

Forecast generation model of municipal solid waste using multiple linear regression	Модель прогнозирования образования твердых бытовых отходов с использованием множественной линейной регрессии
J.A. Araiza-Aguilar 1, *, M.N. Rojas-Valencia 2, R.A. Aguilar-Vera 3 School of Environmental Engineering, University of Science and Arts of Chiapas, North beltway, Lajas Maciel, Tuxtla Gutierrez, Chiapas, Mexico 1 Institute of Engineering, National Autonomous University of Mexico, External circuit, University City, Coyoacan delegation, Mexico City, Mexico 2 Institute of Geography, National Autonomous University of Mexico, External circuit, University City, Coyoacan delegation, Mexico City, Mexico	
A B S T R A C T	Аннотация
The objective of this study was to develop a forecast model to determine the rate of generation of municipal solid waste in the municipalities of the Cuenca del Cañón del Sumidero, Chiapas, Mexico. Multiple linear regression was used with social and demographic explanatory variables. The compiled database consisted of 9 variables with 118 specific data per variable, which were analyzed using a multicollinearity test to select the most important ones. Initially, different regression models were generated, but only 2 of them were considered useful, because they used few predictors that were statistically significant. The most important variables to predict the rate of waste generation in the study area were the population of each municipality, the migration and the population density. Although other variables, such as daily per capita income and average schooling are very important, they do not seem to have an effect on the response variable in this study. The model with the highest parsimony resulted in an adjusted coefficient of 0.975, an average absolute percentage error of 7.70, an average absolute deviation of 0.16 and an average root square error of 0.19, showing a high influence on the phenomenon studied and a good predictive capacity.	Целью данного исследования было разработать прогнозную модель для определения скорости образования твердых бытовых отходов в муниципалитетах Куэнка-дель-Каньон-дель-Сумидеро, штат Чьяпас, Мексика. Множественная линейная регрессия была использована с социальными и демографическими пояснительными переменными. Скомпилированная база данных состояла из 9 переменных со 118 конкретными данными на переменную, которые были проанализированы с использованием теста мультиколлинеарности для выбора наиболее важных. Первоначально были созданы разные регрессионные модели, но только 2 из них были признаны полезными, поскольку в них использовалось мало предикторов, которые были статистически значимыми. Наиболее важными переменными для прогнозирования скорости образования отходов в районе исследования были численность населения каждого муниципалитета, миграция и плотность населения. Хотя другие переменные, такие как ежедневный доход на душу населения и среднее школьное образование, очень важны, они, похоже, не влияют на переменную реакции в этом исследовании. Модель с самым высоким скупостью привела к скорректированному коэффициенту 0,975, средней абсолютной процентной ошибке 7,70, среднему абсолютному отклонению 0,16 и средней ошибке корневого квадрата 0,19, демонстрируя высокое влияние на изучаемое явление и хорошую прогнозирующую способность,
INTRODUCTION	Введение
Because of its high management cost, the amount of Municipal Solid Waste (MSW) generated in population settlements is a significant factor for the provision of public services. According to Intharathirat et al. (2015); Keser et al. (2012); Khan et al. (2016), the amount of MSW and its composition vary	Из-за высокой стоимости управления количество твердых бытовых отходов (ТБО), образующихся в населенных пунктах, является важным фактором для предоставления государственных услуг. Согласно Intharathirat et al. (2015); Кесер и соавт. (2012); Хан и соавт. (2016),

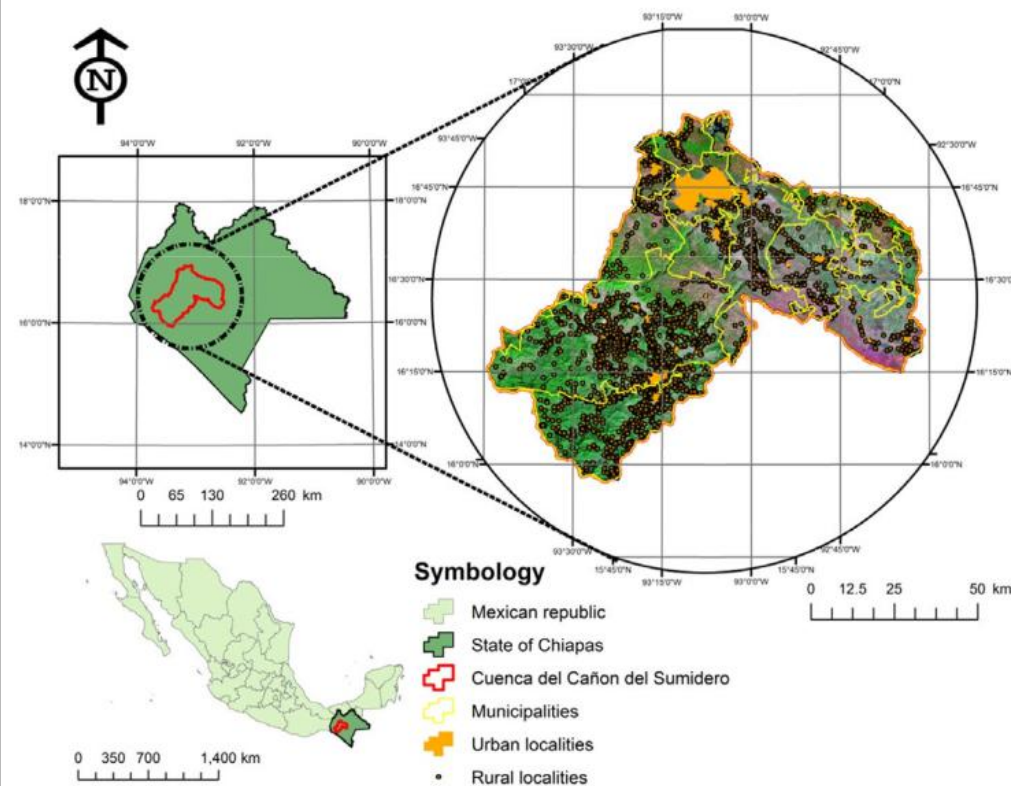
depending on social, environmental and demographic factors. Several researchers have developed models to predict the amount of MSW generated (Mahmood et al., 2018; Kannangara et al., 2018; Pan et al., 2019; Soni et al., 2019), while others analyze the variables that influence their generation and composition (Chhay et al., 2018; Grazhdani, 2016; Liu and Wu, 2010; Liu et al., 2019; Rybová et al., 2018). Unfortunately, due to the social, economic and geographical heterogeneity of the different regions of the world, it is difficult to make inferences or projections with the proposed models, and therefore, the models and their variables have to be adapted to the conditions of other regions, sometimes with little success. Kumar and Samandder (2017) and Shan (2010) reports that some of the difficulties for the adaptation of these models are related to limited or inaccessible information in other countries (databases). In addition, some variables are theoretically valid, but difficult to measure. In other cases, the variables used do not provide information leading to the explanation of the phenomenon, but have to be used, because the model incorporates them. Mexico, this topic has also been addressed, particularly in the center and north of the country (Buenrostro et al., 2001; Márquez et al., 2008; Ojeda et al., 2008; Rodríguez, 2004). However, it is evident that the models proposed are not applicable to the entire national context. According to the OECD (2015), there are notable differences between the central, northern and especially southern regions of Mexico; these include disparities in income, education, access to services, dispersion of localities and other factors, which cause that the consumption patterns, and therefore the amount of MSW, vary greatly. This study presents a model to forecast the generation rate of MSW in the municipalities of the Cuenca del Cañón del Sumidero (CCS), Chiapas State, Mexico. The model considers the information of the most relevant and easily accessible social and demographic variables for the study area, which correspond to statistical data for the years 2010-2015. This model will allow the decision makers of the municipalities of the CCS to determine the quantities of MSW generated, operate properly the waste management systems, and even acquire infrastructure. This study has been carried out in municipalities of the Cuenca del Cañón del Sumidero, Chiapas State, Mexico during 2010 - 2015.

количество ТБО и его состав варьируются в зависимости от социальных, экологических и демографических факторов. Некоторые исследователи разработали модели для прогнозирования количества образующихся ТБО (Махмуд и др., 2018; Каннангара и др., 2018; Пан и др., 2019; Сони и др., 2019), в то время как другие анализируют переменные, которые влияют на их образование и состав (Chhay et al., 2018; Grazhdani, 2016; Liu и Wu, 2010; Liu и др., 2019; Rybová et al., 2018). К сожалению, из-за социальной, экономической и географической неоднородности различных регионов мира с помощью предлагаемых моделей сложно делать выводы или прогнозы, и поэтому модели и их переменные должны быть адаптированы к условиям других регионов. иногда с небольшим успехом. Kumar and Samandder (2017) и Shan (2010) сообщают, что некоторые из трудностей адаптации этих моделей связаны с ограниченной или недоступной информацией в других странах (базы данных). Кроме того, некоторые переменные теоретически допустимы, но их трудно измерить. В других случаях используемые переменные не дают информации, приводящей к объяснению явления, но должны использоваться, потому что модель включает их. В Мексике эта тема также затрагивалась, особенно в центре и на севере страны (Buenrostro et al., 2001; Márquez et al., 2008; Ojeda et al., 2008; Rodríguez, 2004). Однако очевидно, что предложенные модели не применимы ко всему национальному контексту. По данным ОЭСР (2015), между центральными, северными и особенно южными регионами Мексики существуют заметные различия; К ним относятся различия в доходах, образовании, доступе к услугам, рассеяние населенных пунктов и другие факторы, которые приводят к тому, что структуры потребления и, следовательно, количество ТБО значительно различаются. В этом исследовании представлена модель для прогнозирования скорости образования ТБО в муниципалитетах Куэнка-дель-Каньон-дель-Сумидеро (CCS), штат Чьяпас, Мексика. Модель учитывает информацию о наиболее актуальных и легко доступных социальных и демографических переменных для области исследования, которые соответствуют статистическим данным за 2010-2015 годы. Эта модель позволит лицам, принимающим решения в муниципалитетах УХУ, определять количество образующихся ТБО, правильно работать с системами управления отходами и даже приобретать инфраструктуру. Это исследование было проведено в

	муниципалитетах Куэнка-дель-Каньон-дель-Сумидеро, штат Чьяпас, Мексика в течение 2010 - 2015 гг.
MATERIALS AND METHODS	МАТЕРИАЛЫ И МЕТОДЫ
Description of the study area and context	Описание области исследования и контекста
The CCS is located in the State of Chiapas, in the southeast of Mexico, between the coordinates 15° 56' 55" and 16° 57' 26" North Latitude, and 92° 30' 44" and 93° 44' 35" west longitude (Fig. 1). The CCS has 24 municipalities and 2,847 localities; 2,816 localities are rural while 31 are urban. 83% of the population of the study area lives in urban areas (INEGI, 2010). The degree of dispersion is high, especially in the rural localities farthest from the municipal seat.	CCS находится в штате Чьяпас, на юго-востоке Мексики, между координатами 15 ° 56 '55 "и 16 ° 57' 26" северной широты и 92 ° 30 '44 "и 93 ° 44 ° 35" западное долготы (рис. 1). CCS имеет 24 муниципалитета и 2847 населенных пунктов; 2816 населенных пунктов являются сельскими, а 31 городскими. 83% населения исследуемой территории проживает в городских районах (INEGI, 2010). Степень рассеивания высока, особенно в сельских населенных пунктах, наиболее удаленных от муниципальных районов.
Development of the model	Разработка модели
This study uses a multiple linear regression (MLR) model to obtain the generation rates of MSW. Because of their versatility and well-founded theory, MLR models have been widely used in various scientific fields. Their main disadvantage is the preparation of the database (Pires et al., 2008). The hypothesis to use the MLR in this study is based on the effect of the explanatory variables (social and demographic variables) on the response variable (generation rate of MSW). The linear function is shown in Eq. 1.	В этом исследовании используется модель множественной линейной регрессии (MLR) для получения показателей генерации ТБО. Благодаря своей универсальности и обоснованной теории, модели MLR широко используются в различных научных областях. Их основным недостатком является подготовка базы данных (Pires et al., 2008). Гипотеза об использовании MLR в этом исследовании основана на влиянии объясняющих переменных (социальных и демографических переменных) на переменную реакции (скорость образования ТБО). Линейная функция показана в формуле. 1.
$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (1)$	
Where, Y is the response variable, X i (1, 2, 3 ... k) are the explanatory variables, β i (1, 2, 3 ... k) are the regression coefficients and ε is the residual error.	Где Y - переменная отклика, X i (1, 2, 3 ... k) - объясняющие переменные, β i (1, 2, 3 ... k) - коэффициенты регрессии, а ε - остаточная ошибка.
According to Agirre (2006), the MLR is based on two assumptions: i) the explanatory variables must be independent, i.e., free of multicollinearity and ii) the dependent variable must be normally distributed, with zero mean and constant variance. In order to determine the regression coefficients, the least squares method, which is based on minimizing the sum of squared errors (SSE), using Eq. 2.	Согласно Agirre (2006), MLR основывается на двух предположениях: i) объясняющие переменные должны быть независимыми, то есть свободными от мультиколлинеарности, и ii) зависимая переменная должна быть нормально распределена с нулевым средним и постоянной дисперсией. Чтобы определить коэффициенты регрессии, метод наименьших квадратов, который основан на минимизации суммы квадратов ошибок (SSE), используя формулу. 2.
$SSE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (2)$	
Where Y i is the value of each observation and \hat{Y}_i is the predicted value.	Где Y i - значение каждого наблюдения, а \hat{Y}_i - прогнозируемое

Theoretically, low SSE values reflect a better fit of the regression model (Kumar and Samandder, 2017). In order to determine the best regression model (most parsimony), the statistical significance of the explanatory variables and the general model were analyzed. The analysis of the explanatory variables was performed with the t-test, while the degree of adjustment and usefulness of the proposed model was performed by evaluating the F-test and the value of R^2_{adj} using Eqs. 3 and 4, respectively.

Fig. 1: Geographic location of the study area in Cuenca del Cañón de Sumidero in Mexico



значение. Теоретически, низкие значения SSE отражают лучшее соответствие регрессионной модели (Kumar and Samandder, 2017). Чтобы определить лучшую регрессионную модель (наиболее экономную), была проанализирована статистическая значимость объясняющих переменных и общей модели. Анализ объясняющих переменных был выполнен с помощью t-критерия, в то время как степень корректировки и полезность предложенной модели была выполнена путем оценки F-критерия и значения R^2_{adj} , используя уравнения. 3 и 4 соответственно.

Рис. 1: Географическое положение учебного района в Куэнка-дель-Каньон-де-Сумидеро в Мексике

$F = \frac{(SS_{YY} - SSE) / k}{SSE / [n - (k + 1)]} \quad (3)$ $R^2_{adj} = 1 - \left[\frac{(n-1)}{n - (k+1)} \right] (1 - R^2) \quad (4)$	
<p>Where $SS_{YY} = \sum (Y_i - \bar{Y})^2$ represents the sum of the squares of the difference of the observed data (Y_i) and the average of the data (\bar{Y}); k is the number of explanatory variables included in the model; n is the sample size; and R^2 is the coefficient of determination. The value of R^2 was not considered to measure the explanatory power of the regression model, because its value increases when adding more explanatory variables, and it can be a deceptive measure (Chang et al., 2007).</p>	<p>Где $SS_{YY} = \sum (Y_i - \bar{Y})^2$ представляет собой сумму квадратов разности наблюдаемых данных (Y_i) и среднего значения данных (\bar{Y}); k - количество объясняющих переменных, включенных в модель; n - размер выборки; и R^2 - коэффициент детерминации. Значение R^2 не рассматривалось как для измерения объяснительной силы регрессионной модели, поскольку его значение увеличивается при добавлении дополнительных объясняющих переменных, и это может быть обманчивой мерой (Chang et al., 2007).</p>
<p>Data collection</p>	<p>Сбор информации</p>
<p>According to Beigl et al. (2008) and Kolekar et al. (2016), the methods of data collection depend on the scale of the study. In investigations carried out at household or locality levels, the acquisition of information is usually carried out through surveys or interviews; while at district or country scales, the information comes from a database registered by government agencies. This study was made at district scale and therefore the study area includes several municipalities. MSW generation was obtained from SEMANH (2013), the studies by Alvarado et al. (2009) and Araiza et al. (2015). The social and demographic information (explanatory variables) was obtained from CONAPO (2017) and INEGI (2010). The compiled information allowed the elaboration of a database of 9 variables, with 118 specific data per variable, coming from all the municipalities of the state of Chiapas (Table 1). The inferences of the proposed model were made on the municipalities of the CCS. This database was analyzed with the MINITAB software version 16.</p>	<p>Согласно Beigl et al. (2008) и Kolekar et al. (2016), методы сбора данных зависят от масштаба исследования. В исследованиях, проводимых на уровне домохозяйств или населенных пунктов, получение информации обычно осуществляется посредством опросов или интервью; в то время как в масштабах района или страны информация поступает из базы данных, зарегистрированной государственными органами. Это исследование было проведено в масштабе района, и поэтому область исследования включает несколько муниципалитетов. Поколение ТБО было получено от SEMANH (2013), исследования Alvarado et al. (2009) и Araiza et al. (2015). Социальная и демографическая информация (объяснительные переменные) была получена из CONAPO (2017) и INEGI (2010). Собранные данные позволили создать базу данных из 9 переменных, со 118 конкретными данными на переменную, поступивших из всех муниципалитетов штата Чьяпас (Таблица 1). Сделанные выводы из предложенной модели были сделаны на муниципалитетах CCS. Эта база данных была проанализирована с помощью программного обеспечения MINITAB версии 16.</p>
<p>Exploratory analysis of variables</p>	<p>Исследовательский анализ переменных</p>
<p>An exploratory analysis of the 9 variables used to check the normality of the data was carried out. The test used was Kolmogorov-Smirnov, with a level of</p>	<p>Был проведен предварительный анализ 9 переменных, используемых для проверки нормальности данных. Используемый тест был</p>

<p>significance of $\alpha = 0.05$. This test showed that the variables Y Gen , X Pop , X Pd , X Pbam , X Hgs , X Ces , X Dpi , did not follow a normal distribution, because their p-value was smaller than the α value considered. In order to adjust their values, the variables were transformed with natural logarithms. The variables X As and X Mi were not transformed because their data followed a normal distribution (Table 2).</p>	<p>Колмогорова-Смирнова, с уровнем значимости = 0,05. Этот тест показал, что переменные Y Gen, X Pop, X Pd, X Pbam, X Hgs, X Ces, X Dpi не следовали нормальному распределению, поскольку их значение p было меньше, чем рассматриваемое значение α. Чтобы скорректировать их значения, переменные были преобразованы с натуральными логарифмами. Переменные X As и X Mi не были преобразованы, поскольку их данные следовали нормальному распределению (таблица 2).</p>
<p>Multicollinearity analysis and variable screening</p> <p>An analysis of the explanatory variables was made prior to the selection of the best MLR model. Through a multicollinearity test, some of the variables initially considered were eliminated. Especially, the variance inflation factor (VIF) and the Pearson correlation coefficient (r) were used. Similar to Keser et al. (2012), the r coefficient was used to detect the bivariate association, while the VIF was used to detect the multivariate correlation. Eqs. 5 and 6 describe the tests used.</p>	<p>Анализ мультиколлинеарности и переменный скрининг</p> <p>Анализ объясняющих переменных был сделан до выбора наилучшей модели MLR. С помощью теста на мультиколлинеарность некоторые из первоначально рассмотренных переменных были исключены. В частности, использовались коэффициент инфляции дисперсии (VIF) и коэффициент корреляции Пирсона (r). Подобно Кесеру и соавт. (2012), коэффициент r использовался для обнаружения двумерной ассоциации, в то время как VIF использовался для обнаружения многомерной корреляции. Уравнения. 5 и 6 описывают используемые тесты.</p>
$VIF_k = \frac{1}{(1 - R_k^2)} \quad (5)$ $r = \sqrt{1 - \frac{SSE}{SS_{YY}}} \quad (6)$	
<p>VIF value is calculated using the R^2 of the regression equation; the explanatory variables denoted by k are analyzed as dependent variables, while the others are used as independent variables; thus, VIF is calculated for each explanatory variable k.</p>	<p>Значение VIF рассчитывается с использованием R^2 уравнения регрессии; объясняющие переменные, обозначенные k, анализируются как зависимые переменные, в то время как другие используются как независимые переменные; таким образом, VIF рассчитывается для каждой объясняющей переменной k.</p>
<p>The cut-off value of VIF used in this study was 4. According to Ghinea et al. (2016), when $VIF < 1$, the explanatory variables are not correlated; when $1 < VIF < 5$, the explanatory variables are slightly correlated; and when $VIF > 5$ or 10, the explanatory variables are highly correlated. The value of r indicates the relationship between two variables (positive or negative); its value ranges between -1 and 1. There are no clearly defined cut-off in the literature. Arriaza (2006) indicates that with values of r greater than 0.3, there may be signs of</p>	<p>Предельное значение VIF, использованное в этом исследовании, составило 4. Согласно Ghinea et al. (2016), когда $VIF < 1$, объясняющие переменные не коррелированы; когда $1 < VIF < 5$, объясняющие переменные слегка коррелируют; и когда $VIF > 5$ или 10, объясняющие переменные сильно коррелируют. Значение r указывает на связь между двумя переменными (положительной или отрицательной); его значение находится в диапазоне от -1 до 1. В литературе нет четко определенных</p>

<p>correlation, with values greater than 0.8, there are serious problems of multicollinearity. As in Grazhdani (2016), in this study it was considered that a value of $r \geq 0.6$ (positive or negative), indicates correlation between the explanatory variables. The elimination of explanatory variables was performed in an iterative procedure, i.e., the VIF values were initially determined for the 8 variables; subsequently, the variable with the highest VIF was eliminated and the next iteration with 7 variables was performed. This elimination procedure ended when a VIF cut-off value of 4 was found. Finally, other eliminations were made based on the values of r. Subsequently, 3 explanatory variables were used in the search stage for a better model (of greater parsimony). The first variable selected was X Pop, i.e., the “total population” of each municipality, under the hypothesis that the larger the population, the greater the consumption and thus the greater the amount of MSW generated. The second explanatory variable used was X Pd “population density”, under the premise that dispersion patterns or agglomeration of inhabitants per unit area influences MSW generation. The third variable used was X Pbam “population born in another municipality”, which can be seen as migration, i.e., people who move to other places to seek better living conditions. The process of mobilization of people causes changes in consumption patterns of a new place of settlement. Other models that do not follow the principle of parsimony were also created (more than 3 explanatory variables), but they should not be used to forecast waste generation rates, since they have very low accuracy values and some of their explanatory variables are not significant.</p>	<p>пределов. Арриаза (2006) указывает, что при значениях r, превышающих 0,3, могут быть признаки корреляции, при значениях, превышающих 0,8, возникают серьезные проблемы мультиколлинеарности. Как и в Grazhdani (2016), в этом исследовании считалось, что значение $r \geq 0,6$ (положительное или отрицательное) указывает на корреляцию между объясняющими переменными. Исключение объясняющих переменных было выполнено в итерационной процедуре, то есть значения VIF были первоначально определены для 8 переменных; впоследствии переменная с наибольшим VIF была исключена, и была проведена следующая итерация с 7 переменными. Эта процедура исключения закончилась, когда было найдено значение отсечки VIF 4. Наконец, другие исключения были сделаны на основе значений r. Впоследствии, 3 объясняющие переменные были использованы на этапе поиска для лучшей модели (большей экономии). Первой выбранной переменной был X Pop, то есть «общая численность населения» каждого муниципалитета, согласно предположению, что чем больше население, тем больше потребление и, следовательно, тем больше образуется ТБО. Второй используемой пояснительной переменной была «Pd» «плотность населения», исходя из того, что характер рассеивания или агломерация жителей на единицу площади влияет на выработку ТБО. Третья использованная переменная была X Pbam «население, родившееся в другом муниципалитете», которое можно рассматривать как миграцию, то есть людей, которые переезжают в другие места в поисках лучших условий жизни. Процесс мобилизации людей вызывает изменения в структуре потребления нового места расселения. Были также созданы другие модели, которые не следуют принципу экономии (более 3 объясняющих переменных), но их не следует использовать для прогнозирования показателей образования отходов, поскольку они имеют очень низкие значения точности, а некоторые из их объясняющих переменных не являются значимыми.</p>
<p>Table 1: Description of variables</p>	<p>Таблица 1: Описание переменных</p>

No.	Name of the variable	Symbol	Type of variable	Measure
1	MSW generation	Y_{Gen}	Dependent	Tons/day
2	Population	X_{Pop}	Independent	Inhabitants
3	Population density	X_{Pd}	Independent	Inhabitants/km ²
4	Population born in another municipality	X_{Pbam}	Independent	Inhabitants
5	Average schooling	X_{As}	Independent	Years of study
6	Household with goods and services	X_{Hgs}	Independent	Percent (%)
7	Commercial establishments and services	X_{Ces}	Independent	Number of establishments
8	Daily per capita income	X_{Dpi}	Independent	Mexican pesos/day
9	Marginalization index	X_{Mi}	Independent	Percent (%)

Table 2: Normality test and transformation of variables

Original variable			Transformed variable		
Kolmogorov-Smirnov			Kolmogorov-Smirnov		
	Statistical	p-value		Statistical	p-value
Y_{Gen}	0.338	<0.010	$\ln-Y_{Gen}$	0.050	>0.150
X_{Pop}	0.281	<0.010	$\ln-X_{Pop}$	0.066	>0.150
X_{Pd}	0.271	<0.010	$\ln-X_{Pd}$	0.039	>0.150
X_{Pbam}	0.387	<0.010	$\ln-X_{Pbam}$	0.065	>0.150
X_{As}	0.053	>0.150	X_{As}	---	---
X_{Hgs}	0.183	<0.010	$\ln-X_{Hgs}$	0.054	>0.150
X_{Ces}	0.349	<0.010	$\ln-X_{Ces}$	0.056	>0.150
X_{Dpi}	0.117	<0.010	$\ln-X_{Dpi}$	0.056	>0.150
X_{Mi}	0.054	>0.150	X_{Mi}	---	---

Accuracy of the model and validation

In order to determine the accuracy of the best model found, 3 widely used measures were employed: the Mean Absolute Percentage Error (MAPE), the Mean Absolute Deviation (MAD) and the Root Mean Square Error (RMSE) (Eqs. 7, 8 and 9, respectively). A value of these measures close to zero indicates a high precision of the model (Azadi and karimí 2016).

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| * 100 \quad (7)$$

$$MAD = \frac{1}{n} \sum_{t=1}^n |A_t - F_t| \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (A_t - F_t)^2} \quad (9)$$

In these equations, A_t is the observed value, F_t is the predicted value and n is the sample size. MAPE is expressed in terms of percentage of error, MAD expresses the precision in the units of the data analyzed, and RMSE indicates

Таблица 2: Проверка нормальности и преобразование переменных

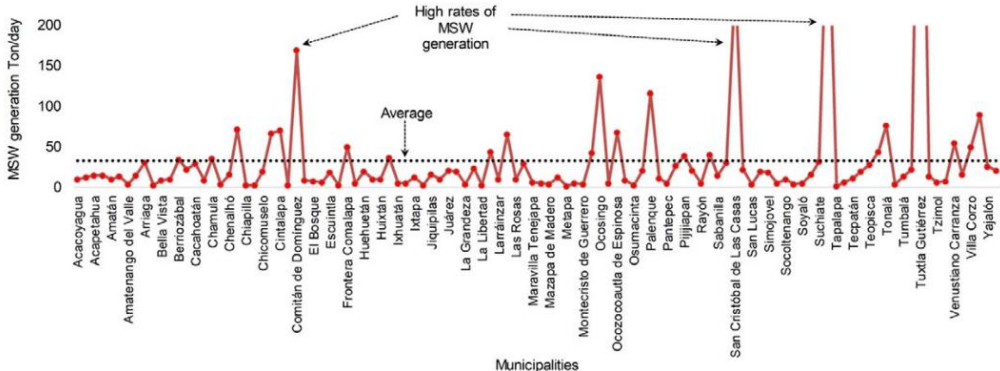
Точность и достоверность модели

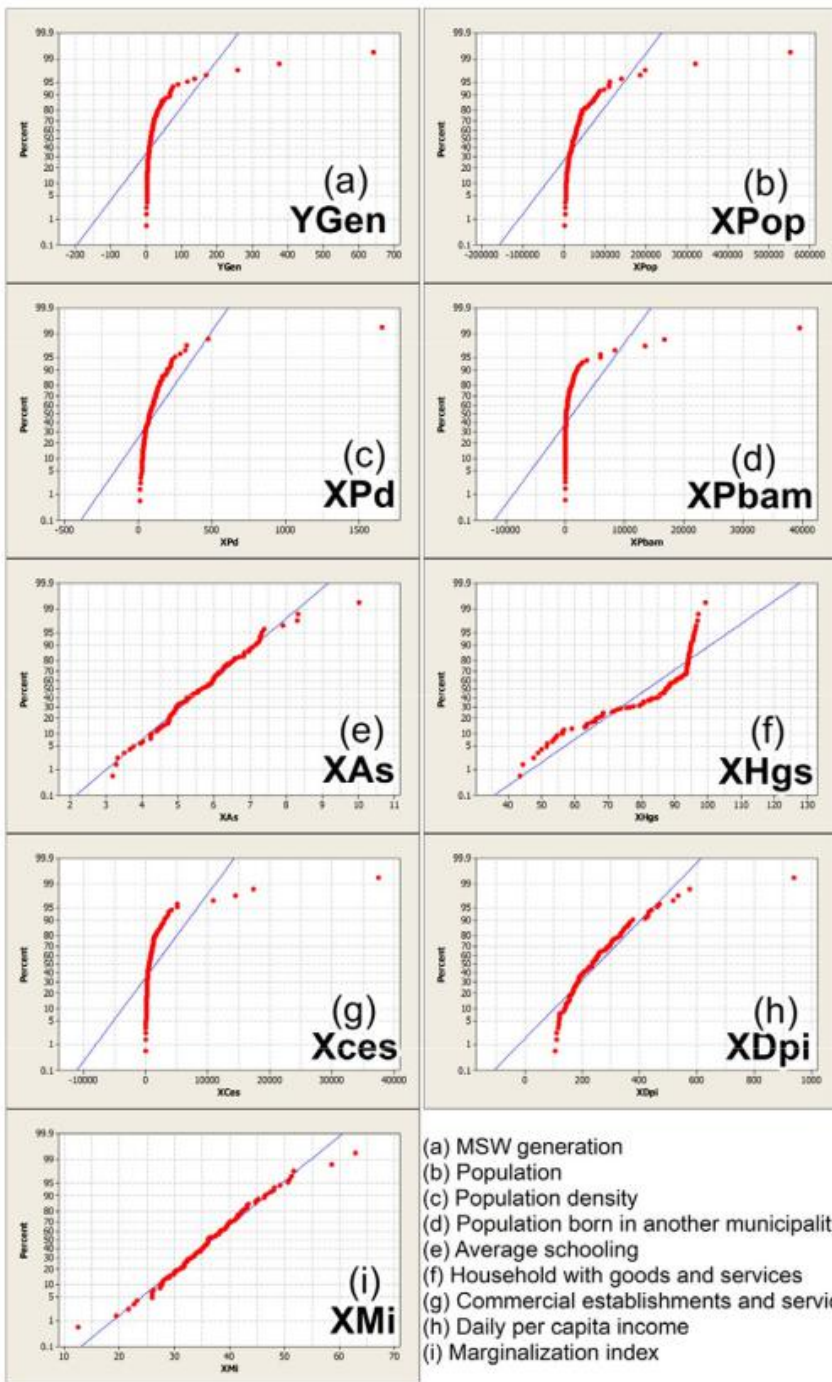
Чтобы определить точность наилучшей найденной модели, были использованы 3 широко используемых показателя: средняя абсолютная ошибка в процентах (MAPE), среднее абсолютное отклонение (MAD) и среднеквадратическая ошибка (RMSE) (уравнения 7, 8). и 9 соответственно). Значение этих показателей, близкое к нулю, указывает на высокую точность модели (Azadi and karimí 2016).

В этих уравнениях A_t - наблюдаемое значение, F_t - прогнозируемое значение, а n - размер выборки. MAPE выражается в процентах от ошибки, MAD выражает точность в единицах анализируемых данных, а

how concentrated the data are around the line of best fit. In order to perform the external validation of the model, the technique called R ² jackknife using Eq. 10. This equation is calculated by systematically eliminating each observation from the data set, estimating the regression equation and determining to what extent the model is able to predict the observation that was removed.	RMSE указывает, насколько сконцентрированы данные вокруг линии наилучшего соответствия. Для того, чтобы выполнить внешнюю проверку модели, метод называется J ² складной нож с использованием уравнения. 10. Это уравнение рассчитывается путем систематического исключения каждого наблюдения из набора данных, оценки уравнения регрессии и определения, в какой степени модель способна предсказать удаленное наблюдение.
$R^2_{jackknife} = 1 - \frac{\sum (y_i - \hat{Y}_{(i)})^2}{\sum (y_i - \bar{Y})^2} \quad (10)$	
The R jackknife coefficient varies between 0 and 100%, larger values suggesting models with greater predictive capacity; $\hat{Y}_{(i)}$ denotes the predicted value for i-th observation obtained when the regression model fits the data with y_i omitted (or removed) from the sample; and \bar{Y} is the simple average of the observed data.	Коэффициент складного ножа R варьируется от 0 до 100%, более высокие значения предполагают наличие моделей с большей прогнозирующей способностью; $\hat{Y}_{(i)}$ обозначает прогнозируемое значение для i-го наблюдения, полученное, когда регрессионная модель сопоставляет данные с y_i , опущенным (или удаленным) из выборки; и \bar{Y} является простым средним из наблюдаемых данных.
Verification of model assumptions	Проверка модельных предположений
The validity of the MLR models is subject to the behavior of the residual errors “ ε ” (difference between observed and predicted values of the dependent variable), particularly their normal distribution, their independence and homoscedasticity (Kumar and Samandder, 2017). The verification of normality was carried out through the Kolmogorov-Smirnov test, with a level of significance of $\alpha = 0.05$. In order to verify the independence of residues, the Durbin-Watson test (d) was applied, looking for values close to 2, because “d” varies between 0 and 4 (Mendenhall and Sincich, 2012). The homoscedasticity assumption was evaluated with the plot of residuals vs predicted, both standardized, looking for a residue behavior that does not fit any known pattern.	Достоверность моделей MLR зависит от поведения остаточных ошибок « ε » (разницы между наблюдаемыми и прогнозируемыми значениями зависимой переменной), в частности их нормального распределения, их независимости и гомоскедастичности (Kumar and Samandder, 2017). Проверка нормальности проводилась с помощью критерия Колмогорова-Смирнова с уровнем значимости $\alpha = 0,05$. Чтобы проверить независимость остатков, был применен критерий Дурбина-Ватсона (d), в котором были найдены значения, близкие к 2, поскольку «d» варьируется от 0 до 4 (Mendenhall and Sincich, 2012). Предположение о гомоскедастичности было оценено с помощью графика зависимости невязок от прогнозируемых, оба стандартизированы, с учетом поведения остатков, которое не соответствует какой-либо известной схеме.
RESULTS AND DISCUSSION	РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ
Statistical analysis of variables	Статистический анализ переменных
The initial exploratory analysis was performed on the response variable Y Gen , which has the behavior shown in Fig. 2. It is observed that some municipalities, which appear to be outliers, show a very high rate of MSW generation.	Первоначальный исследовательский анализ был выполнен на переменной отклика Y Gen, которая имеет поведение, показанное на рис. 2. Наблюдается, что некоторые муниципалитеты, которые выглядят как выбросы, показывают очень высокий уровень образования ТБО.
These atypical values were not eliminated from the analysis because they are	Эти нетипичные значения не были исключены из анализа, потому что

not errors, but rather data that come from the most important municipalities of Chiapas, such as “Tuxtla Gutiérrez, Comitán, San Cristóbal de las Casas and Tapachula”. These municipalities are regional heads, therefore, the number of inhabitants, their patterns of consumption and MSW generation, differ significantly from the rest of the studied area. The normality test of the response variable and of the 8 explanatory variables is shown in Fig. 3. The non-normality of the variables Y_{Gen} , X_{Pop} , X_{Pd} , X_{Pbam} , X_{Hgs} , X_{Ces} and X_{Dpi} , can be seen. For this reason, these variables were transformed using natural logarithms (Fig. 3a, 3b, 3c, 3d, 3f, 3g and 3h).	они не являются ошибками, а скорее данными, полученными из наиболее важных муниципалитетов штата Чьяпас, таких как «Тустла-Гутьеррес, Комитан, Сан-Кристобаль-де-лас-Касас и Тапачула». Эти муниципалитеты являются региональными главами, поэтому количество жителей, их модели потребления и выработка ТБО значительно отличаются от остальной части исследуемой территории. Проверка нормальности переменной отклика и 8 объясняющих переменных показана на рис. 3. Ненормальность переменных Y_{Gen} , X_{Pop} , X_{Pd} , X_{Pbam} , X_{Hgs} , X_{Ces} и X_{Dpi} , может быть виден. По этой причине эти переменные были преобразованы с использованием натуральных логарифмов (рис. 3a, 3b, 3c, 3d, 3f, 3g и 3h).
Forecast model	Прогнозная модель
The coefficients of the MLR model were determined using the Minitab software. Only the explanatory variables that fulfilled the multicollinearity criterion were used. Initially, 2 theoretically valid models were determined; the first one is shown in Eq. 11.	Коэффициенты модели MLR были определены с использованием программного обеспечения Minitab. Использовались только пояснительные переменные, которые удовлетворяли критерию мультиколлинеарности. Первоначально были определены 2 теоретически обоснованные модели; первый показан в формуле 11.
$\ln Y_{Gen} = -8.91 + 1.10 \ln_{X_{Pop}} + 0.0259 \ln_{X_{Pd}} + 0.0688 \ln_{X_{Pbam}} \quad (11)$	
This first model consists of 3 variables, X_{Pop} , X_{Pd} , X_{Pbam} (all transformed). The F-test associated with a variance analysis indicated that the model is statistically valid because p-value < 0.05. This model can thus also be used for forecast purposes. However, it is important to be careful because the explanatory variable X_{Pd} is not statistically significant since the null hypothesis that the coefficient of the variable is equal to zero ($H_0 : \beta_i = 0$) is met. Therefore, the explanatory variable is not related to the dependent variable, i.e., it should not be interpreted.	Эта первая модель состоит из 3 переменных: X_{Pop} , X_{Pd} , X_{Pbam} (все преобразованы). F-критерий, связанный с дисперсионным анализом, показал, что модель является статистически достоверной, поскольку значение $p < 0,05$. Таким образом, эта модель также может быть использована для прогнозирования. Однако важно быть осторожным, потому что объясняющая переменная X_{Pd} не является статистически значимой, поскольку нулевая гипотеза о том, что коэффициент переменной равен нулю ($H_0 : \beta_i = 0$), встречается. Следовательно, объясняющая переменная не связана с зависимой переменной, то есть она не должна интерпретироваться.
The second model is presented in Eq. 12, which consists of 2 explanatory variables “ X_{Pop} and X_{Pbam} ”. Similar to the first model, here also the p-value and the F-test indicate that it is a statistically valid model that can be used for forecasting purposes. Particularly this model is the one of greater parsimony, because it uses only 2 variables.	Вторая модель представлена в формуле. 12, которая состоит из 2 пояснительных переменных « X_{Pop} и X_{Pbam} ». Подобно первой модели, здесь также р-значение и F-критерий показывают, что это статистически достоверная модель, которую можно использовать для целей прогнозирования. В частности, эта модель более экономна, поскольку использует только 2 переменные.
$\ln Y_{Gen} = -8.86 + 1.11 \ln_{X_{Pop}} + 0.0658 \ln_{X_{Pbam}} \quad (12)$	

<p>All the information associated with the analysis of variance is presented in Table 3.</p>	<p>Вся информация, связанная с анализом отклонений, представлена в таблице 3.</p>
<p>The verification of assumptions of the proposed models, especially model 2, is presented in Fig. 4. The probability-probability plot (p-p plot) (Fig. 4a) shows the values of the residuals with a linear pattern indicating normality; additionally, the Kolmogorov- Smirnov value and its associated p-value confirm it (p-value > 0.15). The result of the Durbin-Watson independence test gave a value of 1.979 for model 2, which indicates that the residuals are not correlated. The homoscedasticity test presented in Fig. 4b shows a behavior of the residuals that does not fit any known pattern; therefore, this situation is adequate.</p>	<p>Проверка допущений предложенных моделей, особенно модели 2, представлена на рис. 4. График вероятности-вероятности (p-p график) (рис. 4a) показывает значения невязок с линейной диаграммой, указывающей на нормальность; кроме того, значение Колмогорова-Смирнова и связанное с ним значение р подтверждают это (значение $p > 0,15$). Результат теста независимости Дурбина-Ватсона дал значение 1,979 для модели 2, что указывает на то, что невязки не коррелированы. Тест на гомоскедастичность, представленный на рис. 4, b, показывает поведение остатков, которое не соответствует ни одной из известных схем; следовательно, эта ситуация является адекватной.</p>
<p>Fig. 2: MSW generation rates in the state of Chiapas</p>  <p>The graph displays MSW generation rates for 50 municipalities. The y-axis is labeled 'MSW generation Ton/day' and ranges from 0 to 200. The x-axis lists the municipalities. A dashed horizontal line at approximately 30 Ton/day is labeled 'Average'. Several municipalities show high rates, with peaks around 150-200 Ton/day. Arrows point to these peaks with the label 'High rates of MSW generation'.</p>	<p>Рис. 2: Скорость образования ТБО в штате Чьяпас</p>
<p>Fig. 3: Behavior of the variables analyzed with respect to normality</p>	<p>Рис. 3: Поведение проанализированных переменных относительно нормальности</p>



- (a) MSW generation
- (b) Population
- (c) Population density
- (d) Population born in another municipality
- (e) Average schooling
- (f) Household with goods and services
- (g) Commercial establishments and services
- (h) Daily per capita income
- (i) Marginalization index

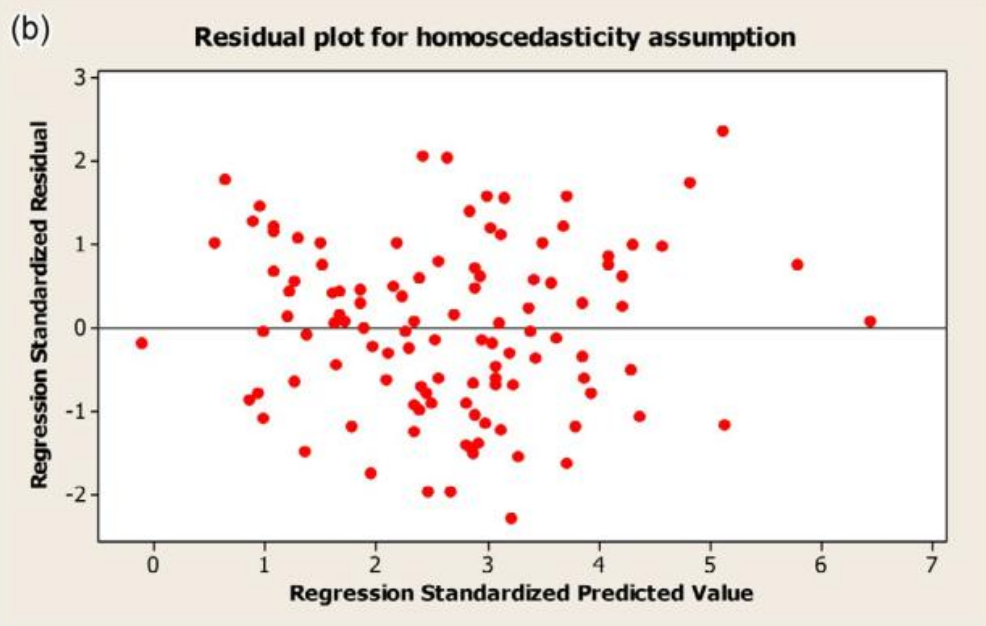
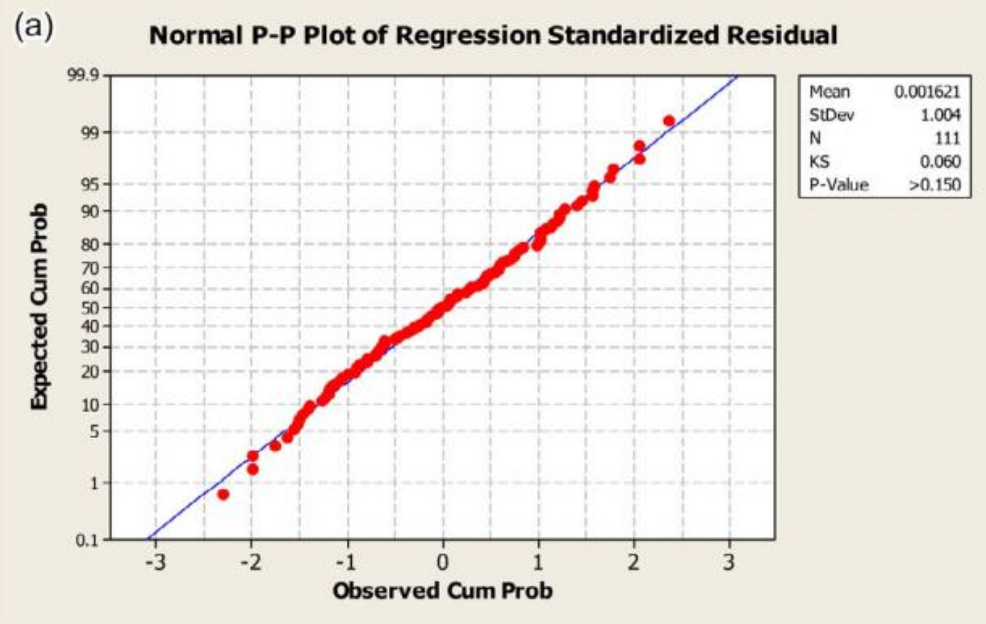
Table 3: Analysis of variance of the proposed models

Model	Source	Degree of freedom (df)	Sum of Squares	Mean square	F	Sig.
1	Regression	3	150.591	50.197	1,449.33	0.000
	Residual	107	3.706	0.035		
	Total	110	154.297			
2	Regression	2	150.549	75.274	2,169.16	0.000
	Residual	108	3.748	0.035		
	Total	110	154.297			

Fig. 4: Verification of model assumptions: (a) Normality of residuals, (b) Independence of residuals

Таблица 3: Анализ дисперсии предлагаемых моделей

Fig. 4: Verification of model assumptions: (a) Normality of residuals, (b) Independence of residuals



On the other hand, the R^2 value of the equations in both models was 0.976, which indicates that 97.6% of the generation rate of MSW Y Gen

С другой стороны, значение R^2 в уравнениях в обеих моделях составляло 0,976, что указывает на то, что 97,6% скорости образования

<p>(transformed) can be explained by the explanatory variables used. It is 2 important to note the gradual decrease of R^2_{adj} (0.975) with respect to R^2, which is due to the adjustment by the introduction of 2 and 3 variables in models 1, 2 and 3, respectively. The high value of R^2 and R^2_{adj} in these models is due to the initial transformation of the explanatory variables, as well as the response variable. Additionally, the data collection carried out in this study influenced these values because they come from a census database, and not from an information survey through interviews. The internal validation of model 2 through MAPE, MAD and RMSE, showed the values of 7.70, 0.16 and 0.19, respectively, which indicates a high precision since the values of these tests are close to 0 (zero). The external validation by R^2 jackknife presented a value of 97.44%. Therefore, model 2 also has a high forecasting capacity.</p>	<p>ТБО Y Gen (в преобразованном виде) можно объяснить используемыми пояснительными переменными. Важно отметить постепенное уменьшение R^2_{adj} (0,975) по отношению к R^2, что связано с корректировкой путем введения 2 и 3 переменных в моделях 1, 2 и 3 соответственно. Высокое значение R^2 и R^2_{adj} в этих моделях обусловлено первоначальным преобразованием объясняющих переменных, а также переменной отклика. Кроме того, сбор данных, проведенный в этом исследовании, повлиял на эти значения, потому что они поступают из базы данных переписи, а не из информационного опроса посредством интервью. Внутренняя проверка модели 2 с помощью MAPE, MAD и RMSE показала значения 7,70, 0,16 и 0,19 соответственно, что указывает на высокую точность, поскольку значения этих тестов близки к 0 (нулю). Внешняя проверка с помощью ножа R^2 составила 97,44%. Следовательно, модель 2 также обладает высокой прогнозирующей способностью.</p>
<p>Non-significant variables</p>	<p>Незначительные переменные</p>
<p>The analysis of the 8 explanatory variables using the VIF test produced the initial elimination of the variables X As, ln-X Hgs, ln-X Ces and ln-X Dpi, since their value was higher than the cut-off of 4. The variables X As and ln-X Dpi have been used mainly in studies at household or locality levels (Khan et al, 2016; Ojeda Table et al., 2008), but in this paper they were used at district level, and the effect of these variables seems not to be important (low correlation with the response variable ln-Y Gen). The variables ln-X Hgs and ln-X Ces were eliminated because they are highly correlated with X Mi, since the latter is a multidimensional indicator that measures deprivation in a population, through variables similar to those eliminated. Finally, through the r test, only X Mi was eliminated, since it was highly correlated with ln-X Pbam, with a coefficient of -0.695, i.e., much higher than the cut-off value of 0.6 (positive or negative); additionally, this variable was less correlated with the ln-Y Gen response variable (Table 4).</p>	<p>Анализ 8 объясняющих переменных с использованием теста VIF позволил вначале исключить переменные X As, ln-X Hgs, ln-X Ces и ln-X Dpi, поскольку их значение было выше, чем пороговое значение 4. переменные X As и ln-X Dpi использовались в основном в исследованиях на уровне домохозяйств или населенных пунктов (Khan et al, 2016; Ojeda Table et al., 2008), но в этой статье они использовались на районном уровне, а влияние эти переменные, кажется, не важны (низкая корреляция с переменной отклика ln-Y Gen). Переменные ln-X Hgs и ln-X Ces были исключены, поскольку они сильно коррелируют с X Mi, поскольку последний является многомерным показателем, который измеряет депривацию в популяции через переменные, подобные тем, которые были исключены. Наконец, с помощью r-теста был исключен только X Mi, поскольку он сильно коррелировал с ln-X Pbam, с коэффициентом -0,695, то есть намного выше, чем значение отсечения 0,6 (положительное или отрицательное); кроме того, эта переменная была менее коррелирована с переменной ответа ln-Y Gen (таблица 4).</p>
<p>Significant variables The transformed variables X Pop, X Pd and X Pbam were used in the search for the best model, since their VIF and r values were below the cut-off values. X Pop has been used in the studies of Azadi and karimí (2016) and Abdoli et al. (2011), as the most important explanatory variable. In this study, Pearson's correlation r-value was 0.985, which</p>	<p>Значимые переменные Преобразованные переменные X Pop, X Pd и X Pbam использовались при поиске лучшей модели, поскольку их значения VIF и r были ниже предельных значений. X Pop использовался в исследованиях Azadi и karimí (2016) и Abdoli et al. (2011), как наиболее важная объясняющая переменная. В этом исследовании r-значение</p>

<p>indicates that it is also the variable most related to the generation of waste, particularly in a positive way, i.e., to a larger population corresponds a greater quantity of MSW. The variable X Pd has been used in few publications. Bel and Mur (2009) use this variable also to obtain the costs associated with waste management. In this study, r-value of 0.161 was obtained, which indicates a poor correlation with the response variable. The analysis of the forecast model 1 indicated that this variable is not statistically significant, and its use must be taken with caution. The X Pbam variable is positively related to the response variable. Its Pearson's correlation coefficient was 0.638. This variable is important in the study area, since it can be concluded that people who move from one municipality to another have different consumption patterns that modify the amounts of MSW. Other explanatory variables mentioned in Kolekar et al. (2016), for instance age, employment status, level of urbanization and environmental variables such as precipitation or temperature, were not used in this study since it is difficult to find a database with information on these variables.</p>	<p>корреляции Пирсона составило 0,985, что указывает на то, что оно также является переменной, наиболее связанной с образованием отходов, особенно в положительном ключе, т. Е. Большему населению соответствует большее количество ТБО. Переменная X Pd использовалась в нескольких публикациях. Бел и Мур (2009) используют эту переменную также для получения затрат, связанных с управлением отходами. В этом исследовании было получено значение r 0,161, что указывает на плохую корреляцию с переменной отклика. Анализ прогнозной модели 1 показал, что эта переменная не является статистически значимой, и ее использование должно приниматься с осторожностью. Переменная X Pbam положительно связана с переменной отклика. Коэффициент корреляции Пирсона составил 0,638. Эта переменная важна в области исследования, поскольку можно сделать вывод, что люди, которые переезжают из одного муниципалитета в другой, имеют разные модели потребления, которые изменяют количество ТБО. Другие объяснительные переменные, упомянутые в Kolekar et al. (2016), например, возраст, статус занятости, уровень урбанизации и переменные среды, такие как осадки или температура, не использовались в этом исследовании, так как трудно найти базу данных с информацией об этих переменных.</p>
<p>Other generated models</p>	<p>Другие сгенерированные модели</p>
<p>Eqs. 13, 14 and 15 show other models generated with the variables initially raised (models 3, 4 and 5 respectively). All these models are statistically significant and are also useful for forecasting purposes, but incorporate explanatory variables that are not significant; therefore, their results are not accurate (Table 5). Additionally, they have low parsimony because they incorporate more than 2 or 3 explanatory variables. Table 6 shows the statistical behavior of the predictors. The p-value and the VIF must be analyzed because they indicate multicollinearity between the variables and also their possible interpretation within the generated model.</p>	<p>Уравнения. 13, 14 и 15 показывают другие модели, созданные с первоначально поднятыми переменными (модели 3, 4 и 5 соответственно). Все эти модели являются статистически значимыми и полезными для целей прогнозирования, но включают в себя пояснительные переменные, которые не являются значимыми; следовательно, их результаты не точны (Таблица 5). Кроме того, они имеют низкую экономию, потому что они включают в себя более 2 или 3 объясняющих переменных. Таблица 6 показывает статистическое поведение предикторов. Значение p и VIF должны быть проанализированы, потому что они указывают мультиколлинеарность между переменными, а также их возможную интерпретацию в сгенерированной модели.</p>

$\ln Y_{Gen} = -8.48 + 1.13 \ln_{X_{Pop}} - 0.0019 \ln_{X_{Pd}} +$ $0.0315 \ln_{X_{Pbam}} - 0.0111 X_{Mi} \quad (13)$	
$\ln Y_{Gen} = -8.42 + 1.13 \ln_{X_{Pop}} - 0.0008 \ln_{X_{Pd}} +$ $0.0324 \ln_{X_{Pbam}} - 0.0070 X_{As} - 0.0119 X_{Mi} \quad (14)$	
$\ln Y_{Gen} = -8.22 + 0.995 \ln_{X_{Pop}} + 0.0039 \ln_{X_{Pd}} +$ $0.0261 \ln_{X_{Pbam}} - 0.0006 X_{As} + 0.133 \ln_{X_{Hgs}} - 0.00525 X_{Mi} \quad (15)$	
Inferences about the municipalities of the study area	Справки о муниципалитетах района исследования
Based on model 2 and its statistical analysis, inferences were made to forecast the generation rate of MSW in the municipalities of the CCS. The forecast was made with the most current data of the variables X Pop and X Pbam , corresponding to the year 2015. Fig. 5 shows MSW generation forecast and its comparison with the original database. In most of the municipalities of the CCS, the generation rate of MSW presented a gradual increase with respect to population growth (variable X Pop), except in Arriaga, Chiapilla, Osumacinta, Suchiapa, Teopisca, Tonalá, Venustiano Carranza and Villaflores, due to the fact that the population of these municipalities did not increase in the 2010-2015 period.	На основе модели 2 и ее статистического анализа были сделаны выводы для прогнозирования скорости образования ТБО в муниципалитетах Куэнка-дель-Каньон-дель-Сумидеро. Прогноз был сделан с использованием самых последних данных переменных X Pop и X Pbam, соответствующих 2015 году. На рис. 5 показан прогноз генерации ТБО и его сравнение с исходной базой данных. В большинстве муниципалитетов Куэнка-дель-Каньон-дель-Сумидеро уровень выработки ТБО демонстрировал постепенное увеличение в отношении роста населения (переменная X Pop), за исключением Арриага, Чьяпилла, Осумасинта, Сучьяпа, Теописка, Тонала, Венустиано Карранса и Вильяфлорес, из-за тот факт, что население этих муниципалитетов не увеличилось в период 2010-2015 гг.
Currently, the study area generates 1,600 tons of MSW/day, of which 74% comes from the regional heads such as Berriozábal, Ocozocoautla de Espinosa, San Cristóbal de las Casas, Tuxtla Gutiérrez and Villaflores.	В настоящее время в районе исследований вырабатывается 1600 тонн ТБО в день, из которых 74% приходится на главные регионы, таких как Берриозабаль, Окозокоаутла-де-Эспиноса, Сан-Кристобаль-де-лас-Касас, Тустла-Гутьеррес и Вильяфлорес.
Table 6: Statistical behavior of predictors in models 3, 4 and 5	Таблица 6: Статистическое поведение предикторов в моделях 3, 4 и 5

Model	Predictor	Coefficient	p-value	VIF	Comments
3	Constant	-8.48	0.000	-----	There is no multicollinearity among predictors, but some of them are not statistically significant (p-value> 0.05).
	$\ln_{X_{Pop}}$	1.13	0.000	1.947	
	$\ln_{X_{Pd}}$	-0.0019	0.937	1.300	
	$\ln_{X_{Pbam}}$	0.0315	0.053	3.556	
	X_{MI}	-0.0111	0.002	2.405	
4	Constant	-8.42	0.000	-----	There is multicollinearity between predictors (VIF> 5) and some of them are not statistically significant (p-value> 0.05).
	$\ln_{X_{Pop}}$	1.13	0.000	1.948	
	$\ln_{X_{Pd}}$	-0.0008	0.975	1.374	
	$\ln_{X_{Pbam}}$	0.0324	0.055	3.789	
	X_{AS}	-0.0070	0.844	5.373	
	X_{MI}	-0.0119	0.026	5.177	
5	Constant	-8.22	0.000	-----	
	$\ln_{X_{Pop}}$	0.995	0.000	5.634	
	$\ln_{X_{Pd}}$	0.0039	0.869	1.377	
	$\ln_{X_{Pbam}}$	0.0261	0.096	3.824	
	X_{AS}	-0.0006	0.986	5.384	
	$\ln_{X_{Hgs}}$	0.133	0.000	6.180	
	X_{MI}	-0.0525	0.309	5.710	

Fig. 5: Forecast of MSW generation rates in the municipalities of the study area

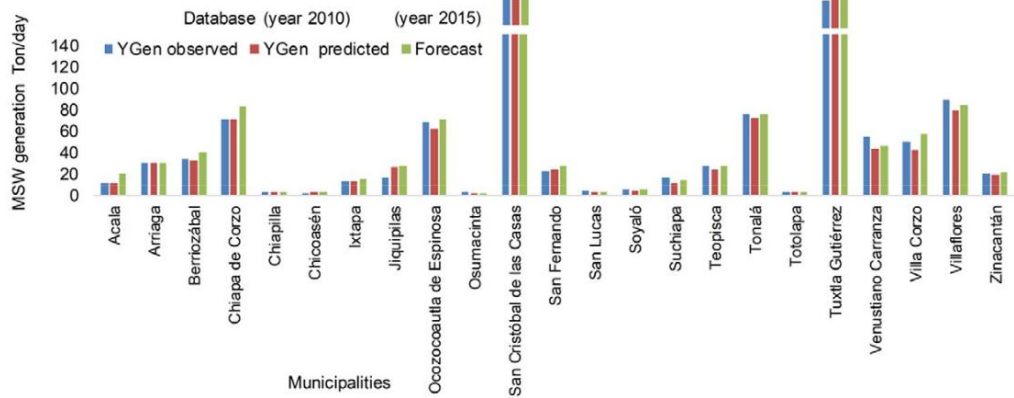


Рис. 5: Прогноз темпов образования ТБО в муниципалитетах района исследования

CONCLUSION

In this study, a forecast model was developed to determine the generation of MSW in the municipalities of the CCS, Chiapas State, Mexico. A MLR was used to obtain the forecast model with social and demographic explanatory variables. Two forecast models were presented and analyzed, with variables that met the multicollinearity test. The most important variables to predict the rate of MSW generation in the study area were the population of each municipality (X_{Pop}), the population born in another municipality (X_{Pbam}) and the population density (X_{Pd}). X_{Pop} is the most influential explanatory variable of waste generation, particularly it is related in a positive way. X_{Pbam} is less related to waste generation. X_{Pd} is the variable that least

Выводы

В этом исследовании была разработана прогнозная модель для определения образования ТБО в муниципалитетах Куэнка-дель-Каньон-дель-Сумидеро, штат Чьяпас, Мексика. MLR был использован для получения модели прогноза с социальными и демографическими пояснительными переменными. Были представлены и проанализированы две прогнозные модели с переменными, которые удовлетворяют критерию мультиколлинеарности. Наиболее важными переменными для прогнозирования скорости образования ТБО в районе исследования были население каждого муниципалитета (X_{Pop}), население, рожденное в другом муниципалитете (X_{Pbam}), и плотность

<p>influences waste generation prediction; in addition, it can present problems of correlation with other explanatory variables. Although other variables, such as daily per capita income (X Dpi) and average schooling (X As), are very important, they do not seem to have an effect on the response variable in this study. The user of this forecast model should use model 2, since it is the one with the highest parsimony (it uses fewer variables); R 2 adj , MAPE, MAD and RMSE values indicated high influence on the explained phenomenon and high forecasting capacity. Additionally, it is important to mention that when using the models proposed for forecasting purposes, it is necessary to make a transformation in the explanatory and response variables (use inverse of natural logarithm). The inferences made on the municipalities of the study area showed that, except in some municipalities, the MSW generation rate usually presented a gradual increase with respect to population growth and with respect to the Number of inhabitants that were born in another entity (migration). Finally, this study can be a solid basis for comparison for future research in the area of study. It is possible to use different mathematical models such as artificial neural network, principal component analysis, time-series analysis, etc., and compare the response variable or the predictors.</p>	<p>населения (X Pd). X Pop - самая влиятельная объяснительная переменная образования отходов, особенно это связано с положительным отношением. X Pbam меньше связан с образованием отходов. X Pd - переменная, которая меньше всего влияет на прогноз образования отходов; Кроме того, это может представлять проблемы корреляции с другими объясняющими переменными. Хотя другие переменные, такие как ежедневный доход на душу населения (X Dpi) и среднее образование (X As), очень важны, они, по-видимому, не влияют на переменную ответа в этом исследовании. Пользователь этой модели прогноза должен использовать модель 2, так как это модель с наибольшей экономией (в ней используется меньше переменных); Значения R 2 adj, MAPE, MAD и RMSE указывают на высокое влияние на объясненное явление и высокую способность прогнозирования. Кроме того, важно отметить, что при использовании моделей, предлагаемых для целей прогнозирования, необходимо произвести преобразование в объясняющую и ответную переменные (используйте обратный натуральный логарифм). Сделанные выводы по муниципалитетам исследуемой территории показали, что, за исключением некоторых муниципалитетов, уровень выработки ТБО обычно представляет собой постепенное увеличение в отношении роста населения и количества жителей, родившихся в другом образовании (миграция). Наконец, это исследование может стать надежной основой для сравнения будущих исследований в этой области. Можно использовать различные математические модели, такие как искусственная нейронная сеть, анализ главных компонентов, анализ временных рядов и т. Д., И сравнивать переменную отклика или предикторы.</p>
<p>ACKNOWLEDGEMENTS</p>	
<p>This study was supported by the Project Support Programme for Research and Technological Innovation [Project UNAM DGAPA-PAPIIT IN105516]. We also appreciate the support of Mexican National Council for Science and Technology (CONACYT) to carry out this work.</p>	
<p>CONFLICT OF INTEREST</p>	
<p>The author declares that there is no conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancy have been completely observed by the authors.</p>	

ABBREVIATIONS	СОКРАЩЕНИЯ
α Level of significance	Уровень значимости
A t Observed value	Наблюдаемое значение
β_i (1,2,3 ... k) Regression coefficients	(1,2,3 ... k) Коэффициенты регрессии
CCS Cuenca del Cañón del Sumidero	Куэнка-дель-Каньон-дель-Сумидеро
d Durbin-Watson test	Тест Дурбина-Ватсона
Eq. Equation	Уравнение
F Fisher test	Тест Фишера
F t Predicted value	Прогнозируемая стоимость
H 0 Null hypothesis	Нулевая гипотеза
k Number of explanatory variables included in the model	Количество объясняющих переменных, включенных в модель
ln-Y Gen Natural logarithm of MSW generation	Натуральный логарифм генерации ТБО
ln-X Pop Natural logarithm of population	Натуральный логарифм населения
ln-X Pd Natural logarithm of population Density	Натуральный логарифм плотности населения
ln-X Pbam Natural logarithm of population born in another municipality	Натуральный логарифм населения, рожденного в другом муниципалитете
ln-X Hgs Natural logarithm of household with goods and services	Натуральный логарифм домашнего хозяйства с товарами и услугами
ln-X Ces Natural logarithm of commercial establishments and services	Натуральный логарифм коммерческих учреждений и услуг
ln-X Dpi Natural logarithm of daily per capita Income	Натуральный логарифм суточного дохода на душу населения
MAD Mean Absolute Deviation	Среднее Абсолютное Отклонение
MAPE Mean absolute percentage error	Средняя абсолютная ошибка в процентах
MLR Multiple Linear Regression	Множественная линейная регрессия
MSW Municipal Solid Waste	Твердые бытовые отходы
n Sample size	Размер образца
p-p plot Probability-probability plot	График вероятности-вероятности
p-value Probability value	Значение вероятности
r Pearson correlation coefficient	Коэффициент корреляции Пирсона
r-value Pearson correlation coefficient	Коэффициент корреляции Пирсона
R 2 Coefficient of determination	Коэффициент детерминации
R 2 adj Adjusted coefficient of determination	Скорректированный коэффициент детерминации
R 2 jackknife Jackknife coefficient of determination	Коэффициент определения складного ножа
RMSE Root Mean Square Error	Средняя квадратическая ошибка
SSE Sum of Squared Errors	Сумма квадратов ошибок
SS Y Y Sum of the squares of the difference of (Y i) and the (Y)	Сумма квадратов разности (Y i) и (Y)
VIF Variance Inflation Factor	Коэффициент инфляции дисперсии
X i (1,2,3 ... k) Explanatory variables	(1,2,3 ... k) Пояснительные переменные

X As Average schooling	Среднее образование
X Ces Commercial establishments and Services	Коммерческие учреждения и Услуги
X Dpi Daily per capita income	Ежедневный доход на душу населения
X Hgs Household with goods and services	Домашнее хозяйство с товарами и услугами
X Mi Marginalization index	Индекс маргинализации
X Pbam Population born in another Municipality	Население, рожденное в другом муниципалитете
X Pd Population density	Плотность населения
X Pop Population	численность населения
Y Average of observed data	Среднее из наблюдаемых данных
Y Gen MSW generation	Образованин ТБО
Y i Value of each individual observation	Ценность каждого отдельного наблюдения
Y \hat{y}_i Predicted value	Прогнозируемое значение