FROST PREDICTION WITH MACHINE LEARNING TECHNIQUES

P. F. Verdes, P. M. Granitto, H. D. Navone, and H. A. Ceccatto

Instituto de Física Rosario (CONICET-UNR)

Bvard. 27 de Febrero 210 Bis, 2000 Rosario, Argentina

ABSTRACT: Frost is the condition that exists when the temperature of the earth's surface and earthbound objects falls below freezing (0°C). These events may have serious consequences on crop production, so actions must be taken to minimize damaging effects. In particular, temperature predictions are of much help in frost protection decisions by providing the hortoculturist with warnings of critical temperatures. Consequently, reliable temperature prediction methods would be of significant economic value. These methods are usually empirical formulae based on significant correlations between the minimum temperature observed towards the end of the night and one or several meteorological variables measured at least several hours before this minimum is reached. These empirical regressions employ antecedent air temperature, various humidity measures, wind speed and cloud cover values. In this work we explore the possibility of developing an empirical prediction system for frost protection of fruits and vegetables in the southern part of the Santa Fe province in Argentina, in the region covered by the agrometeorological station located at Zavalla (33°01′S, 60°53′W). To this end we consider a handful of Machine Learning techniques usually employed in regression and classification problems, including Artificial Neural Networks, Simple Bayes classifiers and *k*-Nearest Neighbors. The results obtained in this preliminary study reveal a very noisy structure of the data that allows for only a slight improvement in performance of some of these more sophisticated nonlinear techniques over the standard (linear) multivariate regression equations.

KEYWORDS: Frost Prediction, Machine Learning, Regression, Classification

INTRODUCTION

Frost is defined in the Glossary of Meteorology as the condition that exists when the temperature of the earth's surface and earthbound objects falls below freezing (0°C). These events are usually classified as *advective* or *radiation* frosts or as a combination of these two types [1]. Advective frosts develop during day or night as a result of a large-scale incursion of cold air, often from polar regions, and are characterized by moderate to strong winds and a well-mixed atmosphere. Radiation frosts occur at night and result from intense, longwave radiation cooling under calm, clear and dry atmospheric conditions. In this case the terrestrial radiation to space is relatively unimpeded because of the absence of clouds and heavy concentration of water vapor, and strong surface inversions develop in the stable atmosphere.

Temperature is one of the primary factors influencing plant growth and its geographical distribution. In particular, frosts may have serious effects on crop production. The minimum temperature below which damage is caused varies with plant species and with growth stage for a given plant [2]. For advective frosts the temperature of the air mass may be well below 0°C and a large-scale protective action is often not feasible. These frosts are an agricultural problem only in high latitude regions and at high elevation elsewhere. Radiation frosts are particularly dangerous for agriculture and horticulture in regions between 15°S and 40°S and between 12°N and 40°N, where relatively infrequent short duration events may cause significant losses. Actions must be taken in the planning and operational stages of agriculture to minimize such damaging effects. A wide range of techniques is used in agriculture, and particularly in horticulture, to optimize the physical crop environment by altering the thermal regime of the air layer near the ground and by reducing long-wave radiation loss from soil and plants [1,2]. Heaters, burning solid, liquid or gaseous fuels, modify the crop environment by internal generation of heat. Wind machines consisting of large fans on towers produce a mixing of cold air within an orchard with the warmer air aloft. They change the sensible heat fluxes at the top of the box. The use of artificial fog is aimed at reducing the longwave radiation loss to the sky. Finally, irrigation with overhead sprinklers adds to the crop surface the fusion heat released from the freezing ice-water film on the plant surfaces. Sprinkler irrigation changes the latent and sensible heat transfer between the crop surface and the air in the box. It must be stressed, however, that most of these active methods for frost protection have significant energy requirements.

Temperature predictions aid in frost protection decisions by providing the hortoculturist with warnings of critical temperatures. Besides the negative environmental impact, the high cost of heating oil and wind machine fuel, and the use of limited water resources are incentives to conserve these resources, and readily available temperature predictions may help to save hours worth of fuel or water consumption. Frost prediction methods have been reviewed by several authors (see, for instance, [3,4]). One usually distinguishes between empirical formulae and semi-empirical or theoretical expressions. Empirical formulae are generally proposed when a high correlation is observed between the minimum temperature observed towards the end of the night and one or several meteorological variables measured at least several hours before this minimum is reached. These empirical regressions employ antecedent air temperature, various humidity measures, wind speed and cloud cover values. These values are obtained at fixed times, often at sunset or at a time when the variables do not change too rapidly. Cellier[4] gives a useful summary of a wide range of empirical equations that employ similar variables. The differences between these equations indicate that many expressions have limited general applicability. In general, such empirical expressions may yield satisfactory results if they are used by competent meteorologists with a good knowledge of local conditions. On the other hand, semi-empirical/theoretical expressions result from a more or less simple analysis of the energy balance and heat transfer processes near the surface. In any case, reliable frost prediction methods to help in the decision-making problem of frost protection would be of significant economic value[5].

In this work we aim to explore the possibility of developing an empirical prediction system for frost protection of fruits and vegetables in the southern part of the Santa Fe province in Argentina, in the region covered by the agrometeorological station located at Zavalla (33°01′S, 60°53′W). To this end we will consider a handful of Machine Learning techniques usually employed in regression and classification problems, including Artificial Neural Networks (ANN), Simple Bayes (SB) classifiers and *k*-Nearest Neighbors (*k*-NN) [6]. The results obtained in this preliminary study point to a very noisy structure of the data, which allows for only a slight improvement in the performance of some of these more sophisticated nonlinear techniques over the standard (linear) multivariate regression equations. Further investigations of the problem should include both the use of the most recent data not available for this study and, most importantly, new predictor variables that could help extracting more accurate nonlinear information. This work is organized as follows. First, in Section 2 we present a brief time-series analysis of the data that will be useful to select suitable temperature predictors. Then, in Section 3 we describe the prediction methods used and in Section 4 we discuss the results obtained. Finally, in Section 5 we draw some conclusions.

2. MINIMUM-TEMPERATURE TIME SERIES ANALYSIS

In order to remove seasonal variations in the minimum-temperature time series we obtained the mean annual wave using all the registers available (daily values between 1973 and 1989). Then, we fit this annual wave using a sinusoidal law, which is the behavior approximately followed by the solar irradiance, and finally subtracted it from the original data (see Fig. 1). To determine whether the residual series had a deterministic component or was purely stochastic in character, we generated 19 surrogates –that is, series with identical Fourier power spectrum but random phases– and tested the null hypothesis that the residual series is simply a linear and stationary Gaussian stochastic process [7]. To this end we used a discriminant statistics based in the prediction error of a nearest-neighbor predictor in pseudophase spaces of different dimensions. If this error were smaller for the original series than for the surrogates, then the hypothesis of a simple stochastic process can be rejected with a 95% level of confidence. However, in this case the prediction error for the original series was smaller than for the surrogates only for embedding dimensions D=2,3, loosing this property for larger D. Such behavior is observed for autocorrelated series, with persistence in their trends, so that it cannot be taken as a clear signature of determinism[8]. Consequently, we assumed that the minimum-temperature record corresponds to some stochastic process and sought for suitable predictor variables as described next.

Besides the minimum-temperature times series, the daily records (taken at 8 p.m.) of four other surface variables (wet and dry bulb temperatures, cloud cover and wind intensity) were also available. Aiming to determine possible predictors, we studied the correlation between minimum temperatures and these surface variables. We finally selected all the variables having n-step correlations with minimum temperatures larger than 0.2, a threshold that optimized the performance of the prediction methods described in the next section. This criterion filtered the following prediction variables: wet and dry bulb temperatures and cloud cover at 8 p.m. of the previous day, and minimum temperatures of the two previous days (the deseasonalized temperature series retains a fairly large persistence, with a 2-step autocorrelation of approximately 0.35). Notice, however, that this analysis is based on linear correlations among the series; a more sophisticated analysis should use a nonlinear criterion based, for instance, on mutual information [9].

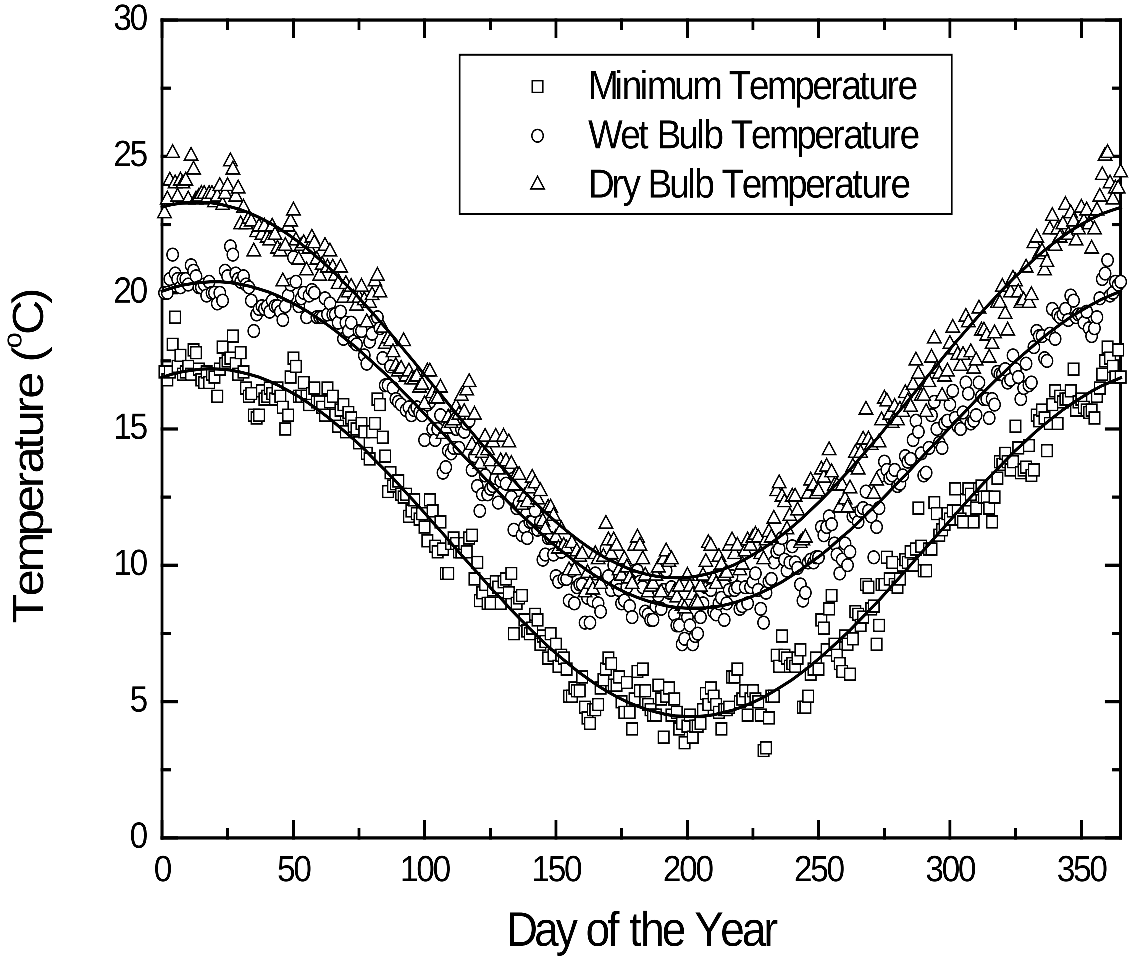

Fig.1: Mean annual waves corresponding to minimum, wet bulb and dry bulb temperatures. The full lines are fits using sinusoidal laws.

3. PREDICTION METHODS

As stated in the introduction, we will consider three different prediction methods to forecast the occurrence of frosts: Artificial Neural Networks (ANN), Simple Bayes (SB) classifiers and *k*-Nearest Neighbors (*k*-NN), using in all cases the same predictor variables selected in the previous section. Furthermore, we will attempt to forecast only whether at night the temperature will drop below 0ºC (frost) or not –a binary classification task– in two different ways: i) by predicting first the minimum temperature values –a standard regression problem– and then transforming the answer to a frost-no frost forecast, and ii) by facing the problem from the outset as a classification task. In the following we briefly describe the techniques used; a good introduction to these methods can be found in [6].

ANN are computational architectures of significant importance in the analysis and prediction of time series[10]. In the last years they have been applied to many problems, including synthetic data from chaotic maps[11], real-world time series associated to atmospheric and climatological phenomena[12], economic indicators[13], etc. In our case we have first trained feedforward ANN having 5 input units (corresponding to the five predictor variables used), 1 output unit (for the temperature forecast), and alternatively 2 to 50 hidden units. After the networks learned the map predictor variables-minimum temperature, using the quadratic error as cost function in the learning process, the output was converted to a binary frost-no frost answer. In the next section we will present results corresponding to the 5:30:1 network, but the performance is only very slightly dependent on the number of hidden units. Alternatively, we trained ANN with softmax activation functions in the output units and cross-entropy cost function to tackle the problem directly as a classification task [6]. The results presented in this case correspond to networks with 2 hidden units. In the following we will identify the results corresponding to these two approaches by NR (for Neural Regressor) and NC (for Neural Classifier) respectively.

SB (also called Naive Bayes) is a simple technique based on a probabilistic description of the data that is optimal for supervised learning if the attributes $\mathbf{x}=(x_1,...,x_d)$ used to classify patterns (predictor variables in our case) are independent. In such a case, the class-conditional probabilities factorizes, $P(\mathbf{x}|C)=\prod_{i=1,d} P(x_i|C)$, and a simple fitting of $P(x_i|C)$ to a normal distribution in most cases produces very good results. Notwithstanding the fact that the independence condition is almost always violated, often the method remains competitive against more sophisticated techniques[14]. Here we will see that in spite of its simplicity SB is very competitive for this problem.

In the *k*-NN method, an instance-based approach, one searches the database for *k* states of the system similar to the current one, and makes a temperature prediction using a suitable combination of the known outcomes corresponding to these similar situations. In our implementation we considered two different ways of finding the past states closest to the present one: i) the standard approach (S*k*, for Standard *k*-NN), in which the metric used to find the nearest neighbors weighs all the predictors equally, and ii) a variant in which we used a metric that weighs every predictor variable according to its correlation with the minimum-temperature time series. Since all records were previously normalized to a series with zero mean and standard deviation $\sigma=1$, this last procedure (called M*k*, for Modified *k*-NN) naturally introduces the consideration of the relative importance of different predictors. In this last case good results are obtained using $k \sim 100$ nearest-neighbor states. However, as we will discuss in the next section, the best results provided by this technique correspond to equal weighing of all predictors and an average over a very large number of neighbors ($k \sim 1000$), which essentially recovers standard multivariate regression results for this problem [15].

In all cases we have used 2601 daily records of minimum temperatures and other surface variables corresponding to the period May-September of years 1973-1989. This record was randomly split in training and test sets, using 75% of the patterns for training and the remaining 25% to forecast frost occurrences. We kept the same ratio between number of frosts to total number of registers in both sets, which corresponds to 1951 records in the learning set –comprising 196 frost events– and 650 records with 65 frost events in the test set. Data were provided by the agrometeorological station located in Zavalla, Argentina (33º01′S, 60º53′W). The results obtained with the different methods above described are presented in the next section.

4. RESULTS

For the evaluation of results we considered the usual contingency table:

		PREDICTED	
		YES	NO
OBSERVED	YES	*a*	*b*
	NO	*c*	*d*

from which one can define the following performance indices:

Probability Of Detection: $POD = a/(a+b)$

False Alarm Ratio: $FAR = c/(a+c)$

Critical Success Index: $CSI = a/(a+b+c)$

Percent Correct: $PC = (a+d)/(a+b+c+d)$

Probability Of False Detection: $POFD = c/(c+d)$

To evaluate the economic impact of protecting the cultivar on the basis of the frost forecasts, an economic measure like, for instance, $E=\alpha POFD+\beta(1\text{-}POD)$, should be considered. Here the coefficients α and β represent, respectively, the cost of protecting the plants and the estimated crop loss due to failing to protect them in case of a frost. Of course, much more sophisticated models can be defined[5], including those that allow for changes in coefficients depending on the growth stage. In all cases, the prediction method should be selected as the one that minimizes *E*. Since we do not have reliable economic information available, in order to provide a simple measure in some of the tables below we quote the values of *E* for arbitrary coefficients α=1 and β=5.

We have initially considered two possible scenarios: the first one corresponding to a strict frost prediction when the temperature is expected to drop below 0°C and the second one to an alarm indication when it is predicted to drop below 1°C. Tables I and II are, respectively, the contingency tables for these two cases, giving both separately the corresponding results for the learning and test sets. In addition, Tables III and IV present the values of the performance indices defined above. In all cases we have included the results obtained with the so-called Persistent Predictor (PP), which corresponds to asserting that tomorrow's minimum temperature will be equal to the one observed today. In spite of its simplicity, for largely autocorrelated series this predictor sets a stringent condition on the performance of successful learning methods and it is used here as a baseline. Notice also that the SB and NC methods have not been applied for the second scenario (alarm indication) due to their binary output (they do not regress temperature values as an intermediate step like the other methods).

The analysis of the above tables indicates that all the classification methods considered largely improve on the persistent predictor. In particular, for the strict frost prediction SB has an edge over ANN and *k*-NN according to the *POD* index, although these two last methods produced substantially fewer false alarms (much smaller *POFD*). It is also interesting to note that setting the threshold for alarm indication T_{alarm}=1°C leads to much better performances of ANN and *k*-NN though, as expected, this slightly deteriorates their *POFD* indices. The general improvement can be appraised from the combined index *E*, but this has to be taken with a grain of salt because of the arbitrary assumption made in defining this quantity.

Table I: Contingency tables for the learning and test sets corresponding to the different methods used, with an alarm threshold T_{alarm}=0°C.

LEARNING SET		PREDICTED			
		$T \le 0$		$T > 0$	
OBSERVED	$T \le 0$	PP	85	PP	111
		SB	155	SB	41
		NR	117	NR	79
		NC	125	NC	71
		M*k*	51	M*k*	145
		S*k*	56	S*k*	140
	$T > 0$	PP	65	PP	1690
		SB	113	SB	1642
		NR	12	NR	1743
		NC	30	NC	1725
		M*k*	2	M*k*	1753
		S*k*	2	S*k*	1753

TEST SET		PREDICTED			
		$T \le 0$		$T > 0$	
OBSERVED	$T \le 0$	PP	32	PP	33
		SB	55	SB	10
		NR	33	NR	32
		NC	39	NC	26
		M*k*	16	M*k*	49
		S*k*	19	S*k*	46
	$T > 0$	PP	38	PP	547
		SB	43	SB	542
		NR	10	NR	575
		NC	15	NC	570
		M*k*	2	M*k*	583
		S*k*	2	S*k*	583

Table II: Same as Table I but using T_{alarm}=1°C.

LEARNING SET		PREDICTED			
		$T \le 1$		$T > 1$	
OBSERVED	$T \le 0$	PP	104	PP	92
		NR	145	NR	51
		M*k*	115	M*k*	81
		S*k*	116	S*k*	80
	$T > 0$	PP	101	PP	1654
		NR	39	NR	1716
		M*k*	20	M*k*	1735
		S*k*	17	S*k*	1738

TEST SET		PREDICTED			
		$T \le 1$		$T > 1$	
OBSERVED	$T \le 0$	PP	35	PP	30
		NR	44	NR	21
		M*k*	40	M*k*	25
		S*k*	41	S*k*	24
	$T > 0$	PP	62	PP	523
		NR	23	NR	562
		M*k*	11	M*k*	574
		S*k*	11	S*k*	574

Table III: Performance indices for T_{alarm}=0°C.

0°C	*POD*	*FAR*	*CSI*	*PC*	*POFD*	*E*
PP	0.49	0.54	0.31	0.89	0.06	2.61
SB	0.85	0.44	0.51	0.92	0.07	0.82
NR	0.51	0.23	0.44	0.94	0.02	2.47
NC	0.60	0.28	0.49	0.94	0.03	2.03
M*k*	0.25	0.11	0.24	0.92	0.003	3.75
S*k*	0.29	0.10	0.28	0.93	0.003	3.55

Table IV: Same as Table III but with T_{alarm}=1°C

1°C	*POD*	*FAR*	*CSI*	*PC*	*POFD*	*E*
PP	0.54	0.64	0.28	0.86	0.11	2.41
NR	0.68	0.34	0.50	0.93	0.04	1.64
M*k*	0.62	0.22	0.53	0.94	0.02	1.92
S*k*	0.63	0.21	0.54	0.95	0.02	1.87

In view of the above-discussed results, which point to an important role of the temperature threshold T_{alarm} for regression methods, we have considered the ROC curves that can be constructed by varying this threshold (see Fig. 1). In practice, it is enough to move T_{alarm} approximately in the range (-2,3)°C to obtain the whole curve in this figure. As can be seen, disregarding small statistical fluctuations NR, S*k* and M*k* have essentially the same performance.

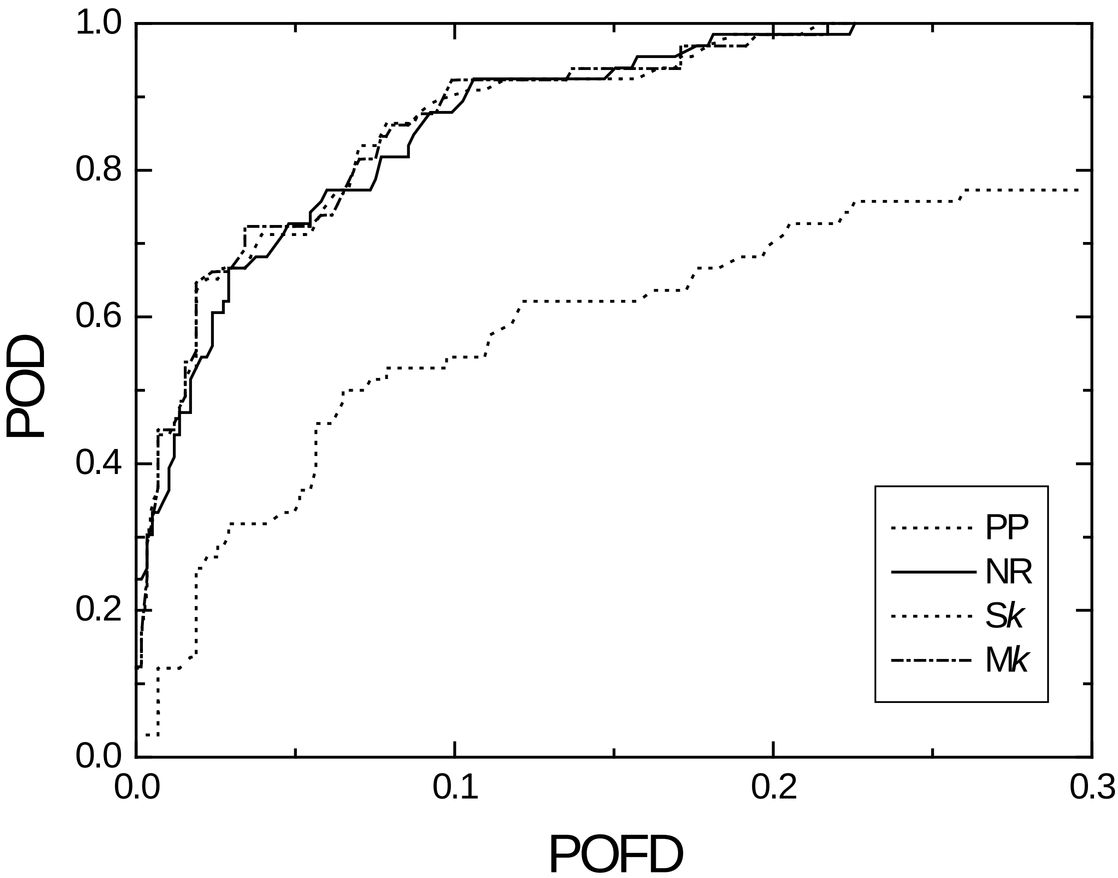

Fig. 2: ROC curves corresponding to the Neural Regressor (NR) and Standard (S*k*) and Modified (M*k*) *k*-Nearest Neighbor methods. For the sake of comparison we also include the results for the Persistent Predictor (PP).

To generate similar ROC curves for classification methods we considered the alternative of introducing a risk function, which modifies the relative importance of the more scarce but damaging frost events against the abundant no-frost pattern examples. For the SB classifier the implementation of this procedure is immediate since, according to the Bayes theorem, it is enough to rescale the class prior probabilities. Similarly, for the NC we chose to introduce in the cross-entropy cost function a coefficient that multiplies the errors coming from undetected frost events, so that for large values of this coefficient the network will fit better these patterns than those corresponding to non-frost events. The ROC curves generated in this way are given in Fig. 3. Notice that the NC performance is given as scattered points that correspond to the results of different network training experiments; the observed dispersion is, however, not important.

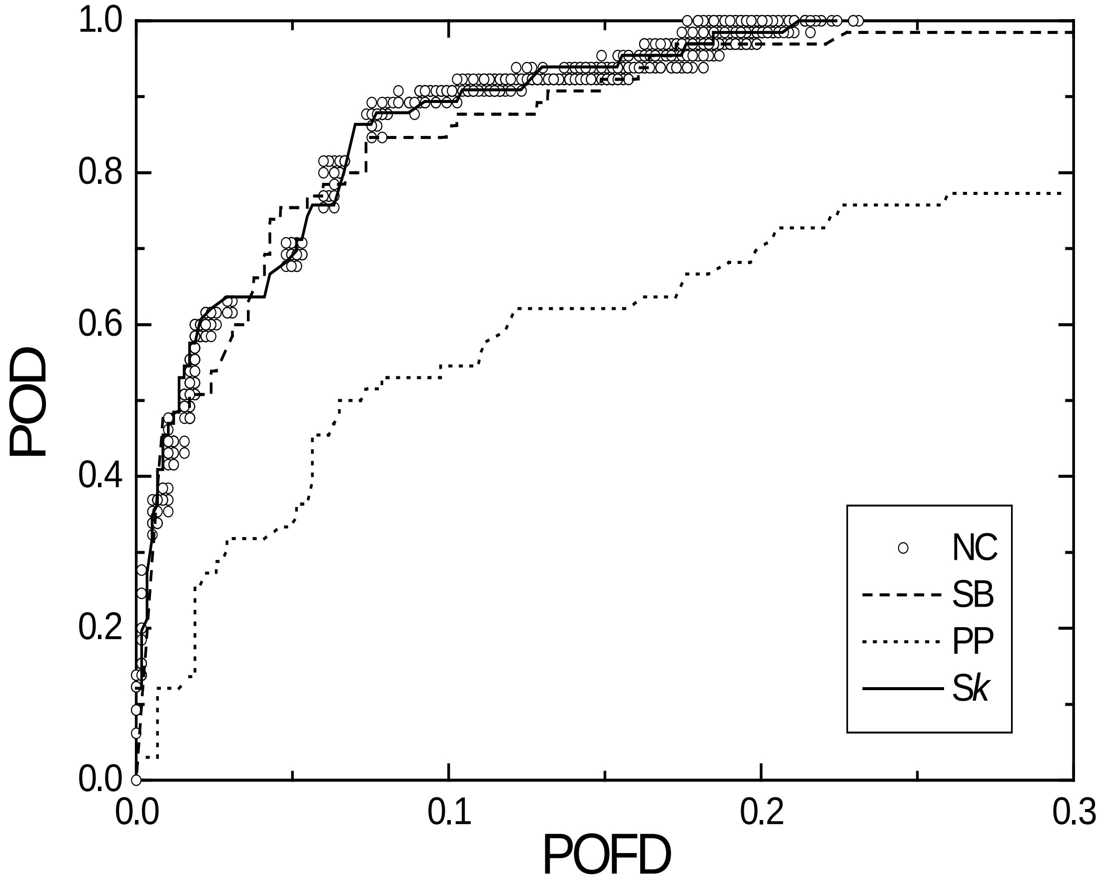

Fig. 3: ROC curve corresponding to the classification methods Neural Classifier (NC) and Simple Bayes (SB). For the sake of comparison we also include the results for the Persistent Predictor (PP) and Standard *k*-Nearest Neighbors (S*k*).

5. Conclusions

We have performed a preliminary study exploring the possibility of developing an empirical prediction system for frost protection of fruits and vegetables in the southern part of the Santa Fe province in Argentina. In this study we applied several Machine Learning techniques usually employed in regression and classification problems, including Artificial Neural Networks, Simple Bayes and *k*-Nearest Neighbors classifiers. The results obtained point to a very noisy structure of the data, which allows for only a slight improvement in the performance of some of these more sophisticated nonlinear techniques over the standard (linear) multivariate regression equations. Further investigations of the problem should include both the use of the most recent data not available for this study and, most importantly, new predictor variables that could help extracting more accurate nonlinear information. Another possibility, currently under investigations, is the modification of the training set for ANN by excluding obvious non-frost patterns, in order to allow the network to focus on the discrimination of close frost-no frost situations. This can be achieved, for instance, by considering the cloud cover and eliminating those days with extreme values that will not allow the occurrence of *radiation* frost. Notice that this could also be used to eliminate the few *advective* frosts included in the database, which can be predicted using different indicators. Finally, it could be interesting to consider ANN with a much larger number of hidden units, which would certainly overfit the training set. This can be exploited to build regression machines by ensembling diverse ANN predictors, thus reducing the error by maximizing the ensemble variance[16].

ACKNOWLEDGEMENTS

We acknowledge M.A. Lara, O. Sacchi and collaborators for providing us with the data used in this work and for useful discussions. This work was partially financed by ANPCyT through the grant PICT98 11-03834.

6. REFERENCES

[1] J.D. Kalma, G.P. Laughlin, J.M. Caprio & P.J.C. Hamer, *The Bioclimatology of Frost: Its ocurrence, Impact and Prediction* (Springer-Verlag, Berlin, 1992).

[2] N.J. Rosenberg, B.L. Blad & S.B. Verma, *Microclimate The Biological Environment* (John Wiley & Sons, New York, 1983).

[3] J.G. George, *Techniques of frost prediction*, in *Techniques of Frost Prediction and Methods of Frost and Cold Protection*, J.G. George & J.F. Gerber, Eds., WMO **2**, 45 (1978).

[4] P. Cellier, *Contribution a la prevision des temperatures minimales nocturnes en conditions de gelees de Printemps. Etude de l'evolution des temperatures de l'air et du sol au cours de la nuit*, Thesis Doct. , INRA (1982).

[5] F.S. Conklin, A.E. Baquet & A.N. Halter, *A Bayesian Simulation Approach for Estimating Value of Information: An Application to Frost Forecasting*, Technical Bulletin Nr. 196, Agricultural Experiment Station, Oregon State University, Corvallis, OR. (May 1977).

[6] C.M. Bishop, *Neural Networks for Pattern Recognition* (Clarendon Press, Oxford, 1995).

[7] T. Schreiber & H. Kantz, *Nonlinear Times Series Analysis* (Cambrigde University Press, 1998).

[8] M. Casdagli & S. Eubank, Eds., *Nonlinear Modeling and Forecasting* (Addison-Wesley, Reedwod City, 1992).

[9] A.M. Fraser & H.L. Swinney, Independent coordinates for strange attractors from mutual information, Physical Review A **33**: 1134-1140 (1986).

[10] A.S. Weigend & N.A. Gershenfeld, Eds., *Time Series Prediction: Forecasting the Future and Understanding the Past* (Addison-Wesley, Reedwod City, 1994).

[11] P.F. Verdes, P.M. Granitto, H.D. Navone & H.A. Ceccatto, *Forecasting Chaotic Time Series: Global vs. Local Methods*, in *Novel Intelligent Automation and Control Systems, Vol. 1*, J.Pfeiffer, Eds, ALFA-NIACS, 129-145 (1998) **.**

[12] H.D. Navone & H.A. Ceccatto, *Predicting Indian Monsoon Rainfall: A Neural Network Approach*, Climate Dynamics **10**: 305-312 (1994).

[13] E.M. Azoff, *Neural Network Time Series Forecasting of Financial Markets* (John Wiley & Sons, Chichester, 1994).

[14] P. Domingos & M. Pazzani, *Beyond Independence: Conditions for the Optimality of the Simple Bayesian Classifier*, Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (Morgan-Kauffman Publishers, San Mateo, 1995).

[15] M.A. Lara, A. Coronel, O. Sacchi & M. Costanzo, *Incidencia de condiciones ambientales sobre heladas en el área rural de Zavalla*, Actas ASADES **95**: 4-11 (1995) (In Spanish); A. Coronel, O. Sacchi, V. Bisaro, M. Costanzo, M.A. Lara & S. Bischoff, *Modelo de regresión aplicado a las temperaturas mínimas registradas en la estación agrometeorologica de Zavalla* (private communication).

[16] H.D. Navone, P.F. Verdes, P.M. Granitto & H.A. Ceccatto, *A New Algorith for Selecting Diverse Members of a Neural Network Ensemble*, Proceedings VI Internacional Congress on Information Engineering, Buenos Aires, Argentina (2000).