PHILIPPINE ENGINEERING JOURNAL
PEJ 2024; Vol. 45, No. 2: 72-93

Received: 06 September 2024
Received in Revised Form: 17 November 2024
Accepted: 09 December 2024

Flood Flow Modeling under Nonstationarity in the Urban Watersheds of Legazpi City, Philippines

Virgil B. Bilaro[1,2+], Guillermo Q. Tabios III[3]
[1]*Agricultural and Biosystems Engineering Department, Bicol University Guinobatan*
[2]*Civil Engineering Graduate Program, Institute of Civil Engineering, College of Engineering, University of the Philippines Diliman, Quezon City 1101, Philippines*
[3]*Institute of Civil Engineering, College of Engineering, University of the Philippines Diliman, Quezon City 1101, Philippines*
Corresponding author: vbbilaro@bicol-u.edu.ph; vbbilaro@up.edu.ph

Abstract – *Traditional flood frequency analysis assumes stationary conditions (i.e. the mean and other statistical properties are unchanging) prevail in the physical and climatological element driving the phenomenon. With climate change and rapid landcover change, this assumption must be reviewed, and new approaches considering nonstationarity may need to be adopted. Long-term rainfall and land cover data were used to reconstruct historical streamflow in three urban watersheds using deterministic and stochastic techniques. The streamflow models were developed with static and time-evolving built-up land cover area to mimic the effect of land cover change due to urbanization. Annual flood maximum series were developed from each streamflow data set and were tested for trends. The models with time-evolving built-up landcover area (deterministic models) and those with urbanization as co-predictors (stochastic models) were able to generate continuous streamflow time series that yielded flood extremes exhibiting nonstationarity. The annual flood maxima were fitted onto stationary and nonstationary Generalized Extreme Value distribution models using Bayesian approach and successively tested for goodness-of-fit and parsimony. All the annual flood series from both deterministic and stochastic models satisfactorily fit both the stationary and nonstationary Generalized Extreme Value distribution, with the stationary models exhibiting better fit for streamflow models of watersheds with static urbanization scenarios; and the nonstationary models exhibiting better fit for streamflow models of watersheds with evolving urbanization scenarios. In terms of parsimony, the stochastically generated flood models are better than those developed from deterministic models as evidenced by the lower Akaike Information Criterion and Bayesian Information Criterion values for all watersheds. The probability of exceedance of floods through some threshold magnitude increases under nonstationary conditions.*

Keywords: *extreme value distribution; flood frequency; nonstationarity; return period; urbanization; land cover change; streamflow model*

I. INTRODUCTION

The Bicol Region, an administrative region covering six provinces with a distinct socio-cultural and economic identity, is located at the southern tip of the island of Luzon, Philippines. Rainfall and tropical cyclones driven by the monsoons and the easterlies define the climate of the Bicol Region [1]. These climatic elements results to a large amount of annual precipitation and frequent occurrence of intense rainfall events leading to flooding. The changing climate and anthropogenic activities - resulting to land use and land cover changes as well as the infrastructure that are changing the landscape - are expected to affect the behaviour of the water cycle in the region. It is expected that the shifting nature of flooding will continue to influence the lives of the people in the region in the future. The City of Legazpi, being one of two regional urban centers in the Bicol Region, is home to many communities vulnerable to flooding.

Climate change and land use change have altered the magnitude and frequency of occurrence of extreme events such as floods. Understanding the nonstationary nature of random hydroclimatic variables could better equip engineering design practitioners and local government planners in addressing the present challenges and future uncertainties in reducing and managing hydrologic disaster risks.

1.1 Nonstationarity in hydroclimatic time series

A time series is stationary if its statistical properties remain constant over time; while one is nonstationary, if it has statistical properties that change with time [2, 3]. Stationarity, or nonstationarity, could be observed on the mean, the variance or any other higher-order moments around the mean of a random variable. Nonstationarity can be caused by various deterministic components in the time series. The most common types are trends, jumps/shifts and periodicities. A trend is a slow but steady increasing or decreasing change in the mean of a time series triggered by an alteration in the hydrologic or climatic environment. A jump is a sudden change in the mean at some time step in the time series due to extreme conditions. Periodicity is the behaviour of a time series wherein similar values are repeated after some elapsed time. For many climatic or hydrological variables, this is usually due to seasonality.

Nonstationarity can be initiated by natural and anthropogenic causes. Large-scale climatic variability and natural disruptions (such as volcanic eruption or a large landslide) are some examples of natural factors resulting to nonstationarity [4]. The sources of human-induced nonstationarity are land use and land cover changes, dams and other flow control structures, and greenhouse gas induced climate change [5]. The natural [sinusoidal] fluctuating pattern of the different drivers of the climatic system result in the periodicity of many hydroclimatic time series whose frequencies could be observed in the seasonal, annual or even decadal time scales. More important, however, is the interaction of the different climate driving phenomena (spanning varying temporal and spatial scales) that results in either attenuated or amplified responses in any or all climatic variables (e.g. precipitation) and hydrologic elements (e.g. streamflow) resulting to an increasing or decreasing trend. This interaction is further exacerbated by the effects of climate change. A trend is the more common manifestation of nonstationarity in hydroclimatic time series [6]. It can be caused by both natural and anthropogenic factors [7]. The hydrology in a certain location will definitely be impacted by changes in the hydroclimatic system brought about by these natural and anthropogenic components. A trend could result in more frequent occurrence of extreme events such as flooding. More frequent occurrences of floods of a certain magnitude could render a water resource infrastructure's design capacity to be compromised resulting in the reduction of its serviceable lifespan at the very least, or a catastrophic failure as the worst-case scenario in the future.

1.2 Analysis of nonstationary extreme events in hydrology

The study of extreme events such as typhoons, floods, droughts and earthquakes gave rise to the development of the extreme value theory [8]. The Gumbel, Frechet and Weibull distribution are three types of extreme value distributions commonly discussed in most hydrology and statistics textbooks [2, 8, 9]. The Generalized Extreme Value (GEV) distribution is a three-parameter distribution that is ideal for flood analysis. Its probability density function (PDF) and cumulative distribution function (CDF) are presented [10] as:

$$f_x(x) = \frac{1}{\alpha}\exp[-(1-k)y - e^{-y}];y = \begin{cases} -k^{-1}\ln\left[1 - k\frac{(x-\beta)}{\alpha}\right], & k \neq 0 \\ \frac{(x-\beta)}{\alpha}, & k = 0 \end{cases}$$

And

$$F_x(x) = \exp[-\exp(-y)]$$

Where $\hat{\alpha}$ - scale parameter estimate
$\hat{\beta}$ - location parameter estimate, and
k – shape parameter estimate

For random variables with time varying distribution, nonstationarity could be modelled by allowing one or more of its parameters to vary with time. In the case of the GEV distribution presented above, a linear trend, say, in the location parameter is offered by Coles [9]. Other more complex models were also mentioned by the same author, such as one in the quadratic form and a change-point model. An exponential function is commonly used for the scale parameter to ensure that its value remains positive for all values of t [9].

Another common way of expressing nonstationarity is through covariates. A covariate is a variable the behaviour of which may be related to the extreme value series of another. As an example, Prosdocimi et al., [11] modelled the location parameter as a function of the 99th percentile of the daily rainfall for a nonstationary flood model. A similar example is presented by Coles [9] using the southern oscillation index (SOI) as a covariate for the location parameter.

These examples illustrate that nonstationarity in extreme value datasets can be represented via models that either represent any or all of the distribution parameters as varying over time or those that employ covariates that vary with time. It can be further stated that time itself could be treated as a covariate and is thus considered as such by Prosdocimi et al., [11] in developing nonstationary flood frequency models in an urbanizing catchment. Alternatively, time could be viewed as a proxy for covariates that are time variant [12] but which cannot be accounted for by physically or quantitatively available data or measurement [13].

1.3 Analysis of the frequency of extremes

The occurrence of extreme natural events such as floods, typhoons, storm surges and droughts is of special interest in water resources engineering. Their magnitude and the regularity by which they happen are important factors that influence the design and construction of critical infrastructure such as dams, bridges, sea walls, and irrigation systems to name a few. Since the above-mentioned natural events involve random variables, these design questions are best solved using probabilistic techniques that deals with extreme value variables. The common approach, however, is to assume stationarity, or to eliminate the influence of nonstationarity. In one of the early works on extreme value statistics, Gumbel [14] points out that the parameters of the extreme values' distribution must remain constant (i.e. must remain stationary in time or space), or the influence (of nonstationarity) must be considered or eliminated. Little attention was paid to time-space variations of hydroclimatic variables in the early works on extreme natural events until the recent focus on climate change and land cover change due to urbanization and other land-use transformations due to anthropogenic causes. This may be attributed to the computational complexity of considering nonstationarity and the lack of long-term data early on [15, 16].

1.3.1 Return period of extremes under stationary conditions

From the earliest stages of the development of the methodologies that use the concepts of return period in water resources engineering design, the condition of stationarity has always been imposed. Although authors such as Gumbel [17] have emphasized the alternative case of nonstationarity, most hydrology textbooks such as [18], [19], [20] or [3] traditionally present the treatment of recurrence interval assuming stationary conditions. The return period is a concept familiar to most civil and water resources engineers. The return period is the average duration of the time interval of occurrence between events equalling or exceeding a certain magnitude [18]. The familiar form of the stationary return period equation as presented by Chow et al., [18] is:

$$\mathrm{E[t]} = \frac{1}{\mathrm{p}} = \mathrm{T}$$

Where p – the probability of occurrence of an event
T – return period

1.3.2 Return period of extremes under nonstationary conditions

Salas & Obeysekera, [21] developed a nonstationary return period metric for hydrologic events varying through time by having their exceedance probability also varying through time. For a probability p_t that has an increasing trend, the expected waiting time (EWT), of the first occurrence of the success event [22], i.e nonstationary return period:

$$\mathrm{T} = \mathrm{E[X]} = \sum_{x=1}^{x_{max}} x f(x) = \sum_{x=1}^{x_{max}} x p_x \prod_{t=1}^{x-1} (1 - p_t) = 1 + \sum_{x=1}^{x_{max}} \prod_{t=1}^{x} (1 - p_t)$$

If the probability p_t has a decreasing trend, there will come a time x when p_t will be zero. i.e. the exceedance event will not occur, or the nonexceedance probability q_t will conversely increase. It may likewise be possible that the event of decreasing magnitude will converge exponentially to a future constant value [21]. In this case, the return period would be of the form:

$$\mathrm{T} = \mathrm{E[X]} = \sum_{x=1}^{\infty} x p_x \prod_{t=1}^{x-1} (1 - p_t)$$

1.4 Nonstationarity in flood frequency analysis

The recurrence of floods under nonstationary conditions has gained considerable interest among hydrologists and water resources engineers [23]. Several studies have been conducted to establish a link between nonstationarity in extreme flood events and an explanatory variable [13] such as land use changes [11], annual maximum rainfall [23, 24, 25] and population trends [24], sea surface temperature [26], El Niño Southern Oscillation [27], time [7, 23, 25, 28, 29], climate indices and reservoir indices [28, 30] among others. Nonstationary flood models have been developed by fitting annual maximum flows [7, 11, 13, 24, 25, 28, 31, 32], partial duration series [7, 11, 27], or seasonal maximum flows [7, 11, 29] to common probability distributions such as the log normal, generalized extreme value [25] and log Pearson III models for the annual maxima, or the Poisson and generalized Pareto [27] for the partial duration series. The parameters for the afore-mentioned distributions are estimated using Maximum likelihood [7, 25, 29], Bayesian approach [25], least-squares based methods [33] or by using a tool such as the generalized additive model for location, scale and shape (GAMLSS) [23, 24, 28, 29, 32] to model the flood time series under nonstationarity. Uncertainty is taken into account by computing the confidence interval for the desired flood quantile with delta [27, 31], bootstrap [31], and profile likelihood [31] method. Meanwhile, [34] used the equidistant cumulative distribution function method and the equivalent reliability method to take into account the influence of model parameter and precipitation projection uncertainty in estimating the design flood under nonstationarity. Goodness-of-fit is typically evaluated using AIC [23], but other metrics have also been employed such as the asymptotic ratio test and Q-Q plots [27].

The study focused on the development of nonstationary flood frequency models for the urban watersheds of Legazpi City. It aims to demonstrate that despite the lack of observational records, flood frequency analysis on urban watersheds can be done by generating or reconstructing historical flood time series based on precipitation and land use and land cover change. The study shows that nonstationary flood frequency models can be developed from the reconstructed historical flood time series with precipitation and land use and land cover as covariates. It also provides a means to evaluate the difference in the frequency probabilities of flooding under stationarity and nonstationarity in the urban watersheds.

II. METHODOLOGY

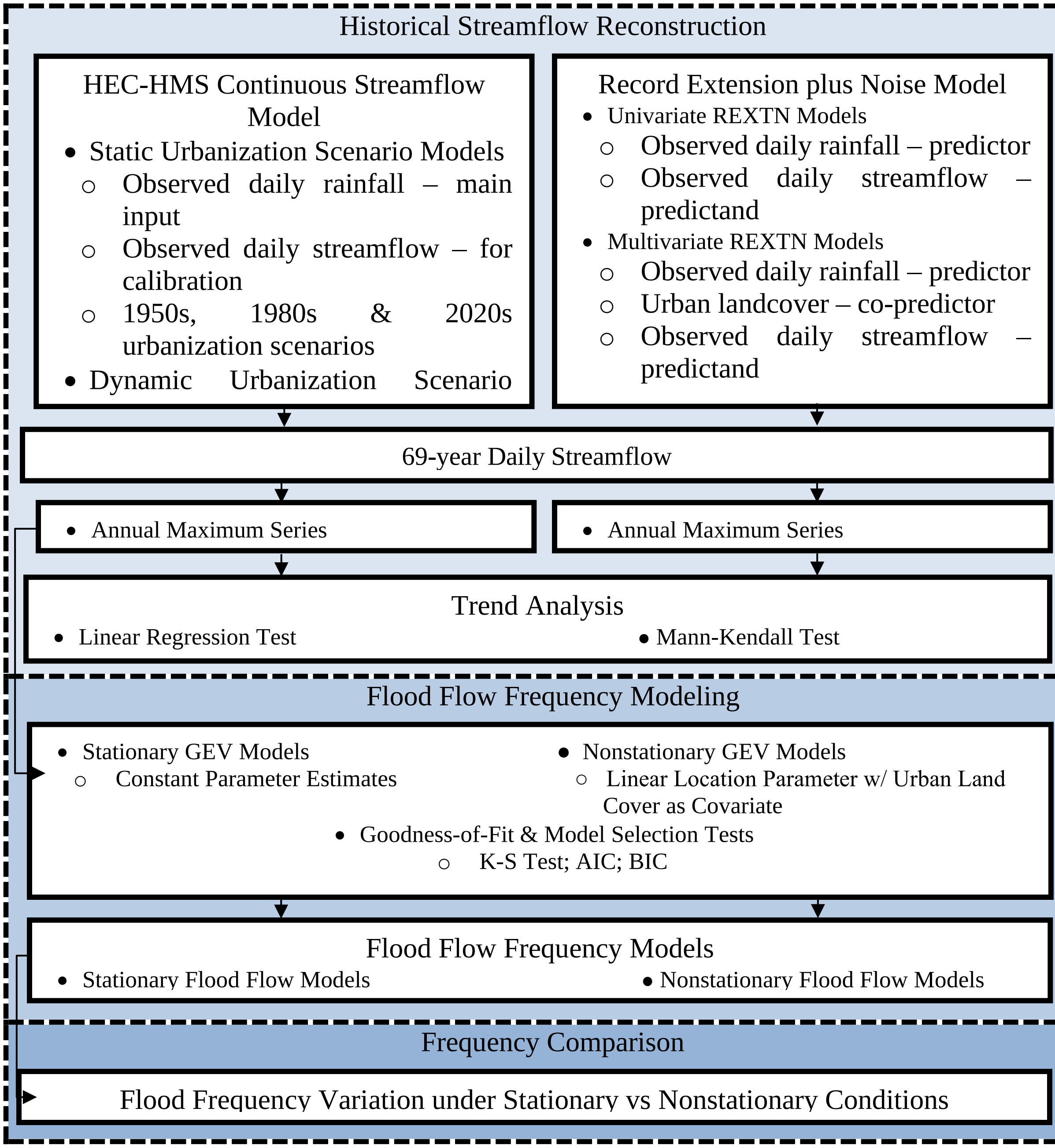


Figure 1 Methodological framework

The methodological framework is represented by the diagram in Figure 1 in which the sequence of steps is connected by arrows illustrating the process from the watershed model development to flood flow frequency analysis. The framework provides an overview the major steps and the associated methodological decisions in the whole research process.

2.1 The study area

Figure 2 shows the selected urban watersheds for this study. The watersheds were selected based on their proximity to an urban center and PAG-ASA monitoring station, and availability of streamflow gauge records. Based on these criteria two ungauged and one gauged urban watersheds were chosen. The ungauged watersheds are named Legazpi subbasin 1 and subbasin 2. Legazpi subbasin 3 is a gauged watershed with a gauging station in Yawa River. Legazpi Subbasin 1 covers the southern sector of the City of Legazpi with an area of 2,236 hectares. It is projected that the southern portion of the city will increasingly become more urbanized in the future since this is the only viable expansion area of the city that is continuously developing and growing more populous. The subbasin is drained by the Sagumayon River on the upstream, which connects to the Macabalo River that drains directly to the Albay Gulf. Legazpi Subbasin 2 covers the main urban and commercial district of the city with an area of 578.47 hectares. A pumping station is constructed at the outlet of this sub-watershed to relieve the city of flood waters that accumulate within the city center during a storm event. Legazpi Subbasin 3 covers the northern part of the urban section of the city with an area of 6,232.04 hectares. The watershed originates from the crater of Mayon volcano, and is the main catchment of its south eastern quadrant. Its main drainage is the Yawa River that flows directly into the Albay Gulf. The Bicol Regional Center that houses most of the government regional offices is located inside the watershed.

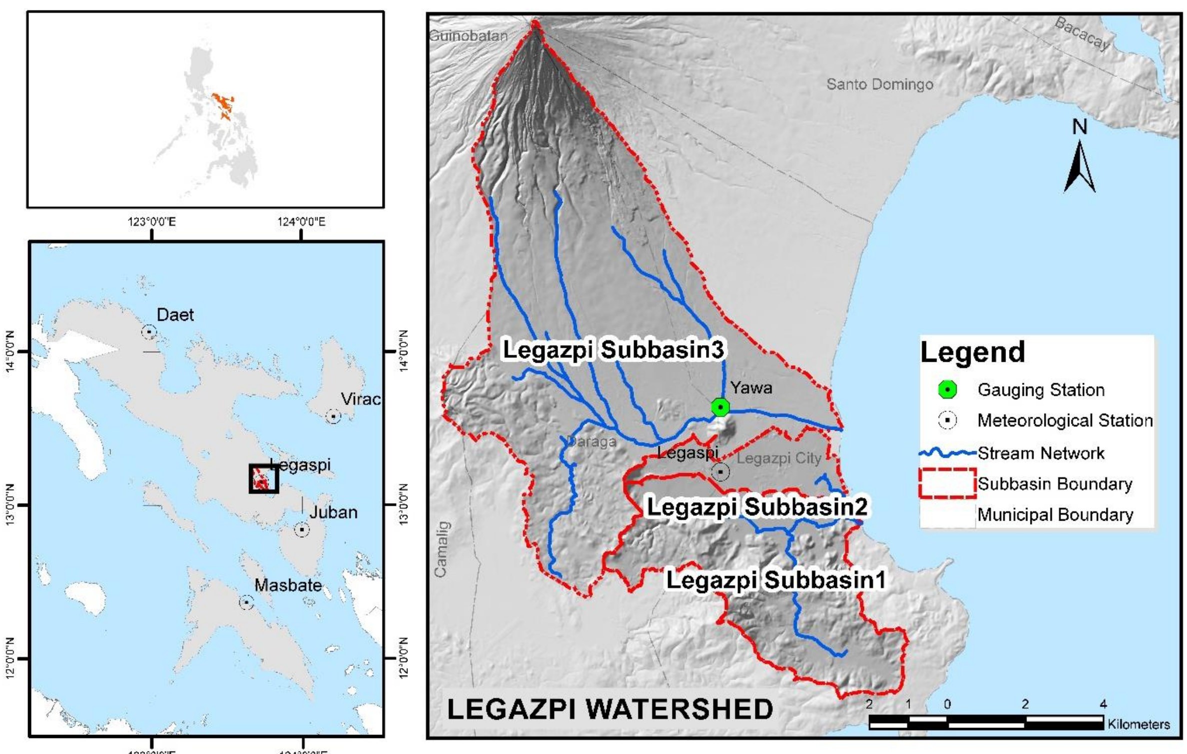


Figure 2 Urban watersheds in the City of Legazpi

2.2 Reconstruction of historical streamflow

Existing streamflow records are short and intermittent in most river monitoring stations in the country. To faithfully capture the nonstationary characteristics of flooding in the selected study area, deterministic and stochastic techniques were used and compared. Actual streamflow records from the Water Projects Division, Bureau of Design of the Department of Public Works and Highways (DPWH) were used to calibrate and validate the models. From the model outputs, a flood time series long enough to be amenable to nonstationary characterization were obtained. The deterministically and stochastically generated series were compared to determine the advantages and disadvantages of the two approaches.

2.2.1 HEC-HMS Continuous Streamflow Model

A basin model of the urban watersheds was set-up in the Hydrologic Modeling System (HEC-HMS) version 4.2 developed, and released in 2016, by the US Army Corps of Engineers Institute for Water Resources Hydrologic Engineering Center (CEIWR-HEC). Using the basin model manager, the watersheds were represented by subbasin elements, which are then connected to a common sink element that represents the water body where all the subbasins drain into. In creating the model domain, the choice of the method to characterize the basin elements' properties is generally driven by its appropriateness to long-term continuous flow simulation. The HEC-HMS program has numerous computational methods that can be used to control the configuration settings of the different hydrologic elements in the basin model [35] but not all are compatible with continuous modelling simulations.

Table 1. Computation methods for HEC-HMS subbasin model

Subbasin Model Component	Method
Area	GIS based analysis
Canopy	Simple Canopy
Surface	Simple Surface
Loss	Soil Moisture Accounting
Transform	Clark Unit Hydrograph
Baseflow	Linear Reservoir

To take into account the increasing urbanization in the watersheds, the whole simulation epoch of 69 years (1951 – 2019) is divided into 5-year periods with the effect of the changing land cover reflected in each period. In particular, the % impervious area is updated for each period based on the prevailing built-up area in the subbasin for the period in question. The built-up area for a particular period is estimated from the regression equations for each subbasin based on Landsat derived land cover maps. Simulations based on static land cover scenarios were also developed wherein % impervious areas corresponding to the 1950s, 1980s and 2020s were used. The calculation methods used in developing the subbasin components of the HEC-HMS models are listed in Table 1.

The meteorologic information needed to set the boundary conditions for how the basin models will perform during continuous simulation were obtained from the observed records of the PAG-ASA Legaspi station for the three urban sub-watersheds of Legazpi. For the ungauged sub-watersheds, since the Legaspi station is located in one of the Legazpi subbasins under consideration, and is very near the other two subbasins, a single meteorologic model based on this station is used in the simulations. Aside from precipitation, maximum and minimum temperatures, relative humidity and wind speed were also used as input data in developing the meteorologic model. The calculation methods used in developing the meteorologic components of the HEC-HMS models are listed in Table 2.

Table 2. Computation methods for HEC-HMS meteorologic model

Meteorologic Model Component	Method
Shortwave Radiation	Bristow Campbell
Longwave Radiation	Satterlund
Precipitation	Gage Weights
Evapotranspiration	Penman Monteith

To improve the numerous parameter estimates of Legazpi subbasin 3, model optimization was conducted by using the historical observation of daily streamflow from Yawa River covering the period of 1980 to 1988. For the ungauged Legazpi subbasins, correction factors were applied based on the optimized parameters of Legazpi subbasin 3. Adjacent Legazpi subbasins (subbasin 1 and 2) where no observed streamflow records are available were calibrated by directly assuming the same optimized parameter values as those of Legazpi subbasin 3 in cases where such approach is acceptable (e.g. parameters that depend on soil characteristics where both gauged and ungauged basins contain similar soil types). For parameters that are not amenable to the previous approach, the values are adjusted by a factor derived from the rate of increase (decrease) between the initial value and the optimized final value of its equivalent parameter in the gauged subbasin.

2.2.2 Record Extension plus Noise Model

The Record Extension plus Noise (REXTN) model of Salas et al., [36] was used as a substitute method to derive the historical streamflow in the urban watersheds. Univariate models were developed using the following equations:

$$\boldsymbol{y_t = \bar{y}_1 + a(y_{t-1} - \bar{y}_1) + b(x_t - \bar{x}_1) + c\varepsilon_t} \qquad \text{(Eq. 1)}$$

Where $\boldsymbol{y_t, x_t}$ -variables representing short record (predictand) and long record (predictor) respectively over the period (N_1) corresponding to the length of the short record series

$\boldsymbol{\bar{y}_1, \bar{x}_1}$ -sample mean of predictand and predictor respectively over period N_1

$\boldsymbol{\varepsilon_t}$ -normally distributed noise with mean 0 and standard deviation 1

And the model parameters can be estimated as:

$$\boldsymbol{\hat{a}} = \frac{s_{y1y1}(1)s_{x1}^2 - s_{x1y1}s_{x1y1}(1)}{s_{x1}^2 s_{y1}^2 - s_{x1y1}^2(1)} \qquad \text{(Eq. 2)}$$

$$\hat{b} = \frac{s_{x1y1} - \hat{a}s_{x1y1}(1)}{s_{x1}^2} \quad \text{(Eq. 3)}$$

$$\hat{c} = \sqrt{s_{y1}^2 - \hat{a}s_{y1y1}(1) - \hat{b}s_{x1y1}} \quad \text{(Eq. 4)}$$

Where s_{wu} - represents the sample lag-0 covariance between variables w_t & u_t

$s_{wu}(1)$ - represents the sample lag-1 covariance between variables w_t & u_t noting that u_t lags one step behind w_t

s_{y1}^2, s_{x1}^2 - sample variance of predictand and predictor respectively over N_1

Meanwhile, the model for multiple explanatory variables has the form:

$$y_t = \bar{y}_1 + a(y_{t-1} - \bar{y}_1) + b_1\left[{x_t}^{(1)} - {\bar{x}_1}^{(1)}\right] + \cdots + b_m\left[{x_t}^{(m)} - {\bar{x}_1}^{(m)}\right] + c\varepsilon_t \quad \text{(Eq. 5)}$$

and the model parameters can be estimated as:

$$\hat{a} = \left[S_{y1y1}(1) - S_{x1y1}^T S_{x1x1}^{-1} S_{x1y1}(1)\right]\left[S_{y1y1}(1) - S_{x1y1}^T(1) S_{x1x1}^{-1} S_{x1y1}(1)\right]^{-1} \quad \text{(Eq. 6)}$$

$$\hat{b} = \left[S_{x1y1}^T - \hat{a}S_{x1y1}^T(1)\right]S_{x1x1}^{-1} \quad \text{(Eq. 7)}$$

$$\hat{c}^2 = S_{y1y1} - \hat{a}S_{y1y1}(1) - \hat{b}S_{x1y1} \quad \text{(Eq. 8)}$$

Where S_{UV} - is the lag-0 covariance matrix between variable vectors U_t & V_t

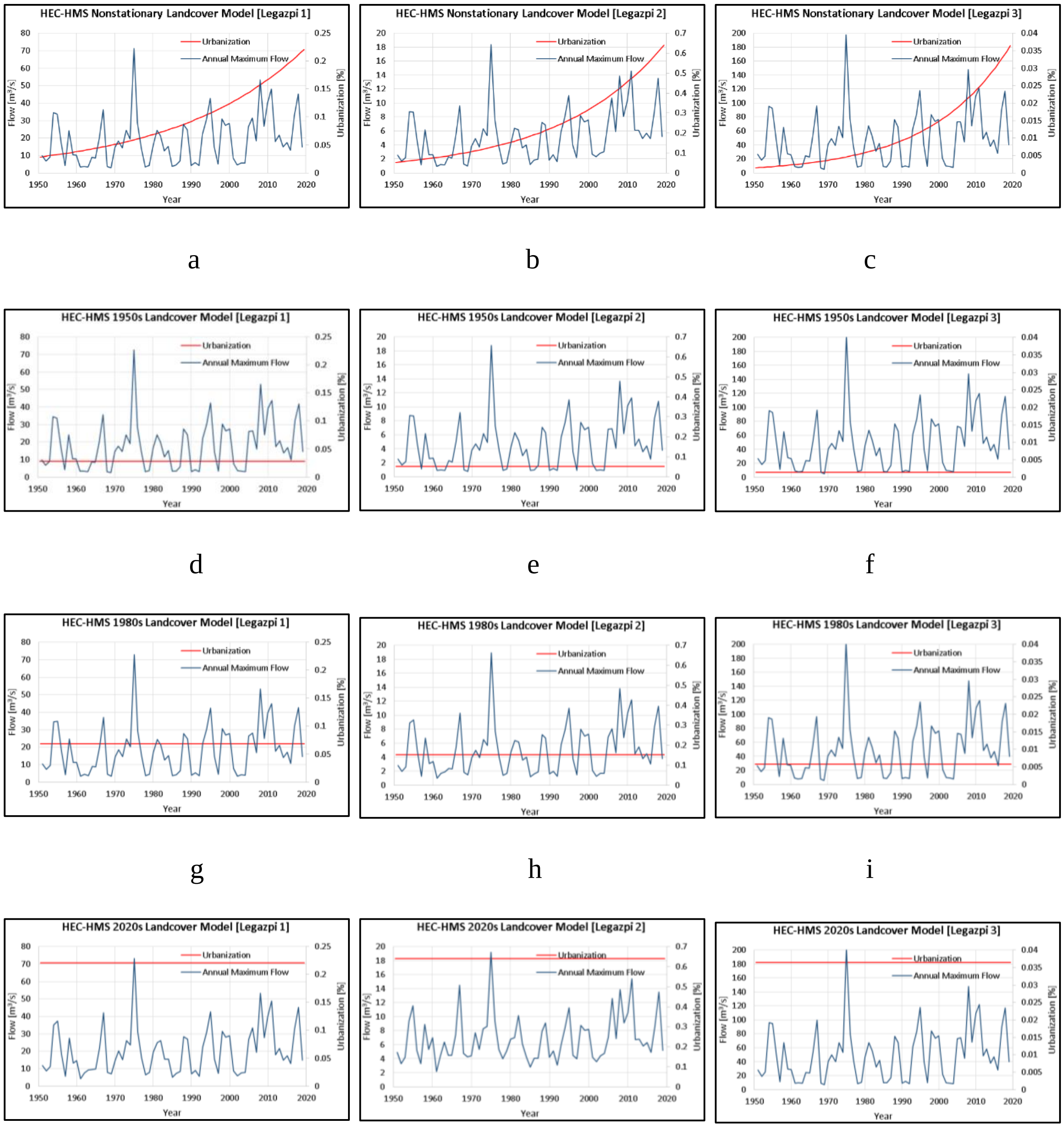
$S_{UV}(1)$ - is the lag-1 covariance matrix between variable vectors U_t & V_{t-1}

The superscript -1 stands for inverse while T stands for transpose.

The rainfall record from the same PAG-ASA stations as in the HEC-HMS models were used as the explanatory variable, x_t for the single variable model. For the multi-variable model, the rainfall record from PAG-ASA and the landcover/ landuse variables were used as the explanatory variables. The short-term streamflow record of Yawa gauging station of the DPWH were used as data for the predictand, y_t. Daily streamflow ensembles of 100 realizations, each spanning from 1950 to 2019, for both single and multiple variable REXTN models were produced to address the uncertainty inherent in the hydrologic model. Each realization is tested for trends, and the proportion of stationary to non-stationary streamflow time series is noted. Using the sum of absolute differences as the objective function [36], the realization with the minimum value is chosen as the best run and is used in the subsequent flood frequency model fitting analyses.

Following the concepts of regionalization [37] estimates based on the subbasins' physical attributes and regional characteristics of channel flow in the project site were used in deriving the necessary statistics such as the mean, covariance and variance of ungauged basins. The mean was estimated from a regionalized scaling factor derived from the ratio of the mean of the observed streamflow of a gauged basin to the basin area. The covariances and variances are derived using the gauged basin's streamflow data and the ungauged basin's landcover data.

2.3 Flood flow frequency models from reconstructed streamflow

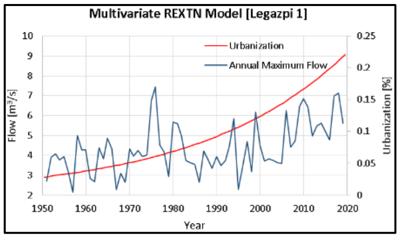


a b c

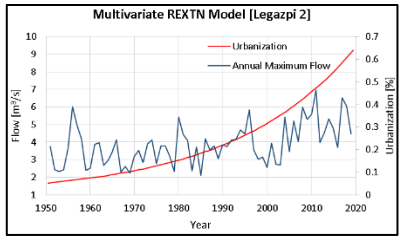
d e f

g h i

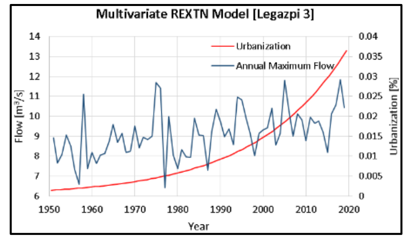
j k l



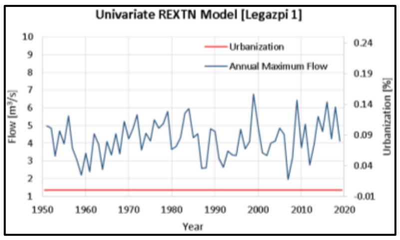
m



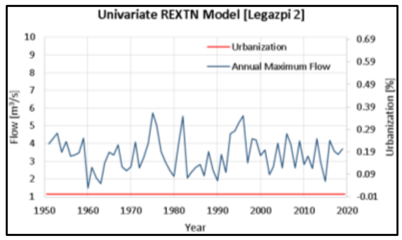
n



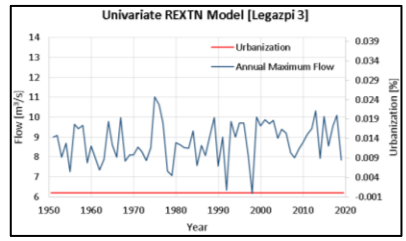
o



p



q



r

Figure 3. Annual maximum streamflow from HEC-HMS (a-l) and REXTN (m-r) models at various urbanization scenarios

From the reconstructed long-term continuous flow models, the highest 1-day discharge in a year were selected to develop the annual maximum flood time series for each watershed. Figure 3 shows the models under different urbanization scenarios. The derived annual maxima series were fitted into stationary and non-stationary extreme value distributions to develop flood frequency models for the urban watersheds. To test for goodness of fit, the Kolmogorov-Smirnov (KS) test was conducted, while the Akaike Information Criterion (AIC) [38], and the Bayesian Information Criterion (BIC) [39] were used to determine the most desirable model.

2.3.1 Stationary GEV distribution models

Following Bayes' rule, the parameters are estimated by multiplying the likelihood function by a prior probability distribution that is subjectively based on previous knowledge, experimental results or prior beliefs regarding the flooding process [40, 41]. A software package for extreme value analysis written in Matlab [42] was used to derive the parameters for the stationary GEV for the annual maximum series. The Bayesian approach is a generalized form of the maximum likelihood method that has been applied numerous times in extreme value analysis [40, 43, 44, 45].

2.3.2 Nonstationary GEV distribution models

Nonstationary models with urbanization as covariates were fitted to the extreme streamflow variables using the above-mentioned extreme value analysis package used for the stationary model. With the extreme flood series obtained from the HEC-HMS and the single and multiple variable REXTN models, nonstationary GEV models were developed with location parameters that are linearly correlated with the urban landcover percentage of the watershed.

2.3.3 Testing for model goodness-of-fit and parsimony

Three tests were conducted to assess and compare the stationary and nonstationary GEV models of the annual flood extremes in the urban watersheds. The Kolmogorov-Smirnov (K-S) test is used to compare if two samples belong to the same distribution. This test serves to confirm if a distribution fits the GEV model. In the Bayesian approach, the simulation is set to generate 6000 realizations, thus the reported results of the K-S test in this study reflects not the test statistic value but instead the rejection rate of the tests for all the generated simulation results [42]. The rejection rate reflects the portion of the output of the simulations wherein the null hypothesis of the K-S test has been rejected. Thus, a lower K-S rejection rate value indicates that a greater portion of the simulation output fits the GEV model. The AIC and BIC are popular selection criteria for model fit and parsimony [46] in hydrologic modelling. The AIC and BIC were used to select between the competing stationary and nonstationary GEV models of the various urbanization scenarios of the HEC-HMS models and the univariate and multivariate REXTN models that best fit the annual flood maxima of the urban watersheds. For both AIC and BIC, the model with the lowest value is selected as the best among competing models. The difference in value however should be significant (>2) for an AIC or BIC rating to be considered superior among competing models.

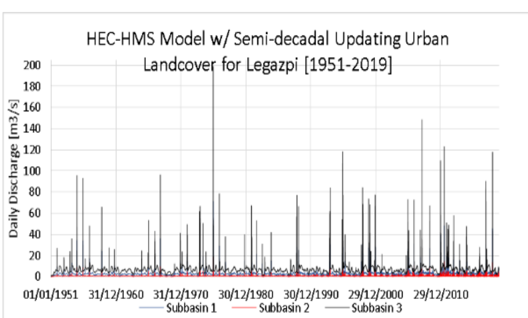
2.4 Frequency modeling and comparison between stationary and nonstationary conditions

The flood magnitude for any return period m is obtained by plugging-in the obtained parameter estimates into the inverse GEV function with $1 - \frac{1}{m}$ as the probability value; or alternatively using the original form of the GEV function equated to $1 - \frac{1}{m}$ [47].

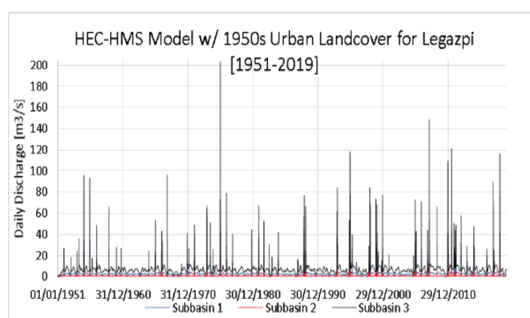
III. RESULTS AND DISCUSSION

3.1 Historical streamflow and trends in annual extreme flows

3.1.1 Simulated historical streamflow



a



b

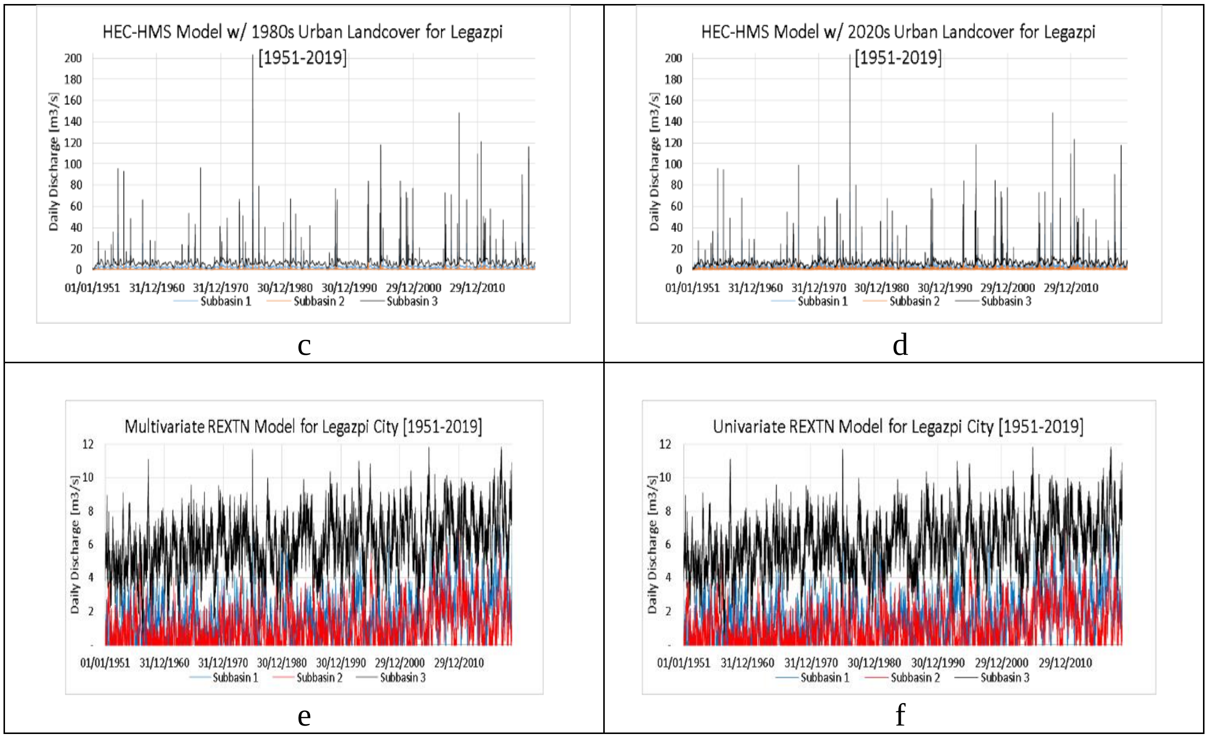


Figure 4. HEC-HMS (a-d) and REXTN (e-f) continuous streamflow simulation outputs

The daily streamflow output, covering the period 1951 to 2019, of the HEC-HMS models and REXTN models are shown in Figure 4. HEC-HMS Models with Nash-Sutcliffe Efficiency (NSE) of 0.339, mean absolute error of 0.2, and root mean square error (RMSE) of 0.3 were developed to generate the historical streamflow for the urbanization scenarios where the % impervious area is being updated every 5 years (a) as well as for scenarios where the % impervious area is not changing for the course of the simulation period (b-d). The evaluation metrics, except for the NSE, show that the model performs very good [48] in simulating the long-term flow of the watershed. The best realization from an ensemble of 100 simulation runs with rainfall and urbanization as predictor variables (e) and rainfall only as predictor variable (f) were used as the reconstructed streamflow series for the REXTN models.

3.1.2 Annual maximum streamflow

Both the deterministic and stochastic models were able to reproduce nonstationary annual maxima of streamflow in the urban watersheds. The HEC-HMS model with % impervious surface that is updated every 5 years, corresponding to the changing built-up area in the watersheds, was able generate nonstationary peak flow; whereas the models with non-changing % impervious surfaces were not able to produce peak flow variables with any significant trends. For the REXTN models, the multi-variable model with rainfall and urbanization as predictors were able to generate a high percentage of streamflow realizations with statistically significant increasing trends. All Legazpi subbasins produced streamflow maxima with increasing trends. The models that produced annual maximum flows with statistically significant trends are presented in Table 3.

Table 3. Watershed models with statistically significant maximum streamflow trends

Watershed	Model	Slope [$m^3/s/yr$]	Trend
Legazpi 1	HEC-HMS – 5yr Urb	0.189	Increasing
	Multivariate REXTN	0.029	Increasing
Legazpi 2	HEC-HMS – 5yr Urb	0.069	Increasing
	Multivariate REXTN	0.028	Increasing
Legazpi 3	HEC-HMS – 5yr Urb	0.466	Increasing
	Multivariate REXTN	0.028	Increasing

3.2 Stationary and nonstationary GEV distribution of annual extreme flows

3.2.1 Parameter estimation of the distribution of streamflow extremes

Table 4 shows the parameters of the stationary and nonstationary GEV distribution for the annual maximum flood of the urban watersheds from various streamflow models. The location parameters of the nonstationary models have linear trends, while the scale and shape parameters have none. The table also shows the goodness-of-fit test results for the said models.

Table 4. Best fitting GEV models for the urban watersheds of Legazpi City

Model	Parameter				Goodness of Fit		
	Location		Scale	Shape	K-S	AIC	BIC
Legazpi Subbasin 1							
HEC-HMS – 5yr Urb (S)	11.36		9.06	0.18	0.84	542.5	549.2
HEC-HMS – 5yrUrb (NS)	10.34	17.64	2.19	0.19	0.91	541.0	550.0
HEC-HMS – 1950 (S)	10.83		9.35	0.17	1.43	543.2	549.9
HEC-HMS – 1950 (NS)	9.60	16.71	2.22	0.18	1.64	543.2	552.2
HEC-HMS – 1980 (S)	11.38		9.15	0.16	1.48	542.2	548.9
HEC-HMS – 1980 (NS)	10.22	15.00	2.20	0.19	1.09	541.8	550.8
HEC-HMS – 2020 (S)	13.41		8.69	0.21	1.51	535.9	542.6
HEC-HMS – 2020 (NS)	11.81	16.24	2.12	0.22	2.39	535.6	544.5
Univariate REXTN (S)	3.82		1.03	-0.19	0.31	209.3	216.0
Univariate REXTN (NS)	5.02	3.53	0.35	-0.38	0.37	209.7	218.7
Multivariate REXTN (S)	3.84		1.09	-0.08	1.01	226.5	233.2
Multivariate REXTN (NS)*	8.67	2.72	0.01	-0.15	1.15	208.4	217.4
Legazpi Subbasin 2							
HEC-HMS – 5yr Urb (S)	3.45		2.46	0.17	2.12	359.0	365.7
HEC-HMS – 5yrUrb (NS)	1.81	7.21	0.74	0.20	1.14	342.5	351.4
HEC-HMS – 1950 (S)	2.81		2.36	0.19	1.67	354.5	361.2
HEC-HMS – 1950 (NS)	2.03	3.93	0.85	0.13	1.29	353.0	361.9
HEC-HMS – 1980 (S)	3.22		2.20	0.21	1.34	350.1	356.8
HEC-HMS – 1980 (NS)	2.51	3.38	0.79	0.20	1.44	348.0	357.0
HEC-HMS – 2020 (S)	5.31		2.20	0.13	1.96	341.4	348.1
HEC-HMS – 2020 (NS)	4.64	3.34	0.76	0.17	2.21	338.4	347.4
Univariate REXTN (S)	2.96		0.88	-0.12	0.87	195.2	201.9

Model	Parameter				Goodness of Fit		
	Location		Scale	Shape	K-S	AIC	BIC
Univariate REXTN (NS)	0.94	2.70	-0.07	-0.18	1.29	197.1	206.1
Multivariate REXTN (S)	3.32		0.96	-0.02	1.85	214.3	221.0
Multivariate REXTN (NS)*	3.66	2.63	-0.18	0.03	0.47	195.3	204.2
Legazpi Subbasin 3							
HEC-HMS – 5yr Urb (S)	29.60		26.41	0.16	1.28	684.6	691.3
HEC-HMS – 5yr Urb (NS)	28.98	11.02	3.24	0.18	1.25	686.6	695.5
HEC-HMS – 1950 (S)	29.88		25.77	0.16	2.45	684.6	691.3
HEC-HMS – 1950 (NS)	28.91	9.83	3.25	0.15	2.58	686.5	695.5
HEC-HMS – 1980 (S)	29.88		25.88	0.17	1.02	685.0	691.7
HEC-HMS – 1980 (NS)	29.40	10.78	3.25	0.16	2.07	686.6	695.5
HEC-HMS – 2020 (S)	30.28		25.62	0.16	1.54	684.7	691.4
HEC-HMS – 2020 (NS)	30.59	8.58	3.26	0.16	1.16	686.7	695.6
Univariate REXTN (S)*	8.39		1.02	-0.28	0.42	199.6	206.3
Univariate REXTN (NS)	20.39	8.26	-0.07	-0.32	0.62	200.1	209.0
Multivariate REXTN (S)	8.89		0.92	-0.11	0.70	200.5	207.2
Multivariate REXTN (NS)	20.58	8.56	0.19	-0.16	0.49	214.9	223.9
(S) – Stationary GEV model (NS) – Nonstationary GEV model * Best fitting model							

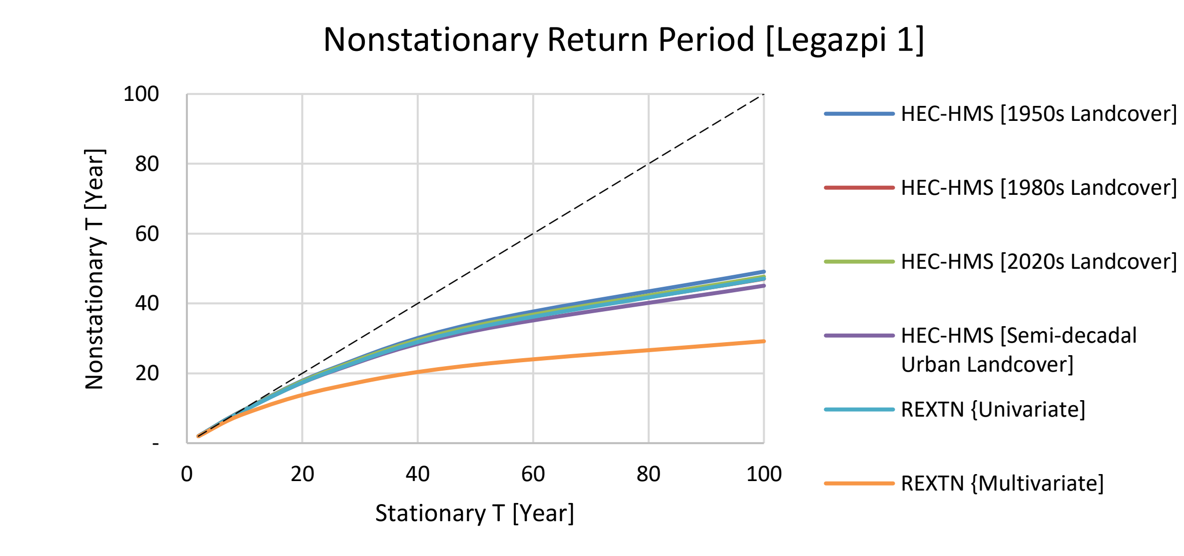
The column for Kolmogorov-Smirnov (K-S) test indicates the rejection rate of the simulation runs; i.e. the proportion of the runs that fit a GEV distribution according to the K-S test. These results demonstrate the appropriateness of taking into consideration nonstationarity in flood modelling of watersheds that are dynamically evolving due to climatic and/or anthropogenic influences. The nonstationary GEV distribution is more appropriate than their stationary counterparts for both physical (Semi-decadal updating HEC-HMS) and stochastic (Multivariate REXTN) flow models that incorporate the effect of shifting landcover, and where the rate of landcover change has attained a certain threshold of significance. The model efficiencies for watersheds where a significant trend in urban area increase is significant have been improved with the nonstationary GEV models. The stationary model is more appropriate for models with static urbanization scenarios.

It should be noted that a significant landcover change alone does not entail a better nonstationary model fit. Among the watersheds, Legazpi subbasin 3 has a comparatively low rate of urbanization, in terms of the modelled increase in built-up area within the watershed. Likewise, the proportion of urban areas in this subbasin relative to other landcover types is lower compared to the other two Legazpi subbasins. In this case, the stationary model is a better fit than its nonstationary counterpart.

3.2.3 Frequency variation under nonstationarity

Figure 5 shows the nonstationary return periods vis-à-vis their stationary counterparts from the HEC-HMS and REXTN models. The diagonal dashed lines across the graphs indicate a one-to-one correspondence between the return periods of flood of certain magnitude in both stationary and nonstationary model scenarios. Thus, the degree of deviation of a certain scenario curve indicates the disparity between the stationary and nonstationary return periods for each model scenario. A curve deviating upward indicates that the nonstationary model is predicting longer return periods (less frequent occurrence) than the stationary model; while a downward deviating curve indicates that the nonstationary model is predicting shorter return periods (more frequent occurrence) than the stationary model. Generally, the models that integrate the influence of built-up land cover expansion over time (i.e. Semi-decadal updating HEC-HMS models and multivariate REXTN models) predicted the shortest nonstationary return periods. This means that nonstationary return periods are smaller in value than the stationary return periods, indicating that the frequency of flood events of certain magnitude are expected to happen more often than previously observed.

The multivariate REXTN model and the HEC-HMS Semi-decadal updating urbanization model have the shortest nonstationary return periods in both subbasins 1 (Figure 5a) and 2 (Figure 5b). Meanwhile, the multivariate REXTN and HEC-HMS Semi-decadal updating urbanization models of subbasin 3 (Figure 5c) do not vary significantly with their univariate and static urbanization scenario counterparts. This is consistent with the goodness-of-fit test results for the third subbasin; in which the stationary GEV model is the appropriate model for annual extreme flows. All HEC-HMS and REXTN models predict nonstationary return periods that are shorter than their stationary counterparts.



a

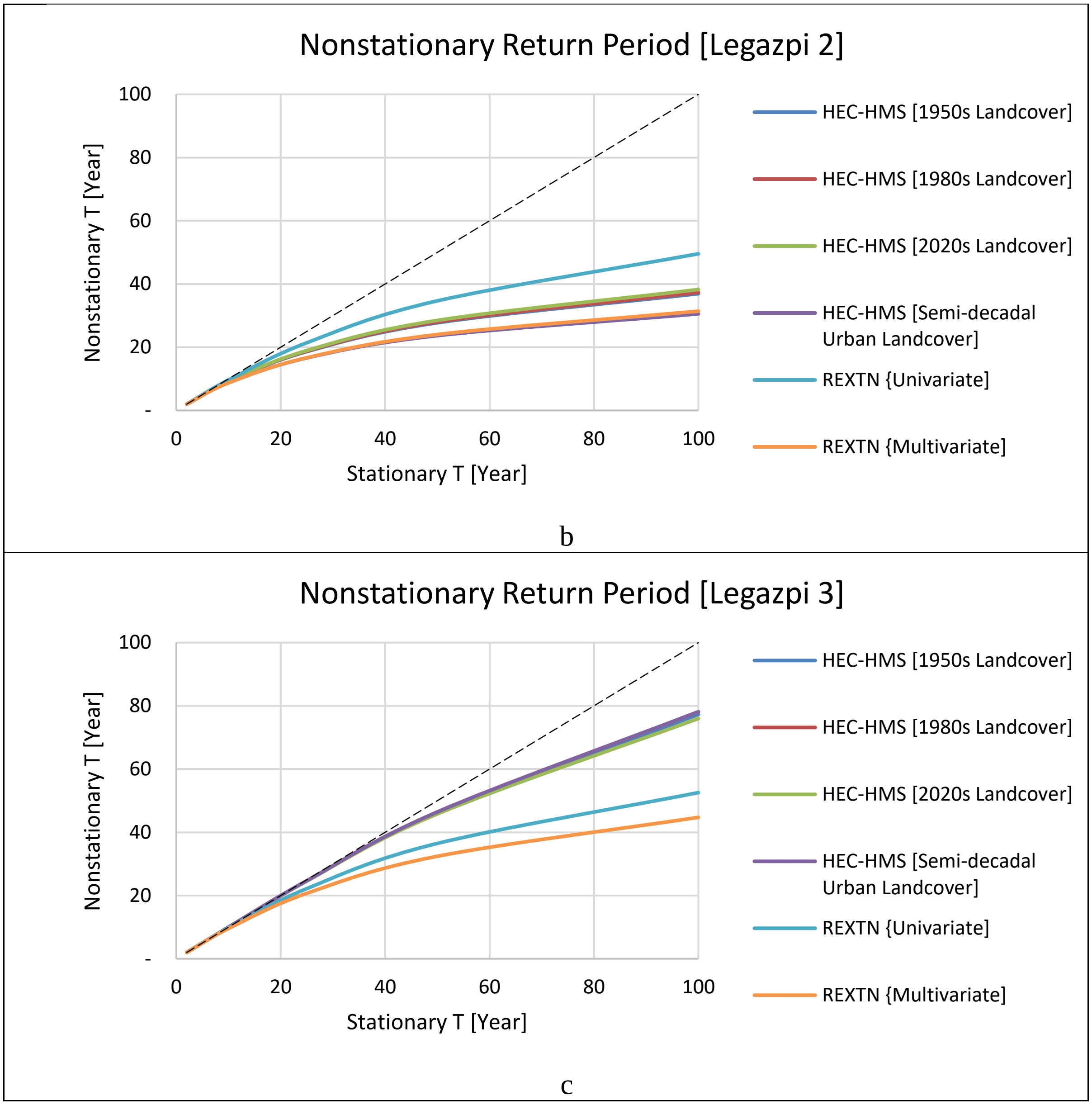


Figure 5. Variation of nonstationary T as a function of the stationary T_0

IV. CONCLUSION

From the analyses of the data inputs and the subsequent development of the above-mentioned models, the following conclusions can be drawn:

Long-term continuous streamflow data can be generated for urban watersheds with limited observational data using deterministic and stochastic techniques with rainfall and built-up land cover as principal driving factors. For the deterministic models, three simulations using static urbanization scenarios and one simulation using a dynamically changing urbanization scheme were developed for each watershed. For the stochastic models, one simulation using only rainfall as predictor and another simulation using both rainfall and urbanization as predictors were developed for each watershed. Trend analyses show that a statistically significant increase can be observed on the annual maximum streamflow for the models with dynamically changing urbanization scenarios. All static urbanization scenario simulations, for both deterministic and stochastic models, produced annual maximum streamflow time series that do not have significant trends.

Stationary and nonstationary GEV models for the flood frequency of three urban watersheds were developed using the Bayesian approach, with all models exhibiting excellent goodness-of-fit based on the Kolmogorov-Smirnov test. The annual maxima of all the models fall within the 90% confidence interval. In terms of efficiency, the AIC and BIC metrics show that the nonstationary GEV model is appropriate for watersheds with significantly increasing urbanization trends, however, if the urban landcover is not significantly large, the stationary model is still more suitable. All the stochastic models performed better than the deterministic models in terms of their AIC and BIC values.

Nonstationary models with urbanization as covariate are predicting flood extremes to be more frequent than the stationary models. The degree by which the nonstationary recurrence interval varies with the stationary recurrence interval become more pronounced for flood quantiles of greater magnitudes. Models that integrate a dynamically changing urban land cover generally have shorter flood recurrence intervals than models with static urban land cover.

The methodologies applied in the study could be used in other urban watersheds of significance in the Philippines to properly assess the effect of urban development in their hydrologic response. In watersheds where urbanization is progressing at significant rates, the nonstationary framework should be employed in quantifying the frequency of floods and assessing the risks involved in the planning and design of critical infrastructure inside the watershed. In developing continuous streamflow models of watersheds, it is recommended that time-varying parameters be used to take into account the effect of anthropogenically induced changes at the watershed in cases where these factors are known to significantly affect the flow regime of critical water bodies inside the watershed.

V. ACKNOWLEDGMENT

This study was funded by the DOST-SEI through the Engineering Research and Development for Technology (ERDT) Graduate Scholarship Program. Grateful acknowledgment is also extended to DOST-PAGASA and the Water Projects Division, Bureau of Design of the DPWH for the climatological and river discharge data.

References:

[1] Puentebella VP. 1992. Climatology of the Bicol Region [Philippines]. PCARRD Book Series.

[2] Maity R. 2018. Statistical methods in hydrology and hydroclimatology. In Springer. https://doi.org/10.1007/978-981-10-8779-0

[3] Wang LK, Yang CT (Eds.). 2014. Modern water resources engineering. Humana Press. https://doi.org/10.1007/978-1-62703-595-8

[4] Salas JD. 1993. Hydrologic time series. In D. R. Maidment (Ed.), Handbook of Hydrology. McGraw-Hill, Inc. First, p. 19.5.

[5] Sivapalan M, Samuel JM. 2009. Transcending limitations of stationarity and the return period process-based approach to flood estimation and risk assessment. Hydrological Processes. 23:1671-1675. https://doi.org/10.1002/hyp

[6] Cohn TA, Lins HF. 2005. Nature' s style: Naturally trendy. Geophysical Research Letters. 32(September): 1-5. https://doi.org/10.1029/2005GL024476

[7] Strupczewski WG, Singh VP, Feluch W. 2001. Non-stationary approach to at-site flood frequency modelling I. Maximum likelihood estimation. Journal of Hydrology. 248(1-4):123-142. https://doi.org/10.1016/S0022-1694(01)00397-3

[8] Kottegoda NT, Rosso R. 2008. Applied statistics for civil and environmental engineers. 2nd ed. John Wiley & Sons Ltd.

[9] Coles S. 2001. An introduction to statistical modeling of extreme values. Springer-Verlag. https://doi.org/10.1007/978-1-4471-3675-0

[10] Hosking JRM, Wallis JR. 1997. Regional frequency analysis: An approach based on L-Moments. Cambridge University Press.

[11] Prosdocimi I, Kjeldsen TR, Miller JD. 2015. Detection and attribution of urbanization effect on flood extremes using nonstationary flood-frequency models. Water Resources Research. 51:4244-4262. https://doi.org/10.1002/2015WR017065.

[12] Katz RW, Parlange MB, Naveau, P. 2002. Statistics of extremes in hydrology. Advances in Water Resources. 25:1287-1304.

[13] Serago J M, Vogel RM. 2018. Parsimonious nonstationary flood frequency analysis. Advances in Water Resources. 112(May 2017):1-16. https://doi.org/10.1016/j.advwatres.2017.11.026

[14] Gumbel EJ. 1958. Statistics of extremes. Columbia University Press.

[15] Obeysekera J, Salas JD. 2016. Frequency of recurrent extremes under nonstationarity. Journal of Hydrologic Engineering. 21(5):1-9. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001339

[16] Sveinsson OGB, Salas JD, Boes DC, Roger AP Sr. 2003. Modeling the dynamics of long-term variability of hydroclimatic processes. Journal of Hydrometeorology. 4:489-505.

[17] Gumbel EJ. 1941. The return period of flood flows. The Annals of Mathematical Statistics. 12:163-190.

[18] Chow VT, Maidment DR, Mays LW. 1988. Applied hydrology. McGraw-Hill, Inc.

[19] Maidment DR. 1993. Handbook of hydrology. DR Maidment, 1st ed. McGraw-Hill, Inc.

[20] Bedient PB, Huber WC, Vieux BE. 2013. Hydrology and floodplain analysis. 5th ed. Pearson Education Limited.

[21] Salas, JD, Obeysekera J. 2014. Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. Journal of Hydrologic Engineering. 19(3):554-568. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000820

[22] Cooley D. 2013. Return periods and return levels under climate change. AghaKouchak A, Easterling D, Hsu K, Schubert S. (Ed.). Extremes in a Changing Climate. Vol. 65. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-4479-0

[23] Chen M, Papadikis K, Jun C, Macdonald N. 2023. Linear, nonlinear, parametric and nonparametric regression models for nonstationary flood frequency analysis. Journal of Hydrology. 616(3):128772. https://doi.org/10.1016/j.jhydrol.2022.128772

[24] Villarini G, Smith JA, Serinaldi F, Bales J, Bates PD, Krajewski WF. 2009. Flood frequency analysis for nonstationary annual peak records in an urban drainage basin. Advances in Water Resources. 32(8): 1255-1266. https://doi.org/10.1016/j.advwatres.2009.05.003

[25] Šraj M, Viglione A, Parajka J, Blöschl G. 2016. The influence of non-stationarity in extreme hydrological events on flood frequency estimation. Journal of Hydrology and Hydromechanics. 64(4): 426-437. https://doi.org/10.1515/johh-2016-0032

[26] Stedinger JR, Griffis VW. 2011. Getting from here to where? Flood frequency analysis and climate. Journal of the American Water Resources Association. 47(3):506-513. https://doi.org/10.1111/j.1752-1688.2011.00545.x

[27] Silva AT, Naghettini M, Portela MM. 2015. On some aspects of peaks-over-threshold modeling of floods under nonstationarity using climate covariates. Stochastic Environmental Research and Risk Assessment. https://doi.org/10.1007/s00477-015-1072-y

[28] Lopez J, Frances F. 2013. Non-stationary flood frequency analysis in continental Spanish rivers, using climate and reservoir indices as external covariates. Hydrology and Earth System Sciences. 17:3189-3203. https://doi.org/10.5194/hess-17-3189-2013

[29] Debele SE, Strupczewski WG, Bogdanowicz E. 2017. A comparison of three approaches to non-stationary flood frequency analysis. Acta Geophysica. 65(4):863-883. https://doi.org/10.1007/s11600-017-0071-4

[30] Wang M, Jiang S, Ren L, Xu CY, Shi P, Yuan S, Liu Y, Fang X. 2022. Nonstationary flood and low flow frequency analysis in the upper reaches of Huaihe River Basin, China, using climatic variables and reservoir index as covariates. Journal of Hydrology. 612(Part C):128266. https://doi.org/10.1016/j.jhydrol.2022.128266

[31] Obeysekera J, Salas JD. 2014. Quantifying the uncertainty of design floods under nonstationary conditions. Journal of Hydrologic Engineering. 19(7):1438-1446. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000931

[32] Hecht JS, Barth NA, Ryberg KR, Gregory AE. 2022. Simulation experiments comparing nonstationary design-flood adjustments based on observed annual peak flows in the conterminous United States. Journal of Hydrology X. 17:100115. https://doi.org/10.1016/j.hydroa.2021.100115

[33] McEachran ZP, Karwan DL, Sebestyen SD, Slesak RA, Ng GHC. 2021. Nonstationary flood-frequency analysis to assess effects of harvest and cover type conversion on peak flows at the Marcell Experimental Forest, Minnesota, USA. Journal of Hydrology. 596:126054. https://doi.org/10.1016/j.jhydrol.2021.126054

[34] Hu Y, Liang Z, Peng A, Wang K, Wang J, Li B. 2023. Nonstationary hydrological frequency analysis in light of model parameters and climate projections uncertainty. Journal of Hydrology, 617(Part C): 129120. https://doi.org/10.1016/j.jhydrol.2023.129120

[35] U.S. Army Corps of Engineers, Institute for Water Resources, Hydrologic Engineering Center. 2016. Retrieved http://www.hec.usace.army.mil/software/hec-hms/documentation/HEC-HMS_Users_Manual_4.0.pdf

[36] Salas JD, Tarawneh Z, Biondi F. 2015. A hydrological record extension model for reconstructing stream flows from tree-ring chronologies. Hydrological Processes. 29:544-556. https://doi.org/10.1002/hyp.10160

[37] Katz RW, Parlange MB, Naveau P. 2002. Statistics of extremes in hydrology. Advances in Water Resources. 25:1287-1304.

[38] Akaike H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control. 19(6):716-723. https://doi.org/10.1109/TAC.1974.1100705

[39] Kim H, Kim S, Shin H, Heo JH. 2017. Appropriate model selection methods for nonstationary generalized extreme value models. Journal of Hydrology. 547:557-574. https://doi.org/10.1016/j.jhydrol.2017.02.005

[40] Kottegoda NT, Rosso R. 2008. Applied statistics for civil and environmental engineers. 2nd ed. John Wiley & Sons Ltd.

[41] Sadegh M, Vrugt JA. 2014. Approximate Bayesian computation using Markov Chain Monte Carlo simulation: DREAM(ABC). Water Resources Research. 50:6767-6787. https://doi.org/10.1002/2014WR015386

[42] Ragno E, Aghakouchak A, Cheng L, Sadegh M. 2019. Process-informed nonstationary extreme value analysis (ProNEVA) user manual. https://doi.org/10.1016/j.advwatres.2019.06.007

[43] El Adlouni S, Ouarda TBMJ, Zhang X, Roy R, Bobee B. 2007. Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. Water Resources Research. 43:1-13. https://doi.org/10.1029/2005WR004545

[44] Martins ES, Stedinger JR. 2000. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. Water Resources Research. 36(3):737–744.

[45] Ragno E, AghaKouchak A, Cheng L, Sadegh M. 2019. A generalized framework for process-informed nonstationary extreme value analysis. Advances in Water Resources. 130:270-282. https://doi.org/10.1016/j.advwatres.2019.06.007

[46] Marco JB, Harboe R, Salas JD. 1993. Stochastic hydrology and its use in water resources systems simulation and optimization. Vol. 237.

[47] Cooley D. 2013. Return periods and return levels under climate change. AghaKouchak A, Easterling D, Hsu K, Schubert S. (Ed.), Extremes in a Changing Climate. Vol. 65. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-4479-0

[48] Moriasi DN, Arnold JG, Van Liew MW, Binger RL, Harmel RD, Veith TL. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE. 50(3):885-900. https://doi.org/10.1234/590