Scientific Papers. Series B, Horticulture. Vol. LXVIII, No. 1, 2024
Print ISSN 2285-5653, CD-ROM ISSN 2285-5661, Online ISSN 2286-1580, ISSN-L 2285-5653

FROST EVENTS FORECAST USING MACHINE LEARNING IN BULGARIA

Boryana TSENOVA, Veska GEORGIEVA, Metodi DINEV

National Institute of Meteorology and Hydrology, 66 Tsarigradsko chosse, Sofia, Bulgaria

Corresponding author email: boryana.tsenova@meteo.bg

Abstract

In the present study a scheme for damaging frost occurrence forecast in Bulgaria is presented. It is based on Random Forest technique and uses the regional numerical weather prediction (NWP) model ALADIN output as predictor. Initially, the statistical model is trained with measured data with three-hours frequency at 5 representative meteorological stations in Bulgaria during April and May for the period 1991-2020. Using parameters from the regional NWP model production as predictors gives possibility to forecast frost probability 72 hours ahead. The performance of the scheme is evaluated. Results for 27 synoptic stations during April (2021-2023) show a probability of detection above 0.85 and a false alarm rate below 0.1 independently of the remoteness of the forecast. Most of the considered cases were correctly discriminated by the scheme as "frost" and "non-frost" cases, which would not be the case if only considering the forecasted minimum temperature. Our results show that frost could be forecasted by the presented scheme 3 days before its occurrence, which should be enough to react to minimize damage caused in the agricultural sector.

Key words*: frost, forecast, Random Forest, Bulgaria.*

INTRODUCTION

Increasing tendencies of the average monthly air temperatures in Bulgaria during the last decades are established. The indicated trends of changing temperature conditions are unfavourable for the fruit species, especially the early flowering drupe species. Warming in winter, and especially in January and February, cause premature development of fruit trees and subsequent frosts cause damage to varying degrees, depending on their development. Damage from late spring frosts is a major limiting factor for fruit production in Bulgaria. Frost events are locale and are difficult to be predicted. They are associated, among other factors, with the type of soil, terrain orientation, and the damage caused to plants depends on its intensity and duration. There are several techniques to minimize damage caused by frost in the agricultural sector (De Melo-Abreu et al., 2016), however, it is necessary to know in advance the likelihood for occurrence of this phenomenon. Many warning systems for frost have been developed based on predictions from numerical weather prediction (NWP) models (Prabha and Hoogenboom 2008), vegetation indices estimated by satellites (Gabbrielli et al., 2022), statistical models (Lee et al., 2016), fuzzy logic (Cadenas et al., 2020), neural networks (Fuentes et al., 2018; Ding et al., 2020) and statistical indices (Anandhi et al., 2013; Rozante et al., 2019). In recent years, significant advancements have been made regarding the application of machine learning in studies related to frost prediction (Diedrichs et al., 2018, Jamei et al., 2015, Kalaiarasi and Maheswari 2020, Diniz et al., 2021, Ismail et al., 2021, Talsma et al., 2022, Rozante et al., 2023 and others).

In the present study the scheme for damaging frost events probability forecast in Bulgaria is presented. The scheme is based on Machine Learning and more precisely Random Forest and the operational NWP model ALADIN (Termonia et al., 2018) output as statistical model predictors. A Random Forest Algorithm is an extremely popular supervised machine learning algorithm that is used for Classification and Regression problems based on the concept of ensemble learning which is a process of combining multiple classifiers to solve a complex problem and improve the performance of the model. All statistical analyses for the present study were performed using R Statistical Software (R Core Team 2021)

METHODOLOGY

Every 3-hours measured data for the months of April and May for a period of 30 years (1991-2020) at 5 meteorological stations in Bulgaria are used to train and to test the statistical model. The stations Kjustendil, Stara Zagora, Kneja, General Toshevo and Dobrich (marked in red in Figure 1) are chosen as being of the most representative for occurrence of damaging frost on orchards causing severe loss on national fruit production during the last decades. The following parameters were determined by the model as damaging frost predictors: air and dew point temperature and relative humidity at 2 m, mean sea level pressure, wind speed at 10 m and the hour of measurement. In the statistical model, cases with frost were determined in the following way: for dates when damaging frost was registered at the stations (frost days), which usually is during the night time, the periods between 9 UTC at the previous day until 6 UTC of the frost day are considered as "frost related cases" (and are designated by 1, while "non-frost related cases" by 0 in the model). The model performance for the test sample of data is relatively high with a Probability of Detection (POD) of 0.74 and a False Alarm Rate (FAR) of 0.03.

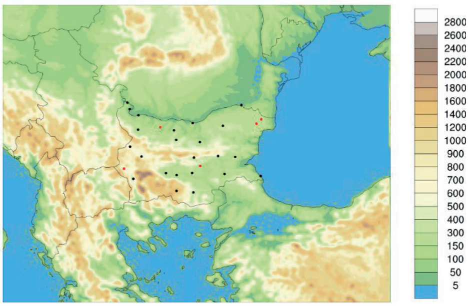

Figure 1. Meteorological stations in Bulgaria from which measurements are used for the present study

Further, for predicting the damaging frost probability, the same parameters are used but taken from the operational NWP model ALADIN. The operational model configuration at NIMH is the following: the integration domain (shown in Figure 1) covers a big part of the Balkan Peninsula, centered on Bulgaria, with a horizontal resolution of 5 km, 105 vertical levels, a time step of 300 s and a forecast range of 72 h. It is run four times daily, at 00, 06, 12 and 18 UTC and it uses the global ARPEGE of Météo-France. To reinforce the impact of night hours forecast (when frost occur) but to take into account also the day hours forecast, the scheme for damaging frost probability was formed in the following subjective way:

- if at least one of the night periods (between 21 and 06 UTC) is determined as a "frost related case" and at least one of the previous day daily periods (between 09 and 18 UTC) is determined as a "frost related case" then the probability of damaging frost occurrence for the day FP =100%;
- if at least one of the night periods (between 21 and 06 UTC) is determined as a "frost related case" and none of the previous day daily periods (between 09 and 18 UTC) is determined as a "frost related case" then the probability of damaging frost occurrence for the day FP = 80%;
- if none of the night periods (between 21 and 06 UTC) is determined as a "frost related case" and at least one of the previous day daily periods (between 09 and 18 UTC) is determined as a "frost related case" then the probability of damaging frost occurrence for the day FP = 50%;
- if none of the night periods (between 21 and 06 UTC) is determined as a "frost related case" and none of the previous day daily periods (between 09 and 18 UTC) is determined as a "frost related case" then the probability of damaging frost occurrence for the day FP = 0%.

RESULTS

Forecast data for the month of April for the 3 years period 2021-2023 were evaluated and verified using the data for damaging frost registered at 27 meteorological stations in Bulgaria shown in Figure 1. For this considered period 133 cases of frost were registered at the considered stations, which are about 10 times less than the "non frost" cases (1296). First, if considering if any damaging frost probability was forecasted (fp>0), Table 1 shows the performance of the scheme for damaging frost detection over the 27 considered synoptic stations. As ALADIN forecast range is 72 h, we have a prediction for 3 days ahead. 1st forecast designates the closest, while 3rd – the farthest. For this study only model run at 06 UTC is considered. It is visible that POD is high, but decreases slightly with the forecast moving away in time (from 0.89 to 0.85 for the 3rd forecast), while FAR is very low, but slightly higher for 1st forecast.

Table 1. April 2021-2023: Probability of Detection (POD) and False Alarm Rate (FAR) of the scheme for damaging frost probability forecast for 1, 2 and 3 days ahead (respectively 1st forecast, 2nd forecast and 3rd forecast)

	1st forecast	2nd forecast	3rd forecast
POD	0.89	0.89	0.86
FAR	0.069	0.066	0.066

From all 133 frost cases, 101 cases (about 76%) were correctly forecasted by all 3 forecasts, while only 4 cases (about 3%) were not predicted by none of the forecasts. 20 cases (about 15%) were correctly forecasted by two of the three forecasts (10 cases are with correct 1st and 2nd forecasts, 4 - with correct 1st and 3rd forecasts, and 6 - with correct 2nd and 3rd forecasts) and 8 cases (about 6%) were correctly forecasted only by one forecast (3 cases with 1st correct forecast, 2 - with 2nd correct forecast and 3 - with 3rd correct forecast).

If considering the determined frost probability FP by the scheme, for about 62% of cases with correct 1st forecast, FP=80% and 38% of these cases are with FP=100%. The cases with correct 2nd forecast were 70% with FP=80% and 30% with FP=100%. 78% of cases with correct 3rd forecast are with FP=80%, while 22% - with FP=100%. There is only one case with FP=50% which in fact was a non-frost case, thus was treated as a false alarm.

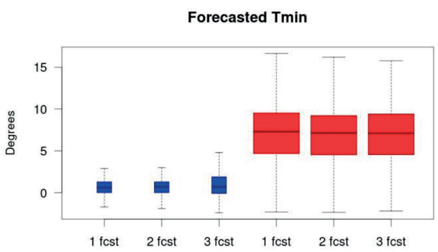

Figure 2. Boxplot of predicted minimum temperature Tmin by 1st, 2nd and 3rd forecasts for cases with (in blue) and without (in red) damaging frost registered

Also, it has to be mentioned that from all 133 frost cases, only for 33 the forecasted minimum temperature is below 0ºC. Figure 2 shows the interval of values of the forecasted minimum temperature for all considered cases. As expected, the median of the forecasted Tmin. for frost cases is considerably lower than this for non-frost cases (about 0.6ºC for the 3 forecasts for “frost cases” and about 7.1ºC for the 3 forecasts for “non-frost cases”). The maximum value of Tmin. for frost cases reaches values above 5ºC, while for some non-frost cases Tmin. is below -3ºC. Our results show that the majority of these cases were correctly discriminated by our scheme as “frost” and “non-frost” cases, which would not be the case if only considering the forecasted Tmin.

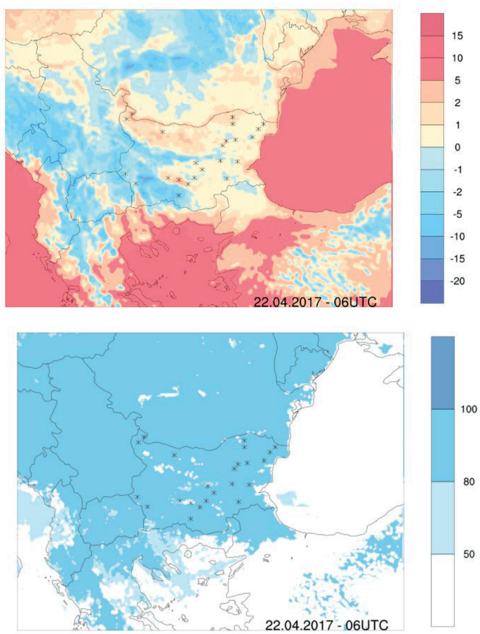

Figure 3. Tmin. forecasted by ALADIN (top panel) and frost probability forecasted by our scheme (bottom panel) for 22/04/2017. Stations with registered damaging frost are indicated with black crosses

Figures 3 and 4 show two case studies with two different frost types. During 22/04/2017 advection frost occurred and was registered at many synoptic stations in Bulgaria. In Figure 3 is shown the forecasted by ALADIN minimum temperature, which is visibly above 0ºC over a big part of Bulgaria. However, the scheme for frost probability forecast gives a probability of 80% over almost whole country, covering all stations with detected frost. During 11/05/2023 a considerably late radiation damaging frost (which is very local and more difficult to be predicted) was registered at two synoptic stations – Dobrich and Rojen. In Figure 3 it is visible that the presented here scheme predicts very correctly the probability of frost occurrence.

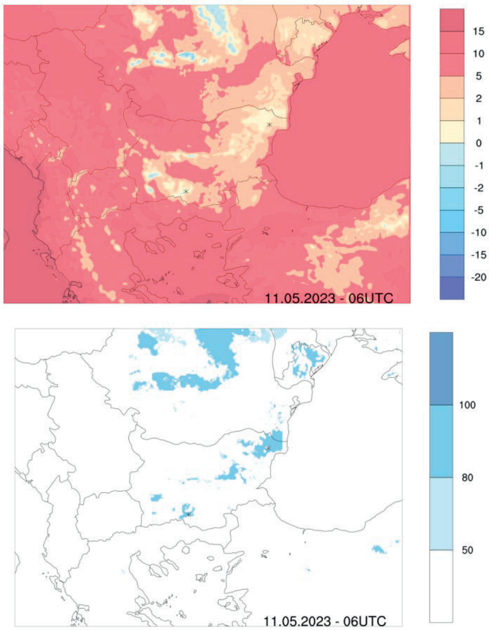

Figure 4. Tmin. forecasted by ALADIN (top panel) and frost probability forecasted by our scheme (bottom panel) for 11/05/2023. Stations with registered damaging frost are indicated with black crosses

CONCLUSIONS

Frosts are a dangerous phenomenon of a meteorological nature, which limits the potential growing season for field and vegetable crops and orchards. Late spring frost often cause annual damage to varying degrees, and during some years, the harvest is totally compromised in some places due to them. This is also related to the tendencies towards an increase in the average monthly temperatures in winter and the earlier start of the vegetation processes. The average multi-year data on the occurrence of the sensitive phases in orchards (Geogieva et al., 2023) are after the average multi-year date of the onset of spring frost, which makes the production of early flowering species risky in many regions of Bulgaria. Timely and accurate specialized forecasting of freezing conditions is the most effective way to limit the consequences of extreme weather events.

In the present study a scheme for damaging frost occurrence forecast based on Random Forest technique and using NWP model output as predictors is presented and its performance is evaluated. Results for 27 synoptic stations during April the last 3 years show a probability of detection above 0.85 and a false alarm rate below 0.1 independently of the remoteness of the forecast. Most "frost cases" were predicted with a frost probability FP=80%, while about a third of them with FP=100%. Our results show that most of the considered cases were correctly discriminated by the scheme as "frost" and "non-frost" cases, which would not be the case if only considering the forecasted Tmin. As there is no significant difference between the closest and the farthest forecast performance, it could be concluded that damaging frost could be relatively well forecasted by the presented here scheme 72 hours (3 days) before its occurrence, which should be enough to react to minimize damage caused by frost in the agricultural sector. Future investigations will be performed to test the forecast performance of the scheme for 10 days ahead. For this aim output from IFS (the model of ECMWF) will be used as statistical model predictor. The presented here scheme will be incorporated in the operational suite as a specialised NWP model post-processing for predicting damaging frost probability occurrence.

REFERENCES

Anandhi, A., Perumal, S., Gowda, P.H., Knapp, M., Hutchinson, S., John Harrington, J., Murray, L., Kirkham, M.B., Rice, C.W., 2013. Long-term spatial and temporal trends in frost indices in Kansas, USA. Clim. Change 120, 169–181. https://doi.org/10.1007/s10584-013-0794-4.

Cadenas, J.M., Garrido, M.C., Martínez-España, R., Guillén-Navarro, M.A., 2020. Making decisions for frost prediction in agricultural crops in a soft computing framework. Comput. Electron. Agric. 175, 105587 https://doi.org/10.1016/j.compag.2020.105587.

De Melo-Abreu, J.P., Villalobos, F.J., Mateos, L., 2016. In: Villalobos, F.J., Fereres, E. (Eds.), Frost

Protection. Principles Of Agronomy For Sustainable Agriculture. Springer International Publishing, pp. 443–457

Ding, L., Noborio, K., Shibuya, K., 2020. Modelling and learning cause-effect — application in frost forecast. Procedia Comput. Sci. 176, 2264–2273. https://doi.org/10.1016/j.procs.2020.09.285.

Diedrichs, A.L., Bromberg, F., Dujovne, D., Brun-Laguna, K., Watteyne, T., 2018. Prediction of frost events using machine learning and IoT sensing devices. IEEE Internet Things J. 5, 4589–4597. https://doi.org/10.1109/JIOT.2018.2867333.

Diniz, E.S., Lorenzon, A.S., Castro, N.L.M., Marcatti, G.E., Santos, O.P., Júnir, J.C., Cavalcante, R.B.L., Fernandes-Filho, E.I., Amaral, C.H., 2021. Forecasting frost risk in forest plantations by the combination of spatial data and machine learning algorithms. Agric. For. Meteorol. 306, 108450. https://doi.org/10.1016/j.agrformet.2021.108450.

Fuentes, M., Campos, C., García-Loyola, S., 2018. Application of artificial neural networks to frost detection in central Chile using the next day minimum air temperature forecast. Chil. J. Agric. Res. 78, 327–338. https://doi.org/10.4067/ S0718-58392018000300327.

Gabbrielli, M., Corti, M., Perfetto, M., Fassa, V., Bechini, L., 2022. Satellite-based frost damage detection in Support of winter cover crops management: a case study on white mustard. Agronomy 12, 2025. https://doi.org/10.3390/agronomy12092025.

Georgieva, V., V. Kazandjiev, 2023. Late spring frosts and their impact to agriculture in Bulgaria, J. of Agriculture and Food, v. 11, pp.166-175.

Ismail, R., Crous, J., Sale, G., Morris, A., Peerbhay, K., 2021. Developing a satellite-based frost risk model for the Southern African commercial forestry landscape. South. For.a J. For. Sci. 83, 10–18. https://doi.org/10.2989/20702620.2020.1759304.

Jamei, J., Ebrahim, M., Emomali, A., 2015. Predicting Late Spring Frost in the Zab Catchment Using Multilayer Perceptron (Mlp) Model, vol. 23. JOURNAL OF GEOGRAPHY AND REGIONAL DEVELOPMENT.

Kalaiarasi, G., Maheswari, S., 2020. Frost filtered scale-invariant feature extraction and multilayer perceptron for hyperspectral image classification. https://doi.org/10.48550/ARXIV.2006.12556.

Lee, H., Chun, J.A., Han, H.-H., Kim, S., 2016. Prediction of frost occurrences using statistical modeling approaches. Adv. Meteorol. 2016, 1–9. https://doi.org/ 10.1155/2016/2075186.

Prabha, T., Hoogenboom, G., 2008. Evaluation of the Weather Research and Forecasting model for two frost events. Comput. Electron. Agric. 64, 234–247. https://doi.org/ 10.1016/j.compag.2008.05.019.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rozante, J.R., Gutierrez, E.R., Silva Dias, P.L., Almeida Fernandes, A., Alvim, D.S., Silva, V.M., 2019. Development of an index for frost prediction: technique and validation. Meteorol. Appl. 1807 https://doi.org/10.1002/met.1807.

Rozante, J. R., E. Ramirez, D. Ramirez, G. Rozante, 2023 Improved frost forecast using machine learning machine, Artificial Intelligence in Geosciences, 164-181,
https://doi.org/10.1016/j.aiig.2023.10.001

Talsma, C., Solander, K.C., Mudunuru, M.K., Crawford, B., Powell, M., 2022. Frost prediction using machine learning and deep neural network models for use on iot sensors. SSRN Electron. J.
https://doi.org/10.2139/ssrn.4032447.

Termonia, P., Fischer, C., Bazile, E., Bouyssel, F., Brožková, R., Bénard, P., Bochenek, B., Degrauwe , D., Derková, M., El Khatib, R., Hamdi, R., Mašek, J., Pottier, P., Pristov, N., Seity, Y., Smolíková, P., Španiel, O., Tudor, M., Wang, Y., Wittmann, C., Joly, A., (2018),

The ALADIN System and its canonical model configurations AROME CY41T1 and ALARO CY40T1, Geosci. Model Dev., 11, 257–281