Comparative evaluation of accuracy of the multiple methods of heatstroke risk model including time-stratified case-crossover analysis

Mio Nemoto (5492804426@edu.k.u-tokyo.ac.jp)
The University of Tokyo

Satoshi Hirabayashi
The Davey Tree Expert Company, SUNY-ESF

Seisho Sato
The University of Tokyo

Akihiro Fujii
Kogakuin University

Hidetoshi Nojiri
Juntendo University School of Medicine

Tomohiko Ihara
The University of Tokyo

Research Article

Keywords: Urban heat island, Risk model, Heatstroke, Time-stratified case-crossover analysis, Generalized additive model, Tokyo, Climate change, Statistical analysis

Posted Date: June 28th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1783241/v1

Abstract

Background

High temperatures in urban areas owing to climate change and urban heat islands have led to an increase in the number of heatstroke patients. To prevent heatstroke, accurate heatstroke patient prediction model should be used to predict and alert people to their risk. However, most previous models have not tested sufficient training data, although susceptibility to heatstroke is likely to be dependent on year-wise trends and is susceptible to training data. We investigated an accurate heatstroke risk model that is robust to the training data. By examining the factors affecting the accuracy and trade-off between the quantity and quality of adding old data to training data, the method of selecting training data for constructing an accurate model was also investigated.

Method

We compared the accuracies of three methods: multiple regression analysis (MR), generalized additive model (GAM), and time-stratified case-crossover analysis (TC). The training data for all combinations from 2012 were tested. By comparing the errors of each method, we identified the error influencing factors in the training data.

Results

The TC errors were the smallest ($p < 0.005$) and much less sensitive to the training data than others. Best accuracy odds ratios were 1.41–1.44. The error was significantly larger when the number of extremely hot days differed between the training and test data in MR and GAM ($p < 0.01$, $p < 0.05$). All three methods tended to increase accuracy up to a certain point and decrease from the middle of the year when adding past years retroactively from the most recent year in training data.

Conclusions

By using the odds ratios produced by TC with low sensitivity to training data, we can develop highly accurate heatstroke risk model that are robust to training data, which has not been possible with previous models. If data are to be included continuously, a model constructed from three or four years of data from the latest one is the most accurate.

1. Background

The impacts of climate change have put human health in jeopardy[1]. High temperatures, especially in urban centers, have worsened in recent years owing to the heat island effect[2]. In cities around the world, there are concerns about the health effects of heat waves[3, 4]. In 2003, a heat wave in France led to 14,800 excess deaths in comparison with the mean mortality during the summers of 2000 to 2002. Most of these deaths occurred in urban areas[5]. In Japan, the number of emergency transports due to heatstroke (hereinafter referred to as heatstroke patients) has been increasing annually and has become a social problem, especially in urban areas[6]. According to the Fire and Disaster Management Agency of the Ministry of Internal Affairs and Communications in 2018, from May to September, there were 95,137 heatstroke patients nationwide, half of which were among seniors aged 65 and older, and of which 7,843 were in Tokyo. On average, there were 63 patients per day with the highest of over 400. To prevent heat stroke, the Ministry of the Environment issued heat stroke alerts based on wet bulb globe temperature (WBGT) and began operating them in cooperation with the Japan Meteorological Agency in 2020[7, 8]. When the WBGT is forecasted over 33 degrees a warning is issued across the nation. However, using WBGT as an alert indicator is inappropriate in two ways. First, WBGT overestimates the magnitude of

humidity more than the actual human body heat balance[9]. Second, different areas have different optimum temperatures[10]. Therefore, it is questionable to use the same WBGT criteria for all the areas. One idea is to predict the number of heatstroke patients in each area to provide warnings that are appropriate to the reality of the given area[11].

Various models have been developed to predict the number of heatstroke patients. Typical models developed in recent years have used outdoor temperature as the model's predictor variable and exponential function methods or non-linear prediction models, such as the generalized linear model (GLM), generalized additive model (GAM), and distributed lag non-linear model (DLNM) as its prediction method. Ravindra et al. explains various applications of GAM to link climatic variability with adverse health outcomes[12]. In a model focused on the assessment of heat-related mortality, it is useful to employ DLNM to account for lag[13]. In previous studies specific to heatstroke, Yamamoto et al. used the highest WBGT to predict heatstroke patients exponentially with daytime population[14]. Sato et al. used GLM with average temperature, average humidity, average wind speed, and solar radiation[15]. Kodera et al. developed a model that takes into account physiological responses[16]. Their model considers the average temperature of the previous three days in predicting indoor patients. The exponentially predicted number of heatstroke patients using Kodera's formula was first reported in a weather forecast in Japan in 2020[17, 18]. In recent years, research using machine learning has emerged, and Ikeda et al. reported that among 55 models, the one using GAM with multiple variables was more accurate than machine learning-based models[19]. Ogata et al. also pointed out the high accuracy of GAM, but also created a hybrid model with XGBoost to prove its high accuracy[20]. However, most of these heatstroke risk model studies used only one or several years of test data and limited training data, making it unclear whether they would work in other years.

In previous studies in Europe and the United States, time-stratified case-crossover analysis (TC) has often been used to derive the relationship between heat-related mortality and temperature. For example, Voorhees et al. calculated the relationship between heat-related mortality and temperature by calculating odds ratios using TC and substituting them into the health impact function[21]. Wilson et al. also used a case-crossover analysis to determine the odds ratio of heat-related deaths and patients in Sydney[22]. However, there are very few examples using TC that focus solely on heatstroke. Fujitani et al[23]. used TC in the relationship between heatstroke patients and temperature in Tottori, although the accuracy of TC as a heatstroke prediction model has not been verified.

In this study, we compared the accuracy of the three methods, multiple regression analysis (MR), GAM, and TC, using the same variables by calculating the number of heatstroke patients in Tokyo, which is the largest built-up urban area in the world[24]. We also examined how to select the training data and test data, which are often easily determined in this modeling process. We used all the combinations of training data to find the most accurate model and detect the factors that influence the root mean squared error (RMSE).

2. Methods

2 – 1 Study area

Tokyo is located in the Kanto region in the eastern part of Japan. It has a population of 13,995,469 as of April 1, 2022, which is the largest in Japan[25]. The east side of Tokyo is plain and some parts of it are adjacent to the ocean. So, it is affected by the moist easterly winds from the ocean. As there are mountains on the west side, it is affected by the dry warm winds over the mountains from the west. Summers are very hot and humid, and data from the past 30 years (1991–2020) shows that the mean relative humidity from June to September is 75%, and the mean maximum temperature in August is 31.3°C[26].

2–2 Meteorological data

It has been reported that meteorological conditions other than temperature, such as humidity, wind speed, and precipitation, have a much smaller impact on the risk of heatstroke than temperature[27]. This is supported by the fact that

most elderly people who die from heatstroke do so indoors[28]. Therefore we did not include variables other than temperature. We used the daily maximum temperature observed at the meteorological observatory (Tokyo) of the Japan Meteorological Agency from June 1 to September 30, 2012–2020[26].

2–3 Heatstroke patients

The daily number of patients transported by ambulance in Tokyo Metropolitan Area provided by the Fire and Disaster Management Agency of the Ministry of Internal Affairs and Communications[29] was used for the heatstroke patient data in our model. Data from June 1 to September 30, 2012–2020, were used.

2–4 Modeling methods

Three methods were employed: multiple regression analysis (MR), generalized additive model (GAM), and time-stratified case crossover analysis (TC).

2-4-1. Multiple regression model

Previous studies have shown that the number of heatstroke patients increases exponentially with temperature[11]. Therefore, we log-transformed the objective variables and attempted a regression. However, because the objective variable contained 0, it was set to $(y+1)$. Because TC described in subsection 2-4-3 controls for each day of the week of each month, we adjusted the conditions by adding the month and day of the week to the predictor variables as well (Eq. [1]).

$$\mathrm{Log}(y+1) = \beta_0 + \beta_1 T_{\max} + \beta_2 Jun + \beta_3 Jul + \beta_4 Aug + \beta_5 Mon + \beta_6 Tue + \beta_7 Wed + \beta_8 Thu + \beta_9 Fri + \beta_{10} Sat \quad [1]$$

where y is the number of heatstroke patients, β is coefficients, $T_{\max}$ is the daily maximum temperature, and each month and day of the week is entered as a predictor variable with 0 or 1. By adding each month to the predictor dummy variable, we also expected to consider the susceptibility to heatstroke, or heat acclimatization, in each month. The predictor variables were adopted using a stepwise method with the smallest Akaike information criterion.

2-4-2. Generalized additive model

A generalized additive model (GAM) is a model in which the linear variables of the generalized linear model can assume not only normal distribution but also other probability distributions as objective variables[15], and are sums of nonlinear functions, as expressed in Eq. [2]. It has the advantage of being able to assume the shapes of the objective and prediction variables without any assumptions[30].

$$y = \beta_0 \sum_{i=1}^{M} f_i\left(x_i\right) \quad [2]$$

where y in Eq. [2] is the predicted number of heatstroke patients, β_0 is a coefficient, and $f_i(x_i)$ is a third-order smoothing spline curve per explanatory variable. x is the predictor variable and i is the number of predictor variables. The daily maximum temperature was used as a spline function, and each month and day of the week was added to the variables as in MR. The predictor variables were adopted in the same way as described in the subsection 2-4-1. Therefore, M in Eq. [2] was determined using the stepwise method that minimizes the Akaike information criterion. A semiparametric model was used with the gam function of the mgcv package in R, version 1.3.1093.

2-4-3. Time-stratified case-crossover analysis

Case-crossover is a study design originated in case-control studies[31] and is intended only for disease-affected individuals[32]. It is useful for rare diseases, diseases that develop acutely and have a short incubation period, and deaths

and cases where the exposure fluctuates in a short period of time and the induction period of the disease because of the short exposure[33]. It uses only the information of the cases in which the event occurs, and it compares the same person at different times instead of comparing different people at the same time[34]. As such, there is no confounding by any characteristics such as sex, age, and prior medical history[35]. It is possible to obtain controls with matching genetic and background information[32]. In the area of environmental epidemiology, case-crossover design is used in the air pollution epidemiology[36], and it has been mainly studied for respiratory diseases. Voorhees et al. used it in heat illness to demonstrate the relationship between heat-related mortality and temperature[21].

In this study, we assumed that the onset of heatstroke occurred only once per day. The maximum temperature on the day when heatstroke occurred (case period) was compared to the maximum temperature on another day when heatstroke did not occur (reference period). The reference period was selected on the same day of the week, in the same year and month as the case period, using the time stratification method, referring to examples from previous studies[23, 37]. The analysis was performed using R version 1.3.1093. Continuous data are required to use the casecross function in the season package[38], but since the heatstroke data are limited to the hot season, there were data gaps. We converted the data set for using the survival package to put it into the coxph function, which performed the same process as for conditional logistic regression provided by the casecross function of season package[39]. Then the resulting value was used into the health impact function in Eq. [3] in our analysis[21].

$$y = r \times \left(e^{\beta \Delta x} - 1\right) \times Pop + y_0 [3]$$

where y is the predicted number of heatstroke patients, r is the average transport rate per 100,000 people per day during the training data period, Pop is the population in Tokyo on January 1st of the test year, 13,951,636 in 2020, 13,857,443 in 2019, and 13,754,059 in 2018[25].[25] β is the log odds ratio obtained from relative risk associated with a change in exposure expressed as a temperature-response function, and x is the daily maximum temperature. Δx is calculated from the increase/decrease based on 27.7°C which is the mode of the daily maximum temperature of the training data. This value was well aligned with previous studies reporting that the number of heatstroke patients started to increase around 27°C in Tokyo[40]. y_0 is the mean value for patients at 27.7°C in each training data, which is used as the baseline value.

2–5 Training and test data selection

We used data from 2018, 2019, and 2020 as test data. To find a robust model for training data, we used training data from 2012 to the year prior to the test data and tried all combinations shown in Table 1. To check all possible combinations as training data, we tested 255 patterns for 2020, 127 patterns for 2019, and 63 patterns for 2018 as training data. For example, when testing the year 2020, we have eight years for training data from 2012 to 2019, so we substituted eight for *n* in Eq. [4] and determined the number of combinations as $Com = 255$.

$$Com = \sum_{k=1}^{n} {}_nC_k [4]$$

Table 1
Training and test data selection

Test data	2018		2019		2020	
Training data	2012–2017		2012–2018		2012–2019	
	# of used years	# of combinations	# of used years	# of combinations	# of used years	# of combinations
	6	1	7	1	8	1
	5	6	6	7	7	8
	4	15	5	21	6	28
	3	20	4	35	5	56
	2	15	3	35	4	70
	1	6	2	21	3	56
			1	7	2	28
					1	8
	Sum	63	Sum	127	Sum	255

2–6 Comparison of three methods

We evaluated the accuracy of the predicted number of heatstroke patients using the root mean square error (RMSE) and mean absolute error (MAE), as expressed by Eqs. [5] and [6]. Smaller value of the indicator indicates higher accuracy. The observed value is y_i and the predicted value is $y_i^{'}$. n is the number of predicted days ($n = 122$). The focus was on the RMSE to create a model with no major deviations between the observed and predicted values.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(y_i - y_i^{'})^2} [5]$$

$$MAE = \frac{1}{n}\sum_{i=0}^{n}|y_i - y_i^{'}| [6]$$

2–7 Analyses of influential factors

By comparing the errors of each method, we identified factors that influence the errors in the training data. As shown in the previous studies, the end of the rainy season, midsummer days (daily maximum temperature of 30°C or higher), extremely hot days (35°C or higher), and tropical nights (daily minimum temperature of 25°C or higher) were considered influential factors for heatstroke patients[20, 40]. We attempted a linear regression analysis using these factors as explanatory variables and RMSE as the objective variable to see if these factors are also determinants of model error. In addition, we examined the trade-off between the quantity of the training data and quality of the constructed models when adding old data to the training data to consider changes in heatstroke diagnostic criteria[41, 42], and social conditions. For example, the percentage of elderly people in the total population is increasing every year, affecting the validity of older data.

3. Result

3 – 1. Prediction for 2020

When validated for all combinations of training data, there were differences in the results between the three methods. In MR and GAM, the errors varied significantly depending on the combination of the training data. The TC showed constant and small errors in most models. The narrow interquartile range (IQR) of 1.1 in TC in Fig. 1 indicates that the results are not significantly influenced by the training data. When compared with the same training data, the TC value was significantly smaller than that for the MR and GAM ($p < 0.001$).

For MR, out of the 255 training data combinations, the model using only 2019 had the smallest RMSE (Table 2). The models including 2019 were not contained in the 99th percentile of the one with the largest error. The least accurate model used only 2016 as the training data. For GAM, of the 255 training data combinations, the models using 2014, 2018, and 2019 showed the smallest RMSE (Table 3). The least accurate model used only the oldest data (2012) as the training data. In case of TC, a model built with only 2019 data showed a large RMSE, whereas including data from several other years in the model resulted in a smaller RMSE (Table 4). Furthermore, we found that the models including 2013 and 2015 showed smaller RMSEs. All the top five models had an odds ratio of 1.41 and β coefficient of 0.35. Figures 2 and 3 show graphs of the predicted number of heatstroke patients for the best model of the three methods and the observed number in 2020.

Table 2
Accuracy of top five models for 2020 by multiple regression

Training data (year)	RMSE	MAE	Chosen prediction variables
2019	25.54	24.05	Jun, Jul, Aug, Mon, Tue, Thu
2018, 19	26.92	24.05	Jun, Jul, Aug, Thu
2015, 19	27.43	24.06	Jun, Jul, Aug
2014, 19	29.9	24.13	Jun, Jul, Aug, Sat
2015, 18, 19	30.67	24.17	Jun, Jul, Aug, Mon, Thu

Table 3
Accuracy of top five models for 2020 by GAM

Training data (year)	RMSE	MAE	Chosen prediction variables
2014, 18, 19	27.66	18.78	Jun, Jul, Aug, Mon, Tue, Thu
2014, 17, 18, 19	27.66	17.77	Jun, Jul, Aug, Thu
2015, 17, 18, 19	28.13	19.02	Jun, Jul, Aug
2014, 15, 18, 19	28.13	18.47	Jun, Jul, Aug, Sat
2017, 18, 19	28.29	20.01	Jun, Jul, Aug, Mon, Thu

Table 4
Accuracy of top five models for 2020 by time-stratified case-crossover analysis

Training data (year)	RMSE	MAE	Odds ratio (95%CI) (/°C)	β(/°C)
2013, 16, 17, 18, 19	25.51	15.48	1.41 (1.40, 1.42)	0.35
2013, 14, 16, 17, 19	25.51	15.48	1.41 (1.40, 1.42)	0.35
2012, 14, 15	25.51	15.49	1.41 (1.40, 1.43)	0.35
2013, 14, 15	25.51	15.49	1.41 (1.40, 1.43)	0.35
2013, 15, 16, 18, 19	25.51	15.5	1.41 (1.41, 1.42)	0.35

3 – 2. Prediction for 2019

As in the model of prediction for 2020, the value for TC was significantly lower than those of the two methods ($p < 0.001$). The error in TC had a narrow IQR of 1.7 (Fig. 4), indicating that it is not easily affected by the training data. Conversely, GAM had a large IQR of 4.8 and high variability. MR had a relatively narrow IQR of 2.7; however, MR, the errors were large overall.

For MR, of 127 patterns of the training data combinations, the most accurate model was that using the single year data, which was 2018 (Table 5). The accuracy of the models, including 2014, 2015, and 2018, was good. The top 80 of the 127 patterns, in order of the smallest RMSE, always included 2018, 2014, and 2015. Conversely, the combinations of 2012, 2013, 2016, and 2017 occupy the bottom of the list. For the GAM, of the 127 patterns of the training data combinations, the pattern using 2015 and 2018 showed the smallest RMSE (Table 6). The RMSEs of the patterns using multiple years improved when 2015 and 2018 were included. For TC, of 127 patterns of the training data combinations, the most accurate model was a model using the single year data, which is 2015, unlike the prediction for 2020. All top five models had odds ratio of 1.42 to 1.44 and β coefficent of 0.35 to 0.36 (Table 7).

Table 5
Accuracy of top five models for predicting 2019 by multiple regression

Training data (year)	RMSE	MAE	Chosen prediction variables
2018	52.69	23.87	Jun, Jul, Aug
2015, 18	52.97	22.73	Jun, Jul, Aug
2015	53.93	22.94	Jun, Jul, Aug
2014, 15, 18	54.6	23.07	Jun, Jul, Aug, Sat
2014, 18	54.93	23.21	Jun, Jul, Aug, Sat

Table 6
Accuracy of top five models for predicting 2019 by GAM

Training data (year)	RMSE	MAE	Chosen prediction variables
2015,2018	47.02	26.24	Jun, Jul, Aug
2015,2016,2018	47.03	24.82	Jun, Jul, Aug
2016,2018	47.04	26.11	Jun, Jul, Aug
2015	47.33	23.03	Jun, Jul, Aug, Sat
2014,2015,2018	47.42	24.6	Jun, Jul, Aug, Sat

Table 7
Accuracy of top five models for predicting 2019 by the time-stratified case-crossover analysis

Training data (year)	RMSE	MAE	Odds ratio (95%CI) (/°C)	β(/°C)
2015	46.75	22.18	1.43 (1.42, 1.47)	0.36
2014, 15	46.75	22.18	1.43 (1.42, 1.46)	0.36
2014	46.76	22.18	1.44 (1.41, 1.47)	0.36
2013, 14	46.79	22.15	1.42 (1.42, 1.46)	0.35
2013	46.81	22.13	1.44 (1.42, 1.46)	0.36

3–3. Prediction for 2018

The value of the TC was significantly smaller than those of the two methods ($p < 0.005$) (Fig. 5). Compared to the predictions for 2019 and 2020, although the IQR in TC increased slightly to 2.8, it had the narrowest range and was least affected by the training data. For MR, the top five patterns always included 2015, whereas the patterns using 2012 occupied the bottom 14 positions. When predicting 2018, the model that included the most recent data did not always show good accuracy (Table 8). For the GAM, the pattern obtained using 2015 was the most accurate. The accuracy was better when 2012 was not included. The most recent training data, 2017, were distributed throughout, with no effect on reducing or increasing accuracy (Table 9). For TC, of the 63 patterns of the training data combinations, the model using data from 2012 and 2013 showed the smallest RMSE with odds ratio of 1.43 and β coefficient of 0.36 (Table 10). Models that use recent training data did not always fit. The top five models had odds ratio of 1.42 to 1.44 and β coefficient of 0.35 to 0.36.

Table 8
Accuracy of top five models for predicting 2018 by multiple regression

Training data (year)	RMSE	MAE	Chosen prediction variables
2015	50.53	25.12	Jun, Jul, Aug
2015, 16	53.51	26.51	Jun, Jul, Aug, Mon
2013, 15	54.15	26.75	Jun, Jul, Aug
2013, 15, 16	54.97	27.35	Jun, Jul, Aug, Mon, Thu
2015, 17	55.07	26.63	Jun, Jul, Aug

Table 9
Accuracy of top five models for predicting 2018 by GAM

Training data (year)	RMSE	MAE	Chosen prediction variables
2015	47.25	27.17	Jun, Jul, Aug
2013, 15	50.73	27.33	Jun, Jul, Aug, Mon
2015, 17	51.37	26.85	Jun, Jul, Aug
2014, 15	52.03	27.74	Jun, Jul, Aug, Mon, Thu
2013, 15, 17	52.69	27.38	Jun, Jul, Aug

Table 10
Accuracy of top five models for predicting 2018 by time-stratified case-crossover analysis

Training data (year)	RMSE	MAE	Odds ratio (95%CI) (/°C)	β (/°C)
2012,13	53.61	29.51	1.43 (1.42, 1.45)	0.36
2012,15	53.64	29.55	1.43 (1.42, 1.45)	0.36
2012	53.71	29.21	1.42 (1.39, 1.46)	0.35
2013	53.77	29.7	1.42 (1.42, 1.46)	0.35
2013,14	53.82	29.74	1.44 (1.42, 1.46)	0.36

3–4. Determinants of RMSE

We examined whether the difference in the date of the end of the rainy season in the test data was related to this error. First, we analyzed whether the data were closer or farther from the test data based on the normal rainy season's end of July 19 and found no significant differences in the MR ($p = 0.51$), GAM ($p = 0.62$), and TC ($p = 0.85$). (Tables S10, S11, S12). Next, we examined whether the difference in the mean temperature after the rainy season between the training data and test data was related to the magnitude of the error. No significant results were obtained for either GAM ($p = 0.22$) or TC ($p = 0.72$); however, significant results were obtained for MR ($p < 0.05$) (Table S13). We then examined the effect of the number of tropical nights, which was not significant (MR, $p = 0.46$; GAM, $p = 0.5$; TC, $p = 0.29$). In contrast, for MR, the variables for the number of extremely hot days ($p < 0.01$) and year ($p < 0.05$) were significant (Table S10). Because the coefficient of the number of extremely hot days is 1.64, we can say that, for each additional day, the RMSE becomes 1.64 times larger. The coefficient for the year is -2.47, which means that the RMSE increases by a factor of 2.47 for each year of older data. For the GAM, only the variable for the number of extremely hot days was significant ($p < 0.05$), and the coefficient was 1.55 (Table S11). For TC, none of the results were significant, indicating that the magnitude of the RMSE was not affected by these factors (Table S12). However, the year may have some influence, not just in one direction, and will therefore be examined in the next section.

3–5. Trade-off between quantity and quality of adding old data

We examined how the RMSE changed as we added one year of the training data to the models at a time. Figure 8 shows a comparison of the RMSE between the number of heatstroke patients in 2018 predicted by the three methods and the actual number of cases. All three methods showed a gradual V-shaped curve; the more the previous data years were added to the training data, the higher the accuracy became until a certain stage from which the accuracy was lowered. The error was smaller if the older data (i.e., 2012 and 2013) were not used; however, the error was larger if the new data (i.e., 2017) are

used alone. The right combination lies between the old and the new data. In the MR and GAM, the model using data from 2015 to 2017 showed the smallest RMSE and MAE values. For TC, the model using data from 2014 to 2017 showed the smallest RMSE and MAE with odds ratio of 1.4 (95% confidence interval (CI): 1.38 and 1.41, $p < 0.001$) and log odds ratio of 0.34 (Tables 13, S7, S8, and S9). As shown in Figs. 6 and 7, the same can be said when 2020 and 2019 are used as test data, but the error in the MR is smaller for newer data (Tables 11, 12, S1, S2, S3, S4, S5, and S6).

Table 11
Accuracy of models for 2020 by three methods when training data was decreased by one year from 2012 to 2019

Training data (year)	**MR**	**GAM**	**TC**	**MR, GAM**	**TC**	
	RMSE			**Chosen prediction variables**	**Odds ratio (95%CI) (/ °C)**	**β (/ °C)**
2012–19	41.46	32.36	27.03	Jun, Jul, Aug, Mon, Wed, Thu	1.39 (1.39, 1.40)	0.33
2013–19	40.05	30.67	26.24	Jun, Jul, Aug, Mon, Wed, Thu	1.4 (1.39, 1.40)	0.34
2014–19	38.39	29.49	25.91	Jun, Jul, Aug, Mon, Sat	1.4 (1.40, 1.41)	0.34
2015–19	37.91	29.03	25.51	Jun, Jul, Aug, Mon, Tue, Wed, Thu	1.41 (1.40, 1.42)	0.35
2016–19	37.82	28.91	25.68	Jun, Jul, Aug, Mon, Thu	1.41 (1.40, 1.42)	0.34
2017–19	32.83	28.29	25.53	Jun, Jul, Aug, Mon, Thu	1.42 (1.40,1.43)	0.35
2018–19	26.92	29.78	26.95	Jun, Jul, Aug, Thu	1.43 (1.42, 1.45)	0.36
2019	25.54	49.46	60.06	Jun, Jul, Aug, Mon, Tue, Thu	1.51 (1.48, 1.53)	0.41

Table 12
Accuracy of models for 2019 by three methods when training data was decreased by one year from 2012 to 2018

Training data (year)	MR	GAM	TC	MR, GAM	TC	
	RMSE			Chosen prediction variables	Odds ratio (95%CI) (/°C)	β (/°C)
2012–18	59.55	52.1	51.21	Jun, Jul, Aug, Mon, Wed, Thu	1.39 (1.38, 1.40)	0.33
2013–18	58.62	50.11	50.69	Jun, Jul, Aug, Mon, Wed, Thu	1.39 (1.38, 1.40)	0.33
2014–18	57.57	48.76	50.44	Jun, Jul, Aug, Mon, Sat	1.39 (1.38, 1.40)	0.33
2015–18	57.76	47.92	51.49	Jun, Jul, Aug, Mon, Wed, Thu	1.38 (1.37, 1.40)	0.33
2016–18	59.24	48.21	53.8	Jun, Jul, Aug, Mon, Wed, Thu	1.37 (1.36, 1.38)	0.31
2017–18	57.15	48.6	53.54	Jun, Jul, Aug, Mon, Wed, Thu	1.37 (1.36, 1.39)	0.31
2018	52.69	50.28	51.92	Jun, Jul, Aug	1.38 (1.37, 1.40)	0.32

Table 13
Accuracy of models for 2018 by three methods when training data was decreased by one year from 2012 to 2017

Training data (year)	MR	GAM	TC	MR, GAM	TC	
	RMSE			Chosen prediction variables	Odds ratio (95%CI) (/°C)	β (/°C)
2012–17	60.79	57.58	57.9	Jun, Jul, Aug, Mon, Sat	1.39 (1.38, 1.40)	0.33
2013–17	57.91	55.6	56.87	Jun, Jul, Aug, Mon, Wed, Thu	1.39 (1.38, 1.41)	0.33
2014–17	57.94	57.14	56.08	Jun, Jul, Aug, Mon, Sat	1.4 (1.39, 1.41)	0.34
2015–17	56.31	55.49	58.18	Jun, Jul, Aug, Mon, Wed	1.39 (1.37, 1.40)	0.33
2016–17	61.2	66.89	65.97	Jun, Jul, Aug, Mon, Wed, Thu	1.35 (1.33, 1.37)	0.3
2017	63.11	64.83	68.93	Jun, Jul, Aug, Wed	1.34 (1.31, 1.36)	0.29

4. Discussion

Although heatstroke patient prediction models have been proposed in previous studies, they are not robust because of the restricted training data. For example, Ogata et al[20]. created a model with excellent accuracy, but it was not robust to training data because it used only one test dataset, 2018, and a fixed three-year period from 2015. However, we found that it is difficult to create robust models that account for all training data in GAM and MR, which are sensitive to training data. Therefore, using the optimal odds produced by using a TC that is less sensitive to training data, a risk model with high overall accuracy can be developed. It can be said that the use of TC can make this difficult task possible.

In addition, although previous studies have shown a correlation between the number of tropical nights and midsummer days and the occurrence of heatstroke[29], these factors did not appear to be factors other than the maximum temperature in the risk model. The error was significantly larger when the number of extremely hot days differed between the training and test data in MR and GAM ($p < 0.01$, $p < 0.05$). When extreme temperatures occur, the prediction model cannot fully predict the number of people who will experience heatstroke, and the errors are likely to be large. Previous studies, either linear or nonlinear, have been less accurate in predicting a sharp increase in the total number of patients with heatstroke. It is conceivable that any discrepancy in the number of extremely hot days from the test data would increase the error[17, 19, 20]. Therefore, including training data with the same trend in the number of extremely hot days may increase the accuracy. In addition, the MR also showed a correlation with the mean temperature after the rainy season ($p < 0.05$). If long-term forecasting can predict an event, training data that have a similar trend to the test year can also be included. Furthermore, the inclusion of older training data significantly decreased the accuracy of MR ($p < 0.05$). Regardless of the method used, we also found that the best trade-off between quantity and quality by adding older data to the training data when using consecutive training data is 3 – 4 years from the latest data. The reasons cited for the ineffectiveness of older data include changes in diagnostic criteria and social conditions. An example of a change in social conditions is that the percentage of

elderly people susceptible to heatstroke is increasing each year[43]. We compared three methods using daily maximum temperature as the predictor variable based on previous research[10]. Future studies should examine other methods and the addition of other predictor variables.

5. Conclusions

Comparing the accuracy of the three methods with the heatstroke patients as the objective variable and the maximum temperature as the predictor variable, we found that the error of TC was significantly smaller than those of the other two methods ($p < 0.005$) and less sensitive to training data. Hence, the most accurate odds ratio of 1.41–1.44 and β coefficient of 0.35–0.36 were found after validation of the three test datasets using all possible training data from the past. If data are to be included continuously, a model constructed from three or four years of data from the latest one is the most accurate.

Abbreviations

MR
multiple regression analysis
GAM
generalized additive model
TC
time-stratified case-crossover analysis
GLM
generalized linear model
DLNM
distributed lag non-linear model
WBGT
wet bulb globe temperature
RMSE
root mean squared error
MAE
mean absolute error
IQR
Interquartile range
CI
confidence interval

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets of meteorological generated and/or analysed during the current study are available in the Japan Meteorological Agency repository, https://www.data.jma.go.jp/obd/stats/etrn/.

The heatstroke patients datasets generated and/or analysed during the current study are available in the Fire and Disaster Management Agency of the Ministry of Internal Affairs and Communications repository, https://www.fdma.go.jp/disaster/heatstroke/post3.html.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by JSPS KAKENHI Grant Numbers 20H03949 and 22H01649, Grant-in-Aid for Scientific Research (B), 18KK0123, Fund for the Promotion of Joint International Research (Fostering Joint International Research (B)).

Authors' contributions

MN conceptualized the study, gathered the data, performed the necessary statistical analysis, and wrote the manuscript. SH contributed to the implementation of the time-stratified case-crossover analysis in R. SS was involved in the creation of the R programming. AF contributed to the modeling of multiple regression analysis. HN provided advice from a medical perspective. TI was involved in the interpretation of the results and structured the manuscript. The authors read and approved the final manuscript.

Acknowledgment

The authors are deeply grateful to Akimasa Hirata Director and Professor Nagoya Institute of Technology. He consulted us a lot to identify patients with heatstroke. We also thank the Microclimatic Design Research Group (MDRG) members, and Editage (www.editage.com) for English language editing.

References

1. McMichael AJ, Woodruff RE, Hales S: **Climate change and human health: present and future risks**. Lancet 2006, **367**(9513):859–869.
2. Haines A, Kovats RS, Campbell-Lendrum D, Corvalan C: **Climate change and human health: impacts, vulnerability and public health**. Public Health 2006, **120**(7):585–596.
3. Kalkstein LS, Greene JS: **An evaluation of climate/mortality relationships in large U.S. cities and the possible impacts of a climate change**. Environ Health Perspect 1997, **105**(1):84–93.
4. Tan J, Zheng Y, Tang X, Guo C, Li L, Song G, Zhen X, Yuan D, Kalkstein AJ, Li F: **The urban heat island and its impact on heat waves and human health in Shanghai**. Int J Biometeorol 2010, **54**(1):75–84.
5. Argaud L, Ferry T, Le QH, Marfisi A, Ciorba D, Achache P, Ducluzeau R, Robert D: **Short- and long-term outcomes of heatstroke following the 2003 heat wave in Lyon, France**. Arch Intern Med 2007, **167**(20):2177–2183.
6. Toosty NT, Hagishima A, Tanaka KI: **Heat health risk assessment analysing heatstroke patients in Fukuoka City, Japan**. PLoS One 2021, **16**(6):e0253011.
7. **Ministry of the Environment, Japan "Heat Illness Prevention Information" https://www.wbgt.env.go.jp**. In.; 2021.12.
8. **Japan Meteorological Agency "Heatstroke Alert" https://www.jma.go.jp**/bosai/information/**heat.html**. In.; 2021.12.
9. T. MTaS: **The Science of WBGT Indicators**. In.

10. Honda Y, Kabuto M, Ono M, Uchiyama I: **Determination of optimum daily maximum temperature using climate data**. Environ Health Prev Med 2007, **12**(5):209–216.
11. Fuse A SS, Fuse R, Araki H, Kim F, Miyauchi M and Yokota H.: **Predicting the number of emergency transporters for heat stroke from meteorological data**. Journal of the Japanese Society of Emergency Medicine 2014, **25**(10):757–765.
12. Ravindra K, Rattan P, Mor S, Aggarwal AN: **Generalized additive models: Building evidence of air pollution, climate change and human health**. Environ Int 2019, **132**:104987.
13. Honda Y, Kondo M, McGregor G, Kim H, Guo YL, Hijioka Y, Yoshikawa M, Oka K, Takano S, Hales S *et al*: **Heat-related mortality risk model for climate change impact projection**. Environ Health Prev Med 2014, **19**(1):56–63.
14. Yamamoto M KM, Taifu T, Hanaoka W, and Mochida A: **Evaluation of heatstroke risk based on climate analysis results by WRF using pseudo-warming method**: **Comparison between Tokyo and Sendai in the** 2050**s**. In: *2017*: Annual Research Presentation of the Japanese Society of Wind Engineering: 101–102.
15. Sato T, Kusaka H, Hino H: **Quantitative assessment of the contribution of meteorological variables to the prediction of the number of heat stroke patients for Tokyo**. *Sola* 2020.
16. Kodera S, Nishimura T, Rashed EA, Hasegawa K, Takeuchi I, Egawa R, Hirata A: **Estimation of heat-related morbidity from weather data: A computational study in three prefectures of Japan over 2013–2018**. Environ Int 2019, **130**:104907.
17. Nishimura T, Rashed EA, Kodera S, Shirakami H, Kawaguchi R, Watanabe K, Nemoto M, Hirata A: **Social implementation and intervention with estimated morbidity of heat-related illnesses from weather data: A case study from Nagoya City, Japan**. Sustainable Cities and Society 2021, **74**:103203.
18. **TBS [Good luck!] information in [Good weather]** (**2019**.**10**–**2021**.**3**) **https**://**www**.**tbs**.**co**.**jp**/guttoluck-tbs/. In.
19. Ikeda T, Kusaka H: **Development of models for predicting the number of patients with heatstroke on the next day considering heat acclimatization**. Journal of the Meteorological Society of Japan Ser II 2021.
20. Ogata S, Takegami M, Ozaki T, Nakashima T, Onozuka D, Murata S, Nakaoku Y, Suzuki K, Hagihara A, Noguchi T *et al*: **Heatstroke predictions by machine learning, weather information, and an all-population registry for 12-hour heatstroke alerts**. Nat Commun 2021, **12**(1):4575.
21. Voorhees AS, Fann N, Fulcher C, Dolwick P, Hubbell B, Bierwagen B, Morefield P: **Climate change-related temperature impacts on warm season heat mortality: a proof-of-concept methodology using BenMAP**. Environ Sci Technol 2011, **45**(4):1450–1457.
22. Wilson LA, Morgan GG, Hanigan IC, Johnston FH, Abu-Rayya H, Broome R, Gaskin C, Jalaludin B: **The impact of heat on mortality and morbidity in the Greater Metropolitan Sydney Region: a case crossover analysis**. Environ Health 2013, **12**:98.
23. Fujitani Y, Otani S, Majbauddin A, Amano H, Masumoto T, Kurozawa Y: **Impact of Maximum Air Temperature on Ambulance Transports Owing to Heat Stroke During Spring and Summer in Tottori Prefecture, Japan: A Time-stratified Case-crossover Analysis**. Yonago Acta Med 2019, **62**(1):47–52.
24. **Dermographia World Urban Areas 17th Annual Edition**: **2021 06**. In.
25. **Tokyo statistical data "Tokyo statistical year book" https**://**www**.**toukei**.**metro**.**tokyo**.**lg**.jp/jsuikei/**js-index2**.**htm**. In.; 2021.12.
26. **Japan Meteorological Agency "Search past weather data"**. **Tokyo**: **Japan https**://**www**.**data**.**jma**.**go**.**jp**/obd/stats/etrn/**index**.**php**. In.; 2021.12.
27. Honda Y: **Direct health risks due to climate change: Heat-related morbidity and mortality**. J Natl Inst Public Health 2020, **69**(5):412–417.
28. Shibata Yasue KE, and Matsubara Naoki.: **A Study of the Actual Indoor Thermal Environments and the Consciousness and Behavior for the Prevention to Heat Disorders in the Elderly –Potential for a Cooling Effect of Thermal**

Environment in Residential Buildings Due to Recognition (Visualization) of the Indoor Sensible Temperature during the Summer–. Journal of the Japan Society of Biometeorology 2018, **55**(1):33–50.

29. Fire and Disaster Management Agency of the Ministry of Internal Affairs and Communications" Disaster information", Emergency heatstroke carriers by prefecture. https://www.fdma.go.jp/disaster/heatstroke/post3.html. In.
30. Hastie TJ, Tibshirani RJ: Generalized additive models: Routledge; 2017.
31. Maclure M: The case-crossover design: a method for studying transient effects on the risk of acute events. Am J Epidemiol 1991, **133**(2):144–153.
32. Shiosakai K KT: Case Crossover Study. Pharmaceutical Epidemiology 2014, **18**(2):90–94.
33. Fisher JA, Puett RC, Laden F, Wellenius GA, Sapkota A, Liao D, Yanosky JD, Carter-Pokras O, He X, Hart JE: Case-crossover analysis of short-term particulate matter exposures and stroke in the health professionals follow-up study. Environ Int 2019, **124**:153–160.
34. Basu R, Dominici F, Samet JM: Temperature and mortality among the elderly in the United States: a comparison of epidemiologic methods. Epidemiology 2005, **16**(1):58–66.
35. Lohsoonthorn V, Rattananupong T, Wynne K, Thomas C, Chahal HS, Berhane HY, Mostofsky E, Wuttithai N, Gelaye B: Immediate risk of myocardial infarction following physical exertion, tea, and coffee: A case-crossover study in Thailand. PLoS One 2019, **14**(1):e0210959.
36. Nitta H, Yamazaki S, Omori T, Sato T: An introduction to epidemiologic and statistical methods useful in environmental epidemiology. J Epidemiol 2010, **20**(3):177–184.
37. Basu R, Feng WY, Ostro BD: Characterizing temperature and mortality in nine California counties. Epidemiology 2008, **19**(1):138–145.
38. Barnett AG, Dobson AJ: Analysing seasonal health data, vol. 30: Springer; 2010.
39. Zhang Z: Case-crossover design and its implementation in R. Ann Transl Med 2016, **4**(18):341.
40. Hoshi A IH, and Murayama K.: Characteristics of heatstroke in Tokyo and Chiba. Journal of the Japanese Journal of Biometeorology 2007, **44**(1):3–11.
41. Kondo Y, Hifumi T, Shimazaki J, Oda Y, Shiraishi SI, Hayashida K, Fukuda T, Wakasugi M, Kanda J, Moriya T *et al*: Comparison between the Bouchama and Japanese Association for Acute Medicine Heatstroke Criteria with Regard to the Diagnosis and Prediction of Mortality of Heatstroke Patients: A Multicenter Observational Study. Int J Environ Res Public Health 2019, **16**(18).
42. Nakai S: Death Statistics for Heat Stroke in Japan — What Statistical Classification Should Be Used? —. Journal of the Japanese Society of Biometeorology 2019, **56**(2):67–75.
43. Ministry of Internal Affairs and Communications "Statistics on the Elderly in Japan" https://www.stat.go.jp/data/topics/pdf/topics129.pdf. In.; 2021.9.

Figures

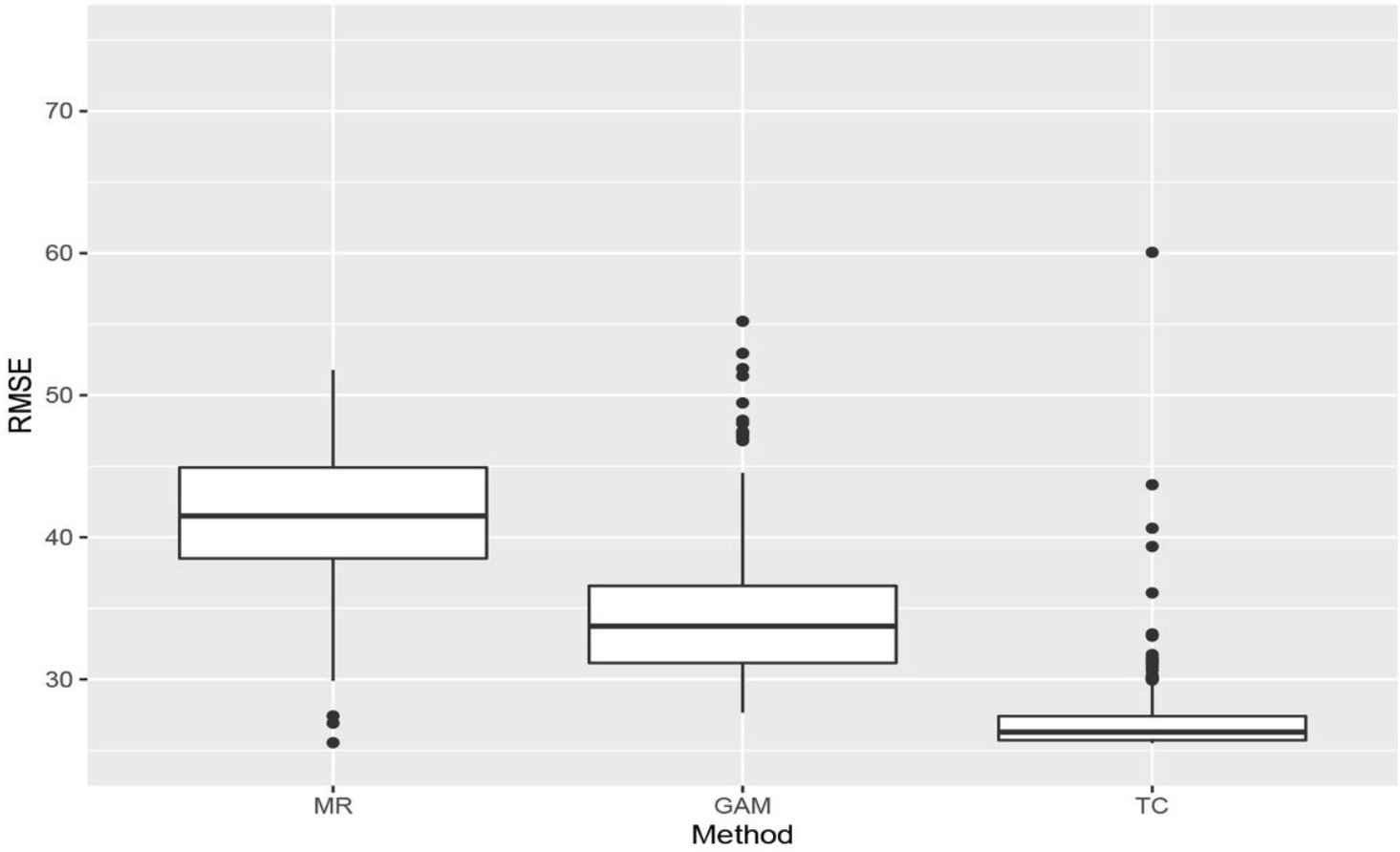

Figure 1

Distribution of RMSE by training data for all combination of 255 cases in 2020.

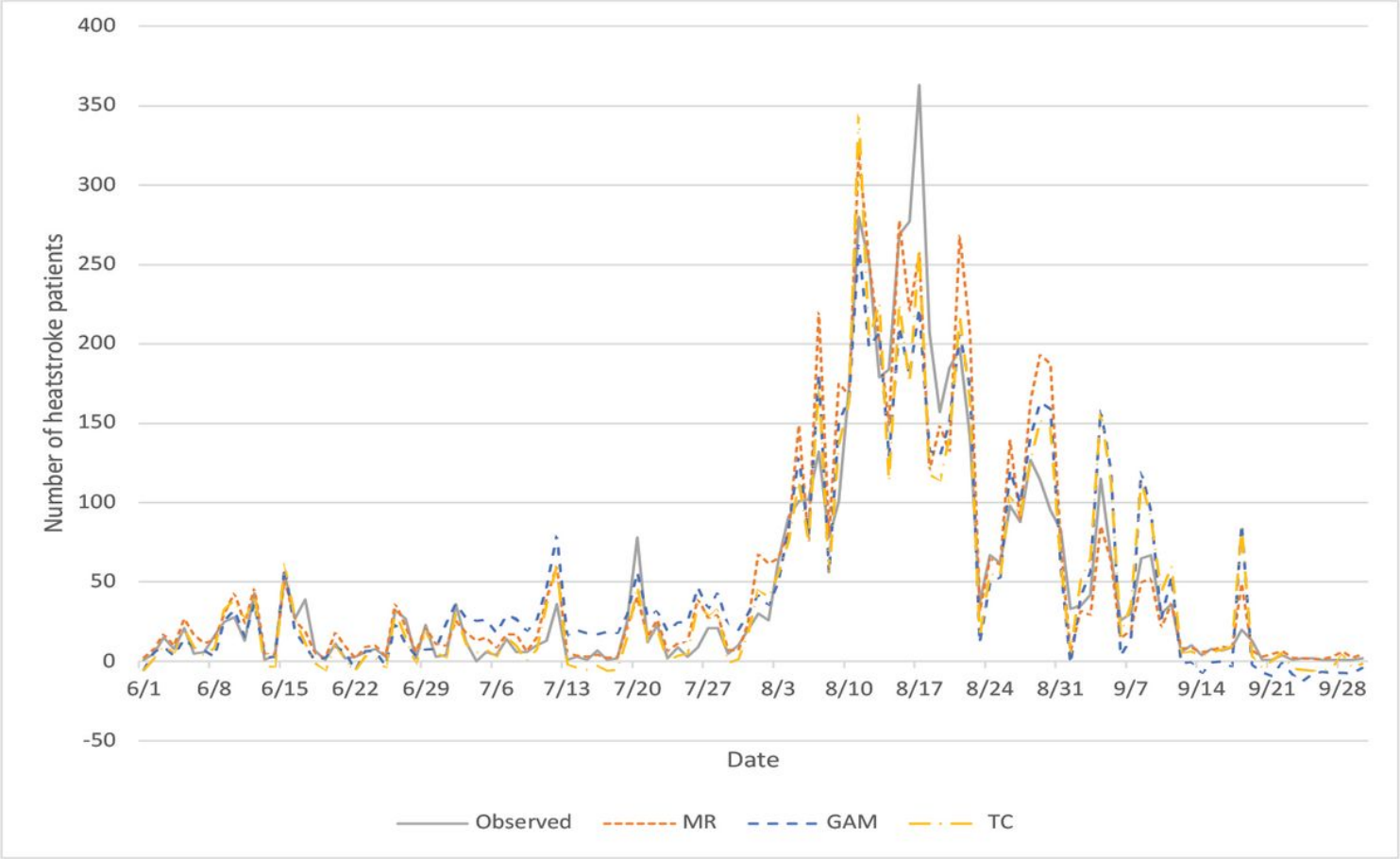

Figure 2

Number of predicted and observed heatstroke patients in 2020 / The most accurate model for each of the three methods was used to predict the number of each day.

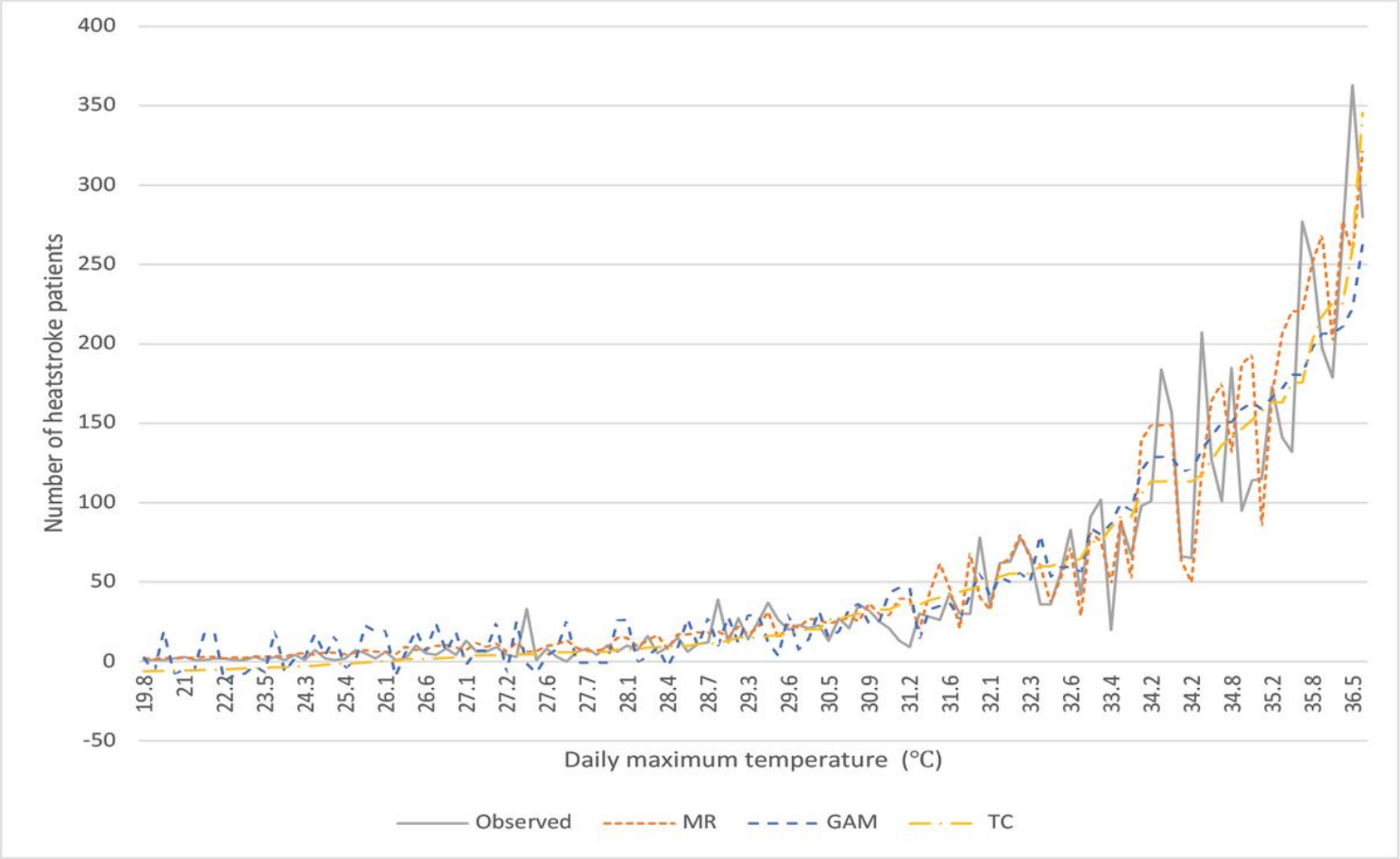

Figure 3

Number of predicted and observed heatstroke patients in 2020 / The most accurate model for each of the three methods was used to predict the number at that daily maximum temperature.

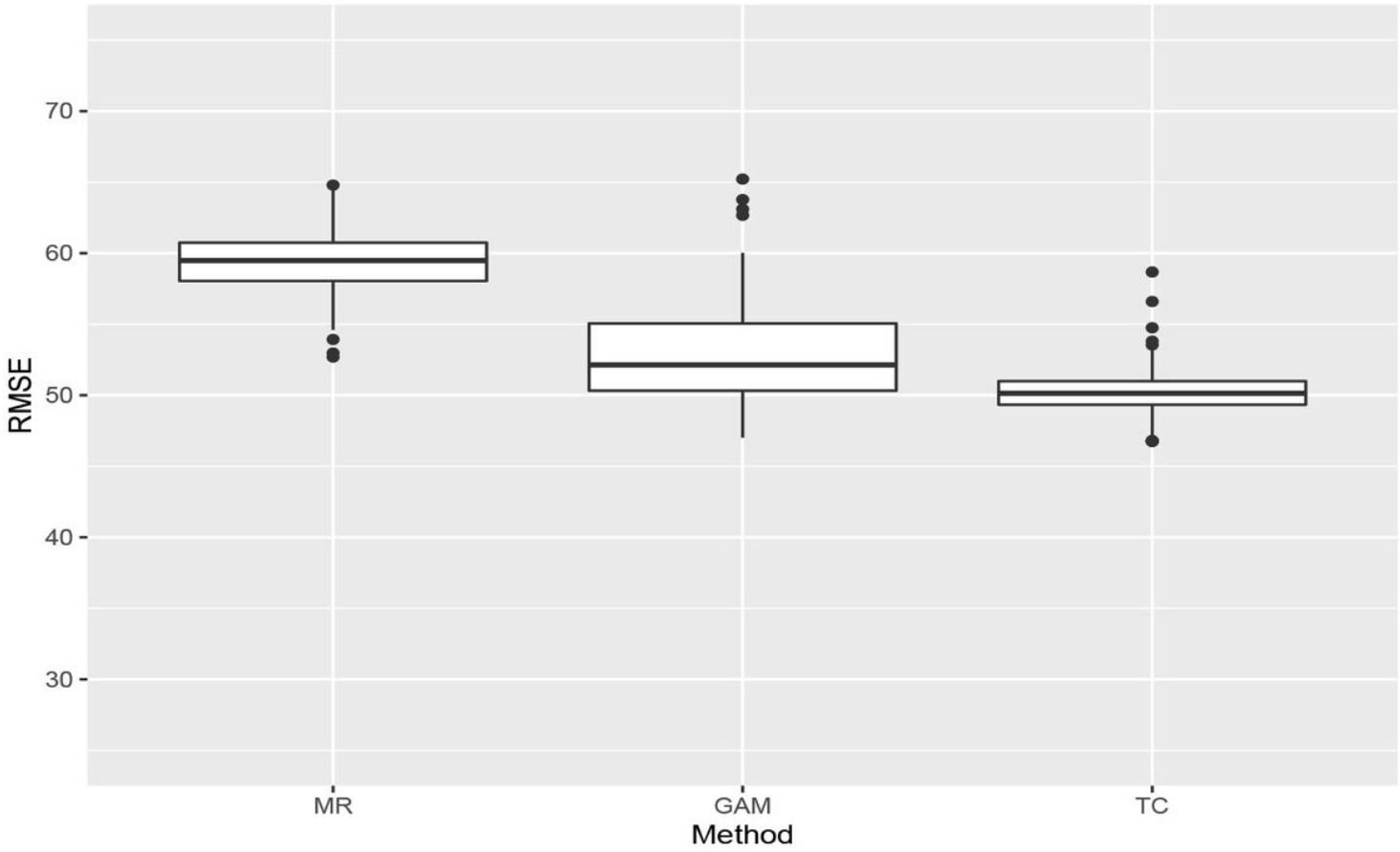

Figure 4

Distribution of RMSE by training data for all combination of 127 cases in 2019

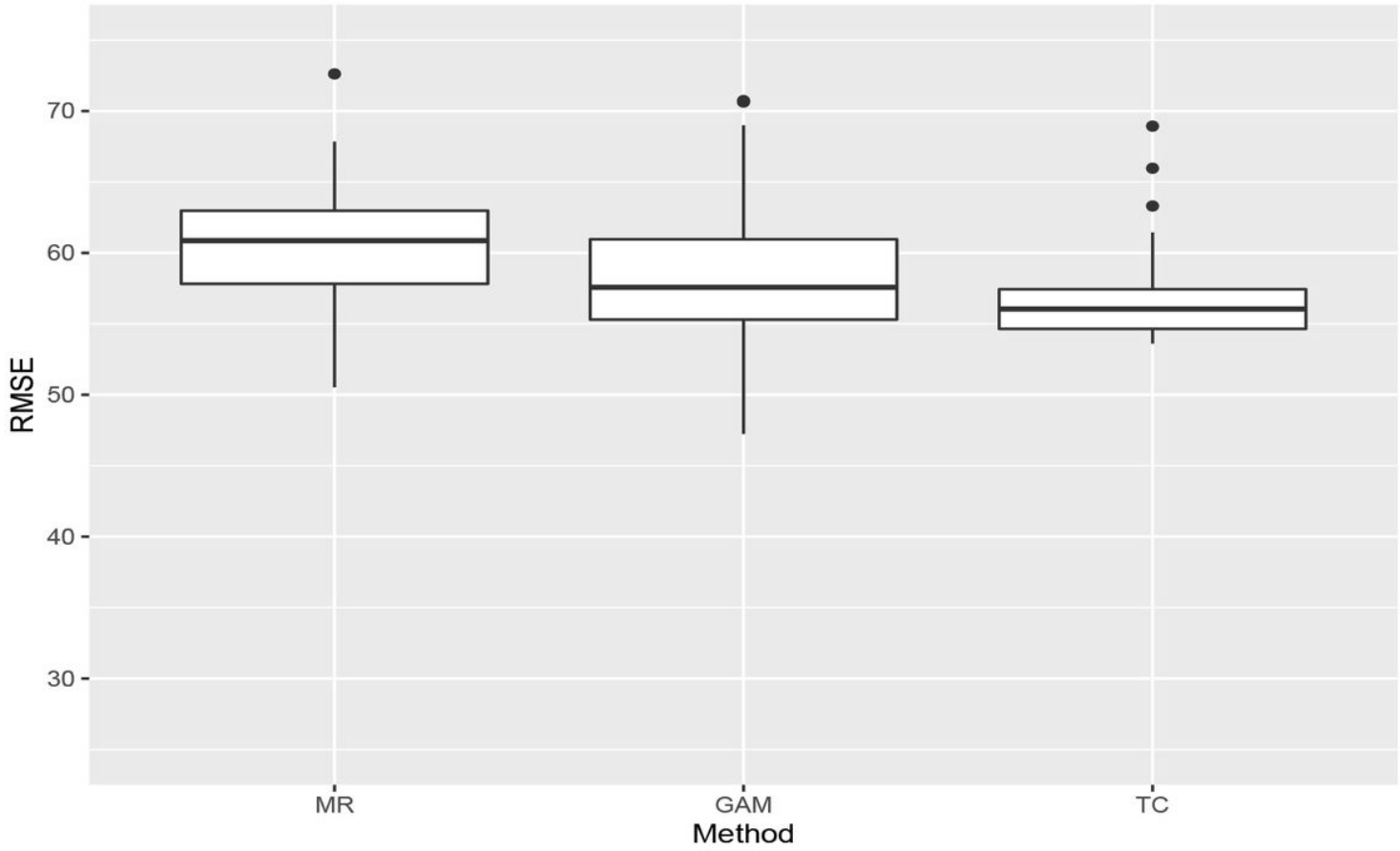

Figure 5

Distribution of RMSE by training data for all combination of 63 cases in 2018

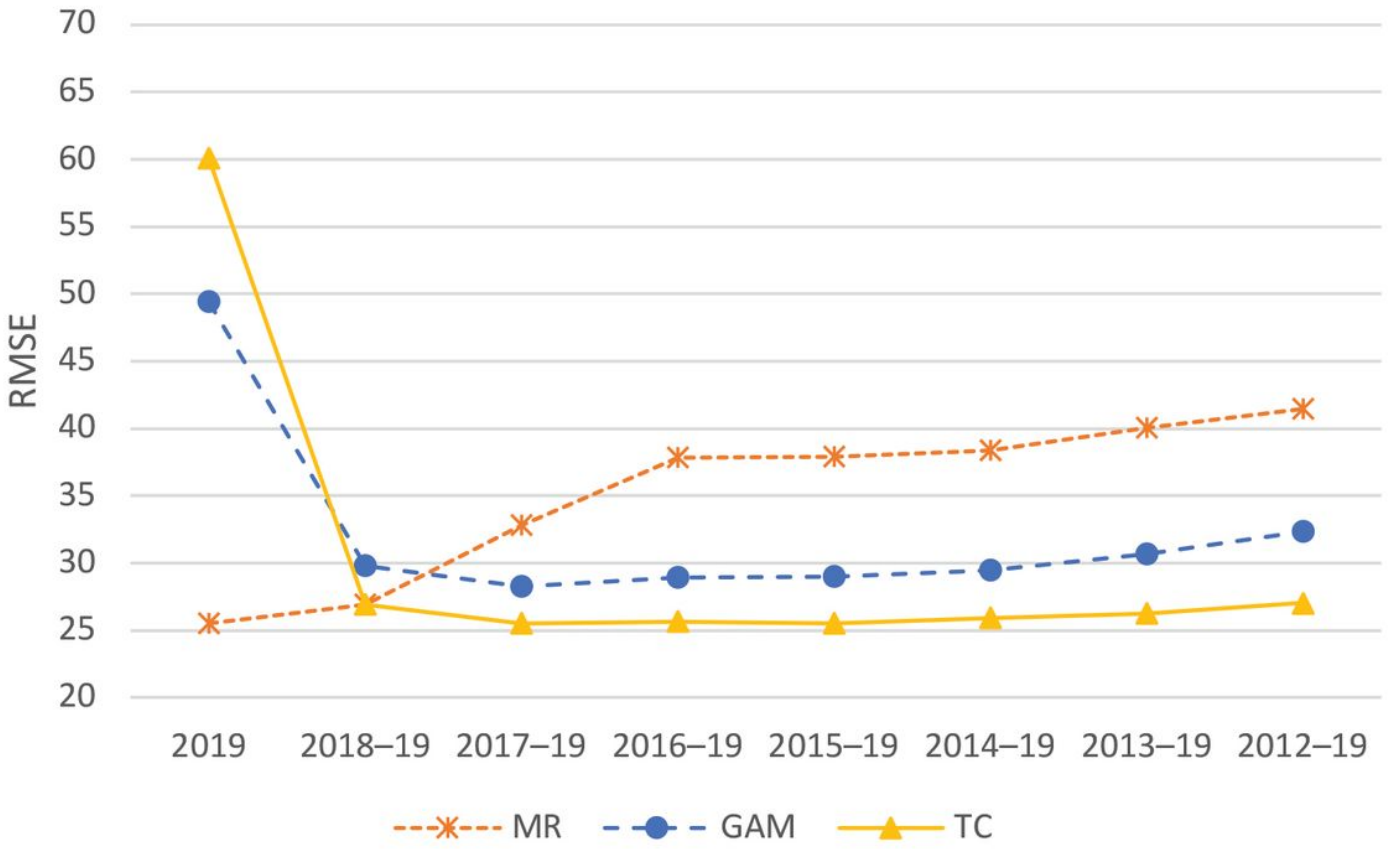

Figure 6

Comparison of the RMSE of the three methods predicting for 2020 / When training data was added by one year from 2012 to 2019.

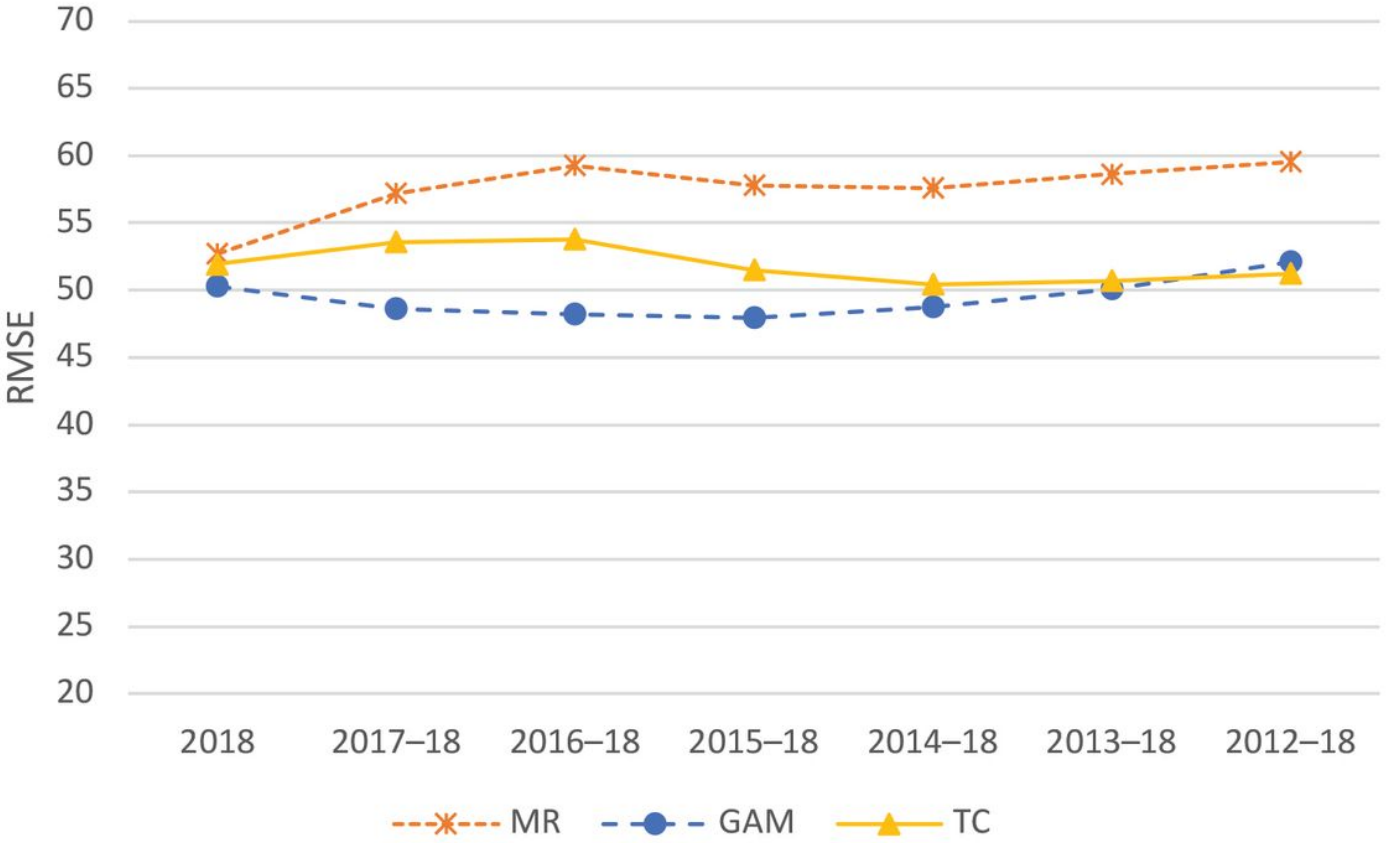

Figure 7

Comparison of the RMSE of the three methods predicting for 2019 / When training data was added by one year from 2012 to 2018.

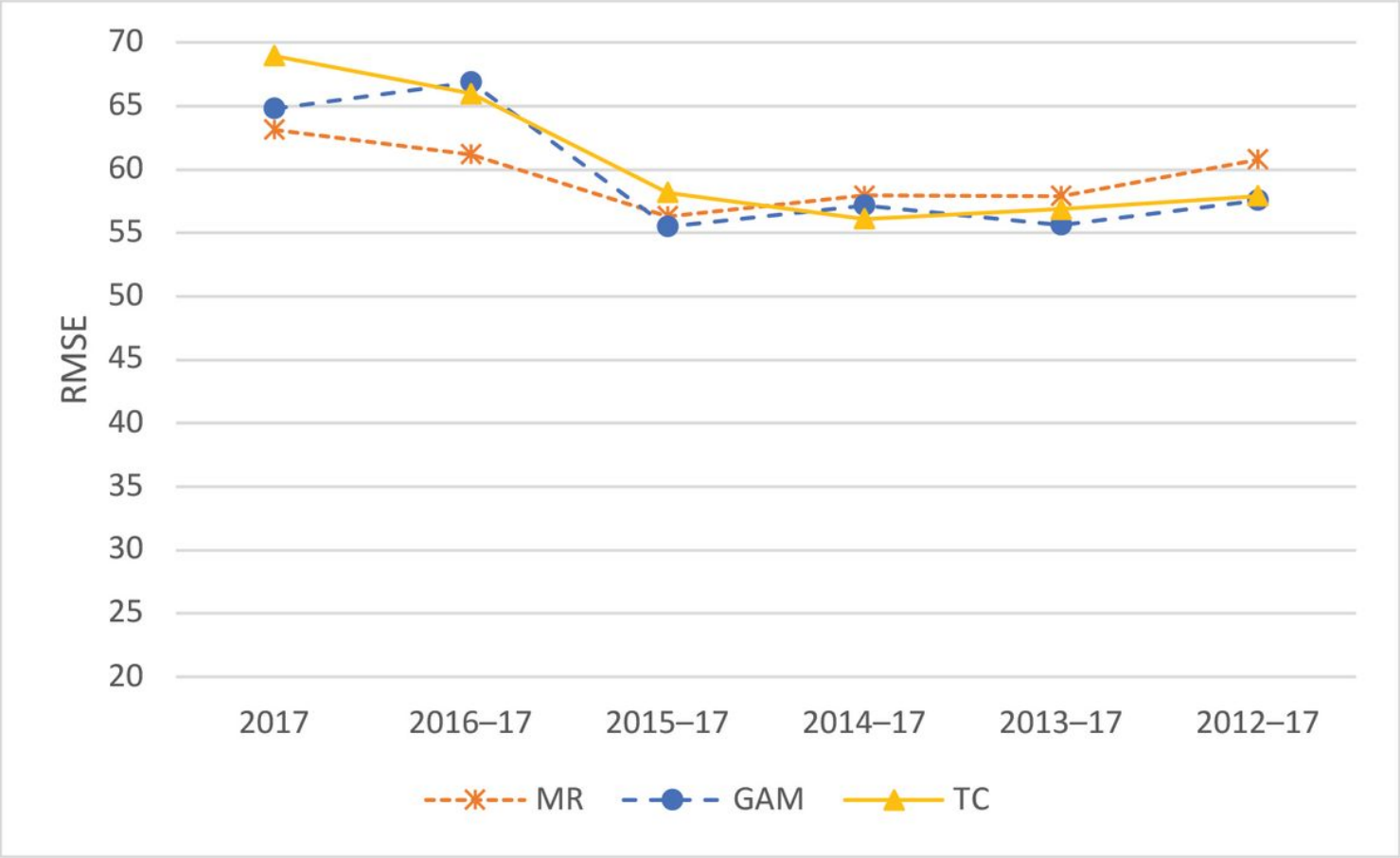

Figure 8

Comparison of the RMSE of the three methods predicting for 2018 / When training data was added by one year from 2012 to 2017.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalMaterial.docx