REVUE DE STATISTIQUE APPLIQUÉE

G. DER MEGREDITCHIAN
J. COIFFIER
H. KERRACHE
J. LEPAS

Schémas de régression et de discrimination pour la prévision locale utilisant les données de la prévision numérique

Revue de statistique appliquée, tome 24, n° 4 (1976), p. 23-34

<http://www.numdam.org/item?id=RSA_1976__24_4_23_0>

SCHEMAS DE REGRESSION ET DE DISCRIMINATION POUR LA PRÉVISION LOCALE UTILISANT LES DONNÉES DE LA PRÉVISION NUMÉRIQUE

G. DER MEGREDITCHIAN, J. COIFFIER, H. KERRACHE, J. LEPAS

INTRODUCTION

Trois orientations principales caractérisent la démarche du prévisionniste en météorologie : la méthode synoptique basée sur la recherche directe des lois empiriques régissant les phénomènes atmosphériques ; la méthode hydrodynamique reposant sur une modélisation de l'atmosphère considérée comme un fluide obéissant aux lois de l'hydrodynamique ; la méthode statistique où le processus de génération des données atmosphériques est assimilé au déroulement d'une épreuve aléatoire.

Nous présentons ici un exemple d'application des méthodes de la statistique multidimensionnelle à la prévision locale météorologique à courte échéance pour le site d'Oran. Toutefois, comme les trois voies possibles pour la prévision loin de s'exclure sont complémentaires, chacune possédant ses avantages et ses inconvénients, nous nous sommes efforcés de mettre à contribution également les deux autres méthodes lors de l'élaboration de notre modèle prévisionnel statistique.

LE FICHIER DES DONNEES

Nous avons utilisé les paramètres météorologiques suivants :

a) des variables météorologiques locales primaires L(j) disponibles au jour j. Ce sont la température minimum $T_{min}(j)$ et maximum $T_{max}(j)$ en 1/10 de degrés Celsius, la durée d'insolation SD(j) en 1/10 d'heure, l'humidité minimum $U_{min}(j)$ et maximum $U_{max}(j)$ en %, la quantité de précipitations RR(j) en 1/10 de mm, les variables indicatrices de l'occurrence des événements pluie, brouillard.

b) des variables météorologiques locales élaborées LE(j) calculées à partir des variables primaires directement mesurées. Ce sont : la différence des températures journalières extrêmes $\Delta T(j) = T_{max}(j) - T_{min}(j)$, l'anomalie climatique de la température moyenne journalière $\delta T(j) = T(j) - \bar{T}(j)$, où $T(j) = [T_{min}(j) + T_{max}(j)]/2$ et $\bar{T}(j)$ est la moyenne climatique de la température pour le jour j, l'énergie totale journalière reçue au sommet de l'atmosphère par unité de surface w(j) en watts/m².

Mots-clés : **Météorologie. Prévision statistique. Régression. Discrimination. Sélection progressive.**

c) des variables synoptiques primaires S(j) décrivant la situation météorologique à l'échelle synoptique. Ce sont les hauteurs du géopotentiel des surfaces isobariques 850 et 500 millibars relevées aux points d'intersection d'une grille approximativement centrée sur Oran et comportant 25 points (cf. fig. 1). Ces données sont fournies par les "analyses" des champs du géopotentiel, c'est-à-dire une opération permettant d'interpoler aux points de grille les valeurs de ces champs relevées aux stations de mesure.

d) des variables synoptiques élaborées SE(j) introduites pour contrebalancer la linéarité du schéma prévisionnel. Ce sont les composantes zonale $U_g(j)$ et méridienne $V_g(j)$ du vent géostrophique aux surfaces isobariques 850 et 500 millibars, le tourbillon relatif $\zeta(j)$ aux surfaces isobariques 850 et 500 millibars, l'advection du tourbillon aux mêmes niveaux $A_\zeta(j)$ et l'advection de température $A_T(j)$ au niveau intermédiaire 675 mb.

Du point de vue prévisionnel nous distinguons également les variables explicatives ou prédicteurs et les variables à expliquer ou prédictands.

MODELES DE PREVISION STATISTIQUE

Soit X(j) le vecteur colonne des n prédicteurs disponibles au jour j :

$$X(j) = \{x_1(j), \ldots, x_i(j), \ldots, x_n(j)\}$$

où $x_i(j)$ désigne la valeur observée du i-ième prédicteur et n le nombre de prédicteurs utilisés. Le vecteur X(j) des prédicteurs est utilisé pour obtenir la valeur prévue $\tilde{y}(j+1)$ du prédictand y (j + 1).

Lors de la réalisation concrète des schémas prévisionnels nous distinguons deux stratégies :

a) La méthode asynchrone (AS) pour laquelle la prévision $\tilde{y}$ (j + 1) de la variable y(j + 1) est effectuée au jour "j" avec une échéance de 24 heures à l'aide d'un vecteur X(j) des prédicteurs du type :

$$X(j) = \{L(j-1), LE(j-1), S(j-1), SE(j-1)\}.$$

b) Le modèle synchrone (SY) pour lequel on compte parmi les prédicteurs les résultats $\tilde{\tilde{S}}(j+1)$ de la prévision hydrodynamique des variables synoptiques. Le vecteur des prédicteurs X(j) utilisé au jour "j" pour la prévision $\tilde{y}(j+1)$ de la variable y(j + 1) est ainsi de la forme :

$$X(j) = \{L(j-1), LE(j-1), \tilde{\tilde{S}}(j+1), \tilde{\tilde{S}}E(j+1)\}\ .$$

Nous pouvons de la sorte prendre en considération les corrélations synchrones entre prédicteurs synoptiques et prédictand. Précisons à cet égard que les liaisons corrélationnelles sont calculées par la méthode dite de la prévision parfaite, autrement dit nous utilisons dans le schéma prévisionnel les corrélations r [S(j + 1), y(j +1)]comme si en fait

$$\tilde{\tilde{S}}(j+1) = S(j+1).$$

Si le prédictand y est une variable continue, nous appliquons l'analyse de régression linéaire multiple. L'équation prévisionnelle optimale au sens de la minimisation de l'erreur quadratique moyenne est de la forme

$$\tilde{y}(j+1) = V_{yX} V_{XX}^{-1} [X(j) - M_x] + m_y,$$

où m_y et M_X sont les espérances mathématiques du prédictand et du vecteur des prédicteurs ($m_y = Ey$, $M_X = EX$), V_{yX} et V_{XX} respectivement les matrices de variances-covariances prédictand-prédicteurs et prédicteurs-prédicteurs $V_{yX} = E[(y - m_y)(X - M_X)']$, $V_{XX} = E[(X - M_X)(X - M_X)']$. En réalité nous utilisons une équation empirique de régression obtenue en remplaçant dans l'équation théorique les paramètres théoriques m_y, M_X, V_{yX}, V_{XX} par leurs analogues empiriques $\hat{m}_y$, $\hat{M}_X$, $\hat{V}_{yX}$, $\hat{V}_{XX}$ calculés sur le fichier d'apprentissage formé par N réalisations du vecteur $\{X(j), y(j+1)\}$.

La qualité de la prévision sur le fichier d'apprentissage peut être exprimée en fonction des paramètres de ce fichier à l'aide du coefficient de corrélation multiple

$$r_n^2 [y, \tilde{y}] = R_{yX} R_{XX}^{-1} R_{Xy}$$

où R_{yX}, R_{XX} sont les matrices de corrélation correspondant aux matrices de variances-covariances V_{yX}, V_{XX}.

Si la variable prédictand est discrète, variable indicatrice de l'occurrence d'un phénomène A codée 1, si A a lieu et 0 si A n'a pas lieu, nous appliquons l'analyse discriminante paramétrique. Dans ce cas la règle de décision optimale est formulée à l'aide de la fonction discriminante

$$g(X) = \frac{p(A) \cdot f_A(X)\, c(\bar{A}/A)}{p(\bar{A}) \cdot f_{\bar{A}}(X)\, c(A/\bar{A})}$$

où $p(A)$ et $p(\bar{A})$ sont les probabilités a priori des événements A et $\bar{A}$, $f_A(x)$ et $f_{\bar{A}}(X)$ les densités de probabilité du vecteur X pour les populations A et $\bar{A}$, $c(\bar{A}/A)$ et $c(A/\bar{A})$ les coûts des erreurs décisionnelles correspondantes.

La règle décisionnelle optimale est alors définie à l'aide du seuil 1 de la façon suivante :

si $g(X) > 1$ on pose $y = 1$, on prévoit l'occurrence de A ;

si $g(X) < 1$ on pose $y = 0$, on prévoit l'occurrence de $\bar{A}$.

Dans le cas des populations gaussiennes il est commode de prendre comme fonction discriminante $w(X) = \ln g(X)$, que l'on compare au seuil zéro. Nous obtenons ainsi une surface discriminante quadratique dans le cas général, mais dans le cas particulier où $V_{XX}(A) = V_{XX}(\bar{A}) = V$ un hyperplan d'équation

$$U(X) = [X - M_X(+)]' V_{XX}^{-1} M_X(-) - S = 0$$

où

$$M_X(-) = M_X(A) - M_X(\bar{A}) \; ; M_X(+) = \frac{1}{2}[M_X(A) + M_X(\bar{A})];$$

$$S = \ln \left[\frac{p(\bar{A})\, c(A/\bar{A})}{p(A)\, c(\bar{A}/A)} \right]$$

le pouvoir discriminant du vecteur des prédicteurs est défini par la distance de Mahalanobis

$$\Delta_n^2 [A,\bar{A}] = M'_X (-)\, V_{XX}^{-1}\, M_X (-),$$

qui nous fournit les probabilités des erreurs décisionnelles :

$$P [A/A] = 1 - \Phi\left[\frac{S}{\Delta} - \frac{\Delta}{2}\right] ; P [\bar{A}/A] = \Phi\left[\frac{S}{\Delta} - \frac{\Delta}{2}\right]$$

$$P [A/\bar{A}] = 1 - \Phi\left[\frac{S}{\Delta} + \frac{\Delta}{2}\right] ; P [\bar{A}/\bar{A}] = \Phi\left[\frac{S}{\Delta} + \frac{\Delta}{2}\right]$$

où $\Phi(z)$ est la fonction de répartition de la variable aléatoire gaussienne centrée réduite.

De même que pour la régression nous utilisons en réalité une fonction discriminante empirique obtenue en remplaçant dans l'équation théorique les paramètres $p(A)$, $p(\bar{A})$, $M_X(A)$, $M_X(\bar{A})$, $V_{XX}(A)$, $V_{XX}(\bar{A})$ par leurs analogues empiriques $\hat{p}(A)$, $\hat{p}(\bar{A})$, $\hat{M}_X(A)$, $\hat{M}_X(\bar{A})$, $\hat{V}_{XX}(A)$, $\hat{V}_{XX}(\bar{A})$ calculés à partir des fichiers d'apprentissage.

C'est pourquoi pour éviter une baisse spectaculaire de la qualité lors de l'application des équations prévisionnelles sur les données d'un fichier test nous devons effectuer judicieusement le choix des prédicteurs en assurant une qualité suffisamment bonne pour un nombre de prédicteurs le plus petit possible. Nous retenons le "meilleur" groupe des k prédicteurs les plus informatifs en imposant que $k \leqslant N$.

La sélection du groupe "optimal" de prédicteurs ne pouvant être réalisée pour $n > 16$ par une analyse exhaustive de tous les groupes de prédicteurs nous avons utilisé une procédure suboptimale de sélection progressive ascendante consistant à trouver successivement le prédicteur apportant le plus à un groupe déjà retenu. La réalisation concrète de cette procédure est facilitée par les formules d'accroissement des critères de qualité de la prévision quand on adjoint au vecteur X à k dimensions des prédicteurs déjà retenus un nouveau prédicteur ξ :

$$r_{k+1}^2 [y,\hat{y}] - r_k^2 [y.\hat{y}] = \frac{[r_{y\xi} - R_{\xi X} R_{XX}^{-1} R_{Xy}]^2}{1 - R_{\xi X} R_{XX}^{-1} R_{X\xi}}$$

$$\Delta_{k+1}^2 [A, \bar{A}] - \Delta_k^2 [A, \bar{A}] = \frac{[m_\xi(-) - V_{\xi X} V_{XX}^{-1} M_X(-)]^2}{\sigma_\xi^2 - V_{\xi X} V_{XX}^{-1} V_{X\xi}} .$$

Ces formules nous permettent au pas $k+1$ de la sélection de remplacer $N - k$ inversions de matrices d'ordre $k+1$ par l'inversion d'une seule matrice d'ordre k. La réalisation de l'algorithme de sélection devient alors suffisamment rapide pour effectuer dans des temps de calcul raisonnables la sélection d'une vingtaine de prédicteurs parmi un millier de prédicteurs initiaux.

La procédure de sélection nous fournit ainsi un ordonnancement préférentiel des prédicteurs potentiels. Il reste alors à déterminer la taille "optimale" du vecteur des prédicteurs. Pour cela nous appliquons les tests de Fischer per-

mettant de vérifier la signification de l'accroissement d'information apporté par l'adjonction de l nouveaux prédicteurs aux k prédicteurs déjà retenus. Les statistiques utilisées sont respectivement :

Pour la régression

$$\hat{\mathcal{F}}_1 = \frac{T - k - l - 1}{l} \frac{\hat{r}^2_{k+1}[y, \tilde{y}] - \hat{r}^2_K[y, \hat{y}]}{1 - \hat{r}^2_{k+1}[y, \tilde{y}]}$$

qui dans l'hypothèse H_o d'un apport informationnel non significatif $r^2_{k+1}[y, \hat{y}] = r^2_k[y, \hat{y}]$ suit asymptotiquement pour des populations gaussiennes une loi de Fischer à l et $T - k - l - 1$ degrés de liberté.

Pour la discrimination

$$\hat{\mathcal{F}}_2 = \frac{T - k - l - 1}{l} \frac{T(A) \cdot T(\bar{A}) [\hat{\Delta}^2_{k+l} - \hat{\Delta}^2_k]}{T(T-2) + T(A)T(\bar{A})\hat{\Delta}^2_k}$$

qui dans l'hypothèse H_o d'un apport informationnel non significatif $\Delta^2_{k+l} = \Delta^2_k$ suit asymptotiquement pour des populations gaussiennes une loi de Fisher à l et $T - k - l - 1$ degrés de liberté .

L'application formelle des tests de Fisher conduisait à retenir un nombre trop élevé de prédicteurs du fait que l'échantillon disponible n'était pas un échantillon au hasard de sorte que les degrés de liberté dans les tests étaient surestimés. Nous avons contourné cette difficulté grâce à l'artifice consistant à remplacer le nombre N de données corrélées par le nombre équivalent $n_{equ}(N)$ de données non corrélées s'exprimant à l'aide de la formule

$$n_{equ}(N) = \frac{(\mathrm{tr}V)^2}{\mathrm{tr}V^2}$$

où V désigne la matrice de variances-covariances caractérisant les intercorrélations entre toutes les données.

Cette formule a été obtenue par l'un des auteurs [4] dans l'hypothèse gaussienne en adoptant une définition de "l'équivalence" entre deux ensembles de variables corrélées et non corrélées basée sur l'identification des fonctions de répartition du coefficient d'anomalicité, paramètre du second degré les caractérisant. L'identité des fonctions de répartition étant comprise au sens de l'égalité des premiers moments, la formule du nombre équivalent de données fictives indépendantes découle immédiatement de la loi de distribution des formes quadratiques gaussiennes.

RESULTATS EXPERIMENTAUX

Passons maintenant à la description des modèles statistiques réalisés à l'Institut Hydrométéorologique d'Oran (projet OMM).

La méthode que nous présentons utilise un fichier d'apprentissage s'étendant sur une période de 7 années, du 1er Janvier 1964 au 31 décembre 1970. Le fichier avait certaines lacunes, on comptait environ 300 données manquantes.

Le fichier test comportait une année de données, du 1er Janvier 1971 au 31 Décembre 1971. La prévision opérationnelle a été à ce jour réalisée sur une période plus courte allant du 24 Janvier 1976 au 39 Avril 1976.

Dans le tableau 1 nous présentons le résultat de la sélection progressive ascendante des prédicteurs tant pour la régression que pour la discrimination sous forme d'un ordonnancement préférentiel des prédicteurs successivement retenus.

Le nombre de prédicteurs jugés significatifs allait de 4 à 10. On remarque la présence prédominante du géopotentiel moyen à 850 mb (Φ850) et une certaine similitude dans la série des prédicteurs sélectionnés tant pour la méthode synchrone, que pour la méthode asynchrone. On constate en outre que l'effet de persistance n'est prédominant que pour l'anomalie de température journalière moyenne.

Après avoir sélectionné les prédicteurs nous avons réalisé concrètement les schémas prévisionnels.

Nous disposons ainsi de plusieurs estimations de la qualité Q de la prévision statistique réalisée dans des conditions différentes. Nous distinguons les indices Q_T (appr) pour une prévision réalisée dans les conditions les plus favorables du fichier d'apprentissage ; nous avons ensuite les indices Q_τ (test) donnant une caractéristique objective de la qualité sur des données indépendantes de celles qui ont servi à établir les équations prévisionnelles, mais pour une taille τ du fichier test généralement bien inférieure à celle T du fichier d'apprentissage, dans notre cas $T \approx 7\tau$. Finalement nous avons voulu nettement séparer des deux premiers indices les indices de qualité Q_θ (opérat.) de la prévision effectuée véritablement dans les conditions de routine opérationnelle.

On a ainsi présenté dans les tableaux 2, 3, 4 les indices Q_T (appr), Q_τ (test), Q_θ (opérat) de qualité de la prévision pour les deux schémas synchrone (SY) et asynchrone (AS) respectivement pour la régression (tableaux 2, 4) et pour la discrimination (tableau 3) ;

Les indices Q de la qualité de la prévision figurant dans ces tableaux sont pour la régression l'erreur moyenne algébrique, l'erreur moyenne absolue, l'écart type de la variable prévue, le coefficient de corrélation entre valeurs prévues et valeurs observées. Pour la discrimination l'indice de qualité est caractérisé par le tableau de contingence des valeurs $n[A_i/A_j]$, où $n[A_i/A_j]$ désigne le nombre de cas où l'on a prévu A_i alors que l'on observait en réalité A_j.

Les résultats de la prévision statistique sont également comparés à ceux de la prévision de persistance (PS) pour laquelle par définition

$$\tilde{y}(j+1) = y(j)$$

. Nous pouvons ainsi juger la qualité des modèles statistiques non seulement par rapport à l'incertitude définie par la climatologie, mais encore par rapport à la cohésion interne des séries chronologiques décrite par la corrélation temporelle entre valeurs successives.

Précisons toutefois que dans le cas de la méthode synchrone (SY) seul Q_θ (opérat.) est un indice réaliste de la qualité du schéma prévisionnel pour le météorologiste.

Bien que le fichier ayant servi à la prévision opérationnelle soit extrêmement limité il semble que l'utilisation des résultats de la prévision numérique pour la méthode synchrone soit bénéfique.

L'examen des tableaux 2, 3, 4 fait apparaître une certaine cohérence globale des résultats obtenus dont la signification ne sera confirmée qu'avec une expérience plus longue de routine opérationnelle.

Les faibles valeurs des indices obtenus pour la prévision opérationnelle même avec la méthode asynchrone laissent à penser que la taille du fichier de routine opérationnelle est bien trop faible et que l'on peut espérer de meilleurs résultats après une année d'expérimentation.

L'estimation objective de la qualité se heurte à deux exigences contradictoires relatives à la taille T du fichier d'apprentissage et τ du fichier test. Pour obtenir des estimations comparables de la qualité on doit avoir $T \approx \tau$, mais pour assurer que les équations prévisionnelles théoriques et empiriques diffèrent le moins possible on doit prendre T aussi grand que possible, donc τ le plus petit possible.

On peut résoudre ce dilemme en appliquant la méthode dite de reconnaissance glissante sur un seul fichier fournissant une estimation de la qualité comparable à celle obtenue sur un fichier test de taille $T - 1$. Le principe de la méthode est le suivant. Soit

$$X(j), y(j+1)$$

les couples des valeurs prédicteurs-prédictands servant à établir les équations prévisionnelles. On élabore une équation prévisionnelle donnant la valeur prévue $\tilde{y}(j+1)$ de $y(j+1)$ à partir d'un fichier d'apprentissage formé en retranchant du fichier initial le couple $X(j), y(j+1)$. De cette façon au lieu d'avoir une équation prévisionnelle que l'on applique T fois nous obtenons T équations appliquées chacune une seule fois.

Nous obtenons ainsi T prévisions effectuées dans les conditions d'un fichier test. Cela nous donne une estimation objective $Q_T(RG)$ de la qualité de la prévision sur un fichier test de taille $T - 1$ comparable à celle T du fichier d'apprentissage.

Nous avons en particulier pour un échantillon au hasard

$$E[Q_T(RG)] = Q_{T-1}\ (\text{test})\ ;\ \mathcal{D}[Q_T(RG)] = 0\left(\frac{1}{T}\right).$$

Nous présentons dans le tableau 5 les résultats de l'estimation de la qualité par la méthode de reconnaissance glissante.

En pratique nous ne calculons pas bien entendu les T équations de régression directement à partir de leurs fichiers respectifs. Nous obtenons d'abord les coefficients de régression pour le fichier global, puis nous corrigeons à chaque fois ces coefficients en fonction des valeurs du couple $X(j), y(J+1)$.

A cette fin on utilise la formule matricielle pour l'inversion d'une matrice B du type $B = A - Xy'$ où $B = (n \times n)$, $X = (n \times 1)$, $y = (n \times 1)$, $A = (n \times n)$.

La matrice inverse s'exprime alors de la façon suivante :

$$B^{-1} = A^{-1} + \frac{A^{-1}X \,.\, y'A^{-1}}{1 - X'A^{-1}y}$$

Cela nous permet de remplacer l'opération coûteuse en temps machine d'inversion de matrice par des opérations algébriques simples.

En conclusion nous soulignerons l'importance de la pratique opérationnelle de prévision statistique en météorologie qui se situe entre la prévision synoptique subjective et la prévision hydrodynamique objective, mais limitée par le modèle d'atmosphère adopté.

L'effort principal doit être ici l'amélioration de la qualité des fichiers et l'élaboration de prédicteurs secondaires plus informatifs.

Les modèles opérationnels réalisés dans de nombreux pays, montrent que c'est là une voie pleine d'avenir.

BIBLIOGRAPHIE

[1] ANDERSON T. – An introduction to multivariate statistical analysis. New York 1957.

[2] DER MEGREDITCHIAN G. – Un nouveau procédé de réalisation des schémas de discrimination et de régression (en russe). *Météorologie et Hydrologie* Moscou 1969 n°7.

[3] DER MEGREDITCHIAN G. – Méthodes statistiques de prévision par classes en météorologie. *La météorologie.* Paris V. 26-1973.

[4] DER MEGREDITCHIAN G. – Sur la définition du nombre de données fictives indépendantes équivalent à un nombre de données corrélées (en russe) *Météorologie et Hydrologie.* Moscou 1969. N° 2.

[5] LACHENBRUCH A. – An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis *Biometrics* 23 N° 4. 1967.

[6] KLEIN W., LEWIS F., MARSHALL F., COLE M. – An operational system for automated prediction of precipitation probability. *International symposium of probability and statistics on the atmospheric sciences* Honolulu 1971.

I – méthode asynchrone								II – méthode synchrone							
Regression						Discrimination		Regression						Discrimination	
Sp		ΔT		δT		Pluie	Brouillard	Sp		ΔT		δT		Pluie	Brouillard
Pred.	Coeff.	Pred.	Coeff.	Pred.	Coeff.	Pred. Coeff.	Pred. Coeff.	Pred.	Coeff.	Pred.	Coeff.	Pred.	Coeff.	Pred. Coeff.	Pred. Coeff.
$\bar{\Phi}_{850}$	7.52	$\bar{\Phi}_{850}$	9.28	δT	10.65	$\bar{\Phi}_{850}$ -0.57	V^2_{500} -0.24	$\bar{\Phi}_{850}$	9.08	$\bar{\Phi}_{850}$	9.26	δT	10.15	$\bar{\Phi}_{850}$ -1.05	$\bar{\Phi}_{850}$ 0.67
U^1_{500}	5.59	U^1_{500}	6.03	ζ_{675}	-3.55	w -0.44	w -0.38	U^1_{500}	6.49	$\bar{\zeta}_{675}$	-2.39	$\bar{\zeta}_{675}$	-1.67	w -0.53	w -0.58
w	5.12	$u_{min.}$	-7.37	AT^1_{675}	5.16	U^1_{500} -1.13	$\bar{\Phi}_{850}$ 0.89	w	4.52	u_{min}	-7.35	V^1_{850}	8.08	U^3_{500} -0.62	V^3_{850} -0.50
AT^1_{675}	-2.51	V^1_{500}	-6.07	$A\zeta^1_{850}$	-3.20	U^2_{500} 0.63	U^2_{850} -0.35	AT^1_{675}	-4.54	V^3_{500}	-6.77	U^1_{500}	-2.46	$A\zeta^1_{850}$ 0.36	U^2_{850} -0.42
a_o	-191.29	AT^3_{675}	-4.40	V^1_{850}	5.37	ΔT -0.15	V^3_{850} -0.74	$\bar{\zeta}_{675}$	-4.12	U^4_{500}	9.24	V^4_{500}	-3.00	AT^3_{675} 0.14	AT^3_{675} -0.24
		a_o	-159.81	V^3_{500}	-3.98	$A\zeta^4_{850}$ 0.18	V^4_{850} 0.65	$A\zeta^1_{500}$	-2.99	U^2_{850}	-7.58	AT^2_{675}	3.00	$A\zeta^1_{500}$ 0.15	V^3_{500} -0.28
				ΔT	5.40	V^4_{500} 0.21	$\bar{\delta T}$ 0.24	a_o	-221.19	$A\zeta^3_{850}$	-4.60	ΔT	5.84	ΔT -0.18	V^4_{850} 0.37
				$u_{min.}$	4.99	$\bar{\Phi}_{500}$ -0.26	$\bar{\Phi}_{500}$ -0.42			V^3_{850}	6.15	$u_{min.}$	5.87	V^3_{850} -0.62	$A\zeta^1_{500}$ -0.17
				a_o	-13.91	b_o 33.16	$u_{min.}$ 0.15			δT	-4.19	$\bar{\Phi}_{500}$	2.89	V^2_{850} 0.61	$u_{min.}$ 0.41
							b_o -10.19			a_o	-149.34	a_o	-148.98	b_o 40.02	ΔT 0.38
															b_o -22.11

Sp : Coefficient d'ensoleillement (%)

ΔT : Amplitude journalière de température (1/10 degré C)

δT : anomalie climatique de la température journalière moyenne (1/10 degré)

U_{min} : humidité minimum journalière (%)

w : énergie solaire totale journalière reçue à la surface du sol par atmosphère complètement claire (Joule/cm^2/jour),

$\bar{\Phi}_1$: géopotentiel moyen

U^k_1 : vent géostrophique

V^k_1 : vent géostrophique

$\bar{\zeta}_1$: tourbillon moyen absolu

$A\zeta^k_1$: advection de tourbillon

AT^k_1 : advection de température

où l indique l'altitude et k le numéro du point sur la grille de la figure 1.

Table 1 : Liste des prédicteurs retenus par la méthode de sélection progressive ascendante et coefficients respectifs dans les équations prévisionnelles.

$$\tilde{y}/\sigma_y = \sum_{k=1}^{n} a_k x_k/\sigma_k + a_o \quad \text{ou} \quad u(X) = \sum_{k=1}^{n} b_k x_k/\sigma_k + b_o .$$

	Fichier d'apprentissage			Fichier test			Prévision opérationnelle		
	AS	SY	PS	AS	SY	PS	AS	SY	PS
Coefficient d'ensoleillement									
Erreur algébrique moyenne	0.0	0.0	0.3	3.1	2.5	-0.1	-	-	-
Erreur absolue moyenne	18.6	17.2	23.6	21.6	18.7	25.0	25.7	27.7	38.8
Ecart-type de la prévision	10.1	14.3	26.2	11.4	16.2	28.3	-	-	-
Coefficient de corrélation	0.41	0.54	0.21	0.38	0.53	0.26	0.28	0.25	-0.10
Amplitude journalière de température									
Erreur algébrique moyenne	0.0	0.0	-0.0	1.5	4.0	7.5	-	-	-
Erreur absolue moyenne	27.3	25.2	38.8	27.4	25.3	35.5	42.1	40.9	58.1
Ecart-type de la prévision	16.1	21.8	38.4	17.8	24.2	36.2	-	-	-
Coefficient de corrélation	0.43	0.56	0.18	0.39	0.56	0.21	0.21	0.30	0.01
Anomalie climatique de la température journalière moyenne									
Erreur algébrique moyenne	-0.0	0.0	-0.1	4.5	3.7	0.0	-	-	-
Erreur absolue moyenne	13.7	13.1	17.8	14.9	14.7	18.0	22.4	18.4	22.8
Ecart-type de la prévision	14.3	21.2	22.4	14.8	15.2	22.6	-	-	-
Coefficient de corrélation	0.53	0.67	0.49	0.56	0.56	0.44	0.29	0.49	0.35

Table 2 : Vérification de la qualité du schéma de régression sur les trois fichiers d'apprentissage, test et opérationnel.

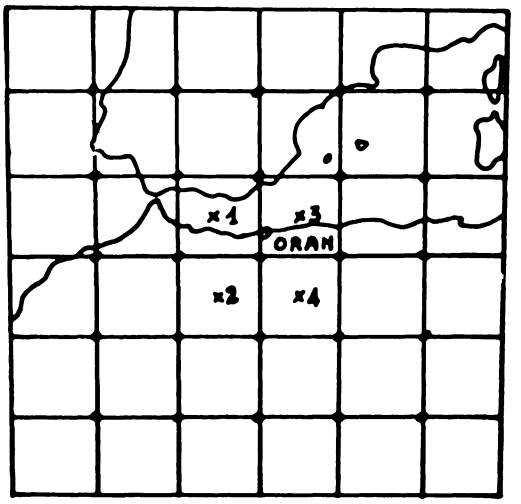

Figure 1 : Grille utilisée pour la prévision locale à Oran.

		Fichier d'apprentissage						Fichier test						Prévision opérationnelle					
		AS		SY		PS		AS		SY		PS		AS		SY		PS	
		Observé						Observé						Observé					
		oui	non	oui	non	oui	non	oui	non	oui	non	oui	non	oui	non	oui	non	oui	non
Pluie prévu	oui	294	463	317	326	146	284	58	67	70	40	37	48	22	31	27	26	13	18
	non	135	1367	121	1495	283	1546	30	180	19	207	51	199	12	21	7	26	21	34
		oui	non	oui	non	oui	non	oui	non	oui	non	oui	non	oui	non	oui	non	oui	non
Brouillard prévu	oui	148	614	153	560	51	170	28	102	27	103	14	22	6	16	7	15	2	7
	non	71	1426	66	1480	168	1870	7	198	6	200	21	278	2	62	1	63	6	71

Table 3 : Vérification de la qualité du schéma discriminant sur les trois fichiers d'apprentissage, test et opérationnel.

	prévision opérationnelle		
	AS	SY	PS
Température minimum			
Erreur absolue moyenne	33,3	31,1	40,9
Coefficient de corrélation	0,43	0,47	0,25
Température maximum			
Erreur absolue moyenne	28,2	23,6	32,2
Coefficient de corrélation	0,46	0,66	0,40
Durée d'ensoleillement			
Erreur absolue moyenne	29,2	31.5	43,9
Coefficient de corrélation	0,44	0,38	0,08

Table 4 : Qualité de la prévision des paramètres primaires (température minimum, maximum, durée d'insolation) pour la prévision opérationnelle.

	T/1	T/2	T/3	T/4	T/5	T/6	T/7
Coefficient d'ensoleillement							
Erreur quadratique moyenne	22,1	22,1	22,1	21,7	22,4	22,2	23,1
Coefficient de corrélation	0,538	0,530	0,538	0,506	0,547	0,507	0,495
Amplitude journalière de température							
Erreur quadratique moyenne	31,3	31,4	32,0	32,4	29,9	32,4	32,3
Coefficient de corrélation	0,559	0,568	0,550	0,560	0,584	0,567	0,553
Anomalie climatique de la température moyenne journalière							
Erreur quadratique moyenne	16,8	17,1	16,6	17,0	17,5	15,7	18,1
Coefficient de corrélation	0,666	0,657	0,688	0,669	0,620	0,726	0,604

Table 5 : Estimation de la qualité de la prévision par la méthode de reconnaissance glissante pour la procédure synchrone SY.