

Improved frost forecast using machine learning methods



José Roberto Rozante ^{a,*}, Enver Ramirez ^a, Diego Ramirez ^b, Gabriela Rozante ^c

^a Center for Weather Forecast and Climate Studies, National Institute for Space Research, Cachoeira Paulista, SP, Brazil

^b University of São Paulo (EEL/USP), Lorena, SP, Brazil

^c São Paulo State University (UNESP), Bauru, SP, Brazil

ARTICLE INFO

Keywords:

Frost
Artificial neural networks
Multilayer perceptron
frost index
Deep learning

ABSTRACT

Frosts are one of the atmospheric phenomena with one of the larger negative effects on the agricultural sector in the southern region of Brazil, therefore, an earlier forecast can minimize their impacts. In the present work, artificial neural networks (ANNs) techniques were applied in order to improve the predicting capabilities of frost events in southern Brazil. In the study, two multilayer perceptron (MLP) ANNs were built, one with ADAM optimizer and the other with SGD. The input parameters MLP-ANNs were numerical predictions of the Eta model. The ANNs were trained using four years (2012–2015), while validation and testing were performed using 2016 and 2017, respectively. An episode of frost that occurred on May 21st, 2018, related to an intense cold air mass, was also utilized to evaluate the performance of the ANNs. The best configurations (topologies and hyperparameters) of the ANNs were identified through experiments, using the highest accuracy obtained during the validation period as a metric. The results of the ANNs with ADAM and SGD optimizers were compared with the predictions of the Eta model. For the case study, an additional comparison against the operational frost index (IG) from the National Institute for Space Research (INPE) was also included. The performance of both ANNs (properly configured) with ADAM and SGD optimizers are comparable one to the other. And both are significantly better compared to the Eta model. The ANNs were able to drastically reduce the underestimation trends of frost events caused by the warm bias of the Eta model. The ANNs also indicated more satisfactory performances when compared to the INPE IG. In general, the ANNs were able to identify deficiencies in Eta predictions, and consequently improve their results. In this sense, the use of ANNs to predict frost events can be a very useful tool in an operational environment.

1. Introduction

Several South American countries, including Brazil, Argentina, and Chile, have a significant contribution from the agricultural sector to their economies (Taboada et al., 2021). However, due to their geographic positioning, these countries are among those struck by frosts, an atmospheric phenomena that may potentially cause serious damage to the food production and agriculture sector. In years characterized by a high frequency of frost, there is a notable reduction in agricultural production, leading to product shortages and, consequently, price increases in both domestic and international markets.

A classic example of this fact was the frost episode that occurred on the night of July 16, 1975 in the southern region of Brazil, which caused a sharp drop in coffee production and an increase of up to 200% in the price of the kilo (Margolis 1979).

The definition of frost depends on the way in which the subject is

approached. Generally speaking, frost is the formation of ice crystals on exposed surfaces, either by freezing dew or by phase change from vapor to ice (Blanc et al., 1963). Several other definitions of the phenomenon are described in the literature, such as: a) occurrence of temperature less than or equal to 0 °C measured in a shelter in a meteorological shelter at a height between 1.25 and 2.0 m (Hogg 1950, 1971); b) occurrence of air temperature below 0 °C, without definition of the type and height of the shelter (Hewett 1971); c) surface temperature below 0 °C, and the existence of a low air temperature that causes damage or death to plants, without reference to surface ice formation (Cunha 1952).

Frosts are local meteorological phenomena, associated, among other factors, with the type of soil, and also with the orientation of the terrain, and the damage caused to crops depends on its intensity and duration (García Pedraza and Vega 1991). Even when they are local phenomena, several authors associate frost occurrences to synoptic patterns. Over South America, the main synoptic systems that influence the occurrence

* Corresponding author.

E-mail address: roberto.rozante@inpe.br (J.R. Rozante).

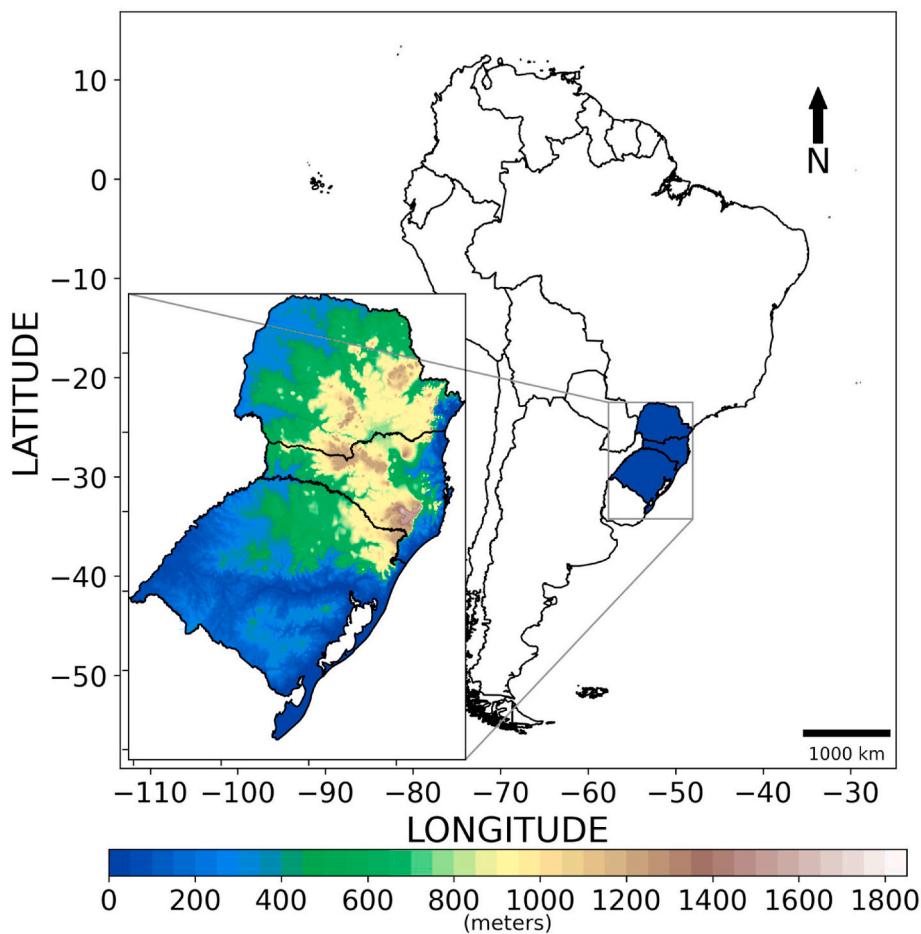


Fig. 1. Study area and topography.

of frosts are: a) flow at high levels of the atmosphere (Satyamurty et al., 1990), b) teleconnections (Fortune and Kousky 1983; Müller and Ambrizzi 2007), c) intensification and positioning of subtropical and polar jets (Müller et al., 2005), d) entry of frontal systems (Parmenter 1976), e) positioning of anticyclones (Hamilton and Tarifa 1978; Rogers and Rohli 1991), f) displacement of Rossby wave trains (Müller and Berri 2007).

There are numerous techniques that can be used 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. Warning systems for frost have been developed/tested based on predictions from numerical weather forecast 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), and neural networks (Fuentes et al., 2018; Ding et al., 2020) and statistical indices (Anandhi et al., 2013; Rozante et al., 2019). Among all the techniques mentioned, neural networks are being successfully applied to several atmospheric phenomena displaying impressive performances. The countries that stand out in the application of machine learning in meteorology around the world are China, the United States of America, Australia, India and Japan, with the most used algorithms being Deep Learning and Artificial Neural Networks (Bochenek and Ustrnul 2022).

To the authors knowledge, the first attempts for frost forecasts using both meteorological observations and neural networks emerged in the late 90s, Robinson and Mort (1997) for Sicilia, Italy; Verdes et al. (2000) for Santa Fé, Argentina; Ovando et al. (2005) for Córdoba, Argentina. Currently, many studies are effectively applying machine learning to predict frost events using different types of neural network architecture,

such as Multilayer Perceptron (Kalaiarasi and Maheswari 2020; Jamei et al., 2015; Diniz et al., 2021), Support Machine Vector (Xu et al., 2021a; Lu et al., 2019; Zendehboudi and Hosseini 2019), Random Forests (Ismail et al., 2021; Diedrichs et al., 2018; Noh et al., 2021), Convolutional Neural Network (Talsma et al., 2022; Wassan et al., 2021).

In recent years, significant advancements have been made regarding the application of machine learning in studies related to frost prediction. Notably, I highlight the work of Lira et al. (2022), in which the authors employed a spatio-temporal neural network architecture, demonstrating substantial improvement over the state-of-the-art methods for frost prediction. Furthermore, the utilization of Convolutional Neural Networks (CNNs) has emerged as an innovation in this research domain. A noteworthy example is the study by Talsma et al. (2023), in which they configured two neural network models: one fully connected and another convolution-based, with the aim of comparing them to a Random Forest (RF) algorithm. The results revealed that the CNN is more effective in capturing variability, reducing overfitting, and enhancing performance when compared to the RF.

Due to the numerous existing algorithms, it is natural to inquire which network architecture suits better for frost forecast. Diniz et al., (2021) compared three machine learning classifiers the Random Forest (RF); Support Vector Machine (SVM) and Multi-layer Perceptron (MLP), and concluded that RF is the most efficient algorithm to indicate frost. Similar results were obtained by Noh et al. (2021) while comparing RF to SVM and logistic regression models. On the other hand, in Xu et al. (2021b) and Zendehboudi and Hosseini (2019) the SVM was the scheme that presented better performance when compared to the other algorithms. It results that the performance of the network depends on several factors such as: data treatment, variables used and tuning of the

Table 1

Example of the input database for the neural network.

n	Data	lat(°)	lon(°)	Tmin(°C)	SLP(hPa)	V(m/s)	Cloud	RH(%)	TOPO(m)	Frost
1	20120501	-30.4	-57.6	4.25	1023.0	3.0	0.0	84.6	51.0	1
...
370	20150901	-25.2	-48.0	15.55	1017.3	6.1	0.9	73.8	73.8	0

algorithm parameters.

In order to minimize the frost-related damage of a country's agricultural production related to frost an earlier forecast is extremely valuable. Therefore, the main objective of the study is to develop an artificial neural network (ANN), properly adjusted to meteorological variables predicted by a numerical model. The obtained network must improve the model forecast and be capable of satisfactorily adding predicting skills for frost events in the southern region of Brazil. The present work is organized as follows: Section 2 presents a brief description of the database used for the study, pre-processing of the data, the region of interest, configuration of the experiments, machine learning algorithms, and the evaluation methods. In Section 3 we present the main results and discussions. In Section 4 we will summarize the main findings in the conclusions section.

2. Methodology

The observed data are measured in a shelter located 2 m above the surface. The typical differences found between the air temperature inside the weather instrument shelter and the air temperature on the grass surface for icy nights is about 5.6 °C (Fagnani and Pinto 1981). Therefore, for the present study, cases of frost occurrence were considered when the observed minimum temperature in the shelter (Tmin) is less than or equal to 6.0 °C. The same Tmin value was also adopted in Rozante et al. (2019). For further discussions about differences between the threshold used in meteorology and agrometeorology check (Rozante et al. (2019), and reference therein) and also (Cunha 1952; Savage 2012).

2.1. Area of interest

The states that compose the southern region of Brazil are Paraná, Santa Catarina and Rio Grande do Sul (Fig. 1), with an area of 576 000 km² (roughly 7.2% of the Brazilian territory), is one of the most important regions for the Brazilian economy. Agriculture is one of the main economic activities in the region and as such is strongly affected by extreme weather events, like intense cold waves (Marengo and Camargo 2008). The subtropical climate is present throughout the region, except in the extreme north of Paraná (Alvares et al., 2013). This type of condition has four well-defined seasons: hot summers, harsh winters and well-distributed rainfall throughout the year. As there is a great variation in temperature throughout the year, the Southern region has the highest range in seasonal thermal amplitude among all Brazilian regions. In winter, polar air masses often occur in the region, causing considerable temperature drops, leading to numerous frost cases in all the three states. The orography of the southern region decreases towards the interior of the continent (east-west direction). As a result, some rivers that are born near the ocean flow inland, emptying into other rivers. It is a region that consists of many plateaus with ancient geological formations, sedimentary and crystalline terrains. The highest regions of the domain are found in the mountains of Santa Catarina, where altitudes reach around 1800 m.

2.2. Database

In the present study, two data sources were used: observed data and numerical weather forecasts.

South American Mapping of Temperature (SAMeT) (Rozante et al.,

2022) is an operational product that combines observed temperature, an estimated lapse rate and a digital elevation GTOPO30 together with the ECMWF reanalysis data to produce a more accurate spatial temperature product at an spatial resolution of 5 km. SAMeT depicts maximum, minimum and average temperature for the whole South American region. The historical dataset runs back to January 1st, 2000 and is updated daily to the present day, the data is made available at <http://ftp.cptec.inpe.br/modelos/tempo/SAMeT>.

The Eta model (Mesinger et al., 1988; Black 1994) is a limited area atmospheric model implemented at INPE since 1996. The model was developed at the University of Belgrade and uses the Arakawa's E grid (Arakawa and Lamb 1977) and the vertical coordinate η , which is appropriate in regions of complex terrain. The domain used comprise the whole South America and part of the adjacent oceans. For the present study, the operational forecasts with spatial resolution of 15 km and 50 vertical levels were used. The initial and lateral boundary conditions come from the Global Forecast System (GFS) analyzes and forecasts, respectively. The Eta model is responsible for providing prognostic meteorological variables used to train, evaluate and test the neural network.

2.3. Data processing

In order to diminish the dimensions of the problem and reduce the computational overhead related to the optimization for the ANNs; a careful extraction of equidistant points at every 40 km for the Southern region from both the SAMeT and Eta model is performed. The selected grid points (370) from the numerical models coincide with those of SAMeT. The occurrences of frost can be described as a binary variable: corresponding to days without frost ($T_{min} > 6.0 \text{ }^{\circ}\text{C} \Rightarrow 0$), and days with frost ($T_{min} \leq 6.0 \text{ }^{\circ}\text{C} \Rightarrow 1$). To classify the days the T_{min} from SAMeT were used.

For the elaboration of the proposed neural network, the values of latitude (lat, °), longitude (lon, °) and altitude (TOPO, m) of the 370 considered points were used as input data, in addition to the 24-h numerical forecasts of selected meteorological variables. The meteorological variables selected were: T_{min} at 2 m (°C), mean sea level pressure (SLP, hPa), relative humidity at 2 m (RH, %), wind magnitude at 10 m (V, m/s), cloudiness (Cloud, no dims). The model's T_{min} is obtained from the diurnal cycle, and the time of occurrence of this T_{min} is used as a reference to extract the other meteorological variables. The meteorological variables are normalized (mean = 0.0 and standard deviation = 1.0) to overcome the difference in units. An example of the data input format for May 1st, 2012 previous to the normalization is shown in Table 1.

2.4. Studied period and frost cases

For the elaboration of the study, the months between May to September from 2012 to 2017 were considered. These months include the transition period (May and September) and the southern Brazil coldest months of the annual cycle (June, July and August). The entire period totalized 902 days, of which 16 days were disregarded due to the absence of data from the Eta model. The year with the highest number of missing data was 2012 with 13 days, 4 in June and 9 in July. The total database was divided in 3 parts, the first (599 days from 2012 to 2015) intended for training the neural network, the second (152 days from 2016) for validation, and the last (151 days from 2017) for testing the

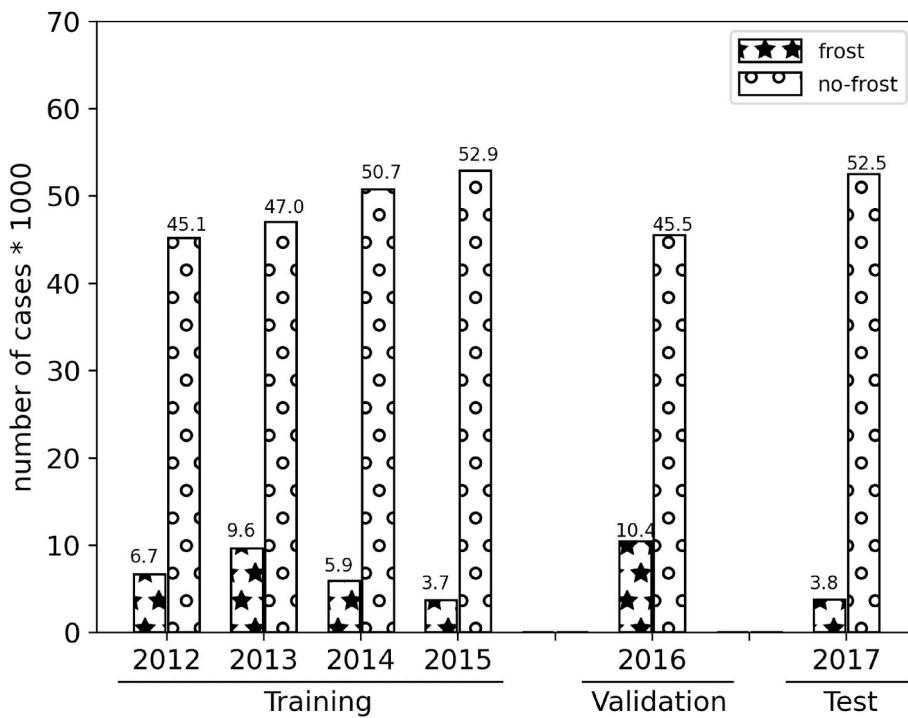


Fig. 2. Number of frost and non-frost events that occurred during training, validation, and test periods.

network. The data used for testing was not previously known by the network.

The total amount of occurrence and non-occurrence of frosts (considering every day and all points in the domain) is reported in Fig. 2. In general, the number of cases of frost occurrence is more than 7 times lower than that of no occurrence. It is also noted that the years used for training and validation, except 2015, presented a significantly higher number of cases of frost when compared to the test period (2017). During the training period, there was a considerable drop in the number of cases from 2013 to 2015. In 2012, the values found do not demonstrate this behavior, probably due to the lack of data already mentioned. Among all the years of training, the 2015 behavior is the closest to the test period. Regarding the validation period, you can check the year with the highest number of frost occurrences in the entire database.

The spatial distribution of the number of cases for frost (Fig. 3) indicates that frosts occurs more frequently in the south of southernmost state of Brazil (Rio Grande do Sul) and on the mountainous regions of Santa Catarina (nearby state), while in the north of Paraná this frequency considerably lower. This pattern distribution is consistent with the results found by Wrege et al. (2018), and confirms a dependence of frost to factors of latitude and altitude. Still in this figure, we can see that the highest number of registered cases (70) occurred in 2016. In 2015 and 2017, the number of frost events is similar (Fig. 2), but in terms of spatial distribution, there is a greater concentration of frost events for the year 2015 (Fig. 3d), when compared to 2017 (Fig. 3f).

2.5. Multi-Layer Perceptron network

The ANN model used was the Multi-Layer Perceptron (MLP). Its architecture basically consists of an input layer, at least one intermediate layer and an output layer. The input layer is a non-computational layer (there are no processing) where the variables are introduced in the network (Rumelhart and James L. 1988). The intermediate and output layers are composed of one or more artificial neurons, and are responsible for all processing in the network. In this model, the architecture used is feedforward, that is, the signal is propagated from the input layer to the output layer. This type of network was developed to solve

regression and classification problems, and uses the supervised learning method through the backpropagation algorithm as training (Rumelhart and McClelland 1987). The formulation description of an example MLP will be presented in matrix form. For a simpler understanding, we are considering a hypothetical neural network with a $3 \times 2 \times 3$ architecture, composed of an input layer, a hidden layer, and an output layer (Fig. 4).

For better understanding, the operation of the network was divided into four stages, which are described below.

- Stage 1 – Initialization:** In this stage, the synapse weights (W_1 and W_2) are initialized. When no information about the weights of the network is available, the common practice consists in, a commonly used method is to initialize the weights randomly, with uniform distribution over a small interval around zero.
- Stage 2 - Input signal propagation:** In this stage, the input patterns (\underline{X} , eq. (1)) are introduced to the ANN in a specific order. Each pattern is propagated forward, layer after layer, until the output pattern is produced (forward phase). This propagation process makes use of an activation function. This function is a non-linear transformation performed on the input data before being sent to the next layer of neurons or finishing it as an output. Mathematical details about the process are presented below.

- i) For hidden layer 1:

$$\underline{v}^1 = W^1 \cdot \underline{X}_b \quad (1)$$

where:

\underline{v}^1 induced field vector $\in R^{2 \times 1}$.

W^1 hidden layer weight matrix $\in R^{2 \times 4}$.

\underline{X} attributes vector (input) $\in R^{3 \times 1}$.

\underline{X}_b biased input vector in the hidden layer (augmented to include a bias) $\underline{X}_b = [1, \underline{X}]^T, \in R^{4 \times 1}$. The bias serves as a parameter to tune the sensibility of each neuron.

Therefore, the output vector from the hidden layer is given by:

$$\underline{O}^1 = F_{a1}(\underline{v}^1) \quad (2)$$

$\underline{O}^1 \in R^{2 \times 1}$, F_{a1} is the activation function

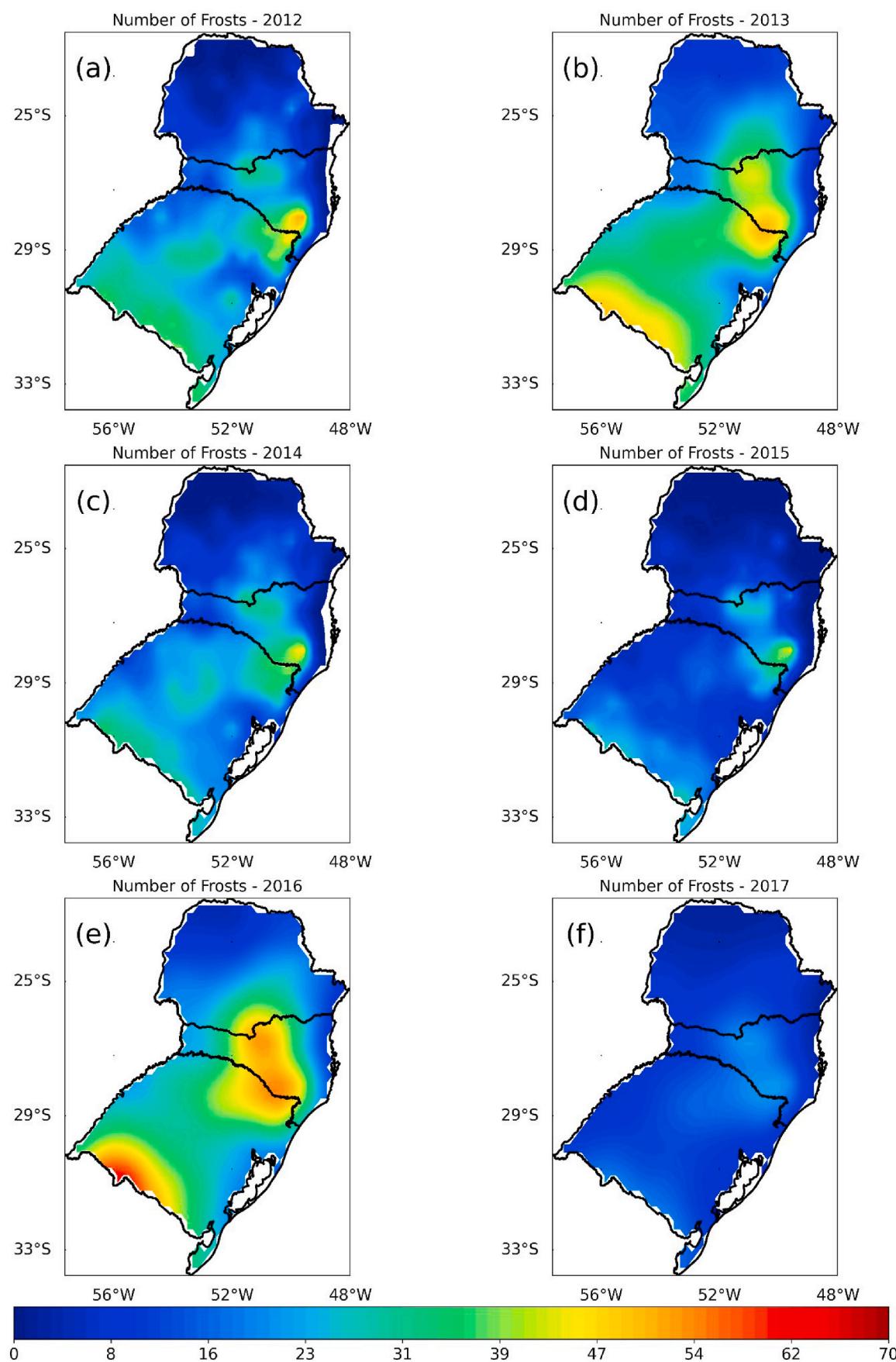


Fig. 3. Spatial distribution of frost events for the years of 2012(a), 2013(b), 2014(c), 2015(d), 2016(e), and 2017(f).

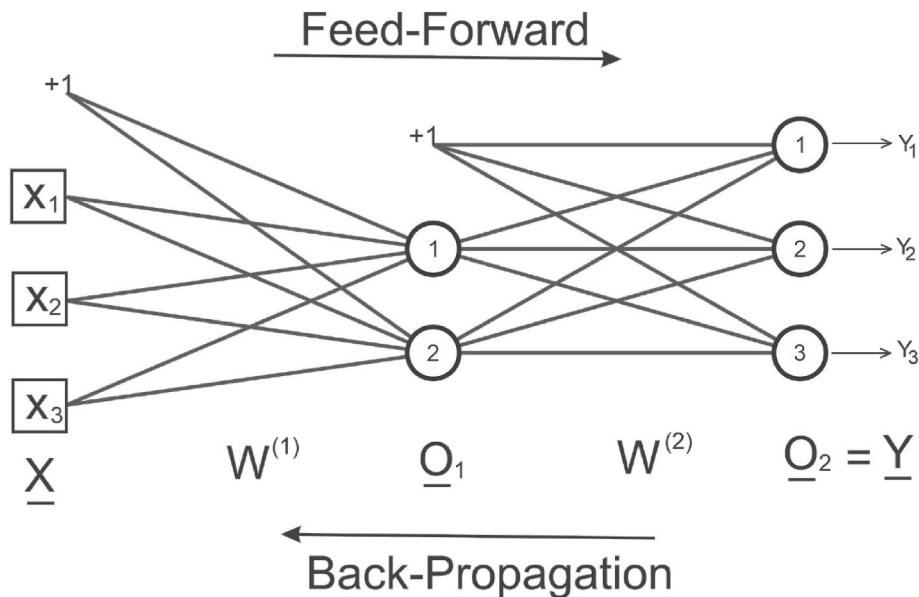


Fig. 4. Architecture of a hypothetical example of a multilayer perceptron neural network.

Table 2

Contingency table with comparative analysis of hits and errors in forecast of occurrence and non-occurrence of a given event.

		OBSERVED		TOTAL
		FROST	NO FROST	
F	FROST	A	B	A + B
	NO FROST	C	D	C = D
R	TOTAL	A + C	B + D	A + B + C + D
E				
C				
A				
S				
T				

Table 3

Statistical indices, formulas and their meanings along with values used for the evaluation (Wilks 2019).

Index	Formula	Values	Meaning
Accuracy	$ACC = \frac{(A + D)}{(A + B + C + D)}$	$0 \leq ACC \leq 1$ Perfect = 1	•Ratio between the total number of hits and the total number of dataset.
Probability of detection	$POD = \frac{A}{(A + C)}$	$0 \leq POD \leq 1$ Perfect = 1	•Proportion of hits given that the event occurred
False alarm rate	$FAR = \frac{B}{(A + B)}$	$0 \leq FAR \leq 1$ Perfect = 0	•Proportion of forecasts of occurrence that did not happen
Critical success index	$CSI = \frac{A}{(A + B + C)}$	$0 \leq CSI \leq 1$ Perfect = 1	•Proportion of hits, omitting cases in which the non-occurrence of the event was correctly forecast
Trade rate	$BIAS = \frac{(A + B)}{(A + C)}$	BIAS > 0; Perfect = 1, BIAS > 1 overestimate BIAS < 1 underestimate	•Ratio of the number of occurrences predicted to the number observed

ii) For the second output layer (output from the network):

$$\underline{v}^2 = W^2 \cdot \underline{O}_b \quad (3)$$

where:

\underline{v}^2 is induced field vector $\in R^{3 \times 1}$.

W^2 output layer weight matrix, $\in R^{3 \times 3}$.

\underline{O}_b is the input vector for the second layer with the bias included $\underline{O}_b = [1, \underline{O}^1]^T, \in R^{3 \times 1}$.

Therefore, the output vector is given by:

$$\underline{O}^2 = y = F_{a2}(\underline{v}^2) \quad (4)$$

$\underline{O}^2 \in R^{3 \times 1}$, F_{a2} is the activation function

c) Step 3 - Error calculation: The output of the network (y , eq. (5))

computed in the previous step is then compared with the desired output. This comparison will generate a value that will determine the network error. This error will be used as a feedback for the connections, which will result in the adjustment of the synaptic weights of each layer in the opposite direction to the propagation of the training signals. The network error calculations are shown below:

$$e_1 = d_1 - y_1$$

$$e_2 = d_2 - y_2$$

$$e_3 = d_3 - y_3$$

$$e = d - y \quad (5)$$

d e y are the vectors of desired value and net result, respectively, both $\in R^{3 \times 1}$.

In our case, as the network is self-associative (desired value is equal to input), then we have:

$$d = \underline{X}$$

$$e = \underline{X} - \underline{Y},$$

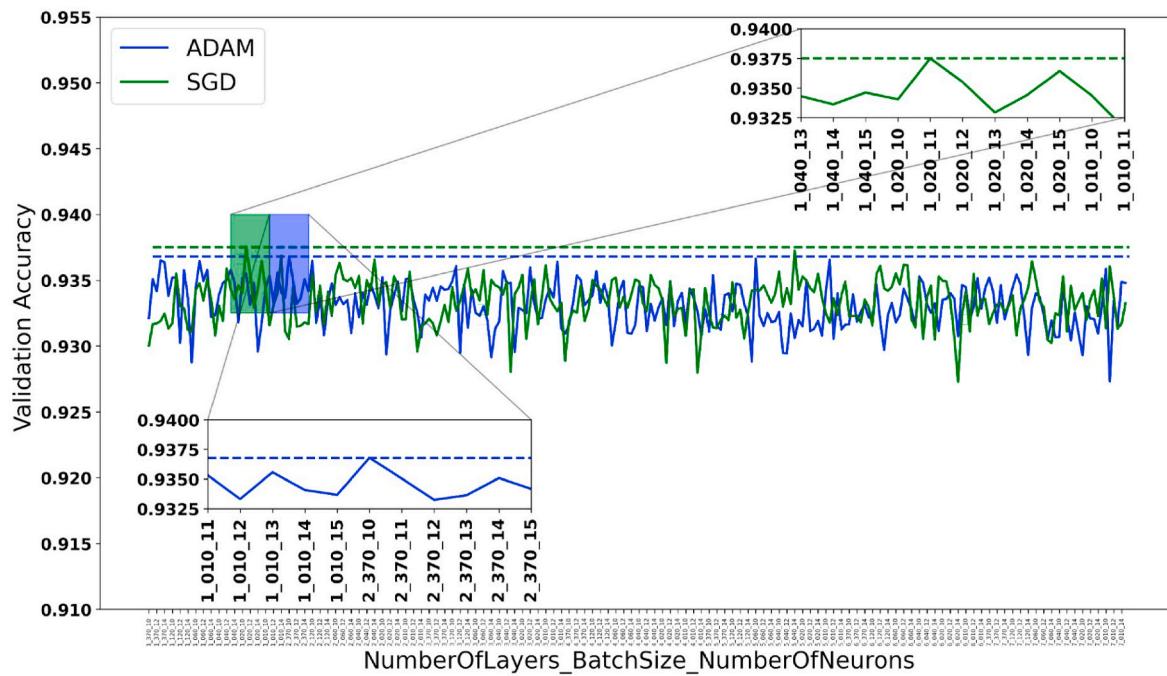


Fig. 5. Accuracy for the validation period for the topology of ANNs with ADAM (blue) and SGD (green) optimizer. The dashed lines indicate the highest accuracy. The X-axis shows different experiments, each labeled with a combination of the topology "NumberOfLayers_BatchSize_NumberOfNeurons". Magnification highlights regions of the graph where the highest accuracy is achieved for both SGD (top right corner) and ADAM (bottom left corner).

$$\in R^{3 \times 1} \quad (6)$$

To eliminate negative values, the total network error in the forward phase is calculated from the sum of the square of the error differences, given by:

$$E_{Total}(W^1, W^2) = \frac{1}{2} \sum_{i=1}^3 e_i^2 \quad (7)$$

c) *Stage 4 - Back propagation:* In this stage, the network training takes place, that is, the interactive adjustments of the weights. Therefore, the weights are adjusted based on the total error (mean squared error). This error is propagated back through the network, starting from the output layer, passing through the hidden layers to the input. The algorithm responsible for training/adjustments is called Back propagation. This algorithm seeks to minimize the quadratic errors from the gradient descent calculation during the iterative process. The generic expression for calculating the gradient and adjusting the weights can be written in a matrix form as:

i) For the second layer we have:

$$\underline{\delta}^2 = diag(\underline{\epsilon}) \cdot [1 - diag(\underline{y}) \cdot \underline{y}] \quad (8)$$

$$diag(\underline{\epsilon}) = \begin{bmatrix} e_1 & 0 & 0 \\ 0 & e_2 & 0 \\ 0 & 0 & e_3 \end{bmatrix} \quad diag(\underline{y}) = \begin{bmatrix} y_1 & 0 & 0 \\ 0 & y_2 & 0 \\ 0 & 0 & y_3 \end{bmatrix} \quad 1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\underline{v}_{\delta} = W^{2T} \cdot \underline{\delta}^2 \quad (9)$$

onde:

$\underline{\delta}^2$ backpropagation gradient $\in R^{3 \times 1}$.

\underline{v}_{δ} gradient of the induced vector is $\in R^{2 \times 1}$.

W^{2T} transpose of the weight matrix $\in R^{3 \times 3}$.

ii) For the (first) hidden layer:

$$\underline{\delta}^1 = diag[1 - diag(O)] \cdot O \cdot \underline{v}_{\delta} \quad (10)$$

From the gradients calculated in layers (1) and (2), we can proceed with synapse weight updates through:

$$W^1 = W^1 + \eta \cdot \underline{\delta}^1 \cdot X_b \quad (11)$$

$$W^2 = W^2 + \eta \cdot \underline{\delta}^2 \cdot O_b \quad (12)$$

The learning rate (η , eq. (11)) is a proportionality constant over the interval $[0 < \eta < 1]$. This rate expresses how fast the network training process is being conducted towards its convergence. The choice of η must be performed with caution to avoid instabilities in the training process.

After updating the weights, the process returns to step 2, that is, a new pattern (X) is presented to the network and propagated again. This procedure occurs several times (epochs) until the total error ($E_{Total}(W^1, W^2)$) of the network reaches values close to zero.

2.6. Hyperparameters and network topology

A challenge for a successful implementation of an ANN is to define values for the hyperparameters, one of which is the network topology, that is closely related to the number of hidden layers and the number of hidden neurons. Determining both the number of hidden layers and neurons is very important and has a great influence on the performance of ANNs (Nitta 2017; Koutsoukas et al., 2017). Although numerous studies address the best way to configure an ANN (Hagiwara 1994; Jin-Yan Li et al., 1995), however, there is still no consensus, since this configuration is heavily dependent on the problem to be addressed. In this sense, it was decided to determine the hyperparameters and topology of the network empirically, that is, using the 'trial and error' method.

The experiments to determine the hyperparameters and topology of the ANN used in the study were trained with 100 epochs (number of times the data is presented to the ANN), using two types of optimizers (an algorithm that minimizes errors and improves performance). The



Fig. 6. Accuracy for the validation period for Hyperparameters of ANNs with ADAM (a) and SGD (b) optimizer. The dashed lines indicate the highest accuracy. The X-axis shows different experiments, each labeled with a combination of the hyperparameters "LearningRate_DecayRate" for ADAM, and "LearningRate_Momentum" for SGD. Magnification highlights regions of the graph where the highest accuracy is achieved.

Table 4

Configurations that presented the highest accuracies considering the validation period.

Experiments	Parameters	ADAM	SGD
Topology	Number of Layers	2	1
	BatchSize	370	20
	Number of Neuron	10	11
Hyperparameters	Learning Rate	0.001	0.004
	Decay Rate	0.7	—
	Momentum	—	0.03

stochastic gradient descent (SGD) is characterized by updating the weights considering only a sample of the data set, which provides a better performance in terms of computational time (Ruder 2016). Adaptive Moment Estimation (ADAM) is also a SGD method, but based on the adaptive estimation of first and second order moments. This method is computationally efficient, with little memory requirements, invariant to diagonal rescaling of gradients, and is well suited for problems that are large in terms of data/parameters (Kingma and Ba 2014). The rectified linear unit (ReLU) activation function was considered for the input and hidden layers. This function returns 0 for all negative values, and the value itself for positive values (Nair and Hinton 2010). For the output layer, the logistic or sigmoid function produces

values in the interval [0, 1].

The highest accuracy obtained in the evaluation of the validation period is used as a criterion to identify which is the best configuration (topology and hyperparameters) for the ANNs with either the SGD or ADAM optimizers, the procedures involved in the identification were divided into two stages.

- a) *Stage 1:* In this stage, efforts were concentrated on defining the most appropriate topology. The experiments were performed varying the number of hidden layers (1–7 layers), sample size (BatchSize) (370, 120, 60, 40, 20, 10), and the number of neurons (10–15 neurons), totaling 504 experiments. For these experiments, the hyperparameters used were the default of each optimizer, that is, ADAM with learning rate and decay of 0.001 and 0.9, respectively, and SGD with learning rate and momentum of 0.010 and 0.0, respectively.
- b) *Step 2:* From the results obtained in step 1, the two optimizers (SGD and ADAM) with the topologies that presented the highest accuracies for the validation period were selected. In these two configurations, the hyperparameters were fine-tuned by varying the learning rate (0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.010) for SGD and ADAM, moment (0.00, 0.01, 0.03, 0.06, 0.09) for SGD, and the decay (0.9, 0.7, 0.5, 0.3, 0.1) for ADAM. Considering that the learning rate was the same for both optimizers, we have a total of 100 experiments.

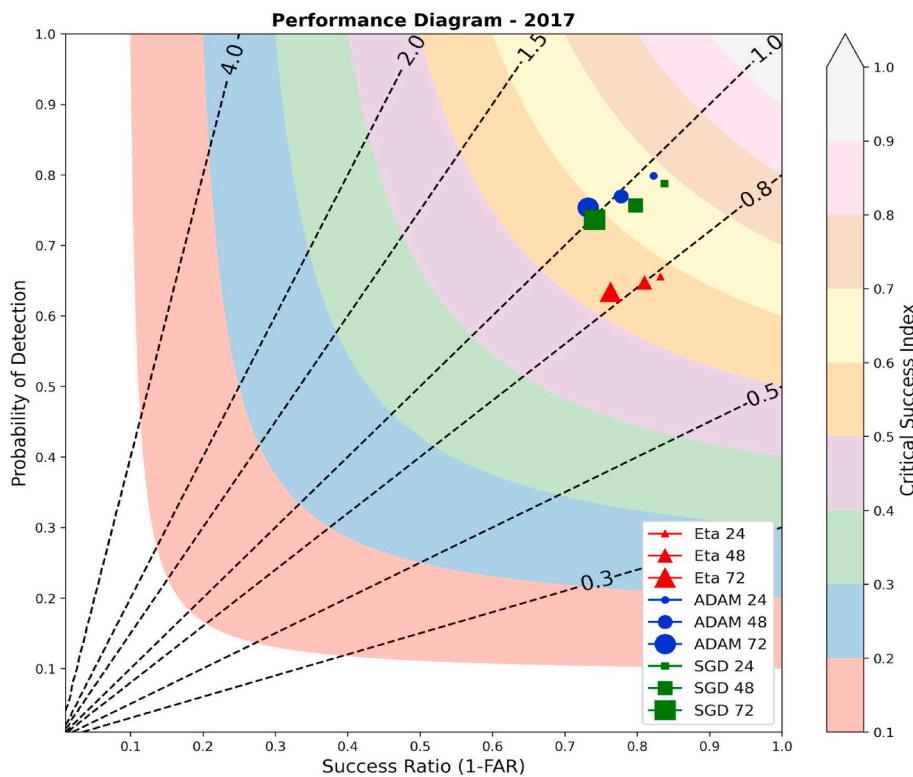


Fig. 7. Performance diagram for the test period (2017) for the Eta (red triangles), ADAM (blue circles), and SGD (green squares) model. As the lead time increases, the size of the geometric figure increases, with the smaller (largest) representing 24 (72) hours lead forecast. The BIAS in the diagram is represented by the dashed lines, while the CSI by the shaded curves.

2.7. Evaluation methods

As the expected results can be evaluated in terms of a binary classification process (occurrence (1) or not (0) of frost events), it was decided to use dichotomous statistical indices obtained from a contingency table (Table 2) to evaluate the performance of algorithms/models. From this table, the quantities "A" and "D" represent the number of correct estimates of either occurrence and non-occurrence of the event, respectively. The quantity "B" denotes the number of times the event was predicted to occur, but did not happen. Analogously, "C" is the number of times the occurrence of the event was observed, but not predicted. The indices obtained from Table 2 used in the evaluation of results are described and presented in Table 3.

3. Results and discussions

The results were discussed in terms of determining the best ANN topology and hyperparameters for two optimizers (ADAM and SGD). And from this determination, the performance of the perceptron networks, together with the predictions of the Eta model were evaluated. The inclusion of Eta in the evaluations occurs in order to verify how much the resulting ANNs are able to improve the performance of the regional model with regard to the prediction of frost events.

3.1. Better ANN configurations

- Topology Determination:** The results of the 504 experiments (detailed in section 2.6) carried out to identify the best configurations for the topology of ANNs with ADAM (blue) and SGD (green) optimizer are shown in Fig. 5. Through this figure, for the ADAM optimizer, the average accuracy of the experiments was 0.9331, while the minimum was found in the experiment (7_010_11) with a value of 0.9273. In the highlighted region on the lower left corner, it can be seen that

the experiment with the highest accuracy (0.9368) was obtained with 2 layers, BatchSize of 370, and 10 neurons (experiment 2_370_10). In the case of the SGD, the average accuracy was 0.9333, while the minimum (0.9273) was verified in the experiment (6_020_14). By expanding the top right corner we can identify that the configuration with one layer, BatchSize of 20, and 11 neurons (1_020_11) was the one that presented the highest accuracy value (0.9375) for the SGD optimizer. In general terms, experiments using the ADAM optimizer present greater variability and slightly lower accuracy than those using SGD.

Analyzing the results regarding neural network topology and neuron quantity, it was found that the best performances were achieved with networks containing fewer neurons in their hidden layers. This discovery suggests that for the specific problem of frost prediction, a more simplified and direct approach may be more effective. Fewer neurons in the hidden layer seem to suffice in capturing the underlying patterns in frost-related meteorological data. However, it's crucial to acknowledge that scientific research is an ever-evolving field, and results may vary based on different datasets, methodological approaches, and specific contexts. A study conducted by Fuentes et. al. (2018), also addressing frost prediction with an MLP neural network, reached a different conclusion. They found that increasing the number of neurons in the network led to improved performance. This implies that, in their dataset and specific context, a more complex approach with more neurons might be necessary to effectively tackle the frost prediction task. This divergence in results underscores the importance of ongoing research and careful consideration of specific conditions in which neural networks are applied. There may be nuances in neural network behavior that depend on data characteristics and prediction objectives. Therefore, while my study favored a simpler approach, the study by Fuentes et. al. (2018) demonstrates that there's no one definitive answer when configuring an MLP neural network for frost prediction. Instead, it is essential to tailor

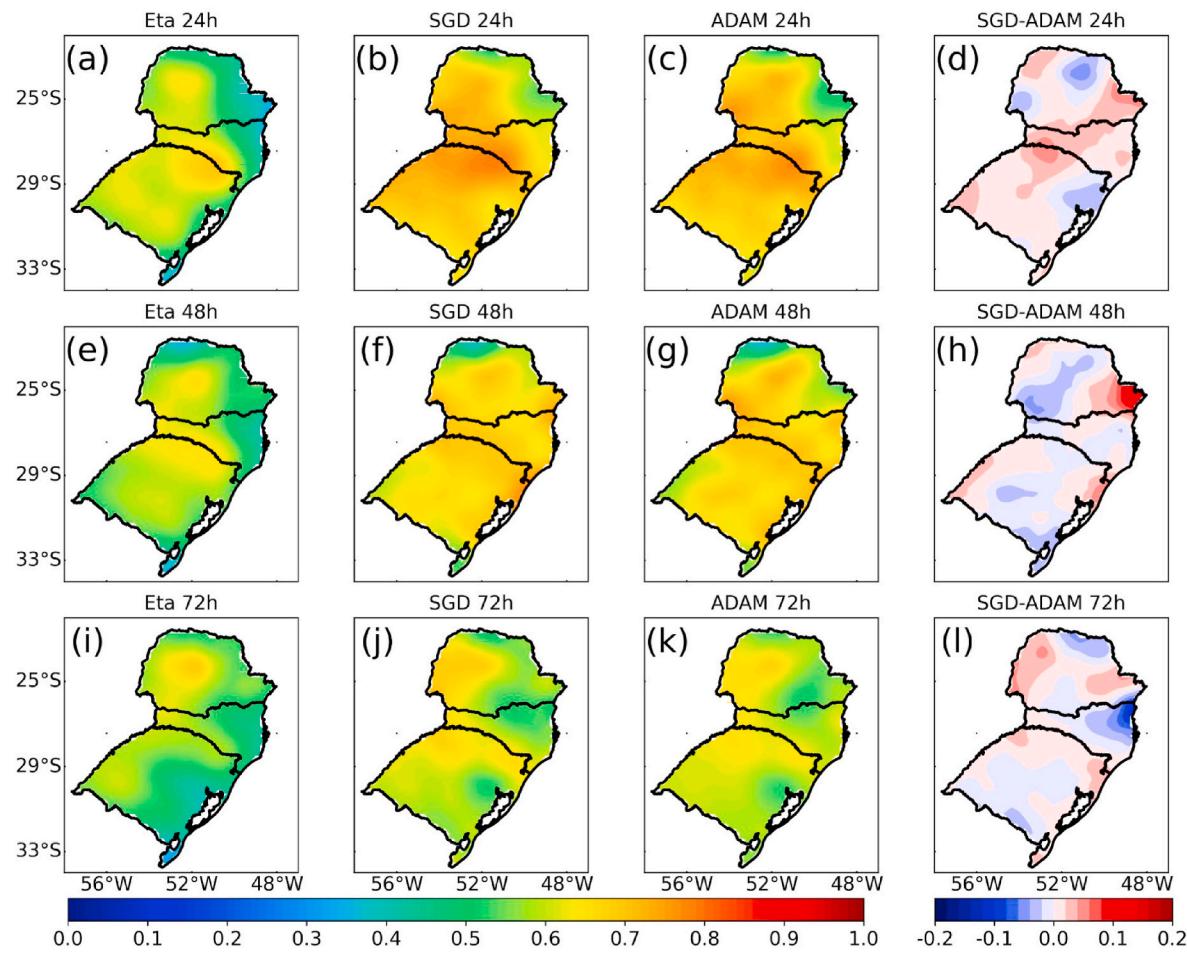


Fig. 8. Spatial distribution of CSI obtained from the year 2017 for Eta (a, e, i), SGD(b, f, j), ADAM(c, g, k), and difference between SGD and ADAM (d, h, l), for 24 (a, b, c, d), 48 (e, f, g, h) and 72-h (i, j, k, l) forecasts.

the approach to the unique nature of each problem and dataset to achieve the best possible results. This dialogue among different studies and approaches is what drives the continuous advancement of science and technology.

ii) *Hyperparameters:* The definition of the best hyperparameters was carried out following the topologies that presented the highest accuracy for the ADAM (2_370_10) and SGD (1_020_11) optimizers (check Topology Determination). The hyperparameters tested were learning rate, momentum, and decay rate, totaling 100 experiments (further details in section 2.6). For the ADAM optimizer, the hyperparameters tested are learning rate and decay rate, while for the SGD the learning rate and momentum. The accuracy of the combinations between the learning rate and the decay rate for ADAM are shown in Fig. 6a. From the figure, it can be seen that the average accuracy of all experiments was 0.9358, with the minimum being 0.9315 (experiment 0.005_0.7). The experiment with the highest accuracy (0.9380) was the configuration with a learning rate of 0.001 and decay 0.7 (0.001_0.7) and a detailed view is presented in the figure. For SGD (Fig. 6b), it was found that the lowest accuracy (0.9326) occurred for the experiment with a learning rate of 0.006 and momentum of 0.00 (0.006_0.00), while the experiment that yielded the highest accuracy (0.9392) among all had a learning rate of 0.004 and momentum of 0.03 (0.004_0.03). This information is most clearly illustrated in the highlighted region of Fig. 6b.

iii) *Synthesis of experiments:* The configurations of the experiments that presented the highest accuracy are summarized in Table 4. For the case of the ADAM optimizer, the best configuration was with 2 layers, BatchSize of 370, 10 neurons, learning rate of 0.001 and decay rate of 0.7. As for the SGD, the most suitable configuration was 1 layer, BatchSize of 20, 11 neurons, learning rate of 0.004 and moment 0.03. Although the computational resources consumed in the hyperparameter experiments were high, the improvements in accuracy were not significant when compared to the default values of each optimizer. Remembering that the activation function used for the input and hidden layers was ReLU and for the output layer Sigmoid, both for the ADAM optimizer and for the SGD. All the results presented below were generated using these activation functions and settings presented in Table 4.

3.2. Evaluation of models

In this section we present the results of the ANNs, configured with the best hyperparameters/topology, applied for the test period (year 2017). The ANNs were obtained with the 24-h forecasts of the Eta model, but for comparison purposes, the 48- and 72-h temperature forecasts were also applied to the same ANNs. Emphasizing that these two lead times did not go through the training and validation process, making these data somehow unknown for the ANNs. The results were approached from two perspectives: one calculating the statistical indices considering all data, both in space and time. And another, spatializing these indices on the domain of interest. In this way, it is possible to quantify and spatially identify the positive and negative points of each

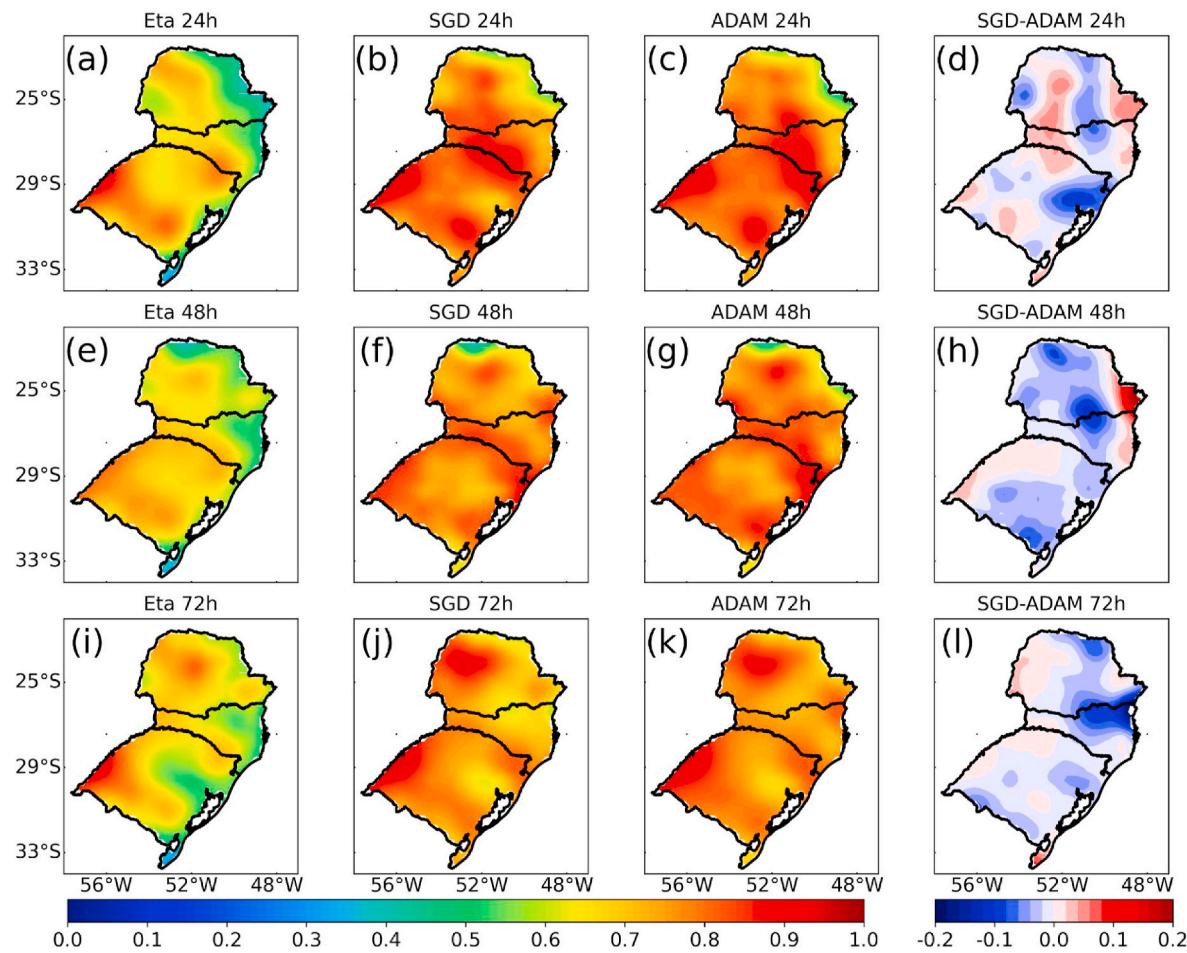


Fig. 9. Spatial distribution of POD obtained from the year 2017 for Eta (a, e, i), SGD (b, f, j), ADAM (c, g, k), and difference between SGD and ADAM (d, h, l), for 24 (a, b, c, d), 48 (e, f, g, h) and 72-h (i, j, k, l) forecasts.

evaluated model.

3.2.1. Space-time evaluation

For this section the performance diagram was used (Roebber 2009). The diagram allows you to explore the geometric relationship between four measures of dichotomous prediction performance: probability of detection (POD), false alarm rate (FAR), critical success index (CSI), and trend rate (BIAS). According to the particularities of this diagram (see Fig. 7), the perfect predictions would be represented in the upper right corner, where POD, CSI and BIAS are equal to 1, and the FAR is equal to 0.

In Fig. 7, the BIAS shows that the Eta model has trends and underestimates by about 20 percent the frost events for all lead times depicted. A warm bias of the Eta model for this region (Rozante et al., 2019) is perhaps the culprit for the underestimation of frost events. In general, both ANNs used are able to significantly improve frost detection as compared to what is indicated by the model. ADAM depicts a mild underestimation for 24 h forecast and overestimation for 72 h forecast, while a perfect BIAS is verified for a 48 h forecast. SGD, on the other hand, indicates slight underestimations for both 24 and 48 h, and at 72 h it presents an optimal BIAS, that is, equal to 1. In terms of BIAS, ADAM indicates more satisfactory results for the 24 and 48 h forecasts.

In general, it is evident that ANNs have the ability to substantially mitigate the bias present in the Eta model, a finding also highlighted in studies conducted by Guarnieri et al. (2006). In these studies, researchers used predictions generated by the Eta model as input for an MLP-type ANN, aiming to predict solar radiation. In an additional study that adopted a similar approach, but with the WRF model, Lima et al.

(2016) also identified the network's capability to reduce the inherent bias in the numerical prediction model. These observations underscore the promising capacity of ANNs to address the discrepancies often encountered in prediction models like Eta and suggest that this approach may serve as a valuable strategy for enhancing the accuracy of meteorological forecasts.

POD analysis shows that ANNs had higher values than Eta, with ADAM slightly higher than SGD. It is also noted that the model showed a slight drop in the probability of detecting the event with the increase in the forecast lead time. ADAM and SGD showed a decrease of 4% (variation between 0.80 and 0.76), while Eta 2% (variation between 0.65 and 0.63). The slightly greater drop seen in the ANNs may be associated with the fact that the ANNs were trained only with the 24 h forecast.

In terms of FAR, both ANNs and Eta presented values very close to each other. For 24 h the values were around 0.17, while for 48 and 72 h they were 0.20 and 0.25, respectively, indicating an increase in the number of false alarms with the increase in the forecast period. It is interesting to note that the ANNs, even considerably increasing the POD values, managed to keep the FAR values low, indicating a potential gain in relation to Eta.

The CSI shows that the ANNs presented performances (between 0.59 and 0.68) and were superior by about 10 percent to those of the Eta (between 0.53 and 0.58) for all lead times. Comparisons between the ANNs show that no significant differences were found for all the lead times evaluated. In general, the performance diagram shows that the ANNs were consistently better than the Eta model in all indexes analyzed and for all forecasted times. Among the ANNs, it can be said that the one with the ADAM optimizer indicated slightly superior

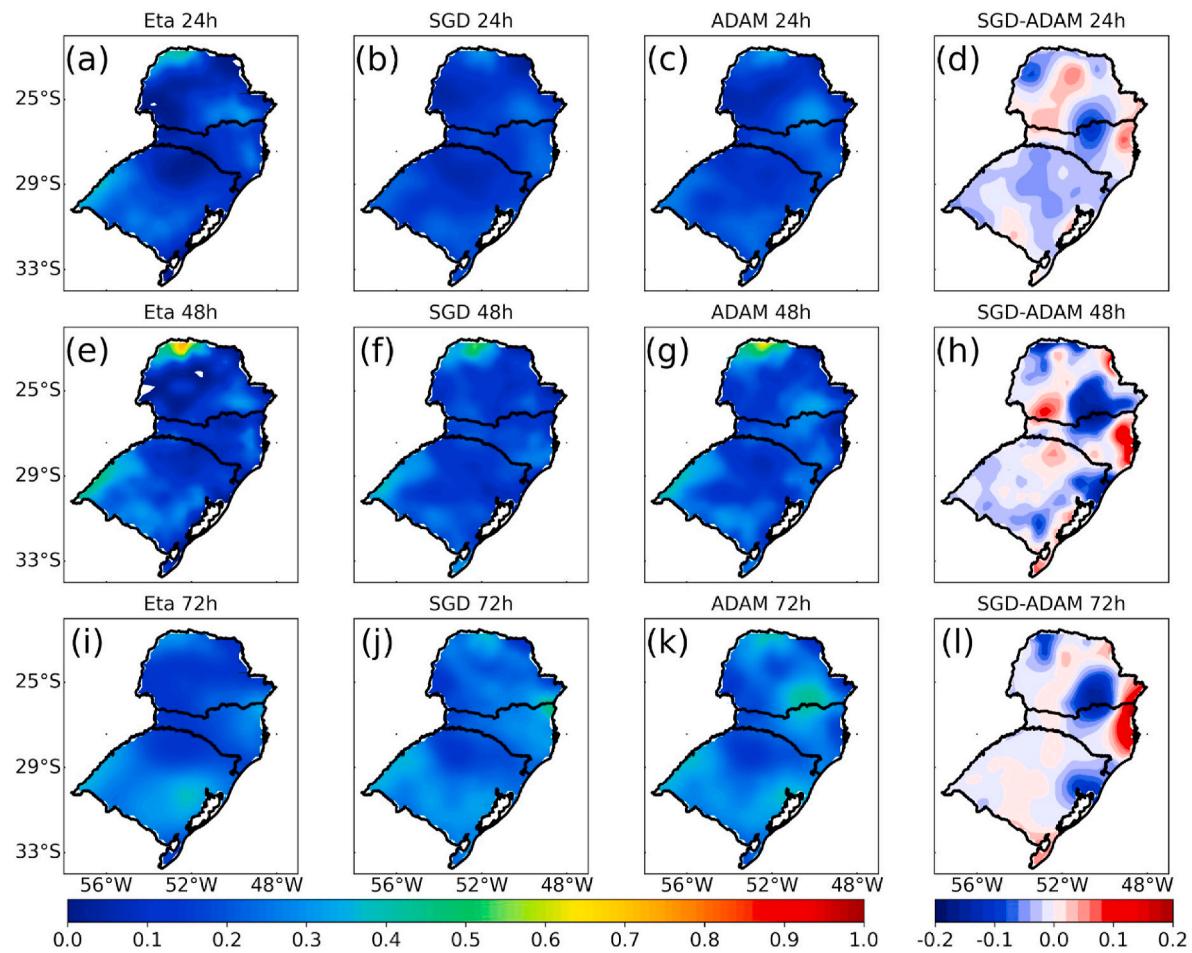


Fig. 10. Spatial distribution of FAR obtained from the year 2017 for Eta (a, e, i), SGD(b, f, j), ADAM(c, g, k), and difference between SGD and ADAM (d, h, l), for 24 (a, b, c, d), 48 (e, f, g, h) and 72-h (i, j, g, l) forecasts.

results, mainly in terms of BIAS and POD.

Fig. 8 shows the spatial distribution of the CSI (year 2017) for the Eta, SGD, ADAM, and their respective differences for the 24, 48, and 72 h forecasts. In the 24-h forecasts, it is observed that the ANNs (Fig. 8b and c) presented superior performances than the Eta (Fig. 8a) for the entire domain. Comparison between the ANNs shows similar patterns, with slight differences in northern Paraná and eastern Santa Catarina (Fig. 8d). In 48 h (Fig. 8e, f, 8g, 8h) a drop in the performance of the three models can be seen, being more pronounced in the ANNs, also seen in Fig. 7 (diagram). Even with this drop, ANNs are still able to provide more realistic forecasts (Fig. 8f and g). Even at this time, the RNAs showed a considerable drop in performance in the north of Paraná. At 72 h (Fig. 8i, j, 8k, 8l) the performance drop behavior for the three models is also verified. In general, ANNs provide more satisfactory results both in terms of domain and in relation to forecast times. The most pronounced performance drop in the ANNs probably occurred due to the lack of training and validation for larger schedules.

The spatial distribution of the POD (year 2017) for the Eta, SGD, ADAM, and their respective differences for the 24, 48, and 72 h forecasts is shown in Fig. 9. It can be seen that for all times, and through the domain, the ANNs are more likely to predict the frost event when it actually occurs. It is also noted that the probability decreases with the increase in the lead time. The northern region of Paraná is where the ANNs have the lowest POD values. As it can be seen via the differences that it is ADAM that has the highest probability of detection. This result was also identified in the Roeber diagram (Fig. 7).

The results for the FAR are shown in Fig. 10. As it can be seen, for 24 h the highest FAR values (between 0.4 and 0.5) are indicated in the

north and south of Paraná, the eastern of Santa Catarina and the southwest of Rio Grande do Sul (Fig. 10a). For these regions, the ANNs (Fig. 10b and c) manage to reduce the number of false alarms, mainly the one that uses the SGD optimizer (Fig. 10d). For the 48-h forecasts (Fig. 10e, f, 10g, 10h) there is a considerable increase in the FAR for the regions mentioned above for the three models. Although the patterns between models are similar, SGD still indicates slightly lower FAR when compared to ADAM (with the exception of the coast of Santa Catarina and southwestern Paraná, see Fig. 10h). For 72 h (Fig. 10i, j, 10k, 10l) the ANNs indicate higher FAR than Eta. In terms of FAR, it was noted that the SGD optimizer was the one that most reduced the cases of false alarm.

The spatial distribution of the BIAS (Fig. 11) indicates that the model underestimates the temperature by more than 20% for almost the whole domain, a considerable underestimation valid for the three lead times studied (24, 48 and 72 h Fig. 11). It can also be noted that for 72 h (Fig. 11i) the areas with underestimations reduce and high values of overestimations appear. The ADAM optimizer for 24 h (Fig. 11c) presents values close to 1, indicating an almost perfect BIAS, while SGD indicates a slight underestimation. ADAM and SGD for 48 and 72 h (Fig. 11f, g, 11j, 11k) increase the overestimation trends both in terms of area and magnitude, being more pronounced in using ADAM.

The superiority of neural networks, especially MLP, compared to numerical prediction models, as observed in this study, has also been confirmed in previous research, such as that of Li et al. (2021). In this study, the authors configured an MLP and fed it with predictions from the Weather Research and Forecasting (WRF) model, obtaining superior results compared to the numerical prediction model. Valverde Ramírez

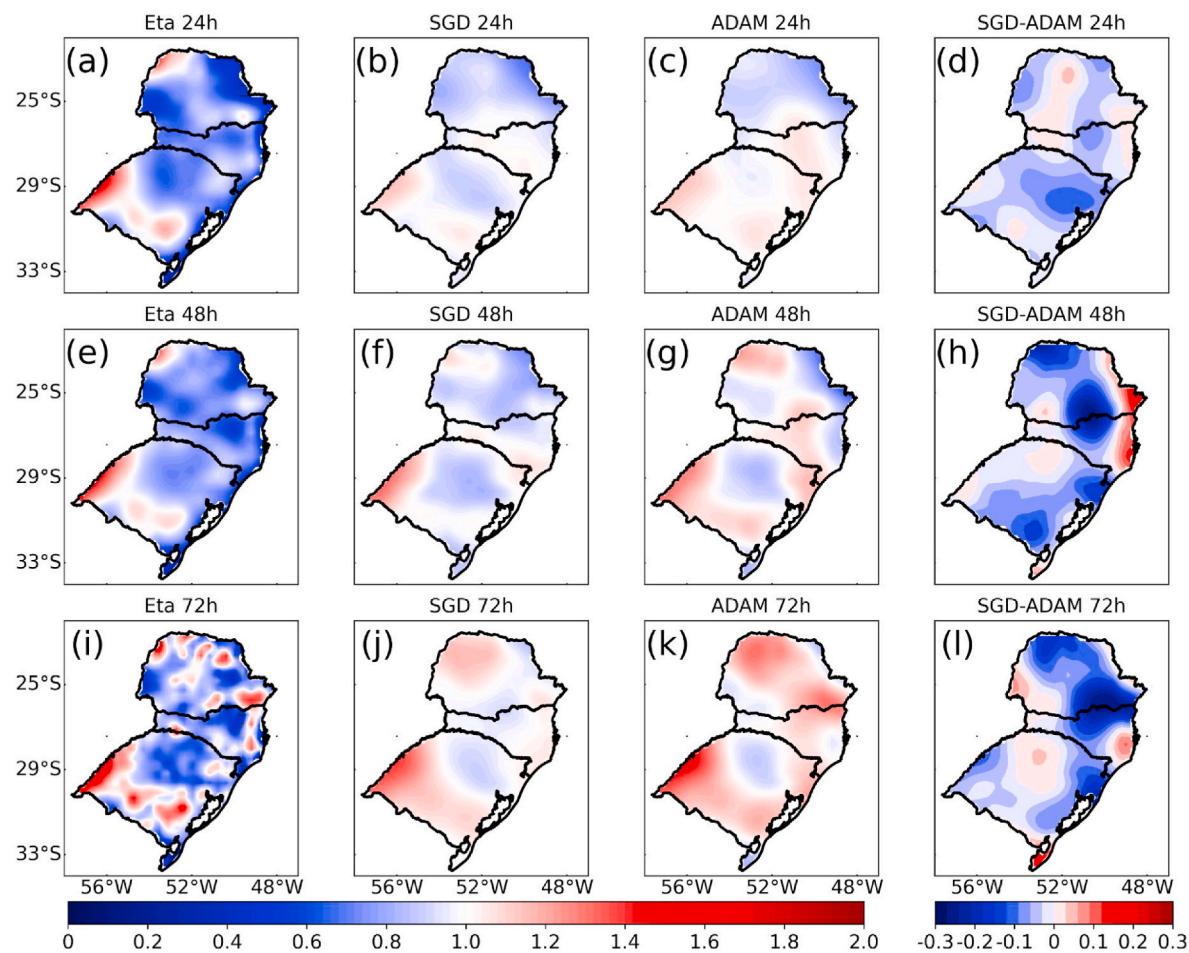


Fig. 11. Spatial distribution of BIAS obtained from the year 2017 for Eta (a, e, i), SGD(b, f, j), ADAM(c, g, k), and difference between SGD and ADAM (d, h, l), for 24 (a, b, c, d), 48 (e, f, g, h) and 72-h (i, j, k, l) forecasts.

et al. (2005), while employing a Neural Network MLP fed with numerical forecasts from the Eta model for the State of São Paulo, Brazil, also demonstrated that the neural network was capable of enhancing the results obtained from the Eta model.

3.3. Case study

In general terms, considering the entire database of the test period, it was found that the ANNs were able to significantly improve the predictions of the numerical model. However, it is still necessary to verify whether these improvements are representative under certain situations. A case study of a very intense cold that hit the South, Southeast and Midwest regions of Brazil on May 21, 2018 was used for the study. This case has already been explored (Rozante et al., 2019; Lucyrio et al., 2020). The intense temperature drop was associated with the passage of an intense polar air mass through the region that caused frost in the three states. Fig. 12 presents the spatial distribution (classified according to SAMEt) of the occurrence of frost (*) and non-frost (o). In Rio Grande do Sul the frosts were concentrated in the western part of the state, while in Santa Catarina and Paraná it covered almost the entire state, except for a narrow strip close to the coast and northwest of Paraná. Among the 370 points located on the domain, most of them (224) were classified as frost, and 146 as non-frost. The Eta model forecasts 24, 48 and 72 h in advance were used as input for the ANNs.

Fig. 13 shows the spatial distribution of the performance of the Eta model, SGD and ADAM for an event that occurred on May 21, 2018. The 24, 48 and 72 h forecasts are displayed. Information about the dichotomous statistical indices are also included. The results from the

numerical model (Fig. 13a, d and 13g) show that most of the frosts in the western strip of Rio Grande do Sul were not predicted, evidencing a warm bias already noted in the evaluations of the test period. In the states of Paraná and Santa Catarina, Eta predicts frost cases relatively well, but with the increase in the forecast lead time, there is an increase in the number of unforeseen frost events, especially for the northern sector of Paraná. For all times the BIAS indicates underestimation trends. In the case of SGD (Fig. 13b, e and 13h), Paraná and Santa Catarina present patterns similar to those of Eta, however, in Rio Grande do Sul there are significant improvements, mainly in the 24 and 72 h forecasts where the CSI values were 0.731 and 0.790, respectively. Like SGD, ADAM (Fig. 13c, f, 13i) is also similar to Eta in the states of Paraná and Santa Catarina, and manages to increase the number of hits in Rio Grande do Sul. In a general context, ADAM and SGD demonstrated similar behaviors with respect to both spatial patterns and performance accuracy. In this specific case, the numerical model displays the highest number of errors, especially in predicting frosts in Rio Grande do Sul. The utilization of ANNs, both with ADAM and SGD optimization, has yielded significant improvements in the results, particularly for 24 and 72-h forecasts.

In a study focusing on the frost index from a statistical perspective, Rozante et al. (2019) developed a frost index (referred to as IG, which stands for 'Índice de Geada' in Portuguese) that significantly enhanced the predictability of frost/non-frost events when compared to Eta. They also used Eta as input data to study the intense frost event that occurred on May 21, 2018, to demonstrate the effectiveness of IG in 72-h forecasts.

Currently, the IG is being generated and made operationally

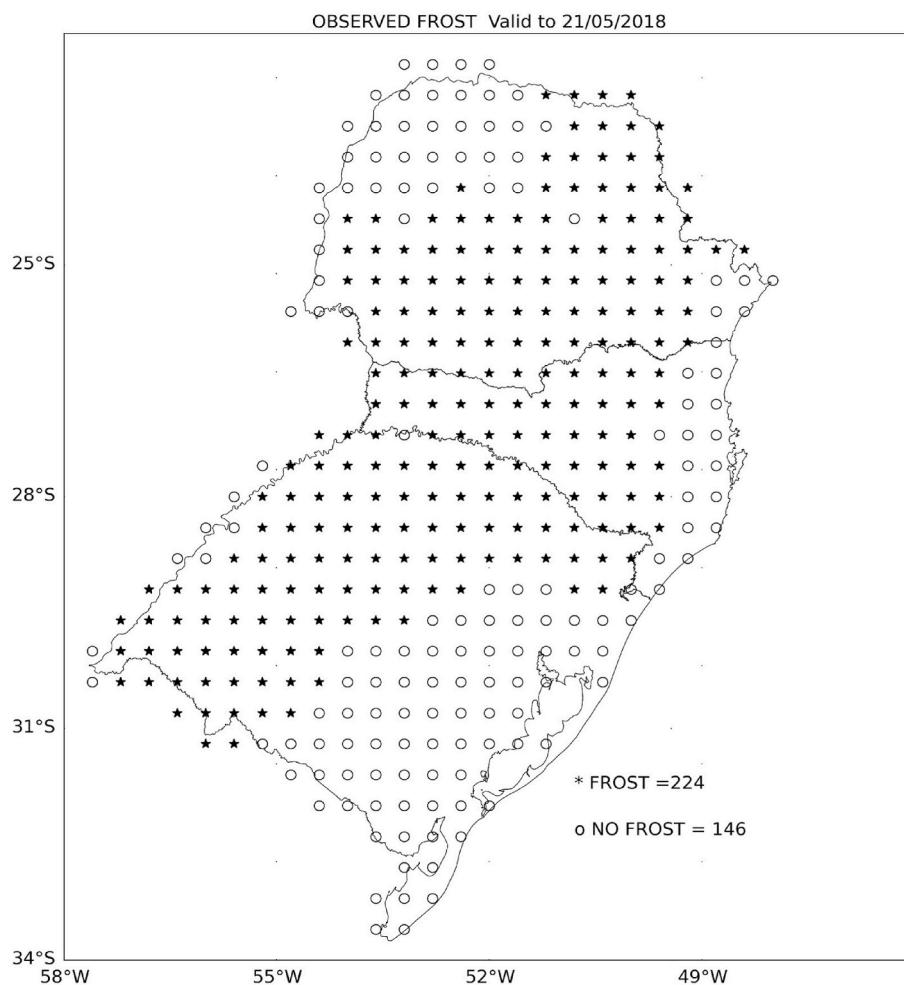


Fig. 12. Spatial distribution of the occurrence of frost (*) and non-frost (o) for the day May 21, 2018.

available on the National Institute for Space Research (INPE) website (<http://tempo.cptec.inpe.br/geada/pt>). In order to verify whether the ANNs could be a possible substitute for the IG, a comparison between these two different methodologies was carried out. The R2 region, defined in Rozante et al. (2019), coincides with the region of this study, and therefore was used in this comparison.

Considering that the IG is generated only for the points where there are observations, it was necessary to extract these points from the linearly spaced grid of the ANNs. For this, the grid cell closest to the existing point of the IG was considered. In this way the comparison and the statistics become more adequate. The results of this comparison and statistics for 72 h in advance are shown in Fig. 14. The worst result to predict frost/non-frost was obtained by using the Eta model alone (Fig. 14a). In this case, the lowest values of CSI (0.587), POD (0.657), and highest FAR (0.154) are obtained, along with an underestimation trend (BIAS = 0.776). Although the IG (Fig. 14d) presents results that are superior to the Eta model, mainly in Rio Grande do Sul, but as they are currently used depicts an inferior skill to those obtained by the ANNs. The IG performed relatively worse in the northern region with an inferior CSI and POD when compared to the ANNs, however, in terms of FAR, the IG performed better than the ADAM but worse than SGD. The comparison between SGD (Fig. 14b) and ADAM (Fig. 14c) indicates that, both in spatial aspects and in statistical terms, the ANN with the SGD optimizer exhibits a slightly superior performance. In a similar manner to this case study, (Garcia 2021) employing 24-h numerical model forecasts applied to machine learning techniques, also demonstrated that artificial neural networks (ANNs) exhibit superior performance compared to IG. And in this case, the model's bias reduction was also

verified, however slight tendencies to underestimate the number of cases were still observed in the neural networks.

4. Conclusions

Through the utilization of an extensive period of numerical forecasts (Eta) and a spatial product representing minimum temperature observations (SAMeT), we have been able to develop two ANNs that operate in conjunction with the numerical model, resulting in an enhanced capability to predict frost occurrences in the Southern Region of Brazil. In a specific case study, the ANNs demonstrated superior performance in comparison to a statistical product (IG) previously developed for frost predictions over a much broader area, currently in use by INPE. These results hold significant relevance as they have the potential to contribute to the minimization of impacts on agriculture and the country's economy. Advanced knowledge of this phenomenon enables farmers to adopt preventive measures, such as artificial misting, irrigation, forced ventilation, and the use of protective coverings, thus reducing production losses and positively influencing product prices.

Among all the configurations (topology, optimizers and hyperparameters) tested in this study, two of them stood as the best: the best ranked was a ANN that used the ADAM optimizer, MLP with 2 hidden layers, BatchSize of 370 and 10 neurons, learning rate of 0.001 and decay rate of 0.7 and the other was the ANN with SGD configured with 1 hidden layers, BatchSize of 20, 11 neurons, learning rate 0.004 and momentum of 0.03. For the construction of the network, the topology impacted the most, whereas the hyperparameters associated with each optimizer method added small corrections. For instance, the increase in

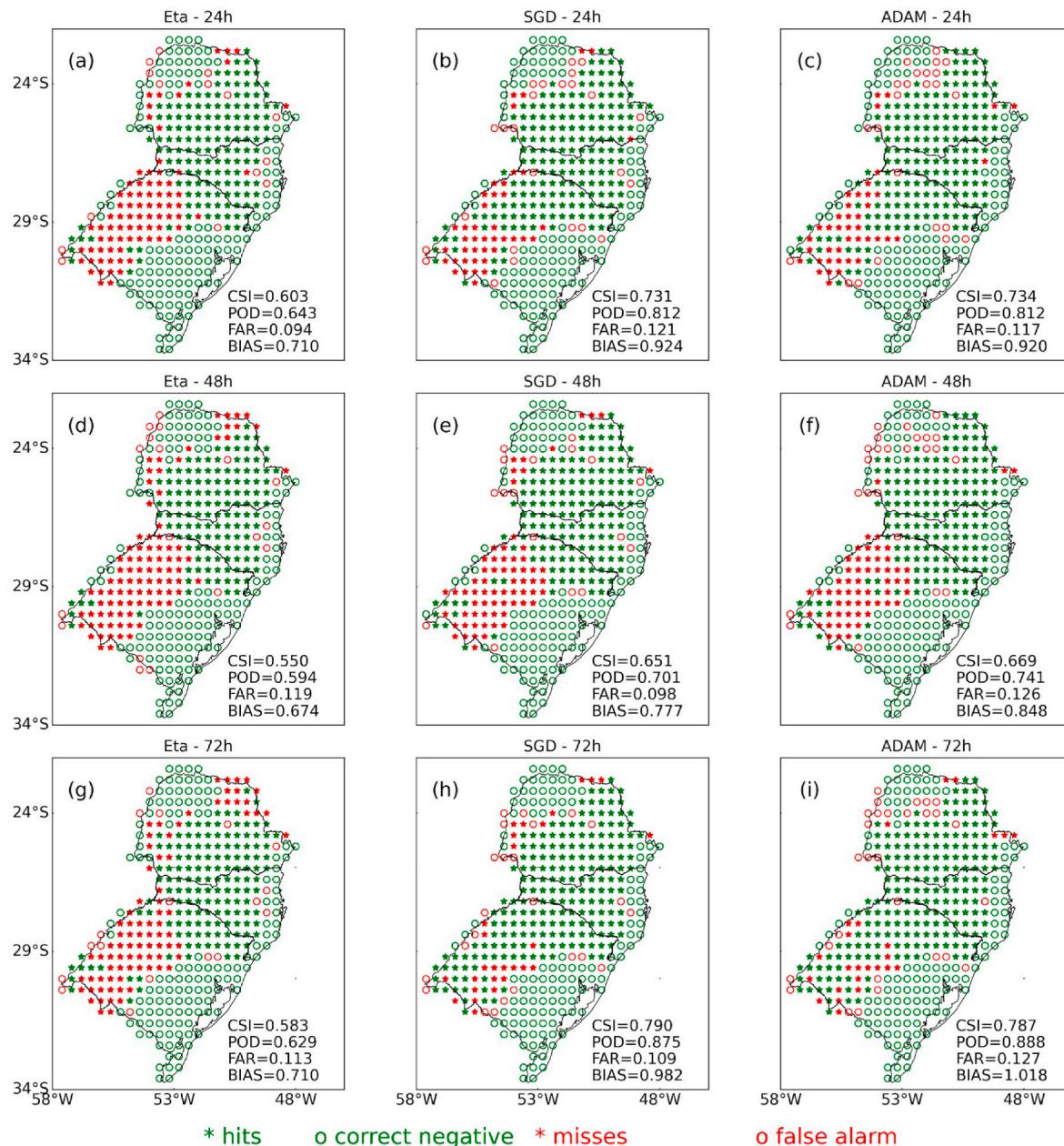


Fig. 13. Spatial distribution of errors (red) and successes (green) of Eta, SGD and ADAM for 24, 48 and 72 h before the event that occurred on May 21/2018. The symbols (*) represent the occurrence of frost, while (o) the non-occurrence.

accuracy using the best hyperparameters was approximately 0.1% for both ADAM and SGD compared to the default values. This demonstrates that the default hyperparameters of each optimizer would already yield satisfactory results for this type of approach.

Another interesting result found in this study was the fact that, although the ANNs were initially trained and evaluated solely for 24-h forecasts, the same ANN configurations were applied to 48 and 72-h forecasts, yielding quite encouraging results. This opens the possibility for future improvements by training the ANNs for all these timeframes. The ANNs demonstrated an ability to recognize and mitigate deficiencies, particularly the warm bias, in the forecasts generated by the numerical model. Consequently, this led to a reduction in the number of frost events that occurred but were not identified by Eta. This observation resulted in the ANNs achieving BIAS values very close to 1 at all forecast times, whereas Eta displayed systematic patterns of underestimations with a BIAS of around 0.8. It is worth noting that the

excellent performance of the ANNs observed in this study can be primarily attributed to the meticulous treatment of input variables, the utilization of an extensive database, and the diligent efforts to identify the most suitable network topologies for the subject under investigation. Additionally, the selection of the threshold used for frost identification, based on the distinctions between weather shelters and open fields, has been shown to be of significant importance.

Based on the results obtained during the test period in 2017, it is concluded that ANNs utilizing both ADAM and SGD optimizers exhibit similar performance. The ANN with ADAM, however, demonstrates slight improvements, particularly in terms of probability of detection and BIAS for the 24-h and 48-h forecast horizons. In comparison to the Eta model's forecasts, the ANNs consistently yield more satisfactory results across all analyzed metrics and forecast timeframes. Regarding the spatial distribution of frost events, the ANNs also outperform the Eta model throughout the entire analyzed domain. The application of ANNs

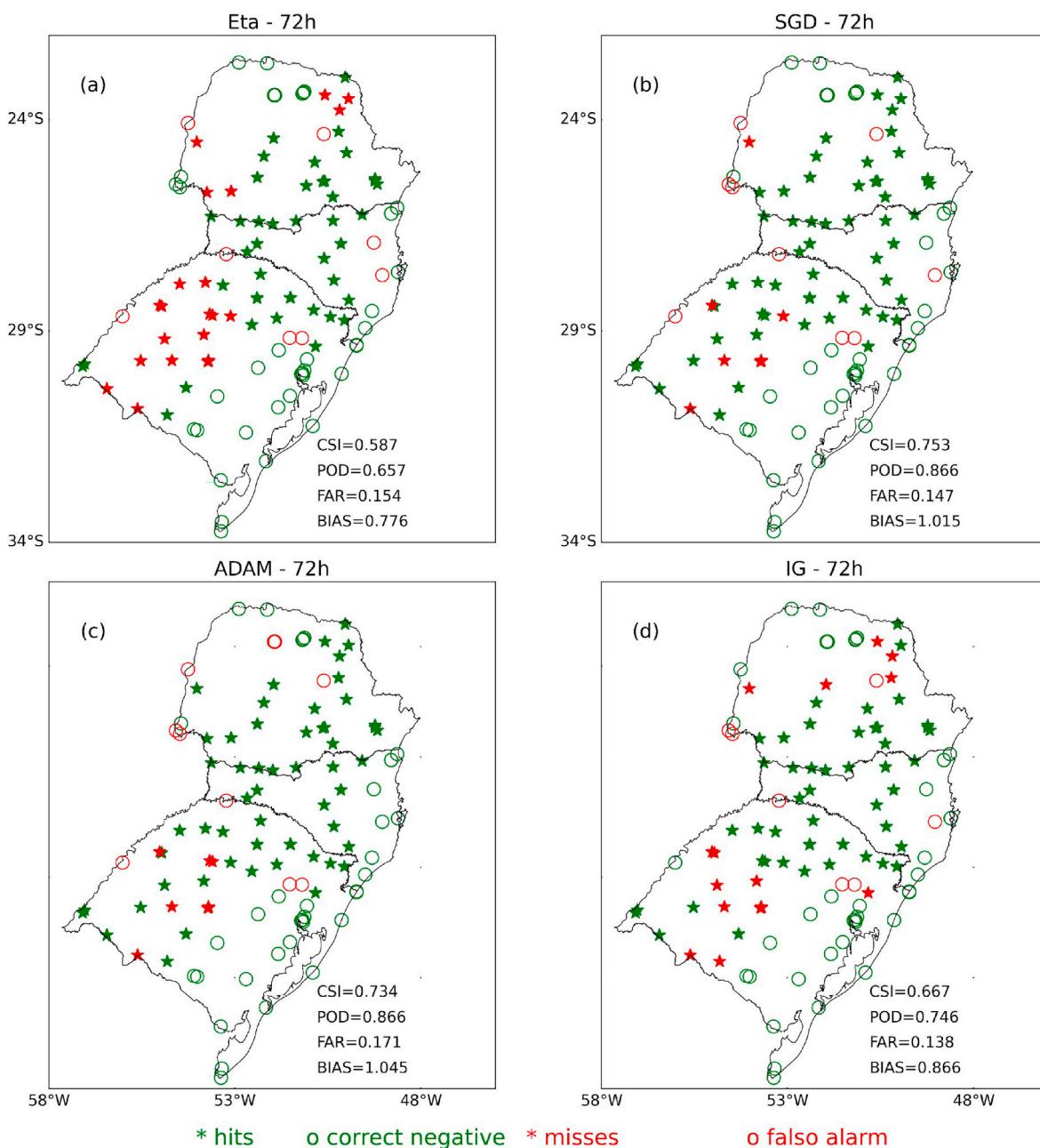


Fig. 14. Spatial distribution of errors (red) and successes (green) of Eta, SGD, ADAM, IG for 72 h before the event that occurred on 05/21/2018. The symbols (*) represent the occurrence of frost, while (o) the non-occurrence.

to the case study on May 21, 2018, validates the significant improvements observed during the test period, particularly in the western region of the state of Rio Grande do Sul. Considering the results from the test period, the case study, and the comparison between ANNs and the IG product (currently operational at INPE), it can be concluded that ANNs serve as highly valuable tools for predicting frost events. Consequently, they have the potential to replace the IG in INPE's operational routine.

CRediT authorship contribution statement

José Roberto Rozante: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Writing – original draft, Writing – review & editing. **Enver Ramirez:** Formal analysis, Writing – original draft, Writing – review & editing. **Diego Ramirez:** Software, Visualization. **Gabriela Rozante:** Software,

Visualization.

Acknowledgments

The authors would like to thank the anonymous reviewers and editorial staff for comments and suggestions that helped to improve the quality of the paper. The present research was performed within the scope of the CNPq's research group: *Clouds, Radiation and boundary layer interactions: from observations to numerical modeling with large volumes of data*. Special thanks to the INPE's Georeferenced Database Program (BIG) that made possible the present research being published.

References

- Alvares, C.A., Stape, J.L., Sentelhas, P.C., de Moraes Gonçalves, J.L., Sparovek, G., 2013. Köppen's climate classification map for Brazil. *Meteorol. Z.* 711–728. <https://doi.org/10.1127/0941-2948/2013/0507>.
- 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>.
- Arakawa, A., Lamb, V.R., 1977. Computational design of the basic dynamical processes of the UCLA general circulation model. In: Chang, J. (Ed.), *Methods in Computational Physics: Advances in Research and Applications*, vol. 17. Elsevier, pp. 173–265 of *General Circulation Models of the Atmosphere*.
- Black, T.L., 1994. The new NMC mesoscale Eta model: description and forecast examples. *Weather Forecast.* 9, 265–278. [https://doi.org/10.1175/1520-0434\(1994\)009<0265:TNNMEM>2.0.CO;2](https://doi.org/10.1175/1520-0434(1994)009<0265:TNNMEM>2.0.CO;2).
- Blanc, M.L., Geslin, H., Holzberg, I.A., Mason, B., 1963. Protection against frost damage. *World Meteorological Organization, Tech. Note 51 (WMO-No. 133.TP.60)*.
- Bochenek, B., Ustrnul, Z., 2022. Machine learning in weather prediction and climate analyses—applications and perspectives. *Atmosphere* 13, 180. <https://doi.org/10.3390/atmos13020180>.
- 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>.
- Cunha, J.M., 1952. Contribuição para o estudo do problema das geadas em Portugal. I.S. A., Lisboa.
- 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.
- 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/IoT.2018.2867333>.
- 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>.
- Diniz, E.S., Lorenzon, A.S., Castro, N.I.M., Marcatti, G.E., Santos, O.P., Júnior, 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>.
- Fagnani, M.A., Pinto, H.S., 1981. Simulação de temperaturas de folhas de cafeeiros em noites sujeitas a geadas de irradiação. In: *Congresso Brasileiro de Agrometeorologia. Rio Grande do Sul, Pelotas*, pp. 139–142.
- Fortune, M.A., Kousky, V.E., 1983. Two severe freezes in Brazil: precursors and synoptic evolution. *Mon. Weather Rev.* 111, 181–196. [https://doi.org/10.1175/1520-0493\(1983\)111<0181:TSFIBP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1983)111<0181:TSFIBP>2.0.CO;2).
- 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>.
- García, J.R.M., 2021. Explorando técnicas de aprendizado de máquina para aprimoramento da previsão de geadas no sul e sudeste do Brasil. In: *Anais do XIII Congresso Brasileiro de Agroinformática (SBIAGRO 2021)*, Congresso Brasileiro de Agroinformática. Sociedade Brasileira de Computação, Brasil, pp. 144–153.
- García Pedraza, L., Vega, J.G., 1991. Las Heladas de Irradiación en España. *Ministerio de Agricultura Pesca y Alimentación, Madrid. Hojas divulgadoras n.º 1/91.* 20pp.
- Guarnieri, R.A., Pereira, E., Chou, S., 2006. Solar radiation forecast using artificial neural networks in south Brazil. In: *Proceedings Of 8 ICSHMO, 8 Th International Conference on Southern Hemisphere Meteorology and Oceanography - 8 ICSHMO. Foz do Iguaçu, Brazil*, pp. 1777–1785.
- Hagiwara, M., 1994. A simple and effective method for removal of hidden units and weights. *Neurocomputing* 6, 207–218. [https://doi.org/10.1016/0925-2312\(94\)90055-8](https://doi.org/10.1016/0925-2312(94)90055-8).
- Hamilton, M.G., Tarifa, JoséR., 1978. Synoptic aspects of a polar outbreak leading to frost in tropical Brazil. *Mon. Weather Rev.* 106, 1545–1556. [https://doi.org/10.1175/1520-0493\(1978\)106<1545:SAOAO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1978)106<1545:SAOAO>2.0.CO;2). July 1972.
- Hewett, E.W., 1971. Preventing Frost Damage to Fruit Trees.
- Hogg, W.H., 1950. Frequency of radiation and wind frosts during spring in Kent. *Meteorol. Mag.*
- Hogg, W.H., 1971. Spring frosts. *Agriculture*.
- 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. Afr. 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>.
- Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization. <https://doi.org/10.48550/ARXIV.1412.6980>.
- Koutsoukas, A., Monaghan, K.J., Li, X., Huan, J., 2017. Deep-learning: investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data. *J. Cheminf.* 9, 42. <https://doi.org/10.1186/s13321-017-0226-y>.
- 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>.
- Li, Jin-Yan, Chow, T.W.S., Yu, Ying-Lin, 1995. The estimation theory and optimization algorithm for the number of hidden units in the higher-order feedforward neural network. In: *Proceedings of ICNN'95 - International Conference on Neural Networks*, vol. 3. *ICNN'95 - International Conference on Neural Networks, Perth, WA, Australia*, pp. 1229–1233. IEEE.
- Li, Y., Lang, J., Ji, L., Zhong, J., Wang, Z., Guo, Y., He, S., 2021. Weather forecasting using ensemble of spatial-temporal attention network and multi-layer perceptron. *Asia-Pac. J. Atmospheric Sci.* 57, 533–546. <https://doi.org/10.1007/s13143-020-00212-3>.
- Lima, F.J.L., Martins, F.R., Pereira, E.B., Lorenz, E., Heinemann, D., 2016. Forecast for surface solar irradiance at the Brazilian Northeastern region using NWP model and artificial neural networks. *Renew. Energy* 87, 807–818. <https://doi.org/10.1016/j.renene.2015.11.005>.
- Lira, H., Martí, L., Sanchez-Pi, N., 2022. A graph neural network with spatio-temporal attention for multi-sources time series data: an application to frost forecast. *Sensors* 22, 1486. <https://doi.org/10.3390/s22041486>.
- Lu, Y., Hu, Y., Li, P., Paw U, K.T., Snyder, R.L., 2019. Prediction of radiation frost using Support vector machines based on micrometeorological data. *Appl. Sci.* 10, 283. <https://doi.org/10.3390/app10010283>.
- Lucyri, V., Nunes, M.D., Reboita, M.S., Lemes, M.D.C.R., 2020. Validação das previsões de três ondas de frio pelo modelo GFS no Centro-Norte do Estado de São Paulo e Triângulo Mineiro. *Rev. Bras. Geogr. Física* 13. <https://doi.org/10.26848/rbgf.v13.5.p1994-2018>, 1994.
- Marengo, J.A., Camargo, C.C., 2008. Surface air temperature trends in Southern Brazil for 1960–2002. *Int. J. Climatol.* 28, 893–904. <https://doi.org/10.1002/joc.1584>.
- Margolis, 1979. Green gold and ice: the impact of frost on the coffee growing region of Northern Paraná, Brazil. *Mass Emergencies* 4.
- Mesinger, F., Janjic, Z.I., Nićković, S., Gavrilov, D., Deaven, D.G., 1988. The step-mountain coordinate: model description and performance for cases of alpine lee cyclogenesis and for a case of an appalachian redevelopment. *Mon. Weather Rev.* 116, 1493–1518. [https://doi.org/10.1175/1520-0493\(1988\)116<1493:TSMCMD>2.0.CO;2](https://doi.org/10.1175/1520-0493(1988)116<1493:TSMCMD>2.0.CO;2).
- Müller, G.V., Ambrizzi, T., 2007. Teleconnection patterns and Rossby wave propagation associated to generalized frosts over southern South America. *Clim. Dynam.* 29, 633–645. <https://doi.org/10.1007/s00382-007-0253-x>.
- Müller, G.V., Berri, G.J., 2007. Atmospheric circulation associated with persistent generalized frosts in central-southern south America. *Mon. Weather Rev.* 135, 1268–1289. <https://doi.org/10.1175/MWR3344.1>.
- Müller, G.V., Ambrizzi, T., Núñez, M.N., 2005. Mean atmospheric circulation leading to generalized frosts in central southern South America. *Theor. Appl. Climatol.* 82, 95–112. <https://doi.org/10.1007/s00704-004-0107-y>.
- Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted Boltzmann machines. In: *Proceedings of the 27th International Conference on Machine Learning*, Haifa, Israel, Johannes Fürnkranz, Thorsten Joachims, pp. 807–814.
- Nitta, T., 2017. Resolution of singularities introduced by hierarchical structure in deep neural networks. *IEEE Transact. Neural Networks Learn. Syst.* 28, 2282–2293. <https://doi.org/10.1109/TNNLS.2016.2580741>.
- Noh, I., Doh, H.-W., Kim, S.-O., Kim, S.-H., Shin, S., Lee, S.-J., 2021. Machine learning-based hourly frost-prediction system optimized for orchards using automatic weather station and digital camera image data. *Atmosphere* 12, 846. <https://doi.org/10.3390/atmos12070846>.
- Orvando, G., Bocco, M., Sayago, S., 2005. Redes NEURONALES para MODELAR predicción de HELADAS. *Agric. Tec. (Santiago)* 65. <https://doi.org/10.4067/S0365-28072005000100007>.
- Parmenter, F.C., 1976. A southern hemisphere cold front passage at the equator. *Bull. Am. Meteorol. Soc.* 57, 1435–1440. [https://doi.org/10.1175/1520-0477\(1976\)057<1435:ASHCFP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1976)057<1435:ASHCFP>2.0.CO;2).
- 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>.
- Robinson, C., Mort, N., 1997. A neural network system for the protection of citrus crops from frost damage. *Comput. Electron. Agric.* 16, 177–187. [https://doi.org/10.1016/S0168-1699\(96\)00037-3](https://doi.org/10.1016/S0168-1699(96)00037-3).
- Roebber, P.J., 2009. Visualizing multiple measures of forecast quality. *Weather Forecast.* 24, 601–608. <https://doi.org/10.1175/2008WAF2222159.1>.
- Rogers, J.C., Rohli, R.V., 1991. Florida citrus freezes and polar anticyclones in the great plains. *J. Clim.* 4, 1103–1113. [https://doi.org/10.1175/1520-0442\(1991\)004<1103:FCFAPA>2.0.CO;2](https://doi.org/10.1175/1520-0442(1991)004<1103:FCFAPA>2.0.CO;2).
- 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., Ramirez, E., Fernandes, A. de A., 2022. A newly developed South American Mapping of Temperature with estimated lapse rate corrections. *Int. J. Climatol.* 42, 2135–2152. <https://doi.org/10.1002/joc.7356>.
- Ruder, S., 2016. An Overview of Gradient Descent Optimization Algorithms. <https://doi.org/10.48550/ARXIV.1609.04747>.
- Rumelhart, D.E., James, L.M., 1988. Parallel Distributed Processing. & PDP Research Group, pp. 354–362.

- Rumelhart, D.E., McClelland, J.L., 1987. Learning internal representations by error propagation. In: Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations. MIT Press, pp. 318–362.
- Satyamurty, P., Da Conceição Ferreira, C., Alonso Gan, M., 1990. Cyclonic vortices over south America. *Tellus Dyn. Meteorol. Oceanogr.* 42, 194–201. <https://doi.org/10.3402/tellusa.v42i1.11870>.
- Savage, M., 2012. Estimation of frost occurrence and duration of frost for a short-grass surface. *S. Afr. J. Plant Soil* 29, 173–181. <https://doi.org/10.1080/02571862.2012.748938>.
- Taboada, M.A., Costantini, A.O., Bustos, M., Bonatti, M., Sieber, S., 2021. Climate change adaptation and the agricultural sector in South American countries: risk, vulnerabilities and opportunities. *Rev. Bras. Ciênc. Solo* 45, e0210072. <https://doi.org/10.36783/18069657rbcs20210072>.
- 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>.
- Talsma, C.J., Solander, K.C., Mudunuru, M.K., Crawford, B., Powell, M.R., 2023. Frost prediction using machine learning and deep neural network models. *Front. Artif. Intell.* 5, 963781 <https://doi.org/10.3389/frai.2022.963781>.
- Valverde Ramírez, M.C., De Campos Velho, H.F., Ferreira, N.J., 2005. Artificial neural network technique for rainfall forecasting applied to the São Paulo region. *J. Hydrol.* 301, 146–162. <https://doi.org/10.1016/j.jhydrol.2004.06.028>.
- Verdes, P.F., Granitto, P., Navone, H., Ceccato, H., 2000. Frost prediction with machine learning techniques. In: Proceedings of the VIth Argentine Congress on Computer Science, pp. 1423–1433. Proceedings of the VIth Argentine Congress on Computer Science, Argentine.
- Wassan, S., Xi, C., Jhanjhi, N., Binte-Imran, L., 2021. Effect of frost on plants, leaves, and forecast of frost events using convolutional neural networks. *Int. J. Distributed Sens. Netw.* 17, 1550147721105377. <https://doi.org/10.1177/1550147721105377>.
- Wilks, D.S., 2019. *Statistical Methods in the Atmospheric Sciences*. Elsevier.
- Wrege, M.S., Fritzsoms, E., Soares, M.T.S., Prela-Pântano, A., Steinmetz, S., Caramori, P. H., Radin, B., Pandolfi, C., 2018. Risco de ocorrência de geadas na região centro-sul do brasil. *Rev. Bras. Climatol.* 22 <https://doi.org/10.5380/abclima.v22i0.57306>.
- Xu, J., Guga, S., Rong, G., Riao, D., Liu, X., Li, K., Zhang, J., 2021a. Estimation of frost hazard for tea tree in zhejiang province based on machine learning. *Agriculture* 11, 607. <https://doi.org/10.3390/agriculture11070607>.
- Xu, J., Guga, S., Rong, G., Riao, D., Liu, X., Li, K., Zhang, J., 2021b. Estimation of frost hazard for tea tree in zhejiang province based on machine learning. *Agriculture* 11, 607. <https://doi.org/10.3390/agriculture11070607>.
- Zendehboudi, A., Hosseini, S.H., 2019. Modeling of the frost deposition by natural convection on horizontal ultra-low-temperature surfaces. *J. Therm. Anal. Calorim.* 137, 2029–2043. <https://doi.org/10.1007/s10973-019-08087-x>.