A class forecasting method for time series

Mauricio Cabrera-Rios[1], Yaleiry García[2], Nanette Xu[2], Ana Sofía Cabrera-Isaza[3], Wesley T Pardo-Sánchez[1], Benisa Irizarry[2], Yeaniled Muñoz[2], Ray López[2], Clara E Isaza[4], and Mauricio Cabrera-Ríos[2]

[1]Affiliation not available
[2]Industrial Engineering Department The Applied Optimization Group, University of Puerto Rico at Mayagüez Mayagüez PR
[3]Residential Center for Educational Opportunities of Mayagüez STEM Specialized High School Mayagüez PR
[4]Biology Department The Applied Optimization Group, University of Puerto Rico at Mayagüez Mayagüez PR

May 24, 2023

A class forecasting method for time series

Yaleiry García[1]

Nanette Xu[1]

Ana Sofía Cabrera-Isaza[2]

Wesley T. Pardo-Sánchez

Benisa Irizarry[1]

Yeaniled Muñoz[1]

Ray López[1]

Clara E. Isaza[3]

Mauricio Cabrera-Ríos[1]

1 Industrial Engineering Department

The Applied Optimization Group

University of Puerto Rico at Mayagüez

Mayagüez PR 00681

2 Residential Center for Educational Opportunities of Mayagüez

STEM Specialized High School

Mayagüez PR 00682

3 Biology Department

The Applied Optimization Group

University of Puerto Rico at Mayagüez

Mayagüez PR 00681

Corresponding author: Mauricio.cabrera@upr.edu

A class forecasting method for time series

Abstract

A time series is a record of the numerical values of a random variable in chronological order. This information organized in time periods is often used to forecast the value of the random variable in a future period or in a series of future periods. Time series forecasting is used across multiple disciplines from Economics to Public Health mainly seeking forecasting accuracy, sometimes at the cost of the transparency of the method used. This work proposes using contiguous, mutually exclusive classes spanning the range of possible values that the random variable can take at each time period. This framework allows computing the historical transition frequency from one class to another even in cases where the time periods are not adjacent to each other. Subsequently, these frequencies are used to build an index to identify the most likely class (or classes) for the forecasted period. The resulting method is transparent, and it outperforms the moving average and the simple exponential smoothing methods in a comparative study of 21 time series. Applications to other time series involving COVID19 and author citations are presented for illustration purposes.

Keywords: Class forecasting, Frequentist approach, Nonparametric method

Introduction

Understanding historical data based on numerical patterns to predict future behavior has been of interest for decades and is the heart of the discipline of forecasting. The large body of literature in this discipline shows how advanced as well as how elusive its main goal remains. Excellent forecasting reviews are easy to find, for example [1], however, looking at the diversity of tools available it is interesting to note that methods that are simple, transparent and effective are well sought after [2)][3]. This is true even when machine learning and artificial intelligence methods have gained a lot of ground as forecasting tools [4] [5].

The idea of using methods that have an intuitive understanding for decision-makers has been the subject of intense discussion [3] where the Occam's razor principle is often invoked [6]. Using a well-established mathematical pattern, such as a fractal, seems to be another direction that can enhance understanding in forecasting, as described in [7], as is the notion of using intervals as opposed to point forecasts [8)].

This work generates time series forecasts based on two key elements: (1) organizing the time-stamped records into continuous and contiguous classes, and (2) computing the relative frequency between classes using different time lags. The aim is a method that can be intuitively understood as "our forecast follows from what has most frequently happened in the past", without requiring the selection of a models or the computation of any parameters. In addition, the use of a binary tree fractal is proposed to create the classes, as explained in a later section of the manuscript.

Method

Consider a time series with M identically sized periods ($m = 1,2,3, \dots, M$). On each time period, the value of a random variable is recorded, y_m. The sample space in which y_m occurs is divided up in K contiguous and mutually exclusive classes ($k = 1,2,3, \dots, K$). Each y_m is then associated to one of the K classes to be expressed as $\breve{y}_m \in \{1,2,3, \dots, K\}$. A number L of previous time periods or lags ($l = 1,2,3, \dots, L$) is chosen by the user to compute the forecast for a particular time period t. To this end, the training matrix T with $M - L$ rows and $L + 1$ columns is built as follows:

$$
\begin{array}{ccccc}
\breve{y}_1 & \cdots & \breve{y}_{L-1} & \breve{y}_L & \breve{y}_{L+1} \\
\breve{y}_2 & \cdots & \breve{y}_L & \breve{y}_{L+1} & \breve{y}_{L+2} \\
\vdots & & \vdots & \vdots & \vdots \\
\breve{y}_L & \cdots & \breve{y}_{2L-2} & \breve{y}_{2L-1} & \breve{y}_{2L} \\
\vdots & & \vdots & \vdots & \vdots \\
\breve{y}_{M-L} & \cdots & \breve{y}_{M-2} & \breve{y}_{M-1} & \breve{y}_M
\end{array}
$$

(M1)

Using training matrix T as a reference, build L matrices each dimensioned $K \times K$. The $l - th$ square matrix tallies the class transitions from the K classes in the $l - th$ column to the $K\ classes$ in the $(L + 1) - th$ column, that is, it computes the class transition frequencies. For the $l - th$ square matrix, the quantity f_{ij}^l denotes the transition frequency from the class in the $i - th$ row to the class of the $j - th$ column. The $l - th$ square matrix is, as follows:

$$
\begin{array}{cc|ccccc}
 & & \multicolumn{5}{c}{\text{Classes in column } L+1} \\
 & & 1 & 2 & 3 & \dots & K \\
\hline
\text{Classes in} & 1 & f_{11}^l & f_{12}^l & f_{13}^l & \dots & f_{1K}^l \\
\text{Column } l & 2 & f_{21}^l & f_{22}^l & f_{23}^l & \dots & f_{2K}^l \\
 & 3 & f_{31}^l & f_{32}^l & f_{33}^l & \dots & f_{3K}^l \\
 & \vdots & \vdots & \vdots & \vdots & & \vdots \\
 & K & f_{K1}^l & f_{K2}^l & f_{K3}^l & \dots & f_{KK}^l
\end{array}
$$

(M2)

Next, each of the L square matrices are modified to compute the class transition relative frequency defined as:

$$
p_{ij}^l = \frac{f_{ij}^l}{\sum_{\forall j} f_{ij}^l}
$$

(E1)

Such that the $l-th$ square matrix becomes:

$$
\begin{array}{cc|ccccc}
 & & \multicolumn{5}{c}{\text{Classes in column } L+1} \\
 & & 1 & 2 & 3 & \cdots & K \\
\hline
\text{Classes in Column } l & 1 & p_{11}^{l} & p_{12}^{l} & p_{13}^{l} & \cdots & p_{1K}^{l} \\
 & 2 & p_{21}^{l} & p_{22}^{l} & p_{23}^{l} & \cdots & p_{2K}^{l} \\
 & 3 & p_{31}^{l} & p_{32}^{l} & p_{33}^{l} & \cdots & p_{3K}^{l} \\
 & \vdots & \vdots & \vdots & \vdots & & \vdots \\
 & K & p_{K1}^{l} & p_{K2}^{l} & p_{K3}^{l} & \cdots & p_{KK}^{l}
\end{array}
$$

(M3)

To produce a forecast for a particular time period t using the classes in last L previous periods $(\breve{y}_{t-1}, \breve{y}_{t-2}, \ldots, \breve{y}_{t-L})$, the analyst first presents the following indices for each class k , D_t^k:

$$D_t^k = \sum\nolimits_{\forall l} p^{l}_{\breve{y}_{t-(L+1-l)}2}$$

(E2)

Finally, the forecasted class is the one with the largest index value, or the classes tied for the largest index value:

$$\breve{y}_t = \left\{ k_0 \middle| max_{\forall k}\{D_t^k\} \right\}$$

(E3)

Comparative Performance Results

Two baseline methods were chosen to compare the performance of the proposed method: moving average and simple exponential smoothing. For time series forecasting, an analyst has generally no control over the number of time periods available, M, and must decide upon how many previous periods, L, are relevant to generate a forecast. For this comparison, M was kept to 20 time periods and L to three lags. For the proposed method, the number of classes, K, was set to five. All of the time series in the comparison were linearly transformed to fall in the range between 0 and 100, so the size of each class is 20 units. Furthermore, in order to describe the complexity of each series, a second-order polynomial is fit and the coefficient of correlation, R^2 is reported. The rationale for this characterization is that time series with complex behavior will correspond to lower values of R^2. **Table 1** shows the comparative results using the number of correct forecasts based on the $M - L = 20 - 3 = 17$ time periods in the training matrix. The baseline methods were used (as their original design intended) to generate single-valued forecasts that were further translated into classes. This procedure allowed comparing them using the $K = 5$ classes scheme. **Figure 1** shows a sample of three time series (Series 1, 15 and 21 in **Table 1**) scaled to fall between 0 and 100, and their corresponding classes. These three series also show the complexity of the series is here measure through the R^2, as described previously.

The results for the simple exponential smoothing method were obtained using Minitab® 19.2020.1 (64-bit), with automatic optimal smoothing weight selection. The results for both the proposed class forecasting method and the moving average were obtained using MS Excel.

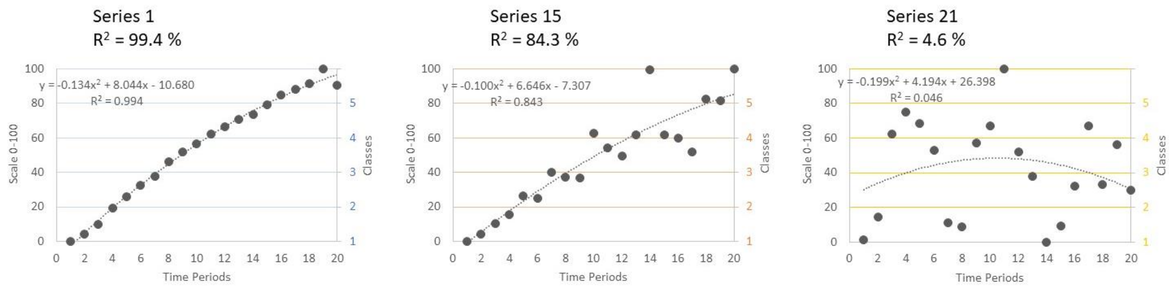

Figure 1. Time series 1, 15 and 21 scaled to fall between 0 and 100 and their respective classification schemes with 5 classes. Increasing behavior complexity is shown with decreasing R^2 value.

Series	R^2 when the series is fitted with a 2nd order polynomial	Number of Time Periods	Number of Lags	Number of Classes	Class Forecasting Method Evaluated using 17 time periods		Exponential Smoothing Evaluated using 17 time periods		Moving Average Evaluated using 17 time periods	
		M	L	K	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
1	0.994	20	3	5	15	2	15	2	9	8
2	0.993	20	3	5	14	3	13	4	11	6
3	0.984	20	3	5	14	3	16	1	9	8
4	0.979	20	3	5	15	2	15	2	10	7
5	0.975	20	3	5	14	3	13	4	10	7
6	0.967	20	3	5	12	5	10	7	8	9
7	0.963	20	3	5	14	3	12	5	9	8
8	0.940	20	3	5	15	2	11	6	8	9
9	0.921	20	3	5	15	2	11	6	8	9
10	0.920	20	3	5	13	4	12	5	10	7
11	0.919	20	3	5	14	3	10	7	9	8
12	0.883	20	3	5	14	3	13	4	10	7
13	0.873	20	3	5	13	4	9	8	7	10
14	0.868	20	3	5	14	3	11	6	9	8
15	0.843	20	3	5	12	5	6	11	6	11
16	0.514	20	3	5	12	5	6	11	7	10
17	0.505	20	3	5	9	8	8	9	8	9
18	0.501	20	3	5	12	5	2	15	3	14
19	0.184	20	3	5	10	7	8	9	7	10
20	0.086	20	3	5	10	7	3	14	6	11
21	0.046	20	3	5	15	2	4	13	2	15
				Minimum	9		2		2	
				First Quartile	12		7		7	
				Median	14		11		8	
				Third Quartile	14.5		13		9.5	
				Maximum	15		16		11	

Table 1. Comparative results of the proposed Class Forecasting Method with Exponential Smoothing and Moving Average. The values with a gray shade indicate the best performance between the three methods. The 21 series are sorted in increasing order of complexity (second column). Descriptive statistics are presented at the bottom of the table.

According to the results in **Table 1**, the class forecasting method outperformed both the simple exponential smoothing and the moving average: it had the largest count of correctly forecasted classes in 20 out of 21 series. Only in series 3, the exponential smoothing obtained a better result, and only in series

1 this method tied the performance of the class forecasting method. These two cases happened when the series´ behavior followed a second order polynomial with R^2 values above 98%. The moving average showed the poorest results in every case. In series with high R^2 values, the performance of the baseline methods declined sharply, while that of the proposed method remained competitive. The descriptive statistics at the bottom of the table show the lesser dispersion of the results in the class forecasting method and summarize the better location of its performance when described by the five-point summary (minimum, first quartile, median, third quartile, and maximum).

The next step included the use of inferential statistics to assess the potential generalization of these results. An ANOVA using the three methods as the fixed levels of a controllable factor and the set of R^2 values as a covariate was carried out on the number of correct class forecasts as a response. Using a Box-Cox transformation (λ=1.35) and a significance of 5%, the residuals passed the Anderson-Darling test for normality (p-value = 6.8%) and a multiple comparisons test for equal variances (p-value = 9.8%). The ANOVA showed that both the controllable factor and the covariate were significant at the stated level (p-values < 0.1%). Subsequently, a Tukey pairwise comparison procedure at 95% confidence was used to determine the ranking of the methods based on their forecasting performance. The results placed the proposed class forecasting method first (95%CI difference of means vs simple exponential smoothing (-4.77,-1.70) and vs moving average (-6.77, -3.70)), followed by the simple exponential smoothing method in second place (95%CI difference of means vs moving average (-3.53, -0.46)) , and the moving average in third place.

The previous ANOVA procedure estimated a standard deviation of 2.07 units in the residuals. A power curve for one-way ANOVA was used to determine that the sample size of 21 series, using said standard deviation value, a difference to be estimated of the size of at least such standard deviation, and a significance of 5%, would provide a power of 81.47%. This was deemed an acceptable probability of correctly rejecting the null hypothesis in the ANOVA-based comparative analysis presented previously.

Application

The proposed classification forecasting method has the potential for application across multiple disciplines. Basing a forecast on historical transition frequencies between classes provides repeatability and transparency. It has been also shown that its forecasting performance is robust and competitive. In this section, the method is applied to analyze a time series related to Covid19 cases in Puerto Rico to further demonstrate its capabilities.

The number of positive Covid19 Cases in Puerto Rico were collected from March 2020 to August 2022 from the Department of Health website [9]. This resulted in $M = 30$ time periods (months). A classification scheme with $K = 8$ classes was adopted while the number of lags was kept to $L = 3$. The training matrix, T, contained $M - L = 30 - 3 = 27$ vectors. A second order polynomial fit gave out a value of R^2=24.65%. For comparison purposes, the same baseline methods were used as in the comparative results section. **Figure 2** shows the original series (top) and the classification results for the three methods using the 27 vectors in the training matrix (bottom). Following the comparison results, the proposed class forecasting method showed a better classification performance (19/27 correct forecasts or $\sim 70\%$) than the baseline methods (exponential smoothing: 10/27 or $\sim 37\%$ and moving average: 5/27 or $\sim 19\%$). Contrasting the original series and the comparison results, it is clear that none of the methods could forecast the spike in Covid19 cases observed on months 22 and 23. After these, however, the proposed class forecasting method performed better than the baseline methods, correctly forecasting 6 out of the following 7 periods (versus 2 and 1 correct forecasts out of 7 for the simple exponential smoothing and the moving average methods respectively).

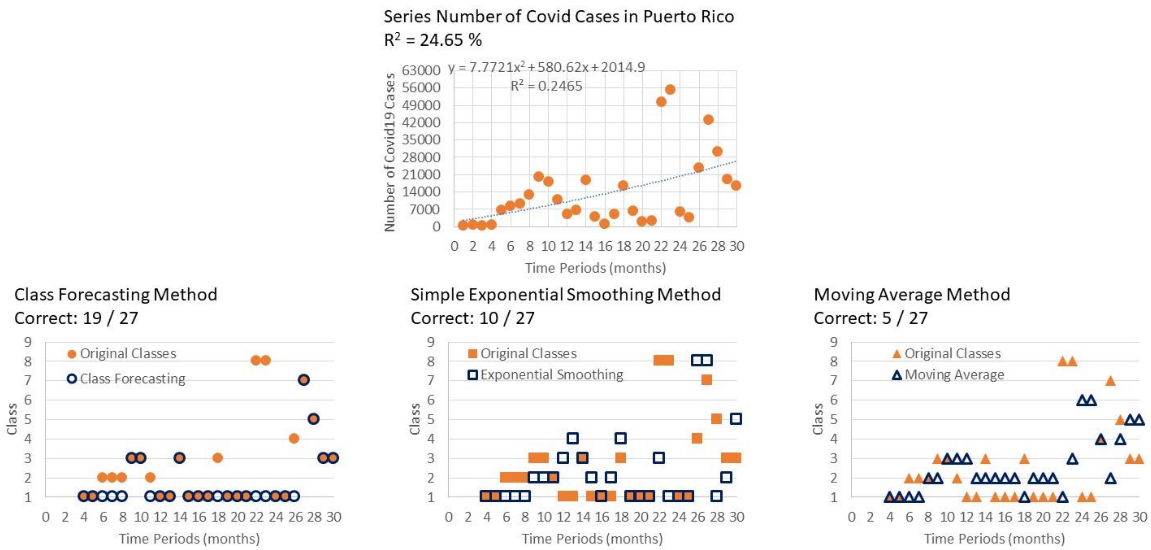

Figure 2. Original time series of Covid19 Cases in Puerto Rico (top). Its R^2 value is 24.65% when fitting a second order polynomial. A comparison of the three methods is presented (bottom): Class Forecasting as proposed, Simple Exponential Smoothing and Moving Average. On each of the the bottom graphs, an empty symbol is a missed forecast. All methods used $M = 30\ months, L = 3\ lags, and\ K = 8\ classes$. Class Forecasting had 19 correct forecasts out of 27, Exponential Smoothing had 10 out of 27, and Moving Average had 5 out of 27.

A fractal scheme for defining the number of classes, K

The proposed Class Forecasting method leaves it to the user to define the number of classes, K. Some users might have a number of classes already defined for their particular application, however, others might not. In the latter case, a possible way involves defining the number of classes by consecutively partitioning the range of the random variable in the time series of interest by half on each step. A fractal tree with $K = 2^H$ categories results from this operation (**Figure 3**). Each partition $h = 1,2,3 ... H$, can be associated to a binary variable, B_h, to indicate if the value of the random variable is in the upper half of the $h-th$ partition ($B_h = 1$) or not ($B_h = 0$). This naturally leads to the notation $B_1B_2B_3 ... B_H$, which conveniently locates the category to which a particular observation belongs. For example, a class 101 would locate a class in the upper half, lower quarter, and upper eight of the range of the random variable in the time series of interest. This scheme provides a meaningful characterization, although it is clear that the number of classes will grow exponentially. The final number of partitions, H, is decided by the user looking for the minimum value that provides class sizes with a desired level of precision. A lower bound for a particular class, k, is accessible through the following expression:

$$\breve{y}_{low}^{k} = y_{min} + (y_{max} - y_{min}) \sum_{h=1}^{H} \left(\frac{B_h}{B^h}\right) \tag{E4}$$

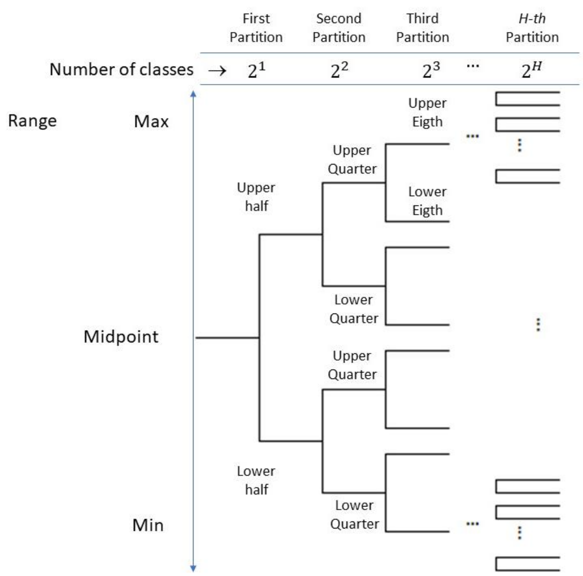

Figure 3. A fractal tree used to determine classes. From left to right, each partition consecutively divides the previous one by half.

Conclusions

Forecasting methods that are transparent in their logic with competitive results that can be explained to people with different backgrounds are in demand. This is especially true after a period like the Covid19 pandemic, when model complexity was not necessarily helpful. In this work, a method to forecast a particular class based on transition frequency is proposed for time series. The method has an intuitive rationale and its competitive performance when compared to moving average and simple exponential smoothing is evidenced in this manuscript. A scheme based on a fractal binary tree is also introduced to help users define a sensible number of classes. Future work in this research line will include coding of the method and widening its evaluation to include different number of lags.

References

[1] F. Petropoulos, D. Apiletti, V. Assimakopoulos, M. Zied Babai, D. K. Barrow, S. Ben Taieb, C. Bergmeir, R. J. Bessa, J. Bijak, J. E. Boylan, J. Browell, C. Carnevale, J. L. Castle, P. Cirillo, M. P. Clements, C.Cordeiro, F. L. Cyrino Oliveira, S. De Baets, A. Dokumentov, J.Ellison, P. Fiszeder, P. H. Franses, D. T. Frazier, M. Gilliland, M. S. Gönül, P. Goodwin, L. Grossi, Y. Grushka-Cockayne, M. Guidolin, M. Guidolin, U. Gunter, X. Guo, R. Guseo, N. Harvey, A. F. Hendry, R. Hollyman, T. Januschowski, J. Jeon, V. Richmond R. Jose, Y. Kang, A. B. Koehler, S. Kolassa, N. Kourentzes, S. Leva, F. Li, K. Litsiou, S. Makridakis, G. M. Martin, A. B. Martinez, S. Meeran, T. Modis, K. Nikolopoulos, D. Önkal, A. Paccagnini, A. Panagiotelis, I. Panapakidis, J.M. Pavía, M. Pedio, D. J. Pedregal, P. Pinson, P. Ramos, D. E. Rapach, J. J. Reade, B. Rostami-Tabar, M. Rubaszek, G. Sermpinis, H. Lin Shang, E. Spiliotis, A. A. Syntetos, P.Dilini Talagala, T.S. Talagala, L. Tashman, D. Thomakos, T. Thorarinsdottir, E. Todini, J. R. Trapero Arenas, X. Wang, R.L. Winkler, A. Yusupova, F. Ziel, **Forecasting: theory and practice** (2022) International Journal of Forecasting 38:3, 705-871, https://doi.org/10.1016/j.ijforecast.2021.11.001.

[2] G. Lee , S-e Yoon, K. Shin Simple epidemic models with segmentation can be better than complex ones (2022) PLoS ONE 17:1, e0262244. https://doi.org/10.1371/journal.pone.0262244

[3] K.C. Green, J.S.Armstrong, Simple versus complex forecasting: The evidence (2015) Journal of Business Research, 68:8, 1678-1685, https://doi.org/10.1016/j.jbusres.2015.03.026

[4] Ricardo P. Masini, Marcelo C. Medeiros, Eduardo F. Mendes, Machine learning advances for time series forecasting (2021) Journal of Economic Surveys, 37:1, 76-111, https://doi.org/10.1111/joes.12429

[5] H. Wang, K.O. Kwok, S. Riley, Forecasting influenza incidence as an ordinal variable using machine learning (2023) Preprint available at medRxiv 2023.02.09.23285705, doi: https://doi.org/10.1101/2023.02.09.23285705

[6] F. Petropoulos, Y. Grushka-Cockayne, E. Siemsen, E. Spiliotis, Wielding Occam's razor: Fast and frugal retail forecasting (2022) Preprint available at http://dx.doi.org/10.2139/ssrn.3792565

[7] Z. Zhou, C. Zhao, Y.Huang, Nonlinear time series analysis of limestone rock failure process, (2023), Measurement, 206, 112259, https://doi.org/10.1016/j.measurement.2022.112259

[8] T.S. Wallsten, Y. Shlomi, C. Nataf, T. Tomlinson, Efficiently encoding and modeling subjective probability distributions for quantitative variables, (2016), Decision, 3:3, 169–189 https://doi.org/10.1037/dec0000047

[9] Puerto Rico Department of Health Website [https://www.salud.gov.pr/CMS/142] Consulted on November 3, 2022