

# Predicción de temperaturas mínimas utilizando otras variables de superficie como predictores

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

*IFIR (Instituto de Física Rosario, CONICET-UNR)*

*Bv. 27 de Febrero 210 bis - (2000) Rosario - Argentina*

*e-mail: verdes@ifir.edu.ar*

La predicción de heladas tiene gran importancia práctica, especialmente durante primavera y otoño. En este trabajo se analizan las series temporales de temperaturas mínimas diarias y otras variables de superficie (temperaturas de bulbo seco y húmedo, humedad relativa, dirección e intensidad del viento y nubosidad) registradas por la estación agrometeorológica de Zavalla (33°01'S/60°53'W) en el período 1973-1990. Se intenta un modelado de la serie de temperaturas mínimas utilizando un método lineal (Predictor Local Lineal) y no lineales (redes neuronales y clasificador de Bayes). Se obtuvieron resultados claramente superiores a los del Predictor Persistente, que es aquél que predice para hoy el valor observado ayer. Se trata de un trabajo exploratorio, aplicado y de importancia económica. Los resultados constituyen un razonable punto de partida para el desarrollo de modelos más sofisticados.

Frost prediction is an important practical issue, specially during fall and spring. In this work we analyse registers of daily minimum temperatures and other surface variables (dry and wet bulb temperatures, relative humidity, wind strength and direction, and nubosity) collected at the agrometeorological station of Zavalla (33°01'S/60°53'W) in the period 1973-1990. We attempt to model the time series of daily minimum temperatures using linear (Local Linear Predictor) and nonlinear strategies (artificial neural networks and Bayes' method). The results are clearly superior to those of the so called Persistent Predictor, for which the current prediction is the last value observed. This is an applied, exploratory work, with implications in economics. The results presented herein constitute a reasonable starting point in the development of more sophisticated models.

## I. INTRODUCCIÓN

Un descenso de la temperatura del aire (medida a pocos centímetros del suelo) por debajo de 0°C constituye una helada. Usualmente se clasifican en radiativas y advectivas, o como combinación de estos dos tipos. Las radiativas ocurren durante la noche, como consecuencia de una pérdida de calor por radiación de onda larga bajo condiciones atmosféricas calmas, claras y secas. Las advectivas, en cambio, son producidas por la incursión de una masa de aire frío y pueden desarrollarse tanto durante el día como la noche<sup>1</sup>. La ocurrencia de una helada puede resultar en un daño a los cultivos, cuya magnitud dependerá de su intensidad y duración, del tipo de cultivo y de su

estado de crecimiento. Si existe congelamiento intracelular, el efecto es inmediatamente fatal. Si, en cambio, la formación de hielo tiene lugar fuera de las células, el daño estará dado por la velocidad y extensión del proceso de deshidratación. Es por este motivo que el conocimiento y predicción de temperaturas mínimas es fundamental para la planificación de las actividades productivas y la eventual protección de los cultivos frente a valores extremos. En este trabajo se intenta predecir la serie temporal de temperaturas mínimas diarias, utilizando registros de la estación agrometeorológica de Zavalla (33°01'S/60°53'W, Provincia de Santa Fe, Argentina) en el período 1973-1990. Se trata de un problema difícil debido al nivel de ruido presente en las observaciones y a la com-

plejidad intrínseca a la dinámica atmosférica. Es por ello que es frecuente en las aplicaciones climáticas encontrar que no se puede mejorar la performance del llamado Predictor Persistente, que es aquél que predice para hoy el valor observado ayer<sup>2</sup>. Si bien extremadamente simple, este método de predicción es razonable debido a la alta persistencia que en general presentan los fenómenos atmosféricos y climáticos, y constituye el piso con el que contrastaremos la eficiencia de los métodos aquí desarrollados.

## II. TESTS DE DETERMINISMO

A los efectos de desestacionalizar el registro de temperaturas mínimas diarias se calculó la onda media anual utilizando todos los datos disponibles (valores diarios entre 1973 y 1990). Los resultados se ajustaron por mínimos cuadrados con una onda sinusoidal dado que la irradiancia solar para cada día del año se puede aproximar por una ley de este tipo, obteniéndose así la serie residual desestacionalizada (Fig. 1).

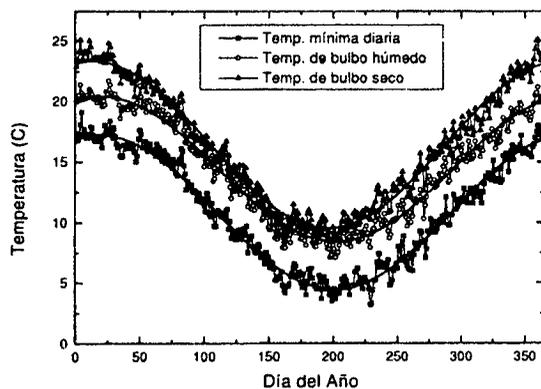


FIG. 1. Ondas medias anuales y sinusoides ajustadas correspondientes.

A fin de determinar el tipo de fenómeno (determinístico o estocástico) que genera esta serie residual se construyeron 19 'surrogates' de la misma (series con idéntico espectro de Fourier). Se utilizó una estadística discriminante basada en el error de predicción de primeros vecinos a medida que se reconstruye el espacio de pseudofases. Si el error de

predicción sobre la serie residual resulta menor que sobre los surrogates, entonces la 'null hypothesis' de un proceso aleatorio Gaussiano, lineal y estacionario, puede ser rechazada con un nivel de significación del 95%. En nuestro caso dicho error resultó menor sobre la serie original que sobre los surrogates sólo para embeddings de dimensión 2 y 3, perdiendo esta propiedad para embeddings de dimensiones mayores. Este resultado es típico de las series autocorrelacionadas, con persistencia en las tendencias, y no constituye una evidencia clara de determinismo<sup>3</sup>. En consecuencia, a efectos de predecir temperaturas mínimas supusimos al registro correspondiente como generado estocásticamente y dependiente de variables aleatorias o predictores.

## III. CORRELACIONES Y BÚSQUEDA DE PREDICTORES

Además de la propia serie de temperaturas mínimas también se dispone del registro diario (tomado a la hora 20) de las siguientes variables de superficie: temperatura de bulbo húmedo, temperatura de bulbo seco, nubosidad e intensidad del viento. Con el objeto de determinar potenciales predictores, se procedió a calcular la correlación a  $n$ -pasos entre la temperatura mínima y las demás variables de superficie (Fig. 2).

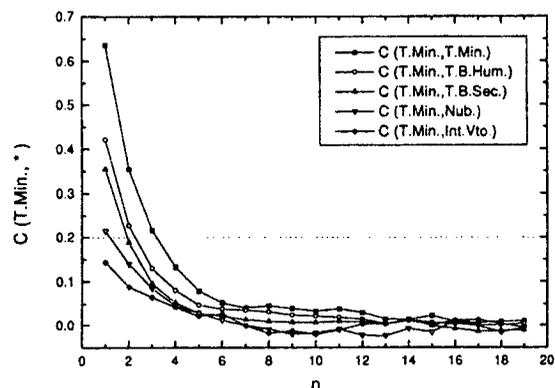


FIG. 2. Correlación a  $n$  pasos entre la temp. mínima y demás variables de superficie.

Se eligió un umbral para desechar predictores; en particular, no se utilizaron variables que muestran una correlación a un paso con la temperatura mínima inferior a 0.2, número que optimizó la performance de los métodos de predicción que se describen en la sección siguiente. Este criterio deja entonces las siguientes variables como predictores: temperatura de bulbo húmedo, de bulbo seco y nubosidad del día anterior. Se incluyó además a la temperatura mínima de los dos días anteriores dado que la serie presenta una autocorrelación relativamente alta a dos pasos ( $\sim 0.35$ ).

#### IV. MÉTODOS DE PREDICCIÓN

Se utilizaron los registros históricos correspondientes al período Mayo-Septiembre de los años 1973-1988 para predecir la ocurrencia de heladas en el mismo período de los años 1989-1990. Se implementaron los siguientes métodos de predicción: Predictor Local Lineal, Redes Neuronales y 'Naive Bayes', usando en todos los casos las mismas variables predictoras.

En el modelo Predictor Local Lineal se buscan los primeros vecinos del estado actual dentro de la base de datos del espacio de predictores, al estilo de<sup>4</sup>. La variante introducida en este trabajo consiste en utilizar una métrica que pondera la importancia relativa de las variables predictoras según su correlación con la temperatura mínima del día siguiente.

Las Redes Neuronales (RN) constituyen modelos computacionales de significativa importancia para el análisis y predicción de series temporales<sup>5</sup>. En los últimos años se han aplicado a distintos sistemas, incluyendo mapas caóticos y series de datos reales tales como datos meteorológicos e indicadores económicos<sup>8,9</sup>. Se entrenaron redes 5:2:1 y 5:5:1 sobre el mapa variables predictoras-temperatura mínima. Más abajo se presentan resultados de la red 5:2:1.

Naive Bayes es una técnica inductiva basada en una descripción probabilística de los datos disponibles para la extracción de conocimiento<sup>6</sup>. Este método de inducción es óptimo para el aprendizaje supervisado si los valores de los atributos de un patrón (en nues-

tro caso: variables predictoras) son independientes dentro de una clase del mismo (las temperaturas mínimas se dividen naturalmente en dos clases: helada y complemento). A pesar de que este postulado es violado la mayoría de las veces, trabajos recientes han demostrado que la extracción de conocimiento a partir de un conjunto de datos usando Naive Bayes es efectiva en la práctica. Más aún, los clasificadores basados en esta técnica han demostrado ser competitivos frente a otras técnicas más sofisticadas en la resolución de problemas reales<sup>7</sup>.

#### V. EVALUACIÓN DE RESULTADOS

Para la evaluación de resultados se utilizaron tablas de contingencia y los coeficientes que se detallan a continuación, de uso frecuente por parte de meteorólogos y estadísticos. Asumiendo la siguiente tabla de contingencia:

		PRED	
		SI	NO
O B S	SI	<i>a</i>	<i>b</i>
	NO	<i>c</i>	<i>d</i>

se definen:

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

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

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

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

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

Se incluye además la medida de error

$$E = 5 b/a + c/a,$$

que da cuenta de la diferencia en la importancia económica entre una helada no predicha y una falsa alarma. El factor 5 fue elegido arbitrariamente; en aplicaciones concretas debería ser fijado en base a un estudio de costos.

#### VI. RESULTADOS

En la Tabla 1 se muestran los resultados obtenidos. Se distinguen dos casos: en

el primero se considera como predicción de helada a una mínima estimada menor que 0°C, y en el segundo se fija un umbral de 1°C como nivel de alarma (predicción de “riesgo de helada”). Puede observarse en la Tabla 1 que se obtuvieron resultados claramente superiores a los del Predictor Persistente. Esta diferencia se ve notablemente reflejada en el índice E, de connotación económica. Nótese también que se pueden obtener resultados muy positivos en el marco de predicción de “riesgo de heladas”. No puede afirmarse que alguno de los métodos sea claramente superior a los demás, si bien Naive Bayes parece mostrar cierta ventaja.

## VII. CONCLUSIONES

La predicción de heladas constituye un problema aplicado de gran importancia en la planificación de actividades frutihortícolas, permitiendo la eventual implementación de medidas de prevención teniendo en cuenta el riesgo económico involucrado. Dada la importancia práctica de este problema, hemos presentado un trabajo de carácter exploratorio en el cual se intentó predecir la ocurrencia de heladas mediante distintos métodos orientados a la extracción de conocimiento a partir de los datos. Se estudió la serie de temperaturas mínimas registradas por la estación agrometeorológica de Zavalla (33°01'S/ 60°53'W, Provincia de Santa Fe, Argentina) entre los años 1973 y 1990. En primer lugar se intentó una reconstrucción del espacio de pseudofases. Ante la falta de evidencia de determinismo en la serie desestacionalizada, se procedió a la búsqueda de predictores. Se intentó un modelado de esta variable utilizando un método lineal (Predictor Local Lineal) y métodos no lineales (Redes Neuronales y clasificador de Bayes), que son técnicas del área de Teoría de Sistemas Dinámicos y Machine Learning. Se obtuvieron resultados claramente superiores a los del Predictor Persistente, concluyéndose que se pueden obtener resultados muy positivos en el marco de predicción de “riesgo de heladas”. No puede afirmarse que alguno de los métodos sea superior a los demás, si bien Naive Bayes parece mostrar cierta ventaja. Los resulta-

dos constituyen un razonable punto de partida para el desarrollo de modelos más sofisticados.

## VIII. AGRADECIMIENTOS

Agradecemos al Dr. Miguel Lara y colaboradores el habernos facilitado los registros históricos y por valiosas discusiones durante la realización de este trabajo.

- 
- <sup>1</sup> Kalma, J.D., Laughlin, G.P., Caprio, J.M. & Hamer, P.J.C. *The Bioclimatology of Frost: Its Occurrence, Impact and Prediction*, Springer-Verlag (1992).
  - <sup>2</sup> Klein, W.H. & Lewis, F., Computer forecasts of maximum and minimum temperatures, *Journal of Applied Meteorology* **9**, 350-359 (1970).
  - <sup>3</sup> Casdagli, M. & Eubank, S. eds. *Nonlinear Modeling and Forecasting*, Santa Fe Institute of Studies in the Sciences of Complexity Proc. vol XII (Redwood City: Addison-Wesley) (1992).
  - <sup>4</sup> Farmer, J.D. & Sidorowich, J.J. *Phys. Rev. Lett.* **59**, 845-848 (1987).
  - <sup>5</sup> Weigend, A.S., Huberman, B.A., & Rumelhart, D.E. *International Journal of Neural Systems* **1**, 193-209 (1990).
  - <sup>6</sup> John, G. & Langley, P. Estimating Continuous Distributions in Bayesian Classifiers. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann Publishers, San Mateo (1995).
  - <sup>7</sup> Domingos, P. & Pazzani, M. Beyond independence: Conditions for the optimality of the simple bayesian classifier. In *Proceedings of the Thirteenth International Conference on Machine Learning*, 105-112, Morgan Kaufmann Publishers (1996).
  - <sup>8</sup> Navone, H.D. & Ceccatto, H.A. *Chaos, Solitons & Fractals* **6**, 383-387 (1995).
  - <sup>9</sup> Navone H.D. & Ceccatto H.A., Predicting Indian monsoon rainfall: a neural network approach. *Climate Dynamics* **10**, 305-312 (1994).
  - <sup>10</sup> Press, W.H., Flannery, B.P., Teukolsky, S.A. & Vetterling, W.T. *Numerical Recipes*, Cambridge University Press (1989).

AJUSTE		Predichas			
		<= 0		> 0	
Observadas	<= 0	PP	111	PP	153
		NB	213	NB	51
		RN	161	RN	103
		PL	143	PL	121
	> 0	PP	154	PP	2030
		NB	189	NB	1995
		RN	69	RN	2115
		PL	52	PL	2132

PREDICCIÓN		Predichas			
		<= 0		> 0	
Observadas	<= 0	PP	13	PP	20
		NB	27	NB	6
		RN	22	RN	11
		PL	19	PL	14
	> 0	PP	20	PP	253
		NB	22	NB	251
		RN	6	RN	267
		PL	6	PL	267

PP: Pred. Persist. RN: Red Neuronal  
NB: Naive Bayes PL: Pred. Local

	PP	NB	RN	PL
POD	0.39	0.82	0.67	0.58
FAR	0.61	0.45	0.21	0.24
CSI	0.25	0.49	0.56	0.49
PC	0.87	0.91	0.94	0.93
POFD	0.07	0.08	0.02	0.02
E	9.23	1.93	2.77	4.00

AJUSTE		Predichas			
		<= 1		> 1	
Observadas	<= 0	PP	137	PP	127
		NB	-	NB	-
		RN	211	RN	53
		PL	184	PL	80
	> 0	PP	238	PP	1946
		NB	-	NB	-
		RN	161	RN	2023
		PL	127	PL	2057

PREDICCIÓN		Predichas			
		<= 1		> 1	
Observadas	<= 0	PP	20	PP	13
		NB	-	NB	-
		RN	28	RN	5
		PL	28	PL	5
	> 0	PP	29	PP	244
		NB	-	NB	-
		RN	16	RN	257
		PL	12	PL	261

	PP	NB	RN	PL
POD	0.61	-	0.85	0.85
FAR	0.59	-	0.36	0.30
CSI	0.32	-	0.57	0.62
PC	0.86	-	0.93	0.94
POFD	0.11	-	0.06	0.04
E	4.70	-	1.46	1.32

TABLA 1