



Statistical Analysis for Long-Term Weather Forecast [†]

Dimitrios Kampolis

School of Applied Arts and Sustainable Design, Hellenic Open University, 26335 Patra, Greece; dkabolis@gmail.com

[†] Presented at the 16th International Conference on Meteorology, Climatology and Atmospheric Physics—COMECAP 2023, Athens, Greece, 25–29 September 2023.

Abstract: A weather forecast is a result of applying science and technology to predict the conditions of the atmosphere in a selected location and time in the future. The main input is a collection of data (atmospheric, land, and ocean), and the resulting output is meteorology (how atmospheric conditions will change). People have been trying to predict the weather by observing nature for thousands of years, but in a scientific way only since the 19th century. First manually and based mostly on changes in barometric pressure, and later in modern times with the contribution of computer-based models (numerical weather prediction). Despite the high inaccuracy of numerical weather prediction beyond 10 days, the interest in long-term weather forecasts is very high due to social reasons—energy sector, civil protection, etc.—and the scientific effort is constant. Here, we propose a statistical weather model for long-term weather forecasts based on weather/climate data time series. We will analyze atmospheric data in 850 hPa for a period of 35 years, resulting in temperature ensembles and temperature deviations for specific periods. Finally, we will contrast the results of the statistical weather model (prediction) with the real data to check the accuracy of the model.

Keywords: long-term weather forecast; seasonal forecast; statistical analysis; temperature deviation; statistical model

1. Introduction

A long-term weather forecast is the Holy Grail of meteorology. Supercomputers and numerical methods [1] are combined to improve weather forecasts beyond 10 days. Despite the applied high science and technology, results are poor due to fundamental obstacles in numerical forecast [2]. Therefore, long-term weather forecast is produced mostly from models based on statistical analysis.

Statistical analysis is the in-depth study of a problem using calculations, tables, and charts in order to produce statistical conclusions [3]. The presentation of the primary statistical material is called classification [4], with geometrical classification and time series classification as the most important subsets.

Time series classification (TSC) [5] is the tracking of the evolution of a variable over time. Time series are described with one or more of the following features:

A long-term trend.

A periodical trend.

Some irregular or random variations.

Statistical analysis of a time series is recommended to describe the regularity that may exist between successive values of a variable having as a purpose to predict the future behavior of that series.

2. Method

The method is based on a time-series analysis of temperature data at 850 hPa. The layer of 850 hPa is selected because it is not affected by surface conditions like the Foehn effect [6], temperature inversion [7], or urban heat island effect [8].



Citation: Kampolis, D. Statistical Analysis for Long-Term Weather Forecast. *Environ. Sci. Proc.* **2023**, *26*, 30. <https://doi.org/10.3390/environsciproc2023026030>

Academic Editors: Konstantinos Moustiris and Panagiotis Nastos

Published: 23 August 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Ten major administrative regions of Greece (Thrace, Macedonia, Epirus, Thessaly, Ionian islands/West Greece, Central Greece/Attica/Euboea, Peloponnese, North Aegean, Crete, and South Aegean) are selected, representing the grids of the study. For each grid, a dataset of daily temperatures in 850 hPa for the last 35 years is downloaded from the Physical Science Laboratory, NOAA [9], and registered in a database.

Since the registration is completed, the database is simplified in 5-day periods and interfaced with the formula of the model.

The formula is a machine learning algorithm based on climate and astronomical cycles—El Nino, La Nina, and solar cycles—and statistical weights, sums, and possibilities. It is described by four main processes:

1. Calculation of average temperatures in periods of 35, 22, 9, and 7 years (A35, A22, A9, and A7 values) (Figure 1);

MARCH	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
March 1-5	-1.3	4.2	2.8	1.3	4.7	3.3	7.9	6.4	-2.6	3.6	6.5	6.6	2.0	7.1	12.2	0.4	3.5	4.1	4.2	6.2	8.6	5.4	9.1	5.2	0.2	2.6	3.8	4.7	4.9	6.7	7.7	3.7	3.0	3.1	-0.7
March 6-10	2.2	1.4	4.6	7.9	-1.1	-3.5	0.4	1.9	-3.5	-0.1	6.8	5.3	0.7	6.5	12.5	1.2	-0.4	1.2	1.0	6.9	5.6	3.6	5.8	-2.8	4.6	5.3	1.5	2.2	6.6	3.8	6.7	8.1	3.1	3.6	-0.1
March 11-15	1.2	2.5	4.9	3.4	0.7	-1.6	2.5	1.5	2.1	3.0	0.8	5.7	3.5	6.6	4.9	2.6	2.0	1.3	2.4	2.4	5.1	0.6	3.6	4.9	1.1	8.8	2.7	1.2	3.6	1.3	6.3	4.6	3.3	4.2	-5.5
March 16-20	3.1	9.6	5.3	6.4	-2.0	2.2	6.0	3.7	1.9	3.1	-3.4	3.0	-0.5	11.0	5.6	-1.2	5.8	7.6	2.1	5.0	8.2	0.1	1.1	7.8	3.7	3.5	6.4	0.7	3.0	5.5	9.3	7.3	3.8	1.5	-2.6
March 21-25	5.1	9.3	8.1	10.5	1.3	10.0	6.7	-1.3	-1.3	-2.6	-3.3	1.0	1.4	11.7	4.0	-3.8	7.2	4.2	7.2	4.6	8.7	1.2	6.1	1.5	7.0	7.4	6.3	5.9	8.8	8.6	6.0	5.6	5.5	2.3	-1.0
March 26-31	2.8	5.5	7.6	9.2	4.3	6.1	5.3	4.5	5.2	2.7	-0.5	5.5	9.6	13.4	-1.1	4.3	9.9	9.3	7.5	5.0	3.7	5.5	7.5	6.2	5.2	8.7	5.8	6.1	4.5	6.2	6.6	3.5	6.1	1.1	5.5
	2.17	5.43	5.55	6.46	1.32	2.76	4.79	2.80	0.29	1.60	1.16	4.52	2.79	9.38	6.35	0.59	4.64	4.61	4.07	5.01	6.66	2.73	5.87	3.81	3.63	6.04	4.74	3.47	5.23	5.36	7.10	5.47	4.13	2.65	-0.73

Figure 1. Average temperatures in 850 hPa.

2. Comparison of last period temperatures with A35 values—0 for lower values and 1 for higher values (B values) (Figure 2);

	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	
March 1-5	0	0	0	0	1	0	1	1	0	0	1	1	0	1	1	0	0	0	0	1	1	1	1	1	0	0	0	1	1	1	1	1	0	0	0	0
March 6-10	0	0	1	1	0	0	0	0	0	0	1	1	0	1	1	0	0	0	0	1	1	1	1	0	1	0	0	1	1	1	1	1	1	0	0	0
March 11-15	0	0	1	1	0	0	0	0	0	1	0	1	1	1	0	0	0	0	0	1	0	1	1	0	1	0	0	1	0	1	1	1	1	1	0	0
March 16-20	0	1	1	1	0	0	1	0	0	0	0	0	1	1	0	1	1	0	1	1	0	1	1	0	0	1	0	1	0	1	1	1	1	1	0	0
March 21-25	1	1	1	0	1	1	0	0	0	0	0	0	1	0	0	1	0	1	0	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	0	0
March 26-31	0	0	1	1	0	1	0	0	0	0	0	0	1	1	0	0	1	1	0	0	0	1	1	0	1	1	1	1	0	1	1	0	1	1	0	0

Figure 2. B values (deviation to 35 years' average temperature).

3. Calculation of average B values in periods of 22, 9, and 7 years (C22, C9, and C7 values—grey) and calculation of corresponding D values (statistical weights) (D22, D9, and D7 values—red) (Figure 3);

0.48	1	0.20	1	0.00	1
0.57	0	0.60	0	0.33	1
0.52	0	0.80	0	0.67	0
0.48	1	0.40	1	0.00	1
0.61	0	0.60	0	0.33	1
0.52	0	0.40	1	0.33	1

Figure 3. C values and D values.

4. Calculation of prediction values based on the formula:

[A35 value] + or – (depending on the D22 value) [ABS of A22 value] + or – (depending on the D9 value) [ABS of A9 value] + or – (depending on the D7 value) [ABS of A22 value]

For each season, the model produces 18 prediction values (6 per month), resulting in a temperature ensemble for each studied region (Figure 4).

Predictions proceed into a second-level analysis resulting in temperature deviations (Figure 5), and they are plotted in weather maps (Figure 6).

The total amount of information (temperature ensembles, temperature deviations, and weather maps) is combined for translating data into text. Short paragraphs are published with reference to possible dates with significant changes in weather circulation.

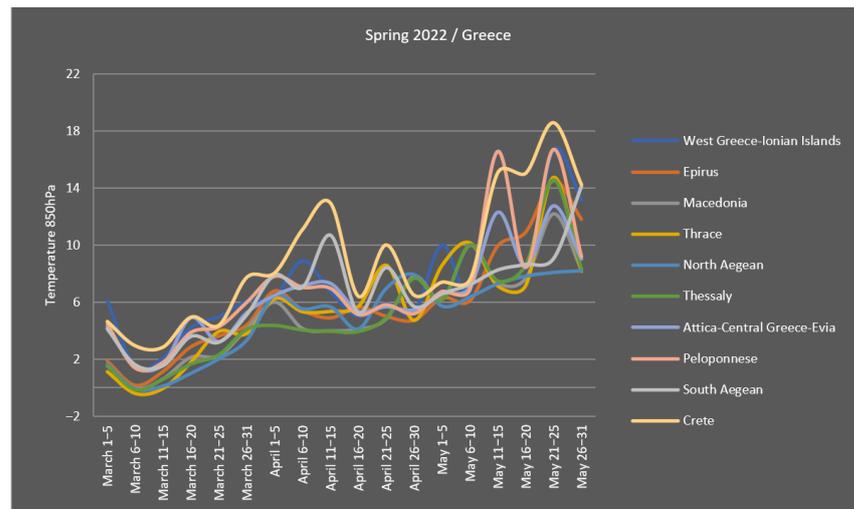


Figure 4. Temperature ensembles for a predicted period [10].

Probability (%)	West Greece	Epirus	Macedonia	Thrace	North Aegean	Thessaly	Central Greece	Peloponnese	South Aegean	Crete
March 1-5	1.1	0.4	-0.5	-0.4	-0.1	-0.5	1.0	1.0	-0.2	-0.8
March 6-10	-1.0	0.4	-0.5	0.5	-0.5	-0.5	-0.6	-1.1	-0.6	-1.7
March 11-15	-0.4	-0.2	-0.3	0.1	-0.1	-0.3	-0.6	-0.6	-0.6	-1.1
March 16-20	-0.2	-0.3	0.4	0.1	-0.1	-0.4	0.9	-0.3	0.4	1.3
March 21-25	-0.2	-0.2	-0.8	-0.3	-0.6	-0.8	-0.7	-0.2	-0.8	-0.9
March 26-31	0.1	0.1	0.1	0.0	-0.9	0.0	0.1	0.1	0.0	0.1
April 1-5	-2.1	1.5	1.4	1.5	1.7	1.6	-2.3	-2.2	-2.4	2.1
April 6-10	2.1	-1.6	-0.1	-0.1	-1.8	-0.1	-2.2	0	-2.4	2.9
April 11-15	-0.1	2.0	-0.8	-0.2	-0.2	-0.8	-0.2	-0.1	2.3	2.9
April 16-20	-2.6	1.8	-0.2	2.0	-2.1	-0.1	-2.7	-2.6	-2.8	-3.6
April 21-25	-3.0	-2.3	0.1	2.0	-2.4	-2.3	-3.0	-3.0	-0.2	-0.3
April 26-30	-4.0	-0.1	-3.2	0.0	-0.1	-3.3	-4.0	-4.1	-0.1	-4.5
May 1-5	-3.8	-3.1	-3.2	0.1	-3.4	-3.2	-3.9	-3.9	-4.1	-4.6
May 6-10	-3.5	-3.0	-0.3	2.8	-0.3	-0.3	-3.7	-3.6	-3.7	-4.4
May 11-15	4.1	-3.5	-0.9	3.6	-0.1	-0.9	1.1	4.0	-0.1	-0.1
May 16-20	-3.9	-3.3	-3.4	-0.7	-3.6	0.2	-4.1	-4.0	-4.3	-0.3
May 21-25	-0.3	-0.4	3.5	-0.2	-3.6	3.6	-4.1	-0.3	-0.3	4.4
May 26-31	-4.2	0.1	-3.5	-3.7	-3.8	-3.6	-4.2	-4.2	-4.3	0.5

Figure 5. Deviation table. Gradient color scale from -30 degrees Celsius (negative deviation: deep blue to light blue) to +30 degrees Celsius (positive deviation: light red to deep red) [10].

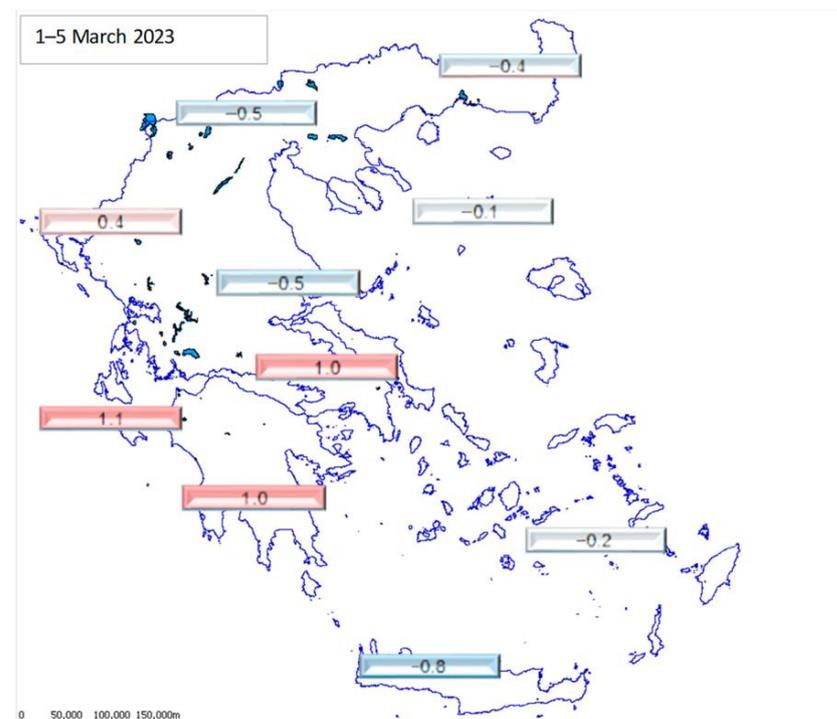


Figure 6. Plotting of deviation (5-day periods). Gradient color scale from -30 degrees Celsius (negative deviation: deep blue to light blue) to +30 degrees Celsius (positive deviation: light red to deep red) [10].

3. Results

We will examine the accuracy of the model for the period 1 December 2021 till 30 November 2022 in terms of deviation trend, separated into four (4) seasons and with respect to the grid of Attica.

We start with the registration of the predicted deviation values and real deviation values downloaded from the Physicals Science Laboratory, NOAA [9]. Since we built the dataset, we proceeded to the transformation of values—1 for positive values and 0 for negative values. Identical values result in 1, and non-identical values in 0 (Figure 7).

Attica	Deviation (P)	Deviation (R)	Prediction	Real	Coefficient
December 1–5	0.0	-0.9 0	0	1	
December 6–10	-1.6	0.1 0	1	0	
December 11–15	-1.3	-0.6 0	0	1	
December 16–20	-1.1	-5.1 0	0	1	
December 21–25	0.0	-4.0 1	0	0	
December 26–31	0.0	2.2 1	1	1	
January 1–5	0.5	5.1 1	1	1	
January 6–10	0.0	2.0 1	1	1	
January 11–15	0.8	-3.2 1	0	0	
January 16–20	0.3	-4.3 1	0	0	
January 21–25	-0.6	-7.1 0	0	1	
January 26–31	-0.7	-4.9 0	0	1	
February 1–5	-1.7	-0.9 0	0	1	
February 6–10	-0.2	-0.8 0	0	1	
February 11–15	0.6	1.9 1	1	1	
February 16–20	0.2	4.6 1	1	1	
February 21–25	-0.5	0.4 0	1	0	
February 26–29	-1.1	-2.5 0	0	1	
March 1–5	-0.1	-5.1 0	0	1	
March 6–10	-1.4	-3.0 0	0	1	
March 11–15	-1.1	-8.3 0	0	1	
March 16–20	0.9	-6.5 1	0	0	
March 21–25	-1.6	-5.8 0	0	1	
March 26–31	-0.6	-0.3 0	0	1	
April 1–5	0.1	2.8 1	1	1	
April 6–10	0.0	3.6 1	1	1	
April 11–15	-1.0	-3.1 0	0	1	
April 16–20	-2.6	-0.9 0	0	1	
April 21–25	-2.7	4.2 0	1	0	
April 26–30	-4.0	3.5 0	1	0	
May 1–5	-4.1	-2.8 0	0	1	
May 6–10	-3.7	-0.8 0	0	1	
May 11–15	-0.1	-1.3 0	0	1	
May 16–20	-4.2	-0.7 0	0	1	
May 21–25	-0.2	3.1 0	1	0	
May 26–31	-4.0	5.7 0	1	0	
June 1–5	-0.1	3.4 0	1	0	
June 6–10	-0.3	1.3 0	1	0	
June 11–15	4.6	-1.3 1	0	0	
June 16–20	0.1	-0.9 1	0	0	
June 21–25	-0.2	2.1 0	1	0	
June 26–30	-5.5	-0.3 0	0	1	
July 1–5	-0.2	1.2 0	1	0	
July 6–10	0.3	-0.9 1	0	0	
July 11–15	-1.6	-1.8 0	0	1	
July 16–20	0.2	0.2 1	1	1	
July 21–25	4.1	1.9 1	1	1	
July 26–31	-6.2	3.2 0	1	0	
August 1–5	-6.5	0.8 0	1	0	
August 6–10	-0.2	0.1 0	1	0	
August 11–15	-6.1	-0.9 0	0	1	
August 16–20	-5.8	2.3 0	1	0	
August 21–25	-5.9	-1.2 0	0	1	
August 26–31	-5.9	-0.5 0	0	1	
September 1–5	-5.6	-0.5 0	0	1	
September 6–10	-5.1	-1.7 0	0	1	
September 11–15	-0.2	-0.9 0	0	1	
September 16–20	-5.1	4.1 0	1	0	
September 21–25	-4.6	0.3 0	1	0	
September 26–30	0.8	2.1 1	1	1	
October 1–5	-4.4	-1.5 0	0	1	
October 6–10	-3.0	-0.2 0	0	1	
October 11–15	-4.0	-1.3 0	0	1	
October 16–20	-4.0	-3.2 0	0	1	
October 21–25	-3.5	-0.9 0	0	1	
October 26–31	-3.3	-1.4 0	0	1	
November 1–5	-2.1	3.1 0	1	0	
November 6–10	-3.0	6.6 0	1	0	
November 11–15	2.5	1.1 1	1	1	
November 16–20	0.0	-1.6 1	0	0	
November 21–25	-2.0	0.4 0	1	0	
November 26–30	-1.9	2.1 0	1	0	

Figure 7. Predicted deviation vs. real deviation and coefficient values (per season). Gradient color scale from -30 degrees Celsius (negative deviation: deep blue to light blue) to +30 degrees Celcius (positive deviation: light red to deep red).

According to the resulting coefficient, the accuracy of each season is the following:

- Winter: 72.2%;
- Spring: 72.2%;
- Summer: 38.9%;
- Autumn: 61.1%.

We observe that winter, spring, and summer meet medium to high scores, while summer meets low scores. Analyzing the accuracy of each month, we have the following results:

- December: 66.7%;
- January: 66.7%;
- February: 83.3%;
- March: 83.3%;
- April: 66.7%;
- May: 66.7%;
- June: 16.7%;
- July: 50%;
- August: 50%;
- September: 66.7%;
- October: 100%;

- November: 16.7%.

Among twelve months, three months meet high scores (February, March, October), five months meet higher-than-average scores (December, January, April, May, September), two months meet average scores (July and August), and two months meet low scores (June and November).

The above results show a weakness in algorithmic function during the summer period, as well as for November. In contrast, from December to May (as well as September), the results show a strong algorithmic functionality, while October meets an exceptional score.

The dysfunctionality of the summer period is under investigation. A mismatch in the order of D and A values in the formula has been detected, it is corrected, and it will be reevaluated in the next seasonal prediction (published on 19 May 2023) [11].

Further research is required, involving more seasons/years of predictions to detect specific dysfunctions of the algorithm and to confirm the good results as well. Additionally, the method should be applied in more grids, and globally if possible, to ensure total application (ongoing). A final and more complicated step is the application of the method with temperature data in 500 hPa and the combination of resulting values in 850 hPa and 500 hPa for a more accurate prediction.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The statistical forecast is published once per season at <https://www.youweather.com> (accessed on 5 May 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Numerical Weather Prediction (Weather Models). Available online: <https://www.weather.gov/media/ajk/brochures/NumericalWeatherPrediction.pdf> (accessed on 12 April 2023).
2. Mailier, P. Can We Trust Long-Range Weather Forecasts? In *Management of Weather and Climate Risk in the Energy Industry*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 227–239.
3. Ali, Z.; Bhaskar, S.B. Basic statistical tools in research and data analysis. *Indian J. Anaesth.* **2016**, *60*, 662–669. [CrossRef] [PubMed]
4. Data classification analysis. Available online: <https://www.ibm.com/docs/en/iis/11.7?topic=dcao-data-classification-analysis> (accessed on 13 April 2023).
5. Fawaz, H.; Forestier, G.; Weber, J.; Idoumghar, L.; Muller, P. Deep learning for time series classification: A review. *Data Min. Knowl. Discov.* **2019**, *33*, 917–963. [CrossRef]
6. Foehn Effect. Available online: <https://www.metoffice.gov.uk/weather/learn-about/weather/types-of-weather/wind/foehn-effect> (accessed on 22 April 2023).
7. What Is a Temperature Inversion? Available online: <https://www.metoffice.gov.uk/weather/learn-about/weather/types-of-weather/temperature/temperature-inversion> (accessed on 22 April 2023).
8. Urban Climate Impacts. Available online: <https://www.metoffice.gov.uk/research/climate/climate-impacts/urban> (accessed on 22 April 2023).
9. Physical Science Laboratory. Available online: <https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html> (accessed on 28 May 2023).
10. YouWeather Statistical Model (YSM)—Άνοιξη 2022 (Ελλάδα). Available online: <https://www.youweather.com/el-gr/article/o-kairos-tin-anoixi-2022-stin-ellada.html> (accessed on 5 May 2023).
11. YouWeather. Available online: <https://www.youweather.com/el-gr/article/ysm-o-kairos-to-kalokairi-2023-stin-ellada.html> (accessed on 5 May 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.