

Joanna WACHNICKA, Katarzyna PALIKOWSKA  
Gdańsk University of Technology (Politechnika Gdańska)

## APPLICATION OF THE GWR MODEL FOR PREDICTING THE ROAD FATALITIES RATE ON THE ROAD NETWORK IN THE NUTS 3 REGIONS IN EUROPE ON THE EXAMPLE OF KUYAVIAN- -POMERANIAN VOIVODESHIP

### Zastosowanie modelu GWR do prognozowania wskaźnika ofiar śmiertelnych na sieci dróg w regionach NUTS 3 w Europie na przykładzie województwa Kujawsko-Pomorskiego

**Abstract:** *The article presents the application of the GWR (Geographically Weighted Regression) model to the description of differences in the level of road traffic safety in individual counties on the example of the Kuyavian-Pomeranian Voivodeship. The GWR model developed for counties, taking into account the diversity of NUTS 3 regions, can be a helpful tool for traffic safety management in voivodships and lower administrative units, and such an approach has not yet been applied.*

**Keywords:** road safety, GWR models, NUTS 3 regions, fatalities

**Streszczenie:** *W artykule przedstawiono zastosowanie modelu geograficznie ważonej regresji (GWR) do opisu różnic w poziomie bezpieczeństwa ruchu drogowego w poszczególnych powiatach na przykładzie województwa kujawsko-pomorskiego. Model GWR opracowany dla powiatów, a uwzględniający zróżnicowanie regionów NUTS 3, może stać się pomocnym narzędziem do zarządzania bezpieczeństwem ruchu drogowego w województwach oraz niższych jednostkach administracyjnych, a takiego podejścia jeszcze nie stosowano.*

**Słowa kluczowe:** bezpieczeństwo ruchu drogowego, modele GWR, regiony NUTS 3, ofiary śmiertelne

## 1. Introduction

Voivodship road safety programs consisted of a diagnosis guaranteeing an objective view of the actual road safety state, a strategy containing a clearly defined goal with the main directions of action and operational programs specifying the tasks to be performed by provincial levels and institutions [2]. On the other hand, local (county and urban) road safety programs are programs implemented by local governments. They were developed in very detail. They included specific solutions and actions in the field of education, supervision, infrastructure and rescue (eg improving a particular intersection, equipping a specific unit with equipment, indicating the exact place of automatic surveillance, etc.). The actions proposed in these programs reflected the local specificity of problems and dangers in the following areas: construction of regional and local road safety structures, road safety education, traffic supervision, road infrastructure and road rescue. The implementation of such a complicated program requires proven and supported by knowledge and results of scientific research tools [3]. The aim of this article is to present the possibility of using the GWR (Geographically Weighted Regression) model as a tool for describing the level of road traffic safety in individual counties of a selected voivodship and the possibility to apply it to more detailed modeling of road safety in particular NUTS 3 regions (counties). Such tools for traffic safety management at the national strategic level, already developed at the Department of Road Engineering, can be used to assess the effectiveness of the measures applied. However, there are still no tools to manage road safety at the strategic level in counties and cities. The argument for the application of a geographically weighted regression (GWR) is the possibility to take into account the geographical location of the counties studied (neighborhood) and the spatial diversity of factors affecting the level of road traffic safety in individual counties. The Kuyavian-Pomeranian Voivodeship was selected for these analyzes, as according to the police report on road accidents for 2017. The Kuyavian-Pomeranian Voivodeship was the second after the Podlaskie voivodeship with the highest death rate per 100 accidents at the level of 16.1. There was also a worried stabilization of the downward trend and stagnation of changes in 2013-2017. Due to the above, an attempt was made to create a GWR model for subregions in this voivodeship.

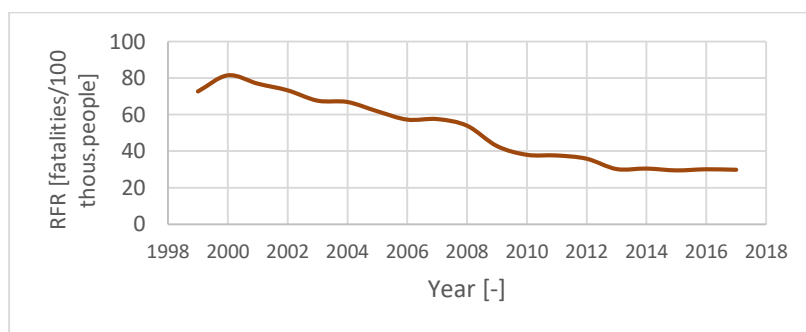


Fig. 1. Change in RFR rate in the province Kuyavian-Pomeranian Voivodeship in years 1999-2017

The Kuyavian-Pomeranian Voivodeship consists of 23 NUT 3 regions (counties), including 4 grodzki counties. In order to interpret the maps in the following drawings, a map with the names of individual provinces is presented below.



**Fig. 2.** Administrative division of the province Kuyavian-Pomeranian

## 2. Research problem

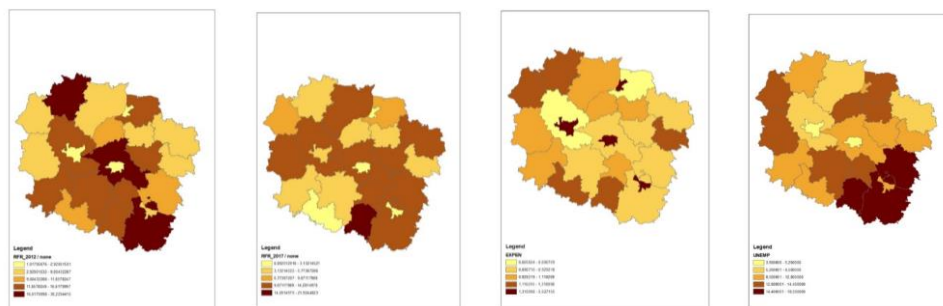
In this article we want to check if the geographical regression model is possible to use, when we want to analyze different factors impact on road safety in selected voivodship. For this purpose we collected data on the number of fatalities and data describing counties in economic, demographic, infrastructural terms and in particular counties of the Kuyavian-Pomeranian Voivodeship. The data came from years 2011-2017. The time range of the data resulted from the fact that data wanted for this article has started to be collected since 2011. Data collected for these analyses: relative death rate of RFR, DP density, relative wages in county PAY, expenditure on county and municipal roads EXPEN, demographic density of county and commune roads in a given county of ROADP, spatial density of county and commune roads ROADL, were collected. spending on municipal roads COASTROAD, length of CYCLE cycling routes, UNEMP unemployment rate, percentage of people living in URBAN cities. Table 1 presents the dependent and independent variables analyzed, together with units and a summary of basic statistics. Due to the fact that this article focuses on counties in the voivodship, we analyzed not only how the level of road safety on the roads of the entire Kuyavian-Pomeranian Voivodeship changed, but also how it looks in individual counties of this voivodship. Figure 3 shows the RFR rate in counties in 2011 and 2017.

Significant improvement of road safety in 2012-2017 took place, among others, in NUTS 3: tucholski, żniński and inowrocławski. A clear decrease in road safety occurred in NUTS 3: radziejewski, nakielski, świecki and brodnicki.

**Table 1**

**List of analyzed variables with basic statistics.**

Variable	Unit	N elements	Average	Minimum	Maximum
DP	people/km <sup>2</sup>	161	36.01	4.48	207.78
PAY	thous zł	161	3.22	2.62	4.38
RFR	fatalities/100 thous people	161	10.54	0.00	32.31
EXPEN	thous zł	161	1.64	0.47	5.55
ROADL	km/km <sup>2</sup>	161	1.06	0.49	3.08
ROADP	km/1000 persons	161	9.14	1.12	19.60
COSTROAD	thous zł/km	138	33.45	7.88	146.13
CYCLE	km	161	31.62	0.00	154.10
UNEMP	%	161	17.72	3.90	28.70
URBAN	%	161	46.25	12.81	100.00



**Fig. 3.** Illustration of the RFR rate in individual counties of the Kuyavian-Pomeranian Voivodeship in 2012 and 2017 and expenditures for county and commune roads EXPEN and UNEMP rates in 2017

In this article we wanted to find the answer whether the recorded changes in the level of security in individual counties can be related to changes in the socio-economic factors studied, which occurred in these counties in the analyzed period 2011-2017. To conduct a responsible policy improving the level of road safety, it is important to answer the question how changes in factors that can be changed (eg. UNEMP unemployment rate, EXPEN or COSTROAD expenditure levels) affect RFR. Spatial diversity of factors potentially affecting the level of road traffic safety is visible in fig. 3. Urban counties have the highest level of EXPEN and the lowest UNEMP rate. The highest unemployment rate are affected



by the south-eastern regions NUTS 3 of the Kuyavian-Pomeranian Voivodeship: radziejowski, włocławski and lipnowski.

### 3. Research method and results

The level of road traffic safety at different counties is diversified (fig. 2). It depends on various factors, including geographical location. In order to research the dependence of the road traffic safety on social, economic and infrastructural factors at county level and taking into account the spatial diversity of variables the Geographically Weighted Regression (GWR) model was applied [1]. The annual total number of fatalities per 1 000 inhabitants (RFR rate) was chosen as a measure of road traffic safety. The classical regression model assumes the spatial stationarity of the studied phenomenon i.e. the nature of the relationship between a set of independent variables (factors) and the dependent variable expressed in the coefficients of one global model, common for each analyzed region. The work [4] showed the need to build separate road traffic safety models for developed and developing countries due to the spatial diversity of socio-economic factors in European countries. In the studied voivodeship the spatial differentiation of factors limit the applicability of a model based on classical linear regression. Presented in tab. 2 values of the Moran *I* statistic lead to the rejection of the null hypothesis about the lack of spatial autocorrelation of individual variables. The geographical location of counties is taken into account by determining geographical coordinates of their centers of gravity.

Table 2

Values of Moran *I* statistic for dependent variable RFR rate and selected independent variables (2017, 23 counties)

Variable	Moran <i>I</i> statistic	p-value	Z-score
RFR	-0.1811	0.0104	-2.56
EXPEN	-0.1637	0.0251	-2.23
UNEMP	0.0785	0.0226	2.28

The obtained p-value values for the variables in tab. 2 are statistically significant ( $< 0.05$ ). A zero value of Moran *I* statistic and Z-score indicate a random spatial pattern. A negative Moran's *I* statistic and negative Z-score value indicate a spatial dispersion of counties (a negative spatial autocorrelation). Counties with the extreme values of the examined variables (RFR, EXPEN) are neighboring less frequently than randomly located regions (fig. 3). Positive values of Moran *I* statistic and Z-score of variable UNEMP indicate a spatial correlation. Counties with high or low unemployment rate are neighboring. In fig. 3 counties from the southern part of the voivodeship form a spatial cluster. GWR model is used to calibrate multiple regression models that allow to describe different relationships that exist at different locations. The basic form of the GW regression model is:

$$y_i = \beta_{i0} + \sum_{k=1}^m \beta_{ik} x_{ik} + \epsilon_i \quad (1)$$

where:

$y_i$  – dependent variable at location  $i$

$x_{ik}$  – independent variable  $k$  ( $k = 1. 2. \dots m$ ) at location  $i$

$m$  – number of independent variables

$\beta_{i0}$  – model estimated coefficient (intercept) at location  $i$

$\beta_{ik}$  – model estimated coefficients  $k$  ( $k = 1. 2. \dots m$ ) at location  $i$

$\xi_i$  – random error term at location  $i$

The GWR method is an extension of the linear regression model, which does not allow to take into account differences between regions.

The key element of the GWR model is the spatial function of weights, which determines spatial relationships between the observed variables:

$$w_{ij} = \begin{cases} e^{-0.5\left(\frac{d_{ij}}{b}\right)^2}, & \text{if } d_{ij} < b \\ 0, & \text{if } d_{ij} \geq b \end{cases} \quad (2)$$

where:

$d_{ij}$  – distance between gravity centers of locations  $i$  and  $j$

$b$  – bandwidth

For the GWR model a regression expression is determined for each county based on the observations in the surrounding counties. Estimation of the parameters of local regression models is made taking into account explanatory variables coming from neighboring counties (their importance decreases as the distance from a given region increases - observations which are nearer to a particular region have a greater weight in the estimation than observations that are further away). The R-CRAN packages `sp` and `spgwr` was used to compute the parameters of the GWR model. The geographical coordinates of the centers of gravity of the counties were transformed into Cartesian coordinates using the Robinson projection. The weights are computed using Gaussian kernel function. The optimal bandwidth was determined using the `gwr.sel` function with the cross validation procedure. At  $b \rightarrow \infty$  weight  $w_{ij} \rightarrow 1$ , which makes the GWR model coincide with the global model (linear, indicated as a reference model). The GWR model (tab. 3) was calibrated on data from 2011-2015. For testing purposes, this model was applied to the forecast of the RFR rate in 2017. Due to the high variability of the RFR rate in individual years at county level, the average values of RFR rate calculated taking into account the values from the previous 2 years, were used in the model. Independent variables PAY and EXPEN in



examined years were expressed as % of the average salary published by ZUS - in order to eliminate the impact of inflation.

The GWR model has favorable the Akaike Information Criterion (AIC) and adjusted  $R^2$  than global reference model (tab. 3). The model with the lowest AIC and the highest adjusted  $R^2$  is the optimal model. Additionally, Moran's  $I$  statistic was used to check if the model residuals are spatially random. For residuals of GWR model the obtained  $p$ -value = 0.239 was not statistically significant. The null hypothesis assuming spatial randomness of model residuals cannot be rejected, which is favorable.

**Table 3**

**GWR Model**

Variable	GWR model coefficients					Global model coefficients
	Min.	1st Qu.	Mediana	3rd Qu.	Max.	
Intersect	-35.10	-25.27	-4.97	1.22	14.52	-8.80
PAY	9.09	18.46	26.19	46.68	54.67	29.17
EXPEN	-2.89	0.01	2.00	2.55	8.07	1.67
ROADL	-3.07	-0.60	0.71	1.49	3.19	-0.34
COSTROAD	-0.15	-0.06	-0.03	-0.02	-0.01	-0.04
UNEMP	-0.22	0.02	0.11	0.22	0.31	0.15
URBAN	-0.23	-0.21	-0.17	-0.13	-0.10	-0.15
RFRPOP	-0.27	-0.19	-0.01	0.31	0.50	0.03
Bandwidth					38.02 km	$\infty$
AIC Criterion					73.93	112.61
Residual sum of squares					16.93	82.37
Adjusted $R^2$					0.96	0.71

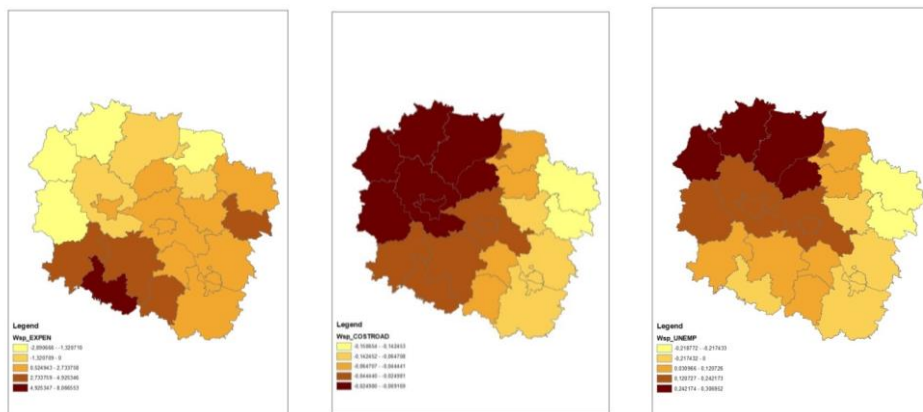
The GWR model coefficients are spatially autocorrelated (tab. 3), which proves the spatially variable influence of input variables on the road traffic safety at county level. The obtained  $p$ -value for the coefficients in tab. 4 are statistically significant and allow to reject the null hypothesis assuming the lack of spatial autocorrelation (spatial randomness). Positive values of the Moran  $I$  statistic and Z-score testify to the spatial proximity of the counties with the extreme values (high and low) of model coefficients. The counties with a strong positive or negative influence of a given variable on the modeled measure (RFR rate) form spatial clusters (are adjacent geographically). The analysis of the spatially differentiated impact of input variables on road traffic safety at county level. based on the variability of the GWR model coefficients. leads to important observations. Each set of the estimated coefficients at  $n$  locations can produce a map of variation (fig. 4) which may give useful information on non-stationarity of the regression relationship. Increasing EXPEN (expenses on county and municipal roads) may reduce RFR in Grudziądzki, Tucholski, Sępoleński and Nakielski Counties (negative values of EXPEN coefficient in fig. 4). However, increasing EXPEN in the Mogileński County and the neighboring counties:

Żniński, Inowrocławski and Radziejowski as well as in the Rypiński County in the east of the voivodship may not positively affect the improvement of road traffic safety.

**Table 4**

**Values of Moran *I* statistic of GWR model coefficients**

Coefficient	Moran <i>I</i> statistic	p-value	Z-score
PAY	0.3109	2.03e-10	6.35
EXPEN	0.1056	4.52e-03	2.84
ROADL	0.1435	5.96e-04	3.43
COSTROAD	0.2139	5.60e-07	5.00
UNEMP	0.2147	1.92e-06	4.76
URBAN	0.3349	1.15e-11	6.79



**Fig. 4.** Spatial diversification of GWR model coefficients in counties of the Kuyavian-Pomeranian Voivodeship

Increasing COSTROAD (expenses on municipal roads) improves safety in all counties. The spatial distribution of the GWR model coefficients (fig. 4) shows that the highest degree of improvement can be achieved in the counties of the eastern part of the Kuyavian-Pomeranian Voivodeship (especially in the Brodnicki and Rypiński Counties). Increasing UNEMP (the unemployment rate) in counties in the north of the voivodship (Sępoleński, Tucholski, Świecki) may affect the increase of the RFR rate (positive value of the coefficient in fig. 4), which means deterioration of road traffic safety. The spatial distribution of RFRPOP coefficients (average value of RFR rate in the previous 3-year period) allows forecasting of a spatially different trend of RFR changes in the following years. Negative values of coefficients in the counties of the north-western and central part of the voivodship indicate a tendency to improve safety in counties with an initially high RFR rate. Positive values of coefficients in the counties of the south-eastern part (fig. 4)





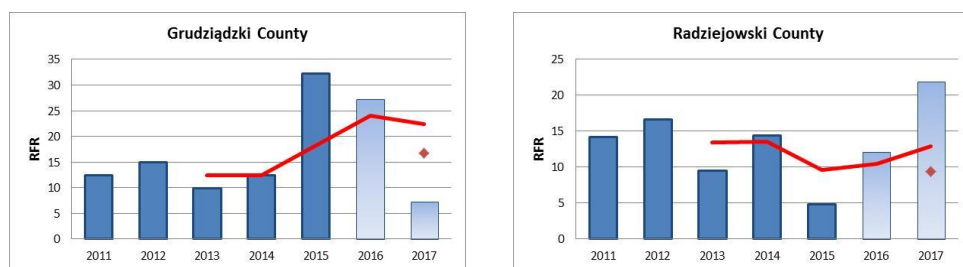
indicate a tendency of unfavorable changes of RFR. Conclusions resulting from the analysis of spatial distribution of GWR model coefficients have been verified for selected counties: Grudziądzki, Brodnicki, Tucholski and Radziejowski. The Grudziądzki and Tucholski Counties differ from the Radziejowski and Brodnicki Counties by the sign of EXPEN and RFRPOP coefficients. Radziejowski and Brodnicki Counties have positive values of the mentioned coefficients, while Grudziądzki and Tucholski Counties have negative values. Negative values of the model coefficients indicate the possibilities of improving road traffic safety in subsequent years.

**Table 5**

**RFR indicator and values of factors in selected counties (2011 and 2017)**

County	RFR		EXPEN		UNEMP		PAY	
	2011	2017	2011	2017	2011	2017	2011	2017
Grudziądzki	12.44	7.44	0.58	0.61	27.6%	14.2%	84%	88%
Brodnicki	7.73	12.70	0.91	0.85	12.6%	6.4%	80%	83%
Tucholski	12.47	4.13	1.32	1.17	19.6%	12.5%	79%	79%
Radziejowski	14.17	21.92	1.21	1.18	23.9%	18.4%	79%	79%

Factor PAY at county level (tab. 5) is expressed as percentage of average salary in the country in year 2011 and 2017. The GWR model was calibrated on data from years 2011-2015 and next was applied for prediction of RFR changes in 2017 (adjusted  $R^2=0.52$ ). The predicted values of RFR in the selected counties are presented in fig. 5. The trend of the RFR (red line in fig. 5) was determined using a moving average of 3 years. The predicted value of the average RFR (red rectangle in fig. 5) concerns 2017. The prediction was computed using the GWR model calibrated on data from years 2011-2015 (RFR values are marked with navy blue rectangles). The RFR from years 2016 and 2017 are marked with light blue rectangles - they are used as a background for assessing the prediction accuracy.



**Fig. 5.** RFR rate in the selected counties of the Kuyavian-Pomeranian Voivodeship (years 2011-2017)

The RFR prediction in 2017 was prepared from the perspective of 2015. In Grudziądzki County real RFR rate in 2015 was 32.31 (fig. 5). The predicted value of RFR in 2017 was 16.72, indicating a significant improvement in road traffic safety over the next



two years. In fact, there has been a greater than expected improvement in safety, which does not adversely affect the assessment of the prediction accuracy. The GWR model predicted correctly a significant reduction in the value of RFR. Improvement of road traffic safety in the Grudziądzki County could have been the result of increased expenses on roads (EXPEN, COSTROAD) and a drop of the unemployment rate in previous years. The prediction of RFR in 2017 for the Radziejowski County, prepared from the perspective of 2015, assumed a twofold increase in RFR. The direction of RFR changes was predicted correctly. The real increase was greater and came from 4.83 in 2015 to 21.92 in 2017.

## 4. Summary

The presented results indicate the legitimacy of using models that take into account the spatial non-stationarity of input factors, including the geographically weighted regression model GWR. Estimated values of global model coefficients may not adequately reflect local differences resulting from the neighborhood of the analyzed territorial units. Thanks to the spatial analysis, it was possible to indicate how individual NUTS 3 regions (counties) could react to changes in the factors included in the model. Analyzing changes in factors affecting RFR at the county level can be useful for taking counteracting measures (additional expenses etc.) in case of undesirable predictions. Though, collecting data on road traffic and exposure on the county level is still difficult. There is no public available data on traffic volumes on county and municipal roads and investments on voivodeship and national roads within a given county. There is no detailed information about the rescue system at the county level also. These and many other shortcomings cause that the creation of descriptive models and then predicting the level of road traffic safety on the road network in a selected county is very difficult and has a relatively significant margin of uncertainty. This paper can be motivating to start collecting more detailed data and sharing them in the form of public accessible data sets. The presented tool in the form of the GWR model has a large cognitive potential and can be helpful for more efficient road traffic safety management at the county level.

## 5. References

1. Fortheringham A.S.: Brunson C., Charlton M.: Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Hoboken, NJ: John Wiley & Sons Ltd, 2002.
2. Jamroz K. et al.: Wojewódzki Program BRD GAMBIT Pomorski. Gdańsk: Fundacja Rozwoju Inżynierii Lądowej, 2002.

3. Jamroz K., Kustra W., Budzynski M., Zukowska J.: Pedestrian protection, speed enforcement and road network structure the key action for implementing Poland's Vision Zero, *Transp. Res. Procedia*, Vol. 14, 2016.
4. Wachnicka J.: Modelowanie miar ryzyka na sieci dróg w regionach. PhD thesis. Politechnika Gdańska, 2017.

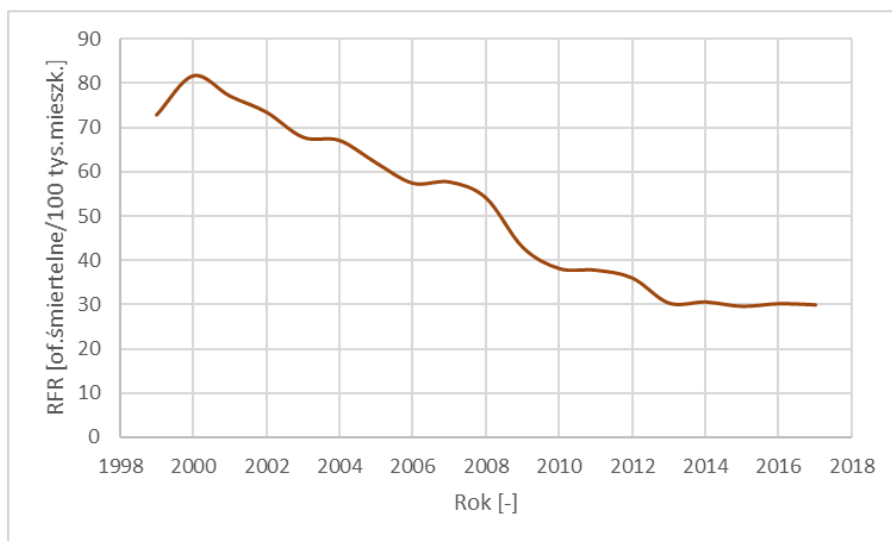
# ZASTOSOWANIE MODELU GWR DO PROGNOZOWANIA WSKAŹNIKA OFIAR ŚMIERTELNYCH NA SIECI DRÓG W REGIONACH NUTS 3 W EUROPIE NA PRZYKŁADZIE WOJEWÓDZTWA KUJAWSKO-POMORSKIEGO

## 1. Wprowadzenie

Wojewódzkie i lokalne programy bezpieczeństwa ruchu drogowego (brd), oprócz programów operacyjnych i sektorowych, stanowiły i stanowią podstawowe narzędzia realizacji krajowych strategii bezpieczeństwa ruchu drogowego [2]. Wagę tych działań podkreślono w programie GAMBIT 2005, zwracając uwagę na działania obszarowe (prowadzone w województwach, powiatach i gminach), które mogą przyczynić się do prawie połowy zamierzonej redukcji ofiar śmiertelnych. Wzorując się na programach krajowych, wypracowano strukturę programu wojewódzkiego i lokalnego. Wojewódzkie programy brd składały się zatem z diagnozy gwarantującej obiektywne spojrzenie na faktyczny stan brd, strategii zawierającej jasno sprecyzowany cel z podaniem głównych kierunków działań oraz programów operacyjnych precyzujących zadania do wykonania przez instytucje i organizacje szczeble wojewódzkiego. Programy lokalne (powiatowe, miejskie) natomiast, realizowane przez lokalne samorządy, były opracowywane z dużym stopniem szczegółowości. Wskazywano w nich na konkretne rozwiązania i działania z zakresu edukacji, nadzoru, infrastruktury i ratownictwa (np. usprawnienie konkretnego skrzyżowania, doposażenie konkretnej jednostki w sprzęt, wskazanie dokładnego miejsca automatycznego nadzoru itp.). Działania proponowane w tych programach odzwierciedlały lokalną specyfikę problemów i zagrożeń w następujących obszarach: budowa regionalnych i lokalnych struktur brd, edukacja brd, nadzór nad ruchem drogowym, infrastruktura drogowa i ratownictwo na drogach. Wdrażanie tak skomplikowanego programu wymaga sprawdzonych oraz popartych wiedzą i wynikami badań naukowych narzędzi [3]. Celem niniejszego artykułu jest przedstawienie możliwości wykorzystania obszarowego modelu geograficznie ważonej regresji GWR jako narzędzia do opisu poziomu bezpieczeństwa ruchu drogowego w poszczególnych powiatach wybranego województwa, a także możliwości zastosowania go do dokładniejszego modelowania poziomu brd w poszczególnych regionach NUTS 3 (powiatach). Takie narzędzia do zarządzania bezpieczeństwem ruchu drogowego na poziomie strategicznym kraju, opracowane już w Katedrze Inżynierii Drogowej, mogą być zastosowane do oceny efektywności zastosowanych działań.



Natomiast nadal brakuje narzędzi do zarządzania bezpieczeństwem ruchu drogowego na poziomie strategicznym w województwach, powiatach i miastach. Argumentem przemawiającym za zastosowaniem geograficznie ważonej regresji jest możliwość uwzględnienia położenia geograficznego badanych powiatów (sąsiedztwa) oraz przestrzennego zróżnicowania czynników wpływających na poziom bezpieczeństwa ruchu drogowego w poszczególnych powiatach. Do niniejszych analiz wybrano województwo kujawsko-pomorskie, gdyż według policyjnego raportu o wypadkach drogowych za rok 2017 było ono drugim po woj. podlaskim województwem o najwyższym wskaźniku zabitych na 100 wypadków, na poziomie 16,1. Zaobserwowano w nim również niepokojące ustabilizowanie się trendu spadkowego i stagnację zmian w latach 2013–2017. Z uwagi na powyższe podjęto próbę stworzenia modelu GWR dla podregionów w tym województwie.



**Rys. 1.** Zmiana wskaźnika RFR w woj. kujawsko-pomorskim na przestrzeni lat 1999–2017

Województwo Kujawsko-Pomorskie składa się z 23 regionów NUTS 3 (powiatów), w tym cztery to powiaty grodzkie. W celu interpretacji mapek na kolejnych rysunkach, poniżej przedstawiono mapkę z nazwami poszczególnych województw.





Rys. 2. Podział administracyjny woj. kujawsko-pomorskiego

## 2. Problem badawczy

W celu sprawdzenia możliwości wykorzystania modelu regresji geograficznej do badania czynników mogących wpływać na poziom bezpieczeństwa ruchu drogowego w poszczególnych powiatach wybranego województwa, zebrano dane o liczbie ofiar śmiertelnych oraz dane charakteryzujące powiaty pod względem ekonomicznym, demograficznym, infrastrukturalnym i w poszczególnych powiatach województwa kujawsko-pomorskiego. Dane pochodziły z lat 2011–2017. Zakres czasowy danych wynikał z tego, że dopiero od roku 2011 zaczęto zbierać dane dot. części charakterystyk, które chciano przeanalizować w niniejszym artykule. Zebrano takie dane jak: względny wskaźnik liczby ofiar śmiertelnych RFR, gęstość zaludnienia DP, względną płacę w powiecie PAY, wydatki na drogi powiatowe i gminne EXPEN, gęstość demograficzną dróg powiatowych i gminnych w danym powiecie ROADP, gęstość przestrzenną dróg powiatowych i gminnych ROADL, wskaźnik wydatków na drogi gminne COASTROAD, długość ścieżek rowerowych CYCLE, stopa bezrobocia UNEMP, procent ludzi mieszkających w miastach URBAN. W tab. 1 zestawiono zmienne zależne i niezależne do analiz wraz z jednostkami i zestawieniem podstawowych statystyk.

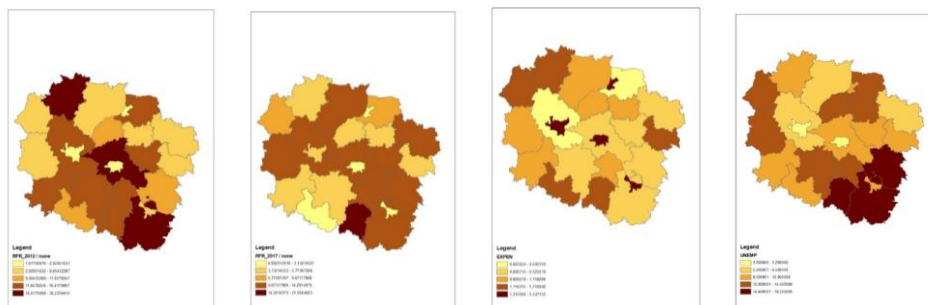
Tabela 1

Zestawienie analizowanych zmiennych wraz z podstawowymi statystykami

Zmienna	Jednostka	N ważnych	Średnia	Minimum	Maksimum
DP	people/km <sup>2</sup>	161	36,01	4,48	207,78
PAY	thous zł	161	3,22	2,62	4,38
RFR	fatalities/100 thous people	161	10,54	0,00	32,31
EXPEN	thous zł	161	1,64	0,47	5,55
ROADL	km/km <sup>2</sup>	161	1,06	0,49	3,08
ROADP	km/1000 persons	161	9,14	1,12	19,60
COSTROAD	thous zł/km	138	33,45	7,88	146,13
CYCLE	km	161	31,62	0,00	154,10
UNEMP	%	161	17,72	3,90	28,70
URBAN	%	161	46,25	12,81	100,00

Z uwagi na to, że w niniejszym artykule skupiono się na powiatach w województwie, przeanalizowano nie tylko, jak się zmienił poziom bezpieczeństwa ruchu drogowego na drogach całego województwa kujawsko-pomorskiego, ale również jak to wygląda w poszczególnych powiatach tego województwa. Na rys. 3 zestawiono ze sobą wskaźniki RFR w powiatach w latach 2011 i 2017. Wynika z niego, że najniższe wartości RFR odnotowywano w powiatach grodzkich (m. Bydgoszcz, m. Toruń, m. Grudziądz i m. Włocławek) i w badanym przedziale lat 2012–2017 poziom bezpieczeństwa ruchu drogowego w tych powiatach utrzymywał się na wysokim poziomie. Znacząca poprawa bezpieczeństwa ruchu drogowego w latach 2012–2017 nastąpiła m.in. w powiatach tucholskim, żnińskim i inowrocławskim. Wyraźny spadek bezpieczeństwa nastąpił w powiatach radziejowskim, nakielskim, świeckim i brodnickim. Celem artykułu jest znalezienie odpowiedzi na pytanie, czy odnotowane zmiany poziomu bezpieczeństwa w poszczególnych powiatach można powiązać ze zmianami badanych czynników społeczno-ekonomicznych, które następowały w tychże powiatach w badanym przedziale lat 2011–2017. Z punktu widzenia prowadzenia odpowiedzialnej polityki zmierzającej do poprawy poziomu bezpieczeństwa ruchu drogowego istotna jest odpowiedź na pytanie, w jaki sposób zmiany czynników, które można kształtować (np. stopa bezrobocia UNEMP, poziomy wydatków EXPEN czy COSTROAD), wpływają na zmiany RFR.





Rys. 3. Ilustracja wskaźnika RFR w poszczególnych powiatach województwa kujawsko-pomorskiego w roku 2012 i 2017 oraz wydatków na drogi powiatowe i gminne EXPEN oraz stopy bezrobocia UNEMP w roku 2017

### 3. Metoda badawcza oraz wyniki

Podjęto próbę budowy modelu opisującego zależność poziomu bezpieczeństwa ruchu drogowego na poziomie powiatu od czynników społecznych, ekonomicznych oraz infrastrukturalnych charakteryzujących powiat oraz uwzględniającego zróżnicowanie przestrzenne zmiennych. Wybrano model regresji ważonej geograficznie GWR (*Geographically Weighted Regression*). Jako miarę poziomu bezpieczeństwa drogowego przyjęto roczną liczbę ofiar śmiertelnych na 1 tys. mieszkańców (RFR). Klasyczna metoda regresji zakłada stacjonarność przestrzenną badanego zjawiska, tzn. wspólny dla każdego analizowanego regionu charakter zależności pomiędzy zmiennymi objaśnianymi a zmienną objaśnianą wyrażony współczynnikami jednego, globalnego modelu. W pracy [4] wykazano konieczność budowy oddzielnych modeli dla krajów rozwiniętych i rozwijających się z uwagi na zróżnicowanie przestrzenne czynników społeczno-gospodarczych w krajach europejskich. Potwierdzeniem istnienia zróżnicowania przestrzennego czynników, a tym samym wskazaniem istnienia ograniczenia stosowalności modelu opartego na klasycznej regresji liniowej, są przedstawione w tab. 1 wartości statystyki  $I$  Morana prowadzące do odrzucenia hipotezy zerowej o braku autokorelacji przestrzennej poszczególnych zmiennych. Położenie geograficzne powiatów uwzględniono poprzez wyznaczenie współrzędnych geograficznych ich środków ciężkości.

Tabela 2

Statystyka  $I$  Morana wybranych zmiennych wejściowych i przyjętej miary brd (dane z 2017 r.)

Zmienna	Statystyka $I$ Morana	p-value	Z-score
RFR	-0,1811	0,0104	-2,56
EXPEN	-0,1637	0,0251	-2,23
UNEMP	0,0785	0,0226	2,28



Uzyskane wartości p-value dla zmiennych w tab. 2 są statystycznie istotne. Ujemne wartości Z-score świadczą o rozproszeniu przestrzennym powiatów o skrajnych wartościach danej zmiennej, tzn. powiaty o wysokich lub niskich wartościach RFR (analogicznie EXPEN) sąsiadują ze sobą rzadziej niż wynikałoby to z losowego rozmieszczenia. Dodatnia wartość Z-score zmiennej UNEMP świadczy o sąsiedztwie przestrzennym powiatów z wysoką lub niską stopą bezrobocia (powiaty podobne tworzą skupiska).

Regresja ważona geograficznie pozwala na uchwycenie w budowanym modelu zmienności współczynników regresji w odniesieniu do każdego regionu. Podstawowa forma modelu GWR:

$$y_i = \beta_{i0} + \sum_{k=1}^m \beta_{ik} x_{ik} + \epsilon_i \quad (1)$$

gdzie:

$y_i$  – zmienna objaśniana dla regionu  $i$

$x_{ik}$  – zmienna objaśniająca  $k$  dla regionu  $i$

$m$  – liczba zmiennych objaśniających w modelu

$\beta_{i0}$  – wyraz wolny dla regionu  $i$

$\beta_{ik}$  – współczynnik regresji odpowiadający zmiennej  $k$  dla regionu  $i$

$\epsilon_i$  – błąd losowy dla regionu  $i$

Metoda GWR stanowi rozszerzenie modelu regresji liniowej, który nie zawsze odpowiednio uwzględnia różnice pomiędzy regionami. Podstawowym elementem metody GWR jest przestrzenna funkcja wag, która określa zależności przestrzenne pomiędzy obserwowanymi zmiennymi. Estymacja parametrów lokalnych modeli regresji dokonywana jest z uwzględnieniem zmiennych objaśniających pochodzących z regionów sąsiednich (ich znaczenie maleje wraz ze wzrostem odległości od danego regionu). Wykorzystano pakiety `sp` i `spgwr`. Elementy macierzy wag zostały wyznaczone za pomocą Gaussian kernel function:

$$w_{ij} = \begin{cases} e^{-0,5\left(\frac{d_{ij}}{b}\right)^2} & \text{dla } d_{ij} < b \\ 0 & \text{dla } d_{ij} \geq b \end{cases} \quad (2)$$

gdzie:

$d_{ij}$  – odległość pomiędzy regionem  $i$  a  $j$

$b$  – szerokość okna

Współrzędne geograficzne środków ciężkości regionów zostały przetransformowane na współrzędne kartezjańskie za pomocą projekcji Robinson (*Robinson Projection*). Optymalną szerokość okna wyznaczono za pomocą funkcji `gwr.sel` z kryterium



walidacji krzyżowej (*cross validation*). Przy  $b \rightarrow \infty$  wagi  $w_{ij} \rightarrow 1$ , co sprawia, że model GWR pokrywa się z modelem globalnym (liniowym, wskazanym jako referencyjny).

Model GWR (tab. 3) zbudowano w oparciu o dane z lat 2011–2015. W celach testowych model ten został zastosowany do prognozy wskaźnika RFR w 2017 r. Z uwagi na wysoką zmienność wskaźnika RFR w poszczególnych latach w modelu zostały użyte wartości średnie RFR obliczone z uwzględnieniem wartości z dwóch lat poprzedzających. Zmienne PAY i EXPEN w poszczególnych latach były wyrażone jako odsetek przeciętnego wynagrodzenia publikowanego przez ZUS w celu wyeliminowania wpływu inflacji.

Tabela 3

## Statystyki modelu GWR

Zmienna	Współczynniki modelu GWR					Współczynniki modelu globalnego
	Min.	1 kwartył	Mediana	3 kwartył	Max	
Wyraz wolny	-35,10	-25,27	-4,97	1,22	14,52	-8,80
PAY	9,09	18,46	26,19	46,68	54,67	29,17
EXPEN	-2,89	0,01	2,00	2,55	8,07	1,67
ROADL	-3,07	-0,60	0,71	1,49	3,19	-0,34
COSTROAD	-0,15	-0,06	-0,03	-0,02	-0,01	-0,04
UNEMP	-0,22	0,02	0,11	0,22	0,31	0,15
URBAN	-0,23	-0,21	-0,17	-0,13	-0,10	-0,15
RFRPOP	-0,27	-0,19	-0,01	0,31	0,50	0,03
Szerokość okna				38,02 km		$\infty$
Kryterium AIC				73,93		112,61
Suma kwadratów reszt modelu				16,93		82,37
Wskaźnik dopasowania $R^2$				0,96		0,71

Model GWR uzyskał korzystniejszą wartość kryterium AIC i wskaźnik dopasowania  $R^2$  niż referencyjny model globalny (tab. 3). Model GWR użyty do przygotowania prognozy RFR w 2017 uzyskał wskaźnik dopasowania  $R^2 = 0,52$ .

Współczynniki modelu GWR wykazują autokorelację przestrzenną (tab. 4), co świadczy o zróżnicowanym przestrzennie wpływie zmiennych wejściowych na poziom brd w poszczególnych powiatach. Uzyskane wartości p-value dla współczynników w tab. 4 są statystycznie istotne i pozwalają na odrzucenie hipotezy zerowej o braku autokorelacji przestrzennej. Dodatkowo wartości Z-score świadczą o sąsiedztwie przestrzennym skrajnych wartości (wysokich i niskich) współczynników odpowiadających poszczególnym powiatom. Powiaty o silnym dodatnim lub ujemnym wpływie danej zmiennej na modelowaną miarę brd sąsiadują ze sobą (tworzą skupiska).

Tabela 4

## Statystyka I Morana współczynników modelu GWR

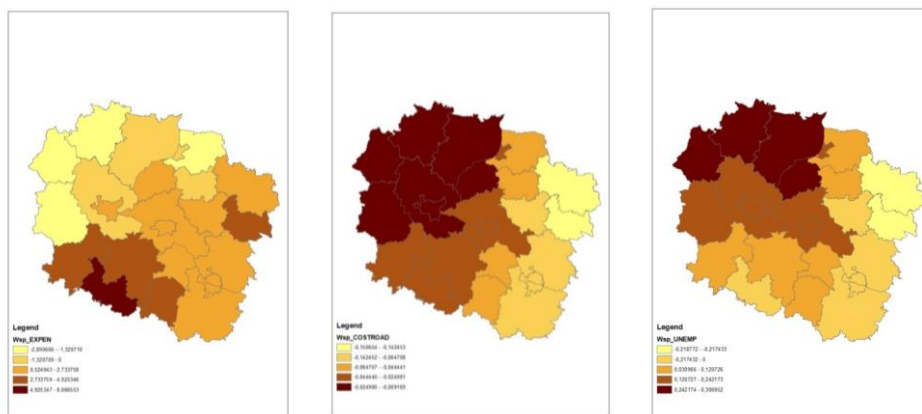
Współczynnik	Statystyka I Morana	p-value	Z-score
PAY	0,3109	2,03e-10	6,35
EXPEN	0,1056	4,52e-03	2,84
ROADL	0,1435	5,96e-04	3,43
COSTROAD	0,2139	5,60e-07	5,00
UNEMP	0,2147	1,92e-06	4,76
URBAN	0,3349	1,15e-11	6,79

Analiza zróżnicowanego przestrzennie wpływu zmiennych wejściowych na poziom brd w poszczególnych powiatach, przeprowadzona na podstawie zmienności współczynników modelu GWR, prowadzi do istotnych spostrzeżeń.

Zwiększenie wydatków na drogi powiatowe i gminne EXPEN może zmniejszyć RFR w powiatach grudziądzkim, tucholskim, sępoleńskim i nakielskim (ujemna wartość współczynnika EXPEN na rys. 4). Natomiast zwiększanie EXPEN w powiecie mogileńskim i w sąsiadujących z nim powiatach żnińskim, inowrocławskim i radziejowskim oraz rypińskim na wschodzie województwa może nie wpłynąć pozytywnie na poprawę bezpieczeństwa ruchu drogowego.

Zwiększanie wydatków na drogi gminne COSTROAD wpływa na poprawę bezpieczeństwa we wszystkich powiatach. Rozkład przestrzenny współczynników modelu GWR (rys. 4) wskazuje, że największy stopień poprawy można osiągnąć w powiatach wschodniej części województwa kujawsko-pomorskiego (szczególnie w powiatach brodnickim i rypińskim). Zwiększenie stopy bezrobocia UNEMP w powiatach na północy województwa (sępoleński, tucholskim, świecki) może wpłynąć na zwiększenie miary RFR (dodatnia wartość współczynnika na rys. 4). Rozkład przestrzenny współczynników RFRPOP (wartość średnia RFR w poprzednim 3-letnim okresie) na rys. 4 pozwala na prognozowanie zróżnicowanego przestrzennie trendu zmian RFR w kolejnych latach. Ujemne wartości współczynników w powiatach północno-zachodniej i środkowej części województwa wskazują na tendencję do poprawy bezpieczeństwa w powiatach o wyjściowo wysokim RFR. Dodatnie wartości współczynników w powiatach części południowo-wschodniej wskazują na tendencję utrzymywania się niekorzystnych wartości RFR. Wnioski wynikające z analizy rozkładu przestrzennego współczynników modelu GWR zostały zweryfikowane dla wybranych powiatów: grudziądzkiego, tucholskiego i radziejowskiego. Powiaty grudziądzki i tucholski różnią się od powiatu radziejowskiego znakiem współczynników EXPEN i RFRPOP. Powiat radziejowski posiada dodatnie wartości wymienionych współczynników, powiaty grudziądzki i tucholski ujemne, świadczące o możliwościach poprawy bezpieczeństwa w latach kolejnych.





Rys. 4. Zróżnicowanie przestrzenne współczynników modelu GWR w powiatach województwa kujawsko-pomorskiego

Tabela 5

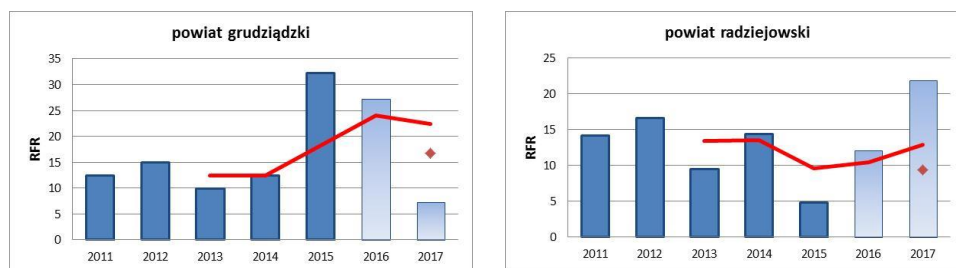
Wartości zmiennych dla wybranych powiatów w latach 2011 i 2017

	RFR		EXPEN		UNEMP		PAY	
	2011	2017	2011	2017	2011	2017	2011	2017
grudziądzki	12,44	7,44	0,58	0,61	27,6%	14,2%	84%	88%
brodnicki	7,73	12,70	0,91	0,85	12,6%	6,4%	80%	83%
tucholski	12,47	4,13	1,32	1,17	19,6%	12,5%	79%	79%
radziejowski	14,17	21,92	1,21	1,18	23,9%	18,4%	79%	79%

Zmienna PAY wyraża odsetek, jaki stanowi średnia płaca w powiecie w stosunku do przeciętnego wynagrodzenia w kraju w danym roku.

Na rys. 5 wyznaczono linię trendu RFR za pomocą średniej ruchomej o okresie trzech lat oraz oznaczono prognozowaną wartość średnią RFR w 2017 (czerwony prostokąt). Prognoza została przygotowana w oparciu o model GWR zbudowany w oparciu o dane z lat 2011–2015 oznaczone granatowymi prostokątami. Lata 2016 i 2017 zostały oznaczone jasnoniebieskimi prostokątami – stanowią tło do oceny prognozowanej wartości RFR w 2017 r.

Prognoza RFR w 2017 r. została przygotowana z perspektywy roku 2015. W powiecie grudziądzkim rzeczywiste RFR w 2015 r. wyniosło 32,31 (rys. 5). Prognozowana wartość RFR w 2017 r. wyniosła 16,72, wskazując na znaczną poprawę bezpieczeństwa w okresie dwóch kolejnych lat. W rzeczywistości nastąpiła większa niż prognozowana poprawa bezpieczeństwa, co nie wpływa negatywnie na ocenę trafności prognozy. Model GWR posłużył do przygotowania prognozy znacznego zmniejszenia wartości RFR. Poprawa bezpieczeństwa w powiecie grudziądzkim mogła być wynikiem zwiększenia wydatków EXPEN, COSTROAD i spadku stopy bezrobocia w latach poprzednich.



**Rys. 5.** Wskaźnik RFR liczby ofiar śmiertelnych w wybranych powiatach województwa kujawsko-pomorskiego w latach 2011–2017

Prognoza dla powiatu radziejowskiego RFR w 2017 r. przygotowana z perspektywy roku 2015 zakładała dwukrotny wzrost RFR. Rzeczywisty wzrost nastąpił z 4,83 w 2015 r. do 21,92 w roku 2017.

## 4. Podsumowanie

Przedstawione wyniki badań wskazują na zasadność stosowania modeli uwzględniających przestrzenną niestacjonarność czynników wejściowych, w tym modelu geograficznej ważonej regresji GWR. Oszacowane wartości współczynników modelu globalnego mogą nie uwzględniać w odpowiedni sposób różnic lokalnych, wynikających z sąsiedztwa analizowanych jednostek terytorialnych. Dzięki przeprowadzonej analizie przestrzennej można było wytypować, jak poszczególne regiony NUTS 3 mogą zareagować na zmiany badanych czynników uwzględnionych w modelu. Analizując zmiany czynników wpływających na RFR w danym województwie, można przewidzieć, jak może się on zmieniać i w przypadku niepożądanego prognoz próbować przeciwdziałać dodatkowymi środkami. Jednocześnie, podczas zbierania danych o powiatach okazało się, że wciąż nie można zestawiać pewnych informacji o systemie transportowym lepiej charakteryzujących m.in. ruch drogowy, a tym samym narażenie. Nie ma ogólnodostępnej informacji o wielkości ruchu na drogach powiatowych i gminnych. Nie ma również danych o inwestycjach na drogach wojewódzkich i krajowych na terenie danego powiatu. Nie ma szczegółowych informacji o systemie ratownictwa na poziomie powiatowym. Te i wiele innych mankamentów powodują, że tworzenie modeli opisowych, a następnie prognozowanie poziomu bezpieczeństwa ruchu drogowego na sieci wszystkich dróg w danym powiecie jest bardzo utrudnione i ma znaczący margines niepewności. Ten artykuł powinien być bodźcem do rozpoczęcia zbierania bardziej szczegółowych danych oraz zestawiania ich w dostępne bazy danych. Zaprezentowane narzędzie w postaci modelu GWR ma w sobie duży potencjał poznawczy i może być pomocne do skuteczniejszego zarządzania brd w poszczególnych powiatach.



## 5. Literatura

1. Fortheringham A.S., Brunson C., Charlton M.: Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Hoboken, NJ: John Wiley & Sons Ltd, 2002.
2. Jamroz K. et al.: Wojewódzki Program BRD GAMBIT Pomorski. Gdańsk: Fundacja Rozwoju Inżynierii Lądowej, 2002.
3. Jamroz K., Kustra W., Budzynski M., Zukowska J.: Pedestrian protection, speed enforcement and road network structure the key action for implementing Poland's Vision Zero, *Transp. Res. Procedia*, Vol. 14, 2016.
4. Wachnicka J.: Modelowanie miar ryzyka na sieci dróg w regionach. Rozprawa doktorska. Politechnika Gdańska, 2017.