Estimating the probability of convection events from statistical analysis of temperature and humidity vertical profiles, shear and helicity

IANIGLA · CONICET · U.N.CUYO · GOBIERNO DE MENDOZA · GOBIERNO DE SAN JUAN

Araneo Diego[1, 2], Silvia Simonelli1[1], Federico Norte[1], Jorge Santos[1, 2] and Alejandro de la Torre[3]

[1] IANIGLA – Programa Regional de Meteorología, CONICET / CCT-Mendoza, Mendoza, Argentina
[2] Instituto de Ciencias Básicas, UNCuyo, Mendoza, Argentina
[3] Facultad de Ingeniera, Universidad Austral, Buenos Aires, Argentina
e-mail: daraneo@mendoza-conicet.gob.ar

UNC ICB NATURALEZA · TECNOLOGÍA · HUMANISMO
Instituto de Ciencias Básicas

OBJECTIVE

The objective of this work is to determine the vertical profile patterns of temperature (T) and dew point (Td) for the Northern of Mendoza Province and to obtain a statistical forecast model of convection using these profiles as predictors.

DATA

T, Td and wind data at 850, 700, 500, 400 and 300 hPa levels corresponding to 272 rawinsondes at 12 UTC from Mendoza-Aero weather station (National Weather Service) were used in order to obtain the profiles and to adjust the forecast model. The rawinsondes from October-March periods of 1987 to 1995 were used for the analysis and 371 rawinsondes from November-March periods of 2006 to 2010 were used to validate the predictions. We defined a Convection Occurrence Index (C) according to the reported observations during 24 hours following the rawinsonde observation (C=1 for convection episodes and C=0 for no convection episodes).

METHODOLOGY

The Principal Component Analysis (PCA) was applied in order to determine the vertical profiles patterns of T and Td according to the following procedure:

$\mathbf{X}_{[10,272]}$ Data matrix. Each column contains the T and Td values at the mentioned levels from the rawindsonde of one corresponding day.

$\widetilde{\mathbf{X}}_{[10,272]}$ Deviations matrix (average rawindsonde is subtracted from each column of $\mathbf{X}$)

$\mathbf{Z} = \widetilde{\mathbf{X}}_\mathbf{s}\mathbf{Q}$ $\mathbf{Z}_{[10\times272]}$ Component Score Matrix

$\mathbf{F} = \mathbf{Q}\mathbf{D}^{1/2}$ $\mathbf{F}_{[10\times272]}$ Component Loading Matrix

$\widetilde{\mathbf{X}}_{\mathbf{s}[10,272]}$ is the matrix resulting of standardize the $\widetilde{\mathbf{X}}_{[10,272]}$ columns $\mathbf{Q}_{[272\times272]}$ and $\mathbf{D}_{[272\times272]}$

are the eigenvector and eigenvalue matrices associated to the correlation matrix for $\widetilde{\mathbf{X}}_{[10,272]}$

columns. Them: $\widetilde{\mathbf{X}}_\mathbf{s} = \mathbf{Z}_\mathbf{s}\mathbf{F}' \Rightarrow \widetilde{\mathbf{x}}_{\mathbf{s}j} = f_{j1}\mathbf{z}_{\mathbf{s}1} + f_{j2}\mathbf{z}_{\mathbf{s}2} + \ldots + f_{j10}\mathbf{z}_{\mathbf{s}10}$

So each rawindsonde can be identified with a number not exceeding 10 loading factors f_{ji} that constitute the weight of each component *j* to represents one single rawinsonde.

Them we define: $\mathbf{C}_{[272x1]}$ the Convection Occurrence Index Vector.

A Logistic Multiple Regression Model is performed between the response vector $\mathbf{C}$ and the Component Loadings $\mathbf{f}_j$ (columns of $\mathbf{F}$) as following:

$\hat{c}_i = \dfrac{e^{w_i}}{1+e^{w_i}}$ Estimated i-th coefficient of $\mathbf{C}$

Where w_i is the i-th element of:

$\mathbf{w} = b_1\mathbf{f}_1 + b_2\mathbf{f}_2 + \ldots + b_n\mathbf{f}_n + b_0[1]_{272}$ or $\mathbf{w} = \mathbf{F}\mathbf{b} + b_0[1]_{272}$

and the coefficients $b_0, b_1, \ldots, b_n$ are fitted by maximum-likelihood.

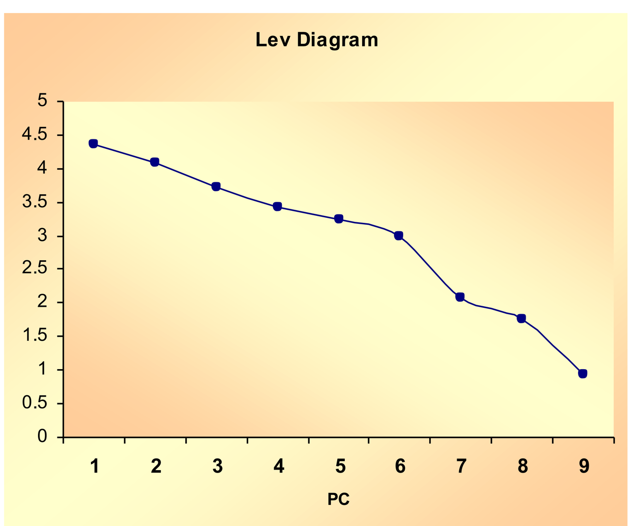
Now, since: $\mathbf{F} = \dfrac{\widetilde{\mathbf{X}}_\mathbf{s}'\mathbf{Z}_\mathbf{s}}{m-1} \Rightarrow \mathbf{w} = \widetilde{\mathbf{X}}_\mathbf{s}'\dfrac{\mathbf{Z}_\mathbf{s}\mathbf{b}}{m-1} + b_0[1]_{326}$

we can define: $\mathbf{A} = \dfrac{\mathbf{Z}_\mathbf{s}\mathbf{b}}{m-1}$ and so $\hat{c} = \dfrac{e^{\widetilde{\mathbf{x}}_\mathbf{s}\cdot\mathbf{A}+b_0}}{1+e^{\widetilde{\mathbf{x}}_\mathbf{s}\cdot\mathbf{A}+b_0}}$

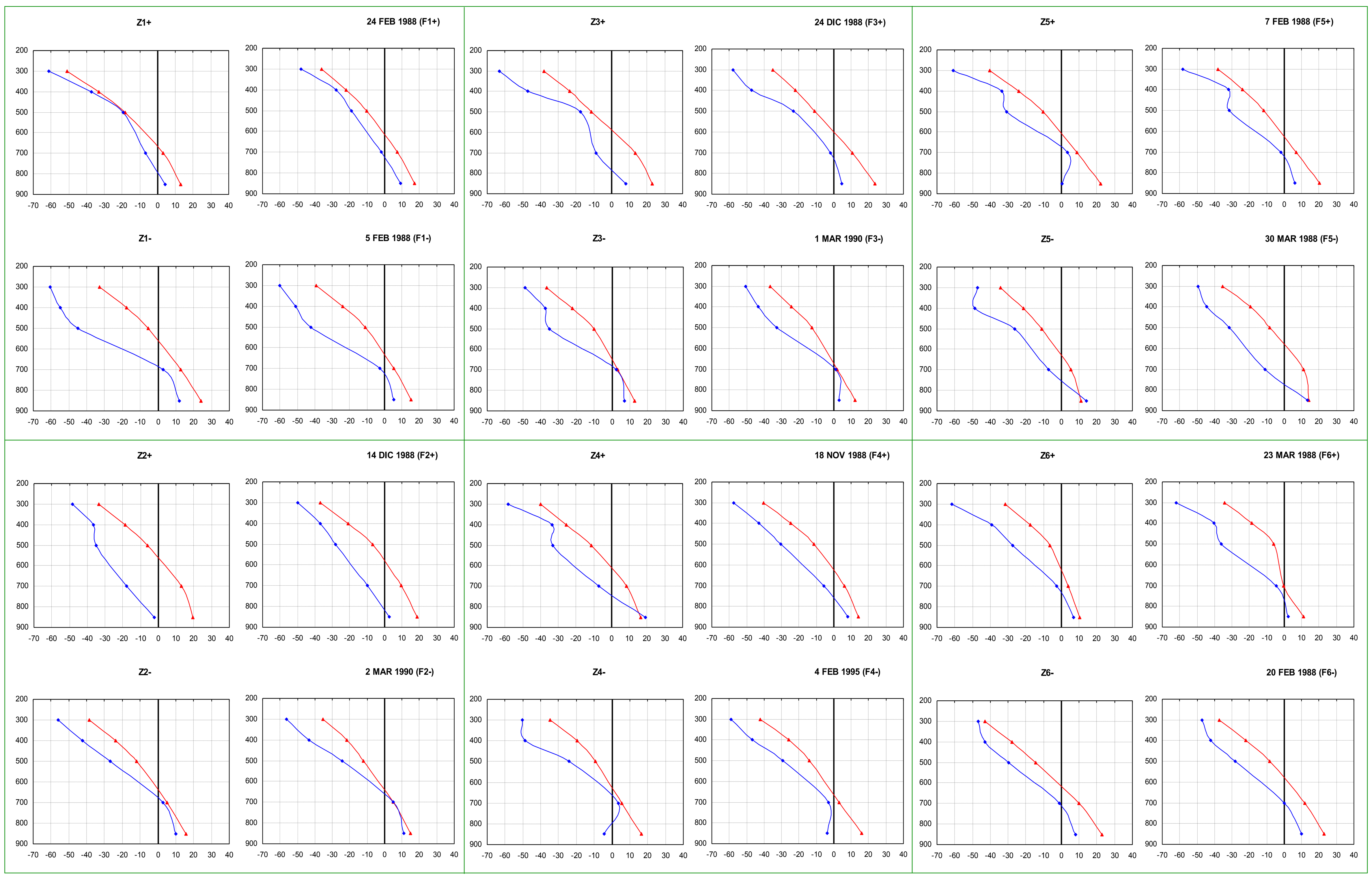
Where $\widetilde{\mathbf{x}}_\mathbf{s}$ is a standardized anomalous rawinsonde of one forecasting day. Also the Helicity and the Shear at several levels can be incorporated as predictors in the Logistic Multiple Regression Model.

PC	Variance	Variance[%]	Cumulative Variance [%]
1	78.6	28.9	28.9
2	59.0	21.7	50.6
3	41.5	15.2	65.8
4	31.0	11.4	77.2
5	25.5	9.4	86.6
6	20.0	7.3	93.9
7	8.1	3.0	96.9
8	5.9	2.2	99.1
9	2.5	0.9	100.0

Variances correspond to the Principal Components, explained percentages and cumulative percentages.



Lev diagram for the Principal Components. The first 6 PCs are notoriously significant.

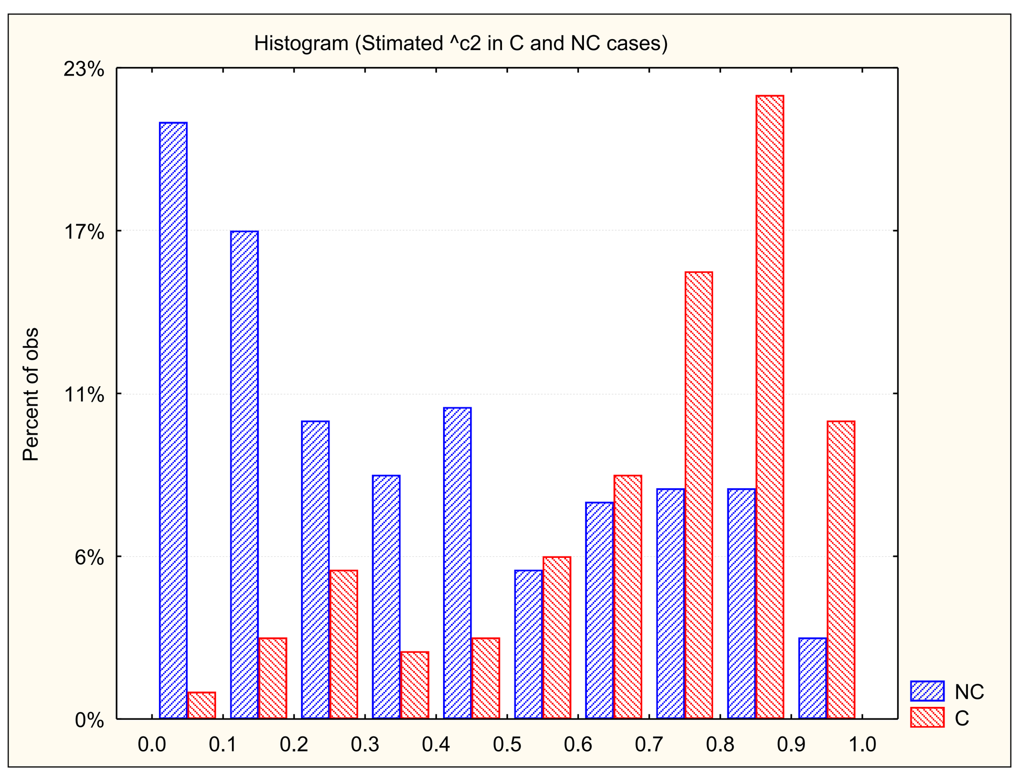


The first 6 PCs (Z) for T (red) and Td (blue) vertical profiles in direct (+) and indirect (-) modes. Real profiles (dates are indicated on each panel) correspond to the maximum and minimum value of the respective Component Loading are also shown at right of each Z panel.

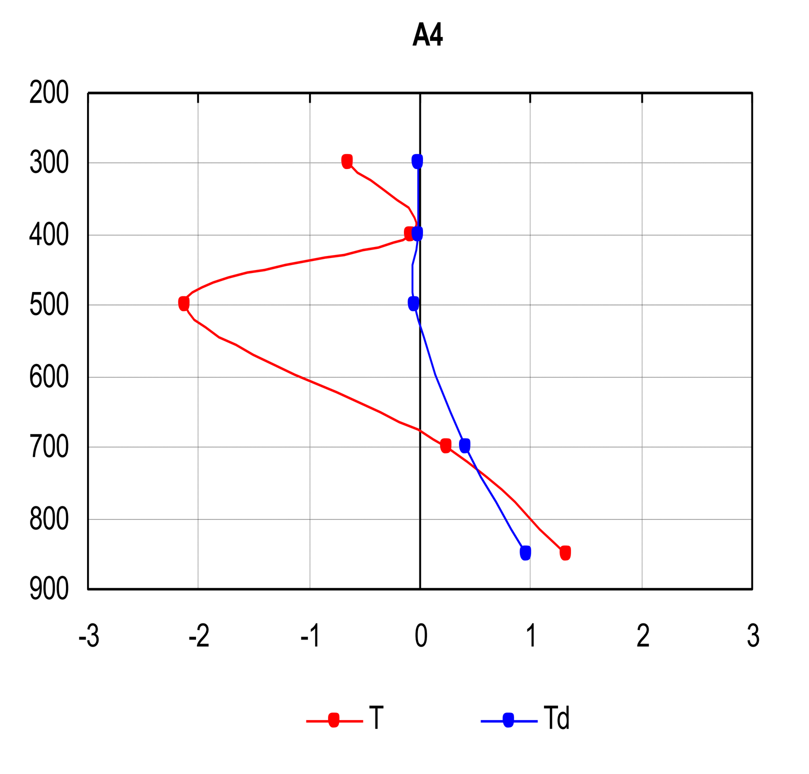
Multiple Logistic Regression Analysis

Predictor	b coefficients						
	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5	Exp. 6	Exp. 7
b_0	0.745334	-0.162803	-0.141094	-0.017351	0.012876	-0.166606	-0.149346
F1	-0.522009	---	---	-0.576897	---	-0.528021	---
F2	**-3.080600**	**-2.886980**	**-2.822430**	**-2.925830**	**-2.800090**	**-3.058030**	**-2.941600**
F3	**1.428180**	**1.666780**	**1.635150**	**1.515580**	**1.518110**	**1.660240**	**1.669280**
F4	**2.682190**	**2.536400**	**2.481940**	**2.400530**	**2.326520**	**2.609400**	**2.557050**
F5	**1.240600**	1.009290	---	0.957268	---	0.975459	---
F6	**-5.255980**	**-4.979410**	**-4.956260**	**-4.909060**	**-4.748800**	**-5.151410**	**-5.055950**
F7	**-3.670430**	**-2.694230**	**-2.594550**	**-3.237570**	**-2.981750**	**-3.063650**	**-2.831490**
F8	-0.254779	---	---	-1.128120	---	-0.946876	---
F9	3.617940	---	---	**4.111450**	**3.892510**	**3.790020**	3.569870
sh_700-850	-0.005868	---	---	---	---	---	---
sh_500-700	-0.042674	---	---	---	---	---	---
sh_400-500	-0.029909	---	---	---	---	---	---
sh_300-400	0.005313	---	---	---	---	---	---
sh_500-850	0.021233	---	---	---	---	---	---
sh_400-700	0.035369	---	---	---	---	---	---
sh_300-500	-0.004502	---	---	---	---	---	---
sh_400-850	-0.040746	---	---	---	---	---	---
sh_300-700	-0.062830	---	---	---	---	---	---
sh_300-850	0.064388	---	---	---	---	---	---
H	**-0.001013**	**-0.000628**	**-0.000641**	---	---	-0.000511	-0.000569
Success [%]	70.1	73.9	70.4	73.6	68.2	73.6	69.5
Surprise [%]	5.7	8.1	9.2	6.5	9.2	7.3	10.0
False alarm [%]	24.3	18.1	20.5	19.9	22.6	19.1	20.5

Multiple Logistic Regression coefficients b (for seven experiments using different predictor combinations) correspond to the predictors: Principal Component Loadings (F_i), shear between levels i and j (sh_i-j), and helicity (H). Significant coefficients at α=0.05 are red highlighted. For the resulting models (predicting "convection" when estimated c> 0.5 and "no convection" when estimated c <0.5) the success, surprise and false alarm percentages are also shown.



Distribution histogram for estimated c values derived from the exp. 2 model in rightful "convection" and "no convection" cases. Most of the estimated c values are <0.5 for "no convection" cases and >0.5 for "convection" cases.



"A" vector correspond to the exp. 4 model (retaining the 9 PCs). It represents an anomalous standardized sounding with a strong T lapse rate between 500 and 850 hPa and very wet air at low levels, which represents a strong instability.

CONCLUTIONS

The PCA resulted useful to obtain the vertical profile patterns of T and Td. The resulting patterns in direct and reverse modes represent real cases. The analysis of T and Td sheds 6 significant components explaining 94% of the system variability.

In the studied cases, over than 70% of the forecast model effectiveness is obtained using only the T and Td profiles (similar to that found in other studies by traditional indices). The effectiveness could be improved by changing the definition of C and the incorporation of the helicity as a predictor.

Profile patterns of T and Td reveals that the probability of convection increases (decreases) with strong (weak) T lapse-rate between low and middle levels of the troposphere and high (low) moisture content in the lower layers.