Paulina Rejda

e-mail: rejdapaulina@gmail.com

Uniwersytet Ekonomiczny we Wrocławiu

Trends and Causes of Child Mortality in African Countries

JEL Classification: C2, C38, N37

DOI: 10.15611/2022.17.6.06

Abstract: The problem of child mortality in African countries is discussed in this article. The goal of this study was to examine the general situation in Africa in terms of economic and living conditions as well as their impact on child mortality in 2000 and 2019 and to select the most appropriate models to characterise this phenomenon. Cluster analysis was used to identify countries with similar socioeconomic characteristics and living conditions. Several regression models were estimated, analysed, and compared. Through the use of spatial models, the geographic context was incorporated into the statistical framework of the regression. The analysis demonstrated a significant improvement in economic and living conditions over the 19-year period. The Spatial Error Model turned out to be the best model for the 2000 data, however, OLS and Negative Binomial also performed well. There was no spatial autocorrelation in 2019 and none of the estimated models provided a good fit for the data. The 2000 models' projections revealed unexpected outcomes. Despite the fact that a rise in life expectancy, access to drinking water, and GDP *per capita* was intended to result in a decrease in the dependent variable, the model estimates predicted that unemployment and the number of HIV-positive children would have the same effect on child mortality.

Keywords: child mortality, African countries, spatial analysis, cluster analysis, regression models.

1. Introduction

Regardless of baseline levels, socioeconomic circumstances, or development tactics, the 20th century saw remarkable reductions in mortality in practically every country on the planet. At the end of the 19th century, decreases were already seen in advanced economies (Ahmad, Lopez, & Inoue, 2000). Significant decreases in developing countries had not occurred until shortly after World War II ended. According to the authors, the initial size of the declines in developing countries was so impressive that it caused speculation in the 1960s and 1970s that the gap in mortality rates between developed and developing countries would reduce greatly by the end of the century. But then, the world became pessimistic about the progress made in reducing mortality. As mentioned by the authors, the concerns led to research on the determinants, patterns, and nature of children's mortality. Researchers examined individual, biological, social, behavioural, demographic, and economic determinants as well as the discrepancies in environmental, cultural, and material living conditions of children that might influence child mortality.

Since 1990, the global death rate of children under the age of 5 has decreased by 59% (Ahmad et al., 2000). According to new mortality estimates, the highest decline was in 2019, when global under-5 deaths fell to 5.2 million from 12.5 million in 1990 (Ahmad et al., 2000). Four years after the Sustainable Development Goals (United Nations [UN], 2015) had been introduced, significant progress in reducing the number of child deaths was made. These objectives include eliminating preventable child mortality, lowering neonatal death rates to less than 12 per 1,000 live births, and lowering child mortality rates to less than 25 per 1,000 live births (Ahmad et al., 2000). The deadline for achieving these goals is 2030. It might be too early to think that the success was a result of the Sustainable Development Goals. However, development in housing, water, sanitation, education, nutrition, and financial security may have contributed to the advancement. It is also possible that the availability of health services contributed to the improvement of the situation. Despite the above-mentioned progress, Sub-Saharan Africa had the highest infant mortality rate in 2019, with 27 deaths per 1,000 live births (Ahmad et al., 2000). According to the authors, a child born in Sub-Saharan Africa was ten times more likely to die in the first month of its life than a child born in a wealthy country. Similarly, sub-Saharan Africa continued to have the world's highest under-5 mortality rate. In 2019, 1 in every 13 children in the region died before reaching the age of 5 (Ahmad et al., 2000).

This article deals with the topic of child mortality in African countries. The goals of this study were to compare the overall situation in Africa regarding the economic and living conditions and their impact on child mortality at the end of the 20th century with the latest year for which data were available, in this case – 2019, and choose the best models to describe this phenomenon.

In order to achieve the above goal, it was decided to examine two detailed problems which were formulated in the form of questions:

- What characterises the countries with the highest and the lowest child mortality rates?
- What are the effects of specific socioeconomic factors on child mortality, and which models best explain this phenomenon?

The data for this research was obtained from the World Bank's World Development Indicators (WDI) database (The World Bank, b.d.). The analysis was carried out using cross-sectional data from 2000 and 2019. Over 40 African countries were analysed using 8 indicators that were chosen based on the possible impact they might have had on the mortality rate of children.

To distinguish countries of similar socioeconomic characteristics and living conditions, a cluster analysis was performed. For the deeper understanding of the dynamics of the relationship between child mortality and the chosen indicators, several regression models were estimated, interpreted, and compared to choose those that best explain the studied phenomenon. The geographic context was incorporated into the regression's statistical framework through the adoption of spatial models.

2. Child mortality in Africa

Since World War II, most developing nations have seen significant reductions in infant and child mortality rates, due to both improved standards of living, and national and international public health programs (Hill & Pebley, 1989). Once these declines were confirmed, considerable optimism arose in the 1960s and 1970s about the chances for a child survival revolution in the Third World. However, as stated by the authors, in the 1980s, such optimism was replaced by a lot of pessimism regarding the progress made in lowering child mortality and the chances for further improvement. According to the estimates of the mortality rate for children under 5 years of age from the UN Inter-agency Group for Child Mortality Estimation (IGME), global under-5 mortality rates have decreased by one-third – from 89 deaths per 1,000 live births in 1990 to 60 deaths per 1,000 live births in 2009 (You, Jones, Hill, Wardlaw, & Chopra, 2010). The authors stated that over the same time span, the number of children under the age of 5 who died has fallen from 124 million in 1990 to 81 million in 2009. The authors claimed this was a strong indication that progress was being made on reducing child mortality in all parts of the world, with many regions having lowered the under-5 mortality rates by 50% or more (You et al., 2010).

Individual, household, and spatially explicit geographical factors all affect the rates of infant and child mortality in West African countries (Balk, Pullum, Storeygard, Greenwell, & Neuman, 2004). Brockerhoff and Hewett (2000) stated that there are significant differences in the likelihood of early childhood survival amongst African ethnic groups. According to the authors, child mortality disparities are closely linked to women's education, ethnic inequalities in household economic position, degree of concentration in the major metropolitan area, and access to and usage of healthcare services.

Harttgen and Misselhorn (2006) used a multilevel model approach to evaluate the effects of individual, household, and cluster socioeconomic factors as well as their relative influences on child mortality and undernutrition in sub-Saharan Africa and South Asia. Wagner et al. (2018) conducted a geospatial analysis on the relationship between armed conflict and child mortality in Africa and discovered that up to 8 years after an armed conflict, even if it was 100 km away, the result was that the cumulative increase in child mortality was 2 to 4 times higher than the contemporaneous increase. Kiros and Hogan (2001) used a Poisson regression model to examine parental education's impact on reducing excess child mortality in Africa in the context of war and famine. Omariba and Boyle (2007) applied a multilevel logistic regression to see if the link between child mortality and family structure differs cross-nationally and over time, with a focus on polygyny, and found that improving maternal education would be highly beneficial to child health. Another study that confirmed the great influence of the mother on the child's well-being was the study by Rawlings and Siddique (2020), who proved that domestic violence affects the mortality of children, focusing mainly on the violence faced by the mother. Research on the relationship between child mortality and HIV infection in Africa was conducted by Newell et al. (2004), who used a Poisson regression model in their analysis.

3. Methodology

3.1. Methods, techniques, and research tools

To better understand the characteristics and causes of child mortality in Africa, several analyses were carried out using various statistical methods.

As the first step, descriptive statistics and correlograms were computed to understand the datasets. A correlogram is a representation of correlation statistics, which specifies the variable quantities' interdependence.

In order to distinguish groups of countries that were similar to each other in terms of the analysed indicators, a cluster analysis was carried out. Grouping was performed with the use of the Euclidean distance and the complete linkage method, also known as the farthest neighbour clustering. Cluster analysis is a term used to describe a family of statistical procedures specifically designed to perform classification across complex datasets. Its purpose is to group objects into clusters so that objects belonging to one group have more in common with each other than with the objects of other clusters. The main goal of the cluster analysis is to divide objects into relatively homogeneous groups based on the many characteristics that define them. It is assumed that the full cluster analysis consists of the following steps (Korzeniewski, 2012):

1) selection of variables and objects,
2) visualisation of variables or objects,
3) normalisation of variables,
4) selection of a measure of distance between the objects,
5) determining the number of clusters,
6) grouping the objects,
7) evaluating the obtained results,
8) characterising, describing, and profiling the created groups.

An important problem in the cluster analysis concerns the method of estimating the optimum number of clusters in the analysed dataset. Most of the ways require subjective analysis. One of the adopted methods is the visual assessment of the hierarchical tree diagram, also known as a dendrogram, based on which it is decided to assume a specific number of *branches* representing the individual clusters (Sneath & Sokal, 1973). In this study, the optimum number of clusters was chosen based on scree plots using the elbow method, i.e., a heuristic which consists of creating a plot of the explained variation as a function of the number of clusters, and then picking the inflexion point of the curve called an 'elbow' which indicates the best number of groups (Joshi & Nalwade, 2013). A scree plot depicts the proportion variance explained as a function of the principal components decreasing in size and is used to visually determine the suitable number of components for further investigation.

To examine the effect of chosen indicators on child mortality, different regression models were estimated. As the first step, multiple linear regression models were estimated using the Ordinary Least Squares (OLS), i.e., a method used for calculating the coefficients of linear regression equations that represent the connection between the independent quantitative variables and the response variable (Dismuke & Lindrooth, 2006).

In some cases, standard OLS regression may produce biased results (Basu, 2019). In the next step, Poisson and Negative Binomial models were created. The Poisson regression model is a generalised linear model with the logarithm as the link function, where the dependent variable has a Poisson distribution. It can be used for modelling count data and contingency tables. Its parameters are estimated using the Maximum Likelihood Estimation (MLE) (Pan & Fang, 2002).

The Negative Binomial model can be used for over-dispersed count data. It is a generalisation of the Poisson model. It does not make the assumption that the variance and mean are equal. The Negative Binomial model has the same mean structure as the Poisson model with an additional parameter to account for over-dispersion (Hilbe, 2011).

Next, to measure the overall spatial autocorrelation in the datasets, Moran's I tests on the OLS residuals were carried out. The Global Moran's I tool measures the similarity of an object to the objects surrounding it using both feature locations and feature values simultaneously (Getis, 2010).

Then, if the test confirmed the presence of autocorrelation, spatial regression models were created. In spatial regression models, the link between areal units is specified exogenously with the use of a weight matrix mimicking the spatial structure and spatial interaction pattern. The specific-to-general approach (Elhorst, 2010) was adopted when creating and selecting models. Starting from the non-spatial model – OLS, going through models containing one type of spatial interaction effect – Spatial Error Model (SEM), Spatial Autoregressive Model (SAR) and Spatial Lag of X model (SLX), and then Spatial Autoregressive Combined model (SAC), Spatial Durbin Model (SDM) and Spatial Durbin Error Model (SDEM), in which the previous models are nested, up to the General Nesting Spatial model (GNS), also known as the Manski model containing all the types of spatial interactions, appropriate models were selected and evaluated (Ward & Gleditsch, 2008).

To determine which model was most likely to provide the best fit for a given dataset, Akaike Information Criterion (AIC) scores were compared. The Akaike information criterion (AIC) estimates the prediction error and thus the relative quality of statistical models, using the maximum likelihood estimation (Bertrand, 1988).

Then, the Likelihood-Ratio tests were performed to check if the more complex models should be chosen over the simpler models, with no spatial interaction effects.

The Likelihood-Ratio test is a statistical method used to compare the goodness of fit of two competing statistical models using the likelihood ratio (Kent, 1982).

3.2. Data source. Selection and description of the variables

The data used in this analysis comes from the World Bank's World Development Indicators (WDI) database and relates to human development. In the beginning, a dataset for the year 2000 was created. It contained 46 indicators for 45 African countries. Indicators were selected that concerned the quality of life, which affects the mortality of children. The dependent variable was the mortality rate of children under 5 years of age. The other indicators were selected based on the possible impact they might have had on the dependent variable. Then, the procedure of removing missing data was performed. Variables containing over 10% missing data were removed to keep as many observations as possible. Islands were removed from the dataset as they technically have no neighbours and therefore would not contribute to the results of the spatial analysis.

The result of this procedure was a table consisting of 8 variables and 45 observations. It contained the following indicators:

- *mortality_under_5* – mortality rate under the age of 5 (per 1,000 live births),
- *life_expectancy* – life expectancy at birth (years),
- *infant_mortality* – infant mortality rate (per 1,000 live births),
- *access_to_drinking_water* – people using at least basic drinking water services (percentage of the population),
- *unemployment* – unemployment rate (percentage of the total labour force),
- *children_with_HIV* – number of children (from 0 to 14 years of age) living with HIV,
- *people_with_HIV* – number of people living with HIV (including children and adults),
- *GDP_per_capita* – GDP per capita (current USD),
- *acc_to_electricity* – people having access to electricity (percentage of the total population).

The table created for the year 2019 was built using the same country codes and indicators as the final table for 2000. It included 3 countries less, due to the problem with data availability.

4. Results

4.1. Correlations and statistics

Figures 1 and 2 show correlations between the variables in 2000 and 2019. In both years a strong positive relationship between child mortality and mortality rate among infants was observed. A high correlation between the response variable and an independent variable is always eligible, but in this case, the number of deaths

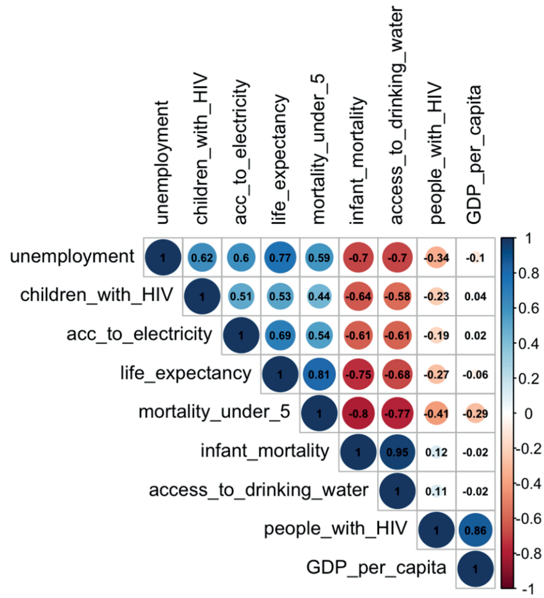

Figure 1. Correlogram (2000)

Source: own elaboration with the use of the R programming language.

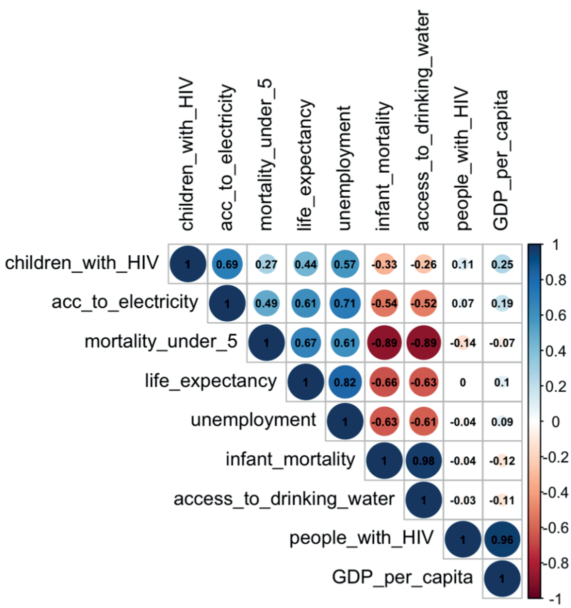

Figure 2. Correlogram (2019)

Source: own elaboration with the use of the R programming language.

of children below 5 years of age also contained the number of infant deaths, so this indicator was deleted from the dataset. There was also a correlation of the same nature between the number of children and the number of people living with HIV. This was the same case as in the previously described variables. The variable *people_with_HIV* was deleted because a strong correlation between independent variables might cause a collinearity issue in future models. In the case of 2019 data, the following variables have been removed: *life_expectancy* (highly correlated with *unemployment*), *infant_mortality*, and *people_with_HIV*.

Tables 1 and 2 contain descriptive statistics that summarise the characteristics of both datasets. The summaries contain minimum and maximum values, percentiles, medians, and means. There is a conclusion that can be drawn out by comparing means – economic and living conditions improved over the examined period of 19 years. Average child mortality dropped by 52% as well as average unemployment and the average number of children living with HIV, which decreased by 11% and 12%, respectively. Average access to drinking water increased by 24%, average access to electricity – by 62%. The highest difference was observed in GDP *per capita*, which in 2019 was 117% higher than in 2000. In the year 2000, the middle 50% of the countries had between 98 and 129 deaths of children per 1,000 live births. In 2019, the range included much smaller values – for every 1,000 live births there were between 43 and 82 child deaths. The mean value of this indicator also decreased. In the analysed period, the maximum child mortality in a given country fell by almost half – from 227 to 117 deaths of children.

Table 1. Descriptive statistics (2000)

Variable	Min	Percentile [25]	Median	Mean	Percentile [75]	Max
life_expectancy	39.40	48.65	50.99	53.12	56.05 7	73.17
mortality_under_5	28	98	138	129.29	169 71	227
access_to_drinking_water	18.10	43.43	55.99	56.39	71.40 28	97.89
unemployment	0.80	3.08	5.64	8.93	12.93 10	35.27
children_with_HIV	100	2 000	13 000	41 501.78	70 000 68k	200 000
GDP_per_capita	124.50	270.54	410.95	959.40	1 007.47 736	7 142.77
acc_to_electricity	0	8.41	20.10	31.35	44.68 23	99.80

Source: own elaboration.

Table 2. Descriptive statistics (2019)

Variable	Min	Percentile [25]	Median	Mean	Percentile [75]	Max
mortality_under_5	11	42.50	57	61.79	82	117
access_to_drinking_water	38.20	60.39	68.37	69.97	82.29	99.89
unemployment	0.50	3.03	5.02	7.93	10.55	28.47
children_with_HIV	100	3 450	10 000	36 738.14	44 000	330 000
GDP_per_capita	228.20	762.66	1 219.52	2 077.21	2 589.68	7 767
acc_to_electricity	8.40	34.37	48.02	50.79	68.54	100

Source: own elaboration.

4.2. Cluster analysis

To get a better understanding of the similarities among the analysed countries, cluster analysis was carried out using the normalised data. Figures 3 and 4 show the scree plots created to make the best choice deciding on the optimum number of clusters. The scree plots allowed for the selection of the number of classes based on the number of clusters for which average distances to the centroid falls suddenly. The numbers were chosen using the *elbow method*. For 2000 there was a visible elbow at $K = 3$, for 2019 the optimum number of clusters was 4.

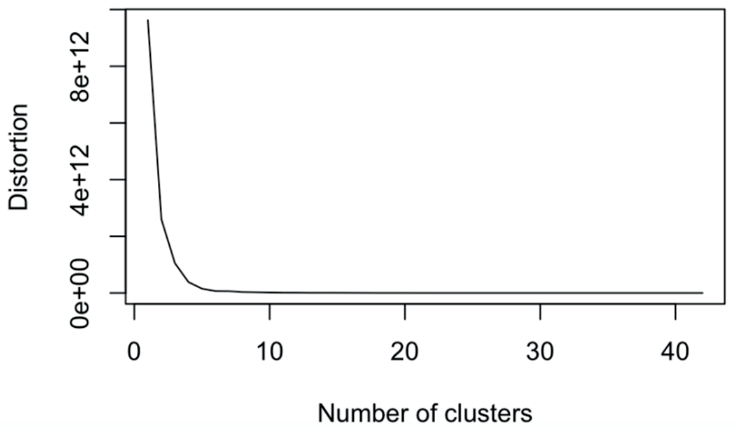

Figure 3. Scree plot (2000)

Source: own elaboration with the use of the R programming language.

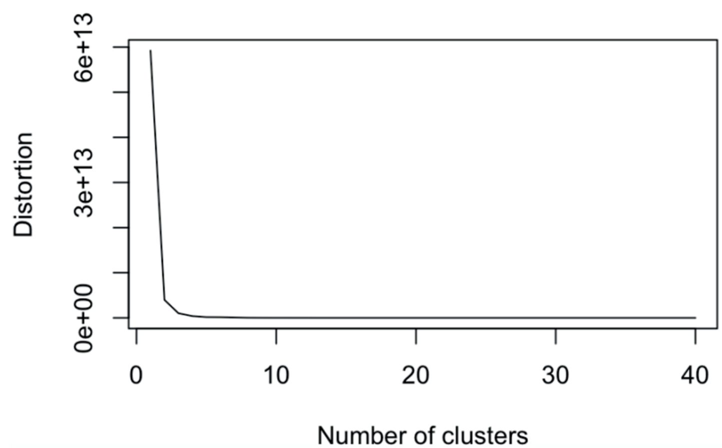

Figure 4. Scree plot (2019)

Source: own elaboration with the use of the R programming language.

As a confirmation of the above conclusions, an R package called NbClust() was used. According to the outcomes, the best numbers of clusters were 3 and 4 for 2000 and 2019, respectively, which confirmed the previous presumption.

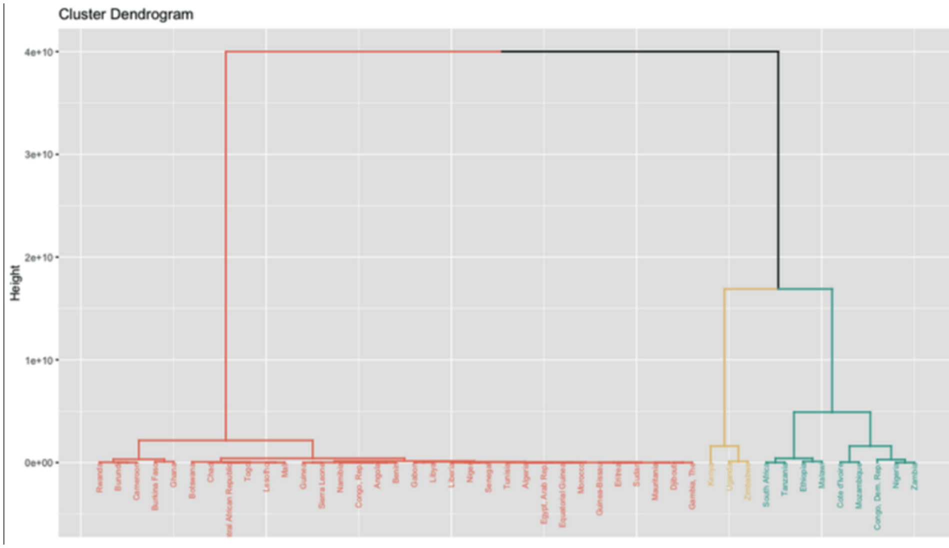

Figure 5. Dendrogram (2000)

Source: own elaboration with the use of the R programming language.

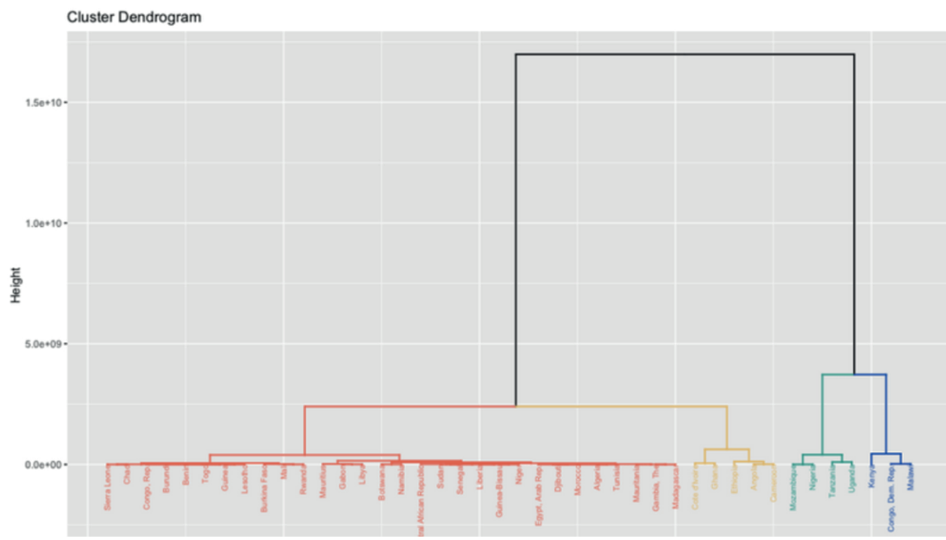

Figure 6. Dendrogram (2019)

Source: own elaboration with the use of the R programming language.

Grouping was performed using the Euclidean distance and the complete linkage method. This method was chosen because it is less prone to noise and outliers, yet it has the ability to split large clusters and prefers globular shapes. Figures 5 and 6 depict dendrograms showing the results of the above-mentioned hierarchical clustering procedures.

Figure 7 shows a map of Africa, where the colours reflect the affiliation of the countries to certain groups. Table 3 contains the mean values of variables in all created clusters.

Figure 7. Map of Africa showing the clustering results (2000)

Source: own elaboration.

The first group (red) included countries having the best living conditions – with the longest life expectancy, the highest GDP *per capita,* the highest percentage of people having access to electricity, and using at least basic drinking water services. The second (yellow) group had the highest child mortality rate and the lowest access to drinking water. Countries belonging to the third (green) cluster had the lowest life expectancy and the lowest electricity availability. Surprisingly, despite the observation of the largest number of children living with HIV, the lowest child mortality rate was observed in the countries belonging to this group.

Table 3. Mean values of variables in the created clusters (2000)

Group	*morta-lity_un-der_5*	*access_to_drin-king_water*	*unemploy-ment*	*children_with_HIV*	*GDP_per_ca-pita*	*acc_to_electricity*	*life_expectancy*
1 (red)	126	60.31	9.7	10 563	1 067.84	34.34	55
2 (yellow)	146	44.62	7.7	109 889	745.80	24.42	49
3 (green)	112	48.55	3.8	176 667	407.47	19.31	47

Source: own elaboration.

Figure 8. Map of Africa showing the clustering results (2019)

Source: own elaboration.

Figure 8 contains a map of Africa, coloured based on the results of the clustering procedure for 2019, and Table 4 presents the mean values of variables in the 4 created clusters. The first group (yellow) was characterised by the lowest unemployment rate and the highest percentage of people having access to electricity. The second cluster (red) contained countries that had a considerably low number of children infected with HIV and quite good living conditions – with the highest access to drinking water, moderately high access to electricity, and the highest GDP *per capita*, which was twice the value of group 3. The third group (blue), on the contrary, included countries with the lowest standard of living concerning life expectancy, GDP *per capita*, and access to water and electricity. The last cluster (green) had the highest child mortality rate and the largest number of children living with HIV. Countries belonging to this group had the lowest average unemployment rate and quite low GDP *per capita* as well as the percentage of people having access to electricity and water.

Table 4. Mean values of variables in the created clusters (2019)

Group	*mortality_under_5*	*access_to_drinking_water*	*unemployment*	*children_with_HIV*	*GDP_per_capita*	*acc_to_electricity*
1 (yellow)	65	65.20	3.9	35 600	1 944.29	61.89
2 (red)	62	72.66	8.9	6 384	2 237.36	51.47
3 (blue)	57	60.65	5.8	80 600	1 141.39	36.82
4 (green)	72	62.84	3.9	122 500	1 155.24	41.00

Source: own elaboration.

Based on a visual assessment of the division of classes, it can be concluded that the division appears to be consistent. It is worth noting that in both years countries like the Democratic Republic of Congo, Zambia and Malawi happened to be in the group that was characterised by the lowest standard of living. Uganda's situation did not change either – in both 2000 and 2019, the country was in a group with the highest number of child deaths per 1,000 live births.

4.3. OLS, Poisson and Negative Binomial regression models

As the first step preceding the models' creation, the normality of the distribution of the explained variable was checked by conducting a Shapiro-Wilk normality test (using the shapiro.test() function from the stats package), which proved that child mortality had a normal distribution for both years, since in both cases the *p*-value was higher than 0.05. Figures 9 and 10 include histograms for 2000 and 2019, respectively.

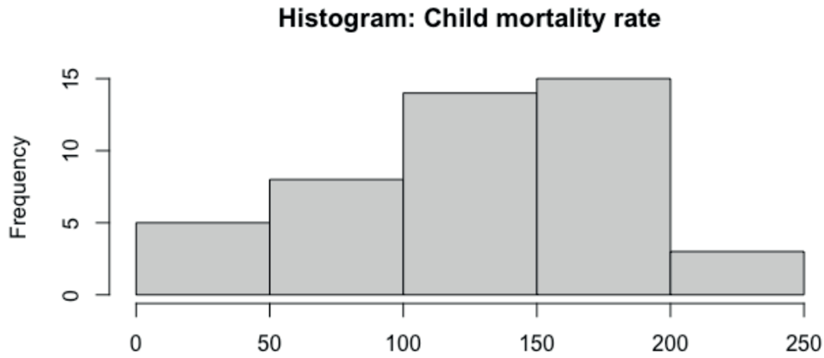

Figure 9. Histogram of the dependent variable (2000)

Source: own elaboration with the use of the R programming language.

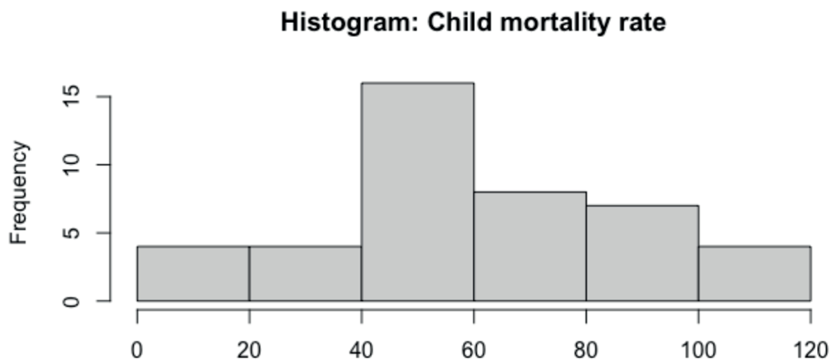

Figure 10. Histogram of the dependent variable (2019)

Source: own elaboration with the use of the R programming language.

The regression models were estimated using the R program. The first linear regression model estimated using the Ordinary Least Squares (OLS) method included all variables from the 2000 dataset. According to the coefficient of determination, R^2, which was equal to 0.83, the fit was quite good. All variables were significant, except for *GDP_per_capita* and *access_to_electricity*.

Two new models were created excluding the above-mentioned variables separately. The R^2 did not change much and *GDP_per_capita* and *access_to_electricity* were still insignificant in the second and third model, respectively.

In the third model, which did not include *access_to_electricity*, the significance of *unemployment* increased compared to the initial model. The R^2 was equal to 0.83, which meant that the independent were drawn:

- if life expectancy at birth increased by 1 year, the mortality rate for children would drop by 4.32 per 1,000 live births, holding other variables constant;
- if unemployment increased by 1, variables explained 83% of the variance in the dependent variable. All the indicators were significant. Table 5 includes the coefficients of the final OLS model. Based on the parameter estimates the following conclusions percentage point, child mortality would decrease by 1.85 deaths per 1,000 live births, holding other variables constant;
- if the number of HIV-positive children increased by 100,000, there would be 28.13 fewer child deaths for every 1,000 live births, holding other variables constant;
- if the percentage of people using at least basic drinking water services increased by 10 percentage points, for every 1,000 live births, 6.29 fewer children would die, holding other variables constant.

Table 5. OLS model coefficients (2000)

	Coefficients	***t* statistic**	***p*-value**
(Constant)	422.5	15.206	< 2e-16
life_expectancy	−4.322	−7.658	2.31e-09
access_to_drinking_water	−0.6292	−2.524	0.015683
unemployment	−1.847	−3.635	0.000784
children_with_HIV	−0.0002813	−4.303	0.000106

Source: own elaboration with the use of the R programming language.

As the next step, the residual analysis was performed. Figure 11 shows the scatter plot of the residuals and the predicted values. Considering the position of the points, there was a suspicion of a non-linear relationship. The studentized Breusch--Pagan test with the test statistic of 1.92 and a p-value of 0.75 proved the absence of heteroscedasticity.

Figures 12 and 13 show the histogram and the Quantile-Quantile (Q-Q) plot for the 2000 data used to check the normality of the residuals. The Q-Q plot indicated that the distribution of the residuals was heavy-tailed. The Shapiro-Wilk test had a test statistic of 0.94886 and a p-value of 0.046, which meant that with the significance level of 0.05, the residuals were not normally distributed.

The OLS model for 2019 data included all the variables. The intercept was highly significant and access to electricity was the only indicator of any significance. If the percentage of people having access to electricity increased by 10 percentage points, for every 1,000 live births, 4 fewer children were expected to die. The overall fit of the model was not very good – the R^2 was equal to 0.47.

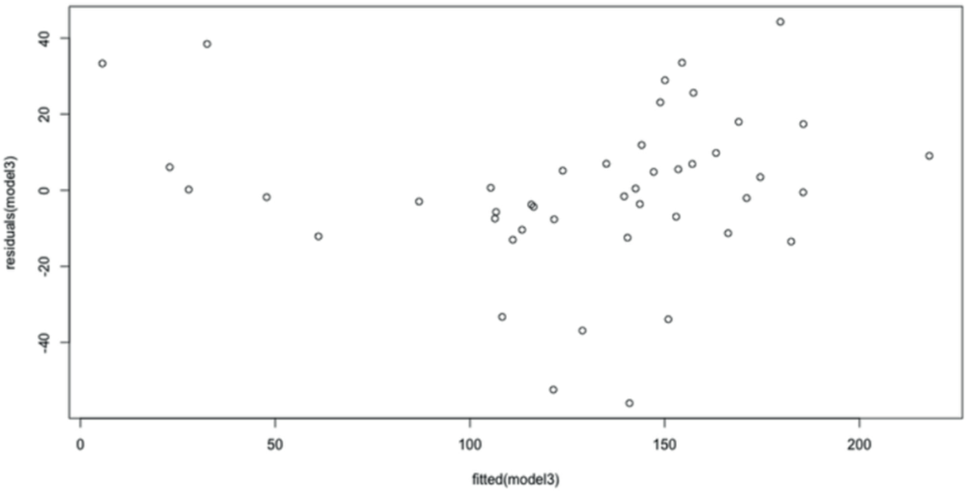

Figure 11. Plot of residuals *versus* fitted values (2000)

Source: own elaboration with the use of the R programming language.

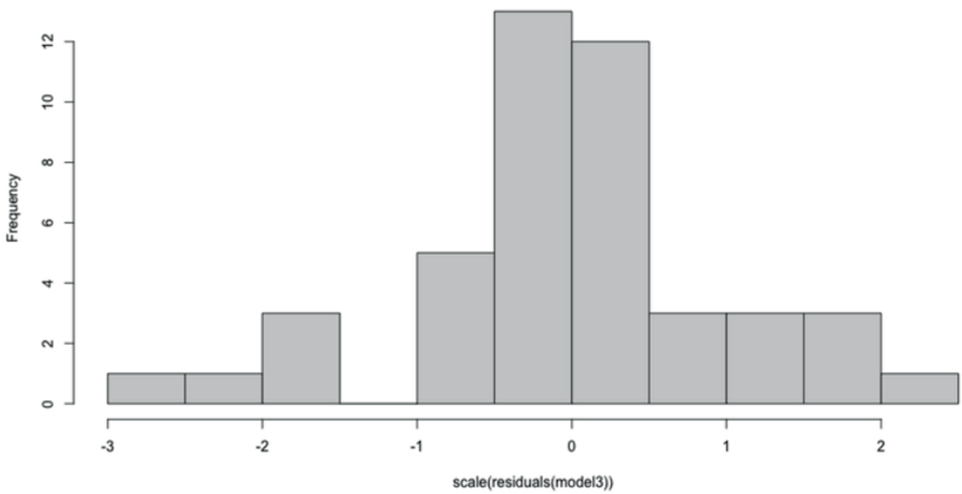

Figure 12. Histogram of the residuals (2000)

Source: own elaboration with the use of the R programming language.

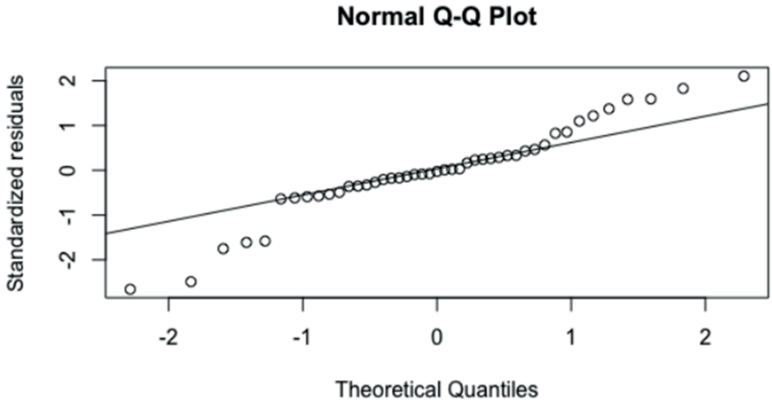

Figure 13. Quantile-Quantile plot (2000)

Source: own elaboration with the use of the R programming language.

Figure 14 depicts the scatter plot of the residuals and the predicted values of the dependent variable. The points seemed to be randomly distributed across the horizontal axis, no specific trends in the distribution of points were noticed. The studentized Breusch-Pagan test with the test statistic of 6.2165 and a p-value of 0.29 was proof of homoscedasticity.

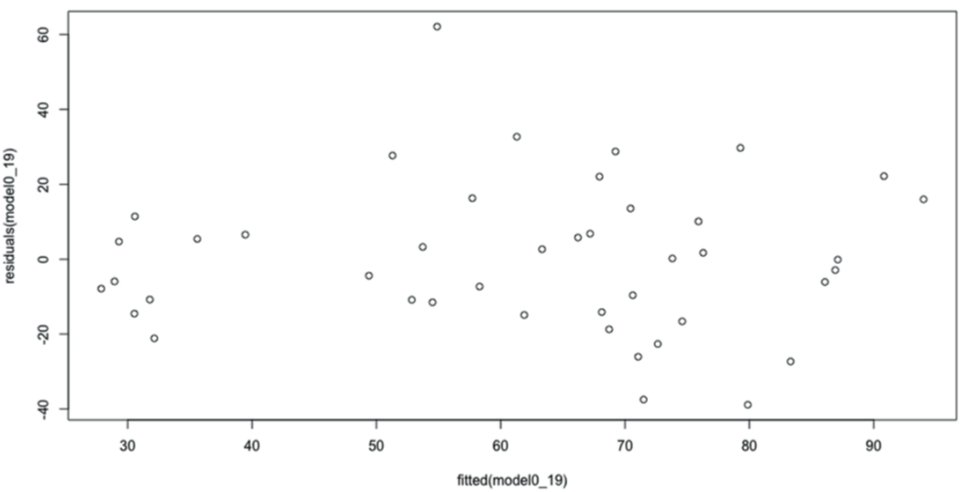

Figure 14. Plot of residuals *versus* fitted values (2019)

Source: own elaboration with the use of the R programming language.

In 2019, the Q-Q plot had lighter tails compared to the plot from the previous period (see Figure 15). The residuals had a normal distribution as confirmed by the

Shapiro-Wilk test with the test statistic of 0.9503 and a p-value of 0.46. Figure 16 depicts the histogram of the residuals.

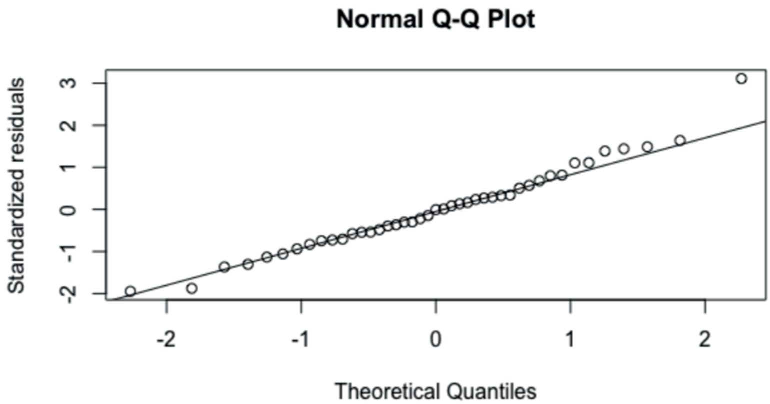

Figure 15. Quantile-Quantile plot (2019)

Source: own elaboration with the use of the R programming language.

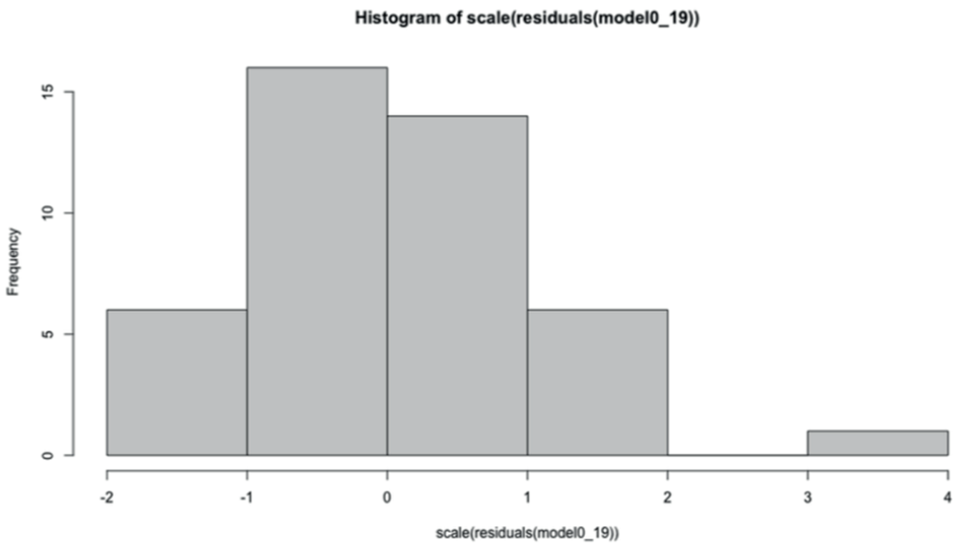

Figure 16. Histogram of the residuals (2019)

Source: own elaboration with the use of the R programming language.

As the next step, the Poisson models were estimated. The original models contained all the indicators. In the case of data from 2000, the second model did not include access to electricity, as it was the only variable that was not significant in the initial model. For 2019, the number of children living with HIV was excluded for the same reason. In both cases, the Akaike Information Criterion (AIC) scores decreased (from 443.63 to 441.72 in 2000 and from 549.81 to 548.53 in 2019), indicating that the modified models fitted the data better.

To see if the variability was higher than expected, the overdispersion tests were conducted. With dispersion estimates greater than 1 (2.9 in 2000 and 6.83 in 2019) and very low *p*-values (0.00043 and 0.00052), the tests confirmed the presence of overdispersion in both models.

Due to the evidence of excessive dispersion, the Negative Binomial regression models were estimated.

In the 2000 model with all the variables, only one indicator (*acc_to_electricity*) was not significant, so a second model was created omitting it. The AIC value for this model was the lowest out of all the previously created Poisson and Negative Binomial models, indicating that the model fitted the data best and proving that there was overdispersion in the final Poisson model. Table 6 includes the coefficients of the final Negative Binomial model. The parameter estimates can be interpreted as follows:

- by increasing life expectancy by one year, we expected to see the mean number of child deaths per 1,000 live births reduced by 4.14%, holding other variables constant;
- a 1 percentage point increase in unemployment was expected to cause a 1.33% decrease of the mean of the response variable, holding other variables constant;
- the number of children living with HIV increased by 100,000 was expected to cause a 21.76% decrease in the mean number of children's deaths per 1,000 live births, holding other variables constant;
- a 10 percentage points increase in access to drinking water was expected to result in a 5% decrease in the mean number of child deaths per 1,000 live births, holding other variables constant;
- GDP *per capita* higher by USD 1,000 was expected to cause a 5.79% decrease of the mean of the dependent variable, holding other variables constant.

Table 6. Negative Binomial model coefficients (2000)

	Coefficients	***p*-value**
(Constant)	7.569	< 2e-16
life_expectancy	−0.0423	< 2e-16
access_to_drinking_water	−0.005015	0.007645
unemployment	−0.01340	0.000653
children_with_HIV	−0.000002176	3.1e-06
GDP_per_capita	−0.00005788	0.44952

Source: own elaboration with the use of the R programming language using the glm.nb() function from the MASS package.

For the 2019 data, two Negative Binomial models were created. The first one included all the indicators and the second one contained the same variables as the final Poisson model. In both cases, the intercept was highly significant and access to electricity was the only significant variable, although its significance was minor. The parameter estimate of *acc_to_electricity* was equal to –0.0073, which meant that a 10 percentage points increase in access to electricity was expected to result in a 7.27% decrease in the mean number of child deaths per 1,000 live births.

4.4. Spatial regression

To gain a better understanding of the phenomena of child mortality that has been examined, the geographical context was included in the regression's statistical framework through the introduction of spatial models. Prior to the models' estimation, the nonnegative matrices **W** were created. The value of 1 was assigned to pairs of neighbouring countries, the rest of the pairs were assigned the value of 0. First, Moran's I tests were performed using the final OLS models and the spatial weights matrix. In the case of the year 2000, the test with a test statistic of 0.16 and a p-value of 0.03 proved that there was a spatial autocorrelation, which did not occur in 2019, for which the test statistic was equal to –0.17 and the p-value was as high as 0.94.

Since the presence of autocorrelation was observed only in the data from 2000, spatial regression models were estimated exclusively for this year.

As the General Nesting Spatial Model is usually overparameterized, it was decided to choose one of the simpler models. Starting with only one spatial interaction, the choice was between the SAR, SLX, and SEM models. Because including lags of the independent variables made little sense in the case of the chosen indicators, the Spatial Lag of X (SLX) model was not considered. In the SAR model, all the indicators were significant, but the p-value for the Likelihood Ratio (LR) test was equal to 0.14, indicating that the simpler model would fit the data better. In the Spatial Error Model (SEM) all of the variables were significant as well. The LR test value was 4.06 with a p-value of 0.04, suggesting that a model with more spatial interactions would provide a significant improvement over the more restricted model.

The Spatial Autoregressive Combined (SAC) model was estimated including both the endogenous interaction effects and the interaction effects among the disturbance terms of different units. The LR tests, comparing the OLS model with the models including one spatial interaction effect, gave the following results:

- the test for OLS *versus* SEM had a p-value of 0.04, which meant that the less restricted model (SEM) should be chosen;
- the test for OLS *versus* SAR had a p-value of 0.14, which meant that the more restricted model (OLS) will provide a better fit;
- the LR test comparing SEM and SAC had a p-value of 0.36, which suggested that it was better to choose the more restricted model (SEM).

The Akaike Information Criterion had the highest score for the OLS model (412.99), the second-highest value was obtained by the SAR model (412.76).

The third place was taken by the SAC model with a value of 412.08. The SEM model had the lowest AIC score, which suggested that it would provide the best fit for the 2000 data. Also, based purely on the Akaike Information Criterion, all of the models with spatial interactions were preferred compared to the final OLS model. Taking into consideration both the Akaike Information Criterion and the Likelihood Ratio tests, the Spatial Error Model was the best choice for the 2000 data.

Theoretically, for the data from 2000, the SEM model could be used, but before proceeding to interpreting the parameters, it is necessary to think about the sense of applying spatial analysis models to this phenomenon. Although the analyses showed the occurrence of spatial autocorrelation, it does not make sense to consider the impact of child mortality in neighboring countries on child mortality in a given country. Just as, most likely, it will not be affected by the living conditions in the neighboring countries.

5. Conclusions

The goal of the study was to analyse the phenomenon of child mortality in African countries and compare the overall situation in Africa regarding the economic and living conditions and their impact on children's mortality rates in 2000 and 2019.

The descriptive statistics showed that there was a significant improvement regarding the economic and living conditions over the considered period of 19 years. Child mortality as well as the number of children living with HIV and unemployment decreased. Significant progress was noticed in the percentage of people having access to electricity and using at least basic drinking water services. GDP *per capita* increased by over 100%.

Surprising information was provided by the correlograms – in 2019, a strong negative relationship between the number of children living with HIV and child mortality was observed.

Cluster analysis helped to distinguish countries of similar socioeconomic characteristics and living conditions. In 2000, three groups were identified. Countries like Kenya, Uganda, and Zimbabwe happened to be in a cluster with the highest child mortality rate and the lowest access to drinking water. South Africa, Tanzania, Ethiopia, Malawi, Côte d'Ivoire, Mozambique, the Democratic Republic of Congo, Nigeria, and Zambia were characterised by the lowest electricity availability, the lowest life expectancy and despite having the largest number of children living with HIV, they also had the lowest children's mortality rate. The rest of the countries were described as those with the best living conditions among African nations. In 2019, Côte d'Ivoire, Ghana, Ethiopia, Angola, and Cameroon had quite good living conditions and a considerably low number of children infected with HIV. They also had the highest average GDP *per capita*, significantly outperforming other clusters in this matter. In both years, Uganda was in a cluster with the highest child mortality. Kenya, the Democratic Republic of Congo, and Malawi were again in the cluster with the worst standard of living. In 2019 Mozambique, Nigeria and Tanzania had the largest number of HIV-positive children. Socioeconomic conditions in these countries were poor, despite the very low average unemployment rate. Other countries were grouped together and characterised by quite good living conditions and the lowest unemployment rate.

Several regression models were computed, analysed, and compared to identify those that best explain the observed phenomenon in order to obtain a better knowledge of the dynamics of the association between child mortality and the chosen variables. For 2000 data, the OLS regression model without GDP *per capita* provided a good fit, but the best results were obtained by the Spatial Error Model that included a spatial interaction effect. In 2019, there was no spatial autocorrelation and none of the estimated models could be considered as an adequate tool to describe the phenomenon of children's mortality in Africa. The models' projections based on the results for 2000 data revealed unexpected outcomes. An increase in life expectancy, access to drinking water, and GDP *per capita* were supposed to result in a decrease in the dependent variable and, what was surprising, unemployment and the number of children living with HIV were supposed to have the same effect on child mortality, according to the model estimates.

Due to the complexity and multidimensionality of the phenomenon, there are still many factors that can be analysed in terms of their potential impact on child mortality in African countries, ranging from economic factors to indicators of overall well-being and quality of life, to variables with a higher level of detail, e.g., regarding maternal traits, habits, and experiences.

The analysis could be repeated using other variables, or other indicators could be added to the study. An interesting idea would be to use panel data that could help in capturing the changes and dynamics of the phenomenon. Although the study showed proof of significant improvement in the economic and living conditions in Africa, there is still a lot of space to explore the subject of factors affecting child mortality in developing countries.

This is an extremely important topic that deserves a lot of attention. The conducted study allowed to identify important aspects that must be resolved in order to introduce effective preventive measures aimed at minimizing child mortality.

References

Ahmad, O. B., Lopez, A., & Inoue, M. (2000). The decline in child mortality: a reappraisal. *Bulletin of the World Health Organization*, *78*(10), 1175–1191. Retrieved from https://apps.who.int/iris/handle/10665/267994

Balk, D., Pullum, T., Storeygard, A., Greenwell, F., & Neuman, M. (2004). A spatial analysis of childhood mortality in West Africa. *Population, Space and Place*, *10*(3), 175–216. https://doi.org/10.1002/psp.328

Basu, D. (2019). Bias of OLS estimators due to exclusion of relevant variables and inclusion of irrelevant variables. *Oxford Bulletin of Economics and Statistics*, *82*(1), 209–234. doi:10.1111/obes.12322

Bertrand, P. (1988). Akaike Information Criterion statistics. *Journal of the Royal Statistical Society Series A*, *151*(3), 567–568. Retrieved from https://ideas.repec.org/a/bla/jorssa/v151y1988i3-p567-568.html

Brockerhoff, M., & Hewett, P. (2000). Inequality of child mortality among ethnic groups in sub-Saharan Africa. *Bulletin of the World Health Organisation*, *78*(1), 30–41. doi: 10.1590/S0042-96862000000100004

Dismuke, C. & Lindrooth, R. (2006). Ordinary least squares. *Methods and Designs for Outcomes Research*, *93*, 93–104. Retrieved from https://www.researchgate.net/publication/292697957_Ordinary_least_squares

Elhorst, P. J. (2010). Applied spatial econometrics: raising the bar. *Spatial Economic Analysis*, *5*(1), 9–28. Retrieved from https://www.tandfonline.com/doi/full/10.1080/17421770903541772

Getis, A. (2010). Spatial autocorrelation. In: M. Fischer, & A. Getis (Eds.), *Handbook of applied spatial analysis* (pp. 255–278). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-03647-7

Harttgen, K., & Misselhorn, M. (2006). *A multilevel approach to explain child mortality and undernutrition in South Asia and Sub-Saharan Africa* (Ibero America Institute for Economic Research (IAI) Discussion Paper No. 152). Retrieved from https://econpapers.repec.org/scripts/redir.pf?u=http%3A%2F%2Fwww2.vwl.wiso.uni-goettingen.de%2Fibero%2Fworking_paper_neu%2FDB152.pdf;h=repec:got:iaidps:152

Hilbe, J. (2011). *Negative binomial regression*. Cambridge: Cambridge University Press.

Hill, K., & Pebley, A. R. (1989). Child mortality in the developing world. *Population and Development Review*, *15*(4), 657–687. Retrieved from https://u.demog.berkeley.edu/~jrw/Biblio/Eprints/%20G-I/hill.pebley.1989_child.mortality.developing.world.pdf

Joshi, K. D., & Nalwade, P. S. (2013). Modified k-means for better initial cluster centres. *International Journal of Computer Science and Mobile Computing*, *2*(7), 219–223. Retrieved from https://www.ijcsmc.com/docs/papers/July2013/V2I7201341.pdf

Kent, J. T. (1982). Robust properties of likelihood ratio tests. *Biometrika*, *69*(1), 19–27. https://doi.org/10.1093/biomet/69.1.19

Kiros, G., & Hogan, D. P. (2001). War, famine and excess child mortality in Africa: the role of parental education. *International Journal of Epidemiology*, *30*(3), 447–455. doi: 10.1093/ije/30.3.447

Korzeniewski, J. (2012). *Metody selekcji zmiennych w analizie skupień. Nowe procedury*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego. http://dx.doi.org/10.18778/7525-695-6

Newell, M. L., Coovadia, H., Cortina-Borja, M., Rollins, N., Gaillard, P., & Dabis, F. (2004). Mortality of infected and uninfected infants born to HIV-infected mothers in Africa: A pooled analysis. *The Lancet*, *364*(9441), 1236–1243. doi: 10.1016/S0140-6736(04)17140-7

Omariba, D. W. R., & Boyle, M. H. (2007). Family structure and child mortality in sub-Saharan Africa: Cross-national effects of polygyny. *Journal of Marriage and Family*, *69*(2), 528–543. https://doi.org/10.1111/j.1741-3737.2007.00381.x

Pan, J. X., & Fang, K. T. (2002). Maximum likelihood estimation. In: *Growth curve models and statistical diagnostics* (pp. 77–158). Springer Series in Statistics. New York: Springer. https://doi.org/10.1007/978-0-387-21812-0_3

Rawlings, S. B., & Siddique, Z. (2020). Domestic violence and child mortality in the developing world. *Oxford Bulletin of Economics and Statistics*, *82*(4), 723–750. https://doi.org/10.1111/obes.12357

Sneath, P. H., & Sokal, R. R. (1973). *Numerical taxonomy. The principles and practice of numerical classification*. San Francisco: W. H. Freeman.

The World Bank. (b.d.). *World development indicators*. Retrieved from https://databank.worldbank.org/source/world-development-indicators

United Nations [UN]. (2015). *Sustainable Development Goals*. Retrieved February 10, 2022 from https://www.un.org/sustainabledevelopment/sustainable-development-goals/

Wagner, Z., Heft-Neal, S., Bhutta, Z. A., Black, R. E., Burke, M., & Bendavid, E. (2018). Armed conflict and child mortality in Africa: A geospatial analysis. *The Lancet*, *392*(10150), 857–865. doi: 10.1016/S0140-6736(18)31437-5

Ward, M. D., & Gleditsch, K.S. (2008). *Spatial regression models.* Series: Quantitative Applications in the Social Science. Thousand Oaks, CA: Sage. https://dx.doi.org/10.4135/9781412985888

You, D., Jones, G., Hill, K., Wardlaw, T., & Chopra, M. (2010). Levels and trends in child mortality, 1990–2009. *The Lancet*, *376*(9745), 931–933. doi: 10.1016/S0140-6736(10)61429-8

Trendy i przyczyny śmiertelności dzieci w krajach afrykańskich

Streszczenie: W artykule omówiono problem śmiertelności dzieci w krajach afrykańskich. Celem pracy było zbadanie ogólnej sytuacji w Afryce pod względem warunków ekonomicznych i życiowych oraz ich wpływu na śmiertelność dzieci w latach 2000 i 2019 oraz wybór najwłaściwszych modeli do scharakteryzowania tego zjawiska. Analiza skupień została wykorzystana do zidentyfikowania krajów o podobnych cechach społeczno-ekonomicznych i warunkach życia. Oszacowano, przeanalizowano i porównano kilka modeli regresji. Dzięki zastosowaniu modeli przestrzennych kontekst geograficzny został włączony do ram statystycznych regresji. Analiza wykazała znaczną poprawę warunków ekonomicznych i życiowych w okresie 19 lat. Model błędu przestrzennego okazał się najlepszym modelem dla danych z 2000 r., jednak MNK i Ujemny Dwumianowy również wypadły dobrze. W 2019 r. nie było autokorelacji przestrzennej i żaden z oszacowanych modeli nie zapewniał dobrego dopasowania do danych. Prognozy modeli z 2000 r. ujawniły nieoczekiwane wyniki. Pomimo tego, że wzrost średniej długości życia, dostępu do wody pitnej i PKB na mieszkańca miał skutkować spadkiem zmiennej zależnej, szacunki modelu przewidywały, że bezrobocie i liczba dzieci zarażonych wirusem HIV będą miały ten sam wpływ na śmiertelność dzieci.

Słowa kluczowe: umieralność dzieci, kraje afrykańskie, analiza przestrzenna, analiza skupień, modele regresji.