\exp(y')}{1 + \exp(y')} = \frac{\exp(b_0 + \sum_{i=1}^{n} b_i \cdot x_i)}{1 + \exp(b_0 + \sum_{i=1}^{n} b_i \cdot x_i)} \qquad (4)$$

where $y = 1$ indicates that the characteristic is present in observation i. In this case, the two outcomes for the predictand were “yes” and “no”, depending on whether the nodes of sewer networks were flooded or not after the occurrence of heavy rainfall events, and the predictors were the parameters included in Table 1.

In addition to the predictors and the predictand, MBLR enables incorporating another term into the analysis, known as frequency, which is an indicator of the number of times that the characteristic to be modelled is present. This concept was adapted to the purpose of this study to express how susceptible the nodes in sewer networks were to flooding. Hence, based on the parameters considered in the methodology so far, the frequency was defined as the ratio of MTI to MLI. This value was expected to provide a measure of the sensitivity of the nodes of sewer networks to reach their full capacity, since it combines two of the main factors favouring the occurrence of floods: accumulation and immediate contribution.

Therefore, flooding was a dichotomous dependent variable (i.e. its presence in a node is either “yes” or “no”) to be estimated using a series of catchment and sewer network continuous independent variables modulated by a frequency term representing the peak flow conditions in the sewer network. Consequently, the application of Eq. (4) yielded a probability indicating how likely a certain node was to be flooded; i.e. the worse the combinations of values in the predictors and the higher ratios of MTI to MLI, the closer the probability (expressed as a decimal) of that node to be 1.

The goodness-of-fit evaluation of MBLR slightly differed from that used for MNLR, due to the particular nature of the predictand in this type of MRA. The quality of MBLR models was assessed through the adjusted deviance R^2 coefficient and the Akaike Information Criterion (AIC) (Akaike 1973), which enabled the comparison of models with different predictors. Furthermore, the Hosmer-Lemeshow test was applied to check whether the deviation between estimated and observed probabilities was unpredictable by the binomial distribution (Hosmer and Lemeshow 2000). This test was found to be more suitable than the Deviance or Pearson tests due to the binary/response/frequency format of the data. The fact that the predictand was a binary outcome made the verification of residuals described for MNLR nonsensical.

The application and testing of the proposed framework enabled detecting which subcatchments and nodes required priority actions in terms of urban drainage planning and management, based on the values of peak runoff and flooding probability obtained through the subsequent application of Eqs. (2), (3) and (4).

Results and discussion: a case study in Espoo, Finland

The proposed methodology was implemented through a case study consisting of an urban catchment located in Espoo, Finland (see (Sillanpää and Koivusalo 2015) for further details). Fig. 2a) shows the spatial arrangement of the sewer network corresponding to this catchment, which was provided by the Helsinki Region Environmental Services Authority HSY and consisted of 75 nodes and 80 conduits, the 79 subcatchments forming the whole catchment area, which covered 10.535 ha, and the relationship between impervious and pervious areas in the catchment, which were delineated from the ortophoto of the study area (Jato-Espino et al. 2017). Fig. 2b) depicts the values of slope in the catchment, which were determined and classified from the Digital Terrain Model of the study area using Geographic Information Systems (GIS) tools (Jato-Espino et al. 2016b).

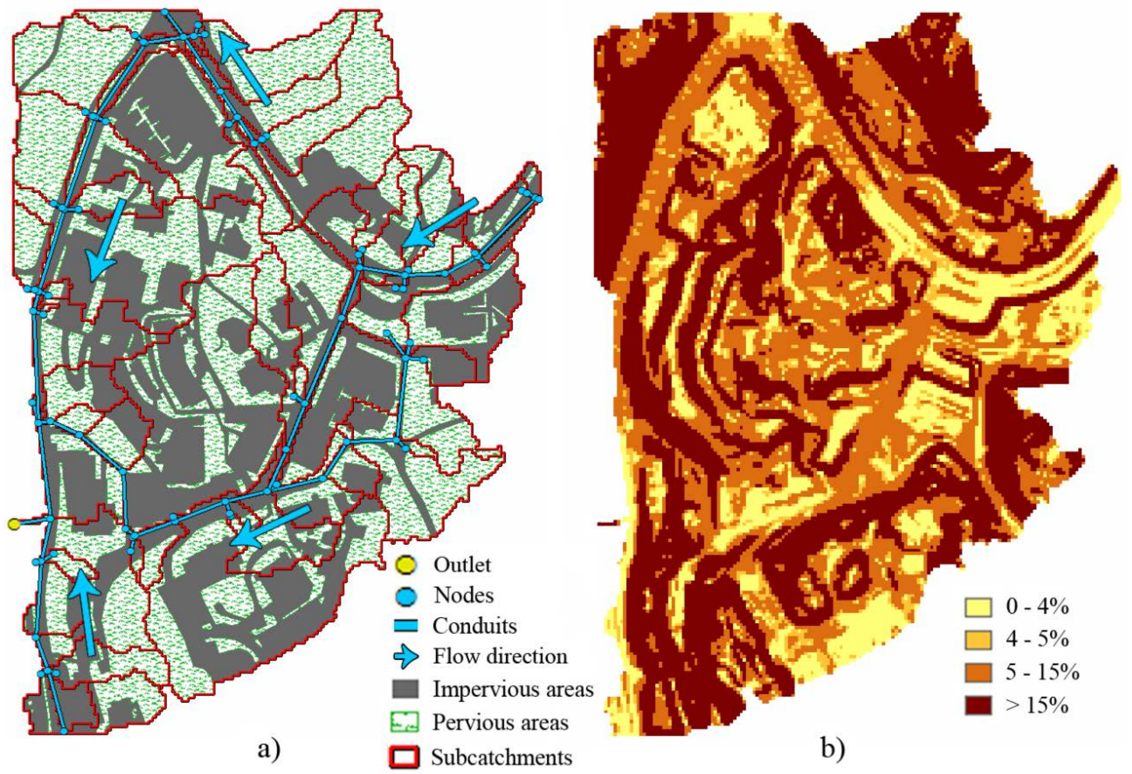

Fig. 2. a) Sewer network, subcatchments and impervious and pervious areas in the study catchment b) Slope (%) in the study catchment

The rainfall events used in this paper were taken from Jato-Espino et al. (2017), who modelled the study catchment in SWMM 5.1.010 (USEPA 2016) using three calibration (CAL 1, CAL 2 and CAL 3) and validation (VAL 1, VAL 2 and VAL 3) rainfall events (Table 2). These simulations reproduced the real hydrographs monitored at the outlet of the catchment with high accuracy, as demonstrated by the goodness-of-fit measures used to test them: Root-Sum Squared Error (RSSE), coefficient of determination (R^2) and Nash–Sutcliffe model efficiency coefficient (E) (Table 2).

The study catchment was re-simulated with the calibrated parameters for different return periods and Climate Change scenarios: RCP4.5 and RCP8.5 (Moss et al. 2008). Table 2 lists the values of duration, depth and intensity associated with four combinations of Climate Change scenario and return period (T) producing floods of different magnitude in the catchment, which were determined through its lag time and the coupling of Intensity–Duration–Frequency (IDF) curves and the Alternating Block Method, respectively.

Table 2. Summary of the rainfall events used to test the proposed methodology. Adapted from Jato-Espino et al. (2017)

Event	Duration (min)	Rainfall depth (mm)	Intensity (mm/h)	RSSE	R^2	E
CAL 1	352	5.0	0.85	81.94	0.91	0.85
CAL 2	686	37.4	3.27	212.81	0.93	0.86
CAL 3	418	12.2	1.75	92.67	0.96	0.93
VAL 1	396	5.2	0.79	42.46	0.97	0.97
VAL 2	288	9.0	1.88	68.26	0.95	0.92
VAL 3	408	23.4	3.44	115.64	0.97	0.96
RCP4.5; T = 5 yr.	106	19.0	10.75	-	-	-
RCP8.5; T = 5 yr.	106	25.6	14.48	-	-	-
RCP4.5; T = 50 yr.	106	31.5	17.84	-	-	-
RCP8.5; T = 25 yr.	106	38.0	21.51	-	-	-

Identification of relevant parameters for peak flow generation in the study catchment

Multiple Linear Regression (MLR) enabled identifying which predictors listed in Table 1 were statistically significant for the generation of peak flow rates and, by extension, determining their degree of contribution to producing high values of Maximum Lateral Inflow (MLI) and Maximum Total Inflow (MTI). MLR models were built stepwise in Minitab 17 (Minitab Inc 2016) to select only those predictors that were statistically significant to explain variations in MLI and MTI at the 95% confidence level (p-value < 0.05).

The results obtained for the modelling of MLI revealed that three parameters were statistically significant for estimating this predictand (p-value < 0.05): $x_{1.1}$ (Subcatchment area), $x_{1.2}$ (Degree of imperviousness in the subcatchment) and $x_{1.4}$ (Average slope in the subcatchment). The most influential predictor for estimating MLI was found to be $x_{1.1}$ with an average contribution of 82.52%, followed by $x_{1.2}$ and $x_{1.4}$ with 6.03% and 2.04%, respectively. Although their weights were different depending on the characteristic of the rainfall events used (Table 2), the contribution of the predictors considered was very similar in all cases. This homogeneity of values under different rainfall events validates the results achieved, since it indicates that the impacts of $x_{1.1}$, $x_{1.2}$ and $x_{1.4}$ on MLI were very similar both when considering common (CAL 1, CAL 2, CAL 3, VAL 1, VAL 2 and VAL 3) and extreme storms. Furthermore, the relationships between these predictors and MLI were logical, because larger subcatchments provide more opportunities to accumulate runoff and both impervious and steep areas facilitate the rapid conveyance of water. Hence, areas devoid of divisions due to drainage network deficiencies, built-up surfaces and topographically problematic sites were found to be more prone to produce high lateral inflows in urban catchments. Consequently, the mitigation of excessive runoff should be approached merging both nature and artificial solutions aimed at vegetating urban surfaces and also ensuring drainage support services, respectively.

Since the methodology was a stepped process in which the values of MTI in the nodes were partially calculated from those of MLI, the latter was included in the MLR models for estimating the former as a single predictor, in addition to those related to the sewer network (Table 1). As a result, x_1 (MLI) emerged as one of the two parameters proving to be statistically significant for predicting MTI, along with $x_{2.3}$ (Cumulative preceding length of conduits). In this case, $x_{2.3}$ was clearly the most important factor influencing the values of MTI in the nodes of the study catchment, with an average contribution of 97.39%. Again, both predictors were positively correlated to the predictand, since they contributed to increased runoff accumulation throughout the sewer network and the catchment, respectively. In fact, x_1 was statistically significant only for the extreme events. However, since they represented the situations in which floods occurred for different combinations of climate scenario and return period, this parameter was concluded to be relevant for the purpose of this research and was therefore not removed from further analyses.

All these results enabled validating the MLR models built for identifying statistically significant parameters for estimating peak flow rates and, therefore, using the information related to their degree of contribution to create MNLR models to predict MLI and MTI with high accuracy.

Modelling of Maximum Lateral Inflow and Maximum Total Inflow in the study catchment

Multiple Non-Linear Regression (MNLR) was used to fit the values of MLI and MTI obtained in all the nodes of the study catchment (Fig. 2a)) through the simulations run in SWMM (Jato-Espino et al. 2017) for the rainfall events that produced floods in the study catchment (Table 2), based on the knowledge acquired from the application of MLR to determine which catchment and sewer network parameters were more relevant for producing peak flow rates. Table 3 and Table 4 summarize the MNLR models determined for predicting both MLI and MTI.

Table 3. Summary of the Multiple Non-Linear Regression (MNLR) models built for the estimation of Maximum Lateral Inflow (MLI, l/s)

Event	Equation	Pred. R^2
RCP4.5; T = 5 yr.	$MLI = (3.21 - 54.85 * x_{1.1} * x_{1.1} + 1.64 * x_{1.1} * x_{1.2} + 1.51 * x_{1.1} * x_{1.4})^{1/0.845}$	0.95
RCP8.5; T = 5 yr.	$MLI = (4.17 - 75.00 * x_{1.1} * x_{1.1} + 2.29 * x_{1.1} * x_{1.2} + 2.04 * x_{1.1} * x_{1.4})^{1/0.845}$	0.96
RCP4.5; T = 50 yr.	$MLI = (4.94 - 93.04 * x_{1.1} * x_{1.1} + 2.90 * x_{1.1} * x_{1.2} + 2.51 * x_{1.1} * x_{1.4})^{1/0.845}$	0.96
RCP8.5; T = 25 yr.	$MLI = (5.77 - 113.55 * x_{1.1} * x_{1.1} + 3.60 * x_{1.1} * x_{1.2} + 3.02 * x_{1.1} * x_{1.4})^{1/0.845}$	0.96

Table 4. Summary of the Multiple Non-Linear Regression (MNLR) models built for the estimation of Maximum Total Inflow (MTI, l/s)

Event	**Equation**	**Pred. R^2**
RCP4.5; T = 5 yr.	$MTI = (6.03 + 9.50 * 10^{-2} * x_{2.3} + 6.39 * 10^{-2} * x_1 - 2.50 * 10^{-5} * x_{2.3} * x_{2.3})^{1/0.682}$	0.96
RCP8.5; T = 5 yr.	$MTI = (7.70 + 1.15 * 10^{-1} * x_{2.3} + 7.11 * 10^{-2} * x_1 - 4.00 * 10^{-5} * x_{2.3} * x_{2.3})^{1/0.682}$	0.95
RCP4.5; T = 50 yr.	$MTI = (8.98 + 1.30 * 10^{-1} * x_{2.3} + 7.54 * 10^{-2} * x_1 - 5.10 * 10^{-5} * x_{2.3} * x_{2.3})^{1/0.682}$	0.93
RCP8.5; T = 25 yr.	$MTI = (10.04 + 1.42 * 10^{-1} * x_{2.3} + 7.90 * 10^{-2} * x_1 - 6.20 * 10^{-5} * x_{2.3} * x_{2.3})^{1/0.682}$	0.92

The high values of predicted R^2 reached for both models, which were always above 0.9, ensured their reliability for making new estimates of MLI and MTI and validated the two-step approach based on the combination of MLR and MNLR. Furthermore, their residuals met the assumptions on which Multiple Regression Analysis (MRA) is based: normality, homoscedasticity and independence. For instance, Fig. 3 provide visual verification of the fulfilment of these assumptions for the worst regression models in Table 3 and Table 4 in terms of goodness-of-fit: RCP4.5; T = 5 yr. (MLI) and RCP8.5; T = 25 yr. (MTI). The approximate straight line in the Q-Q plots ensured the normality of residuals, whilst their random and non-curvilinear distributions around the horizontal axis in the standardized residuals versus fits plots confirmed the homoscedasticity of the regression models. Finally, the residuals versus order plots suggested that there was no serial correlation and their independence could also be assumed.

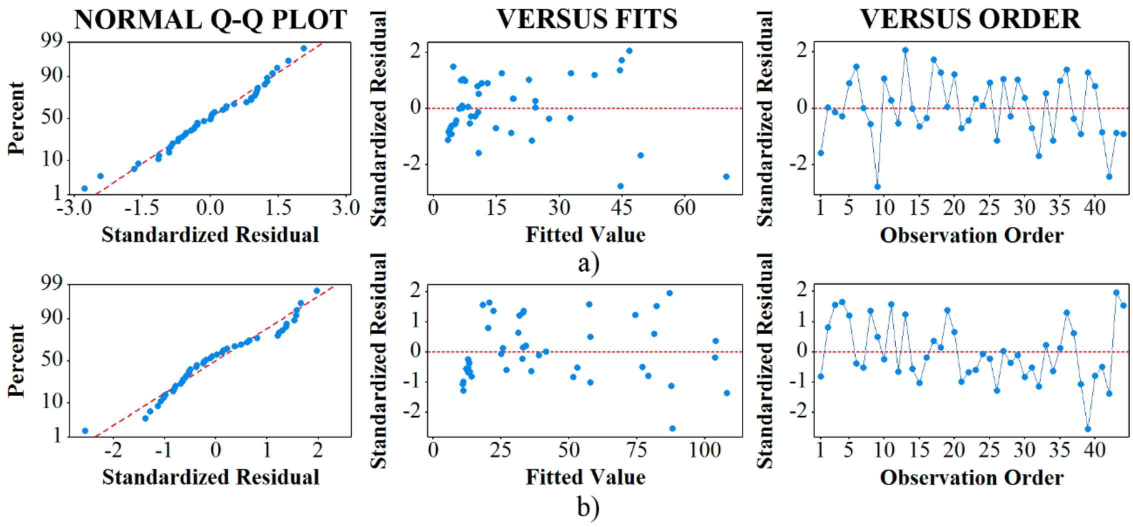

Fig. 3. Residual analyses for the Multiple Non-Linear Regression (MNLR) models built for the estimation of a) Maximum Lateral Inflow (MLI, l/s) for the RCP4.5 scenario and a return period of 5 years and b) Maximum Total Inflow (MTI, l/s) for the RCP8.5 scenario and a return period of 25 years

To facilitate the application of the equations for estimating MLI and MTI shown in Table 3 and Table 4 according to the characteristics of rainfall events, their constant b_0 and weights b_i, b_{ii} and b_{ij} (Eq. (2)) were fitted using MLR again with the intensity (I) of the Climate Change storms listed in Table 2 as predictor. Table 5 collects the equations obtained to predict these weights for both MLI and MTI.

Table 5. Summary of the Multiple Linear Regression (MLR) models built for the estimation of the constant ($\boldsymbol{b_0}$) and weights ($\boldsymbol{b_i}$, $\boldsymbol{b_{ii}}$ and $\boldsymbol{b_{ij}}$) for the Maximum Lateral Inflow (MLI, l/s) and Maximum Total Inflow (MTI, l/s)

Variable	**Equation**	**Pred. R^2**
MLI	$b_{1.0} = 0.692 + 0.238 * I$	1.00
	$b_{1.1*1.1} = 3.675 - 5.434 * I$	1.00
	$b_{1.1*1.2} = -0.329 + 0.182 * I$	1.00
	$b_{1.1*1.4} = 0.008 + 0.140 * I$	1.00
MTI	$b_{2.0} = 2.115 + 0.380 * I$	0.97
	$b_{2.3} = 0.049 + 0.005 * I$	0.96
	$b_1 = 0.046 + 0.002 * I$	0.88
	$b_{2.3*2.3} = 1.111 * 10^{-5} - 3.471 * 10^{-5} * I$	0.98

Again, the MLR models obtained highlighted by the excellent values of predicted R^2 achieved and enabled accepting the assumptions of MRA, as proven in Fig. 4, which depicts the residuals plots associated with the less accurate equations in Table 5: interaction between catchment area and average slope in the subcatchment ($b_{1.1*1.4}$) and b) Maximum Lateral Inflow (MLI) (b_1)

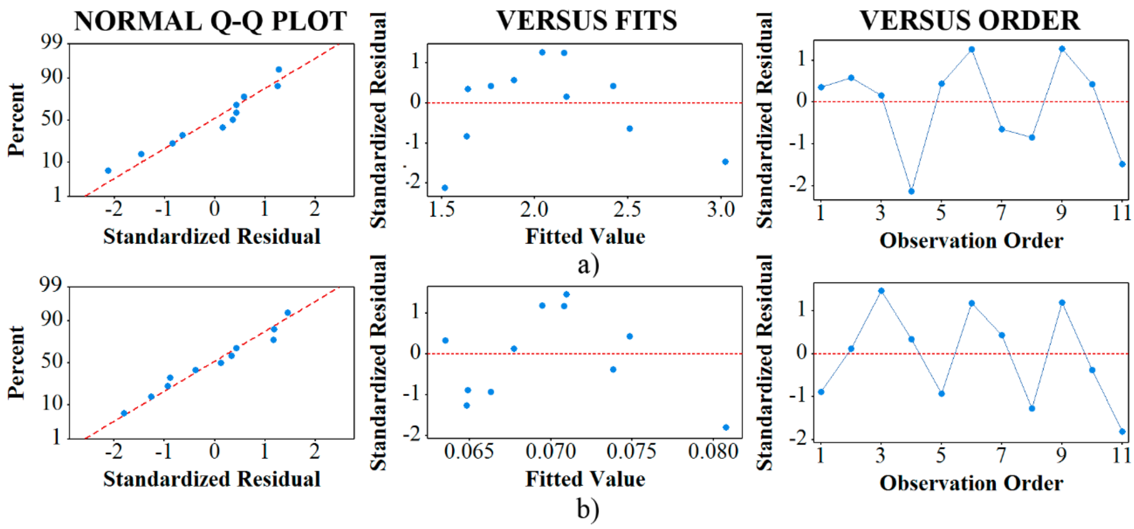

Fig. 4. Residual analyses for the Multiple Linear Regression (MLR) models built for the estimation of weights for a) interaction between subcatchment area and average slope in the subcatchment ($\boldsymbol{b_{1.1*1.4}}$) and b) Maximum Lateral Inflow (MLI) ($\boldsymbol{b_1}$)

The merger of the expressions contained in Table 3 and Table 4 with those determined in Table 5 yielded Eqs. (5) and (6), whose application allowed calculating the values of MLI and MTI in the study catchment from the sole use of easy-to-compute GIS-based factors and the intensity of the rainfall event to be assessed.

$$MLI = [(0.692 + 0.238 * I) + (3.675 - 5.434 * I) * x_{1.1} * x_{1.1} + (-0.329 + 0.182 * I) * x_{1.1} * x_{1.2} + (0.008 + 0.140) * x_{1.1} * x_{1.3}]^{1/0.845} \tag{5}$$

$$MTI = [(2.115 + 0.380 * I) + (0.049 - 0.005 * I) * x_{2.1} + (0.046 + 0.002 * I) * x_1 + (1.111 * 10^{-5} - 3.471 * 10^{-5} * I) * x_{2.1} * x_{2.1}]^{1/0.682} \tag{6}$$

The particularization of Eqs. (5) and (6) to the Climate Change storm events shown in Table 2 produced the values of MLI and MTI represented in Fig. 5 and Fig. 6, respectively. Their comparison with the results obtained through simulation in SWMM resulted in values of R^2 higher than 0.9 in all cases, demonstrating the accuracy of the proposed framework based on the combination of MLR and MNLR to fit the lateral and total peak flow rates in the nodes of the study area due to a series of storms with different intensities and durations, which in turn enable putting a focus on the elements in urban catchments which most contribute to producing flood events and taking water-related actions accordingly.

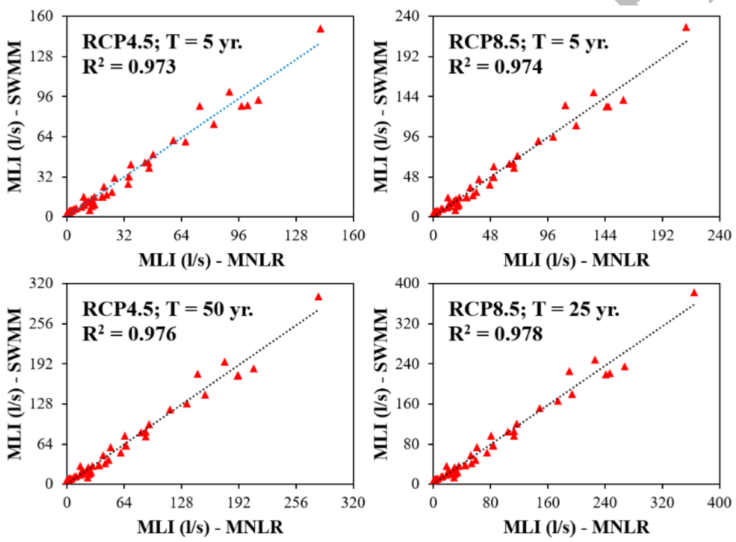

Fig. 5. Fit between the values of Maximum Lateral Inflow (MLI, l/s) obtained through stormwater simulations and those determined using Multiple Non-Linear Regression (MNLR)

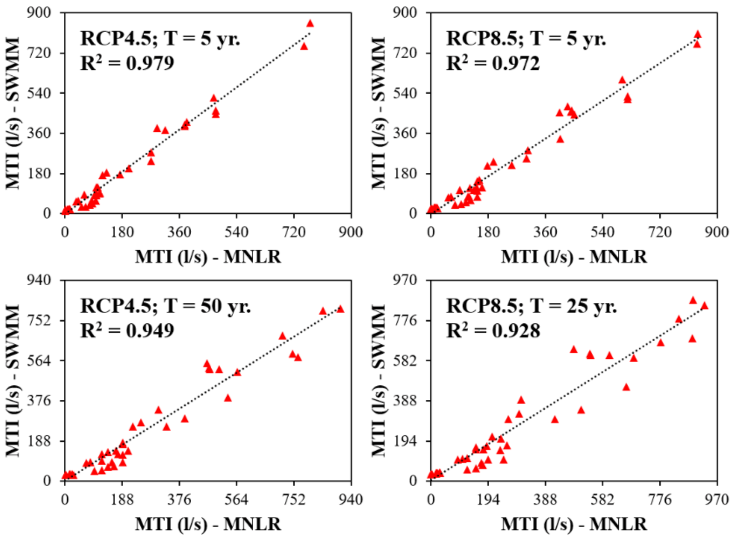

Fig. 6. Fit between the values of Maximum Total Inflow (MTI, l/s) obtained through stormwater simulations and those determined using Multiple Non-Linear Regression (MNLR)

Prediction of flooding probability in the study catchment

The last step to accomplish the implementation of the proposed methodology to the study catchment consisted of combining the parameters listed in Table 1 with the values of MLI and MTI predicted through Eqs. (5) and (6), in order to build Multiple Binary Logistic Regression (MBLR) models for predicting the probability of flooding throughout the sewer network. The equations to estimate y' (see Eq. (4)) for the four rainfall scenarios under consideration are provided in Table 6.

Table 6. Summary of the Multiple Binary Logistic Regression (MBLR) models built for the estimation of Flooding Probability (%)

Event	**Equation**	**Adj. Dev. R^2**	**AIC**	**H-L**
RCP4.5; T = 5 yr.	$y' = -0.49 - 0.26 * x_{1.2} + 0.01 * x_{2.3} + 0.06 * x_{2.4}$	0.83	84.23	0.42
RCP8.5; T = 5 yr.	$y' = -0.17 - 0.26 * x_{1.2} + 0.08 * x_{2.2} + 0.01 * x_{2.3}$	0.81	80.32	0.37
RCP4.5; T = 50 yr.	$y' = -26.41 - 0.13 * x_{1.2} + 0.41 * x_{1.4} + 6.23 * x_{2.1} + 0.11 * x_{2.2} + 18.39 * x_{2.5}$	0.79	88.41	0.07
RCP8.5; T = 25 yr.	$y' = -27.83 + 11.09 * x_{1.1} + 9.79 * x_{2.1} + 22.13 * x_{2.5} - 0.29 * x_{2.9}$	0.74	106.36	0.19

The quality of these models was ensured by the high and low values of adjusted deviance R^2 and Akaike Information Criterion (AIC) reached, which guaranteed the suitability of the predictors included in the equations presented in Table 6 to predict flooding probabilities. Moreover, the results of the Hosmer-Lemeshow (H-L) test for the four MBLR models built yielded p-values above the significance level ($\alpha = 0.05$) in all cases, which further validated their goodness-of-fit. Consequently, the inclusion of the ratio of MTI to MLI as the frequency in MBLR demonstrated to be a key factor to enhance the fit between predicted and simulated probabilities of flooding.

The values obtained for MLI and flooding probability were imported to GIS and mapped as depicted in Fig. 7. MLI was represented according to the subcatchments based on their peak runoff rates, in order to determine their degree of contribution to the nodes to which they flowed, whilst flooding probability was illustrated through the nodes forming the sewer network under study. Since the Climate Change events used to test the methodology provided the ranges of values for which drainage systems will have to be designed in the future to prevent the occurrence of floods in the catchment, this map provided the information required to plan water management strategies to take priority action in vulnerable areas.

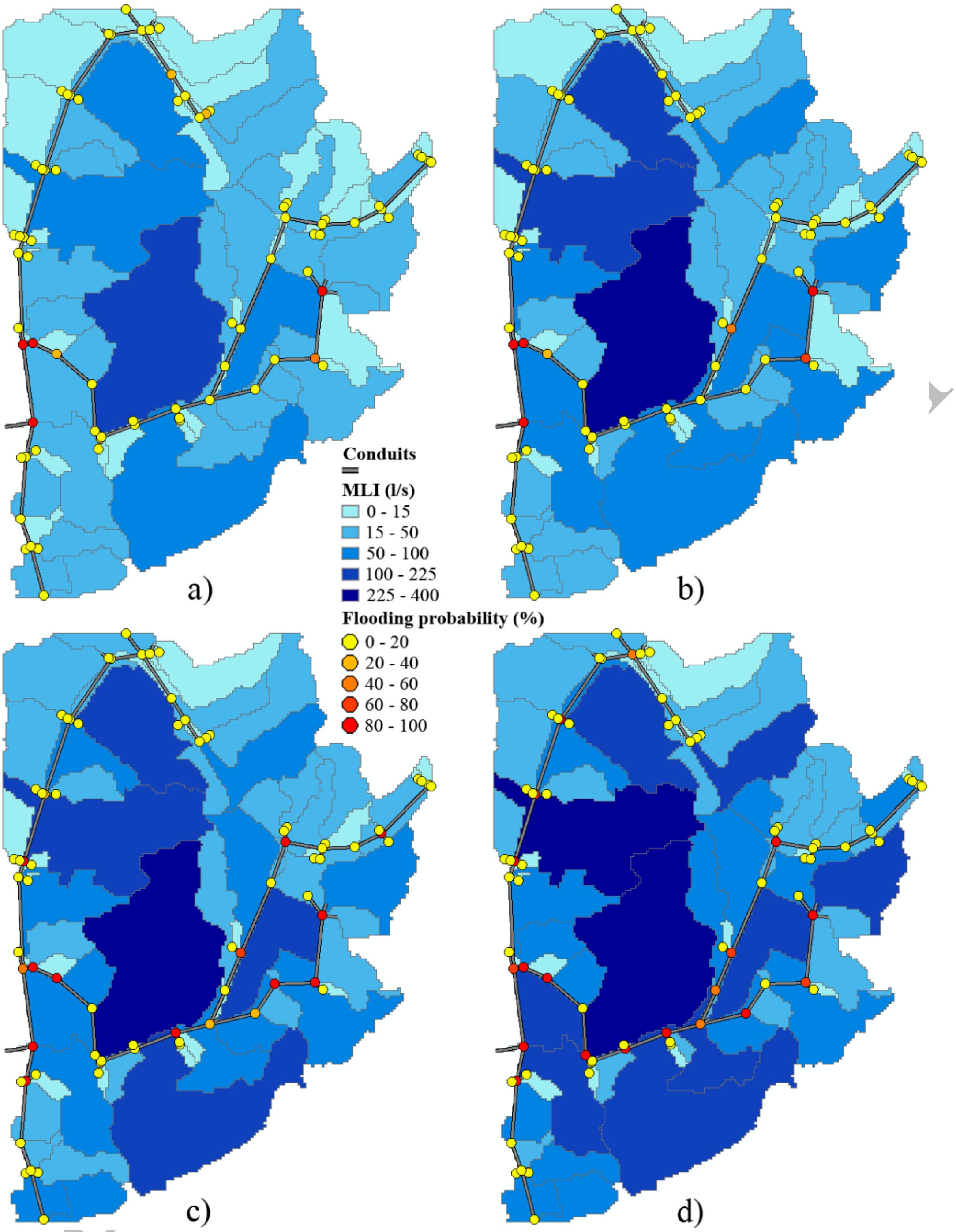

Fig. 7. Maximum Lateral Inflow (MLI, l/s) and Flooding Probability (%) in the subcatchments and nodes in the study area

The results represented in Fig. 7 for short return periods (RCP4.5; T = 5 yr. and RCP8.5; T = 5 yr.) indicated that oversizing the existing sewer network would not result in a relevant improvement of the drainage capacity of the catchment, since aspects like the depth and diameter of its nodes and conduits were not significantly correlated to its flooding susceptibility from a statistical point of view (Table 6). On the contrary, the implementation of Sustainable Drainage Systems (SuDS), also known as BMPs, Low Impact Development (LID) or Water Sensitive Urban Design (WSUD), in those areas with higher values of MLI might decrease the degree of imperviousness of these subcatchments and therefore reduce the amount of lateral inflow received by the nodes of the sewer network too. Di Matteo et al. (2017) and Meerow and Newell (2017) highlighted the importance of the spatial distribution of SuDS to improve urban water-related decision-making processes, in order to maximize their impact by locating them at those sites which most contribute to produce flooding, as illustrated in Fig. 7. In fact, the results presented in Jato-Espino et al. (2016b) demonstrated that the installation of Permeable Pavement Systems at the critical areas shown in Fig. 7 prevented the occurrence of floods in the study catchment when simulating the rainfall scenarios from which these phenomena started to occur in the catchment.

Although the common return periods used to design urban drainage systems range from 2 to 10 years under the assumption of stationarity (Jato-Espino et al. 2016b), Climate Change is expected to accelerate the water cycle in Finland, producing earlier peak flows and increased discharges (Korhonen and Kuusisto 2010). Therefore, exploring the potential consequences derived from storms associated with longer return periods (RCP4.5; T = 50 yr. and RCP8.5; T = 25 yr.) must be a first concern too. According to Table 6, these scenarios would require taking integrated solutions based on extending the capacity of the existing drainage network through larger diameters and invert elevations and smoother slopes, whilst implementing alternative measures to complement its efficiency, such as installing SuDS and/or including new nodes to divide existing subcatchments into smaller areas and reduce high inflow rates in some nodes. The maps illustrated in Fig. 7 can be of great help for focusing on critical areas and optimize the planning and management of resources to prevent floods.

Conclusions

This paper proposed and validated a methodology based on Multiple Regression Analysis (MRA) for assessing flood risk in urban catchments. Multiple Linear Regression (MLR) was applied to select catchment and sewer networks parameters proving to be influential in the occurrence of runoff peaks in urban areas, whilst Multiple Non-Linear Regression (MNLR) and Multiple Binary Logistic Regression (MBLR) models were built to make predictions of Maximum Lateral Inflow (MLI) and Maximum Total Inflow (MTI) in urban catchments and determine the probability of flooding across them, respectively.

The excellent values reached in the MNLR models for the predicted coefficient of determination proved that the combination of catchment and sewer network parameters, especially subcatchment area and cumulative length of preceding conduits, can provide accurate estimates of the maximum peak flow rates in subcatchments and nodes. The subsequent use of MBLR provided high-accuracy prediction models to determine the flooding probability associated with the nodes of sewer-catchments under different extreme rainfall scenarios produced by Climate Change. The results proved that the implementation of Sustainable Drainage Systems (SuDS) might be enough to mitigate floods for the return periods commonly used for urban designs, whilst integrated approaches combining both conventional and alternative water management measures would be required to deal with more extreme scenarios. The fact that these outcomes were based on easy to acquire and/or produce parameters and their relationships were solid in both physical and statistical terms enabled the extrapolation and generalization of the proposed approach to other case studies, since the interpretation and application of MRA is simple and compatible with Geographic Information Systems (GIS).

This methodology is presented as an accessible framework to support the adoption of measures by administrative entities for facilitating their drainage management planning actions and maximizing their impact through their implementation at strategic sites in terms of flood susceptibility. Although the reliability of the results to which it led is not compromised by the location of the study area, further research should consider the application of this methodology to other urban catchments with larger areas, different climate conditions and more complex drainage systems, in order to regionalize the development of prediction models according to the degree of similarity of distinct zones worldwide. The other main future line of action to continue this research should consist of exploring the automation of the proposed methodology through easy-to-use interfaces and/or support tools, so that potential decision-makers and water resources planners without expertise in the statistical techniques considered might apply them by merely providing a series of basic weather and physical inputs.

Acknowledgments

This paper was possible thanks to the research projects SUPRIS-SUReS (Ref. BIA2015-65240-C2-1-R MINECO/FEDER, UE) and SUPRIS-SUPeI (Ref. BIA2015-65240-C2-2-R MINECO/FEDER, UE), financed by the Spanish Ministry of Economy and Competitiveness with funds from the State General Budget (PGE) and the European Regional Development Fund (ERDF). The authors wish to express their gratitude to all the entities that provided the data necessary to develop this research: Helsinki Region Environmental Services Authority HSY, Map Service of Espoo, National Land Survey of Finland, Geological Survey of Finland, EURO-CORDEX and European Climate Assessment & Dataset.

References

Aiken, L. S., West, S. G., and Pitts, S. C. (2003). "Multiple Linear Regression." *Handbook of Psychology*, John Wiley & Sons, Inc., New York (U.S.).

Akaike, H. (1973). "Information theory as an extension of the maximum likelihood principle." *Second International Symposium on Information Theory*, B. N. Petrov, and F. Csaki, eds., Akademiai Kiado, Budapest (Hungary), 267-281.

Ashley, R., Garvin, S., Pasche, E., Vassilopoulos, A., and Zevenbergen, C. (2007). *Advances in Urban Flood Management.* CRC Press, London (U.K.).

Barco, J., Wong, K. M., and Stenstrom, M. K. (2008). "Automatic calibration of the U.S. EPA SWMM model for a large urban catchment." *Journal of Hydraulic Engineering*, 134(4), 466-474.

Beck, N. G., Conley, G., Kanner, L., and Mathias, M. (2017). "An urban runoff model designed to inform stormwater management decisions." *Journal of Environmental Management,* 193 257-269.

Di Matteo, M., Dandy, G. C., and Maier, H. R. (2017). "Multiobjective Optimization of Distributed Stormwater Harvesting Systems." *Journal of Water Resources Planning and Management,* Just Released.

Dongquan, Z., Jining, C., Haozheng, W., Qingyuan, T., Shangbing, C., and Zheng, S. (2009). "GIS-based urban rainfall-runoff modeling using an automatic catchment-discretization approach: A case study in Macau." *Environmental Earth Sciences,* 59(2), 465-472.

Elliott, A. H., and Trowsdale, S. A. (2007). "A review of models for low impact urban stormwater drainage." *Environmental Modelling and Software,* 22(3), 394-405.

Eshtawi, T., Evers, M., Tischbein, B., and Diekkrüger, B. (2016). "Integrated hydrologic modeling as a key for sustainable urban water resources planning." *Water Research,* 101 411-428.

Fisher, R. A. (1925). *Statistical Methods for Research Workers.* Cosmo Publications, Edinburgh (Scotland).

Guan, M., Sillanpää, N., and Koivusalo, H. (2015). "Modelling and assessment of hydrological changes in a developing urban catchment." *Hydrological Processes,* 29(13), 2880-2894.

Hammond, M. J., Chen, A. S., Djordjevic, S., Butler, D., and Mark, O. (2015). "Urban flood impact assessment: A state-of-the-art review." *Urban Water Journal,* 12(1), 14-29.

Hanington, P., To, Q. T., Van, P. D. T., Doan, N. A. V., and Kiem, A. S. (2017). "A hydrological model for interprovincial water resource planning and management: A case study in the Long Xuyen Quadrangle, Mekong Delta, Vietnam." *Journal of Hydrology,* 547 1-9.

Hegger, D. L. T., Driessen, P. P. J., Wiering, M., Van Rijswick, H. F. M. W., Kundzewicz, Z. W., Matczak, P., Crabbé, A., Raadgever, G. T., Bakker, M. H. N., Priest, S. J., Larrue, C., and Ek, K. (2016). "Toward more flood resilience: Is a diversification of flood risk management strategies the way forward?" *Ecology and Society,* 21(4), 52.

Hoornweg, D., Freire, M., Lee, M. J., Perinaz Bhada-Tata, P., and Yuen, B. (2011). *Cities and Climate Change: Responding to an Urgent Agenda.* World Bank Publications, Washington, D.C. (U.S.).

Hosmer, D. W., and Lemeshow, S. (2000). *Applied Logistic Regression.* John Wiley & Sons, New York (U.S.).

Huntington, T. G. (2006). "Evidence for intensification of the global water cycle: Review and synthesis." *Journal of Hydrology,* 319(1-4), 83-95.

Huong, H. T. L., and Pathirana, A. (2013). "Urbanization and climate change impacts on future urban flooding in Can Tho city, Vietnam." *Hydrology and Earth System Sciences*, 17(1), 379-394.

Jato-Espino, D., Charlesworth, S. M., Bayon, J. R., and Warwick, F. (2016a). "Rainfall–Runoff Simulations to Assess the Potential of SuDS for Mitigating Flooding in Highly Urbanized Catchments." *International Journal of Environmental Research and Public Health*, 13(1), 149.

Jato-Espino, D., Sillanpää, N., Charlesworth, S. M., and Andrés-Doménech, I. (2017). "A simulation-optimization methodology to model urban catchments under non-stationary extreme rainfall events." *Environmental Modelling and Software*, In Press.

Jato-Espino, D., Sillanpää, N., Charlesworth, S. M., and Andrés-Doménech, I. (2016b). "Coupling GIS with Stormwater Modelling for the Location Prioritization and Hydrological Simulation of Permeable Pavements in Urban Catchments." *Water (Switzerland)*, 8(10), 451.

Knebl, M. R., Yang, Z. -., Hutchison, K., and Maidment, D. R. (2005). "Regional scale flood modeling using NEXRAD rainfall, GIS, and HEC-HMS/ RAS: A case study for the San Antonio River Basin Summer 2002 storm event." *Journal of Environmental Management*, 75(4 SPEC. ISS.), 325-336.

Korhonen, J., and Kuusisto, E. (2010). "Long-term changes in the discharge regime in Finland." *Hydrology Research*, 41(3-4), 253-268.

Meerow, S., and Newell, J. P. (2017). "Spatial planning for multifunctional green infrastructure: Growing resilience in Detroit." *Landscape and Urban Planning*, 159 62-75.

Minitab Inc. (2016). "Minitab® 17." https://www.minitab.com/en-us/products/minitab/ (11/09, 2016).

Moss, R., Babiker, M., Brinkman, S., Calvo, E., Carter, T., Edmonds, J., Elgizouli, I., Emori, S., Erda, L., Hibbard, K., Jones, R., Kainuma, M., Kelleher, J., Lamarque, J. F., Manning, M., Matthews, B., Meehl, J., Meyer, L., Mitchell, J., Nakicenovic, N., O'Neill, B., Pichs, R., Riahi, K., Rose, S., Runci, P., Stouffer, R., van Vuuren, D., Weyant, J., Wilbanks, T., van Ypersele, J. P., and Zurek, M. (2008). "III. "Representative Concentration Pathways"." *Towards New Scenarios for Analysis of Emissions, Climate Change, Impacts, and Response Strategies*, Intergovernmental Panel on Climate Change (IPCC), Geneva (Switzerland), 5-25.

Osbourne, J. W., and Waters, E. (2002). "Four Assumptions of Multiple Regression That Researchers Should Always Test." *Practical Assessment, Research & Evaluation*, 8 1-7.

Sillanpää, N., and Koivusalo, H. (2015). "Impacts of urban development on runoff event characteristics and unit hydrographs across warm and cold seasons in high latitudes." *Journal of Hydrology*, 521 328-340.

Smith, K., and Ward, R. (1998). *Floods: Physical Processes and Human Impacts.* John Wiley & Sons, Chichester (U.K.).

Tingsanchali, T. (2012). "Urban flood disaster management." *3rd International Science, Social Science, Engineering and Energy Conference 2011, I-SEEC 2011*, 25-37.

USEPA. (2016). "Storm Water Management Model (SWMM) - Version 5.1.011 with Low Impact Development (LID) Controls." https://www.epa.gov/water-research/storm-water-management-model-swmm (11/09, 2016).

Yao, L., Chen, L., and Wei, W. (2017). "Exploring the linkage between urban flood risk and spatial patterns in small urbanized catchments of Beijing, China." *International Journal of Environmental Research and Public Health*, 14(3).