



Schweizerische Meteorologische Anstalt  
Institut suisse de météorologie  
Istituto svizzero di meteorologia  
Swiss Meteorological Institute

No. 174

**Projet No 113**  
**Réseaux de neurones en météorologie:**  
**Une perspective d'application.**

Jacques Ambühl

(présenté à l'EPFL, XIIIème Cours Postgrade en Informatique Technique  
1992, Réseaux de neurones biologiques et artificiels)

Juin 1993

Arbeitsberichte der SMA  
Rapports de travail de l'ISM  
Rapporti di lavoro dell'ISM  
Working Reports of the SMI

© SMA, Publikationen, CH-8044 Zürich

Schweizerische Meteorologische Anstalt  
Krähbühlstrasse 58, Postfach  
CH-8044 Zürich

Tel. (01) 256 91 11, Fax (01) 256 92 78, Telex 81 73 73 met ch

**Projet No 113**  
**Réseaux de neurones en météorologie:**  
**Une perspective d'application.**

Jacques Ambühl

(présenté à l'EPFL, XIIIème Cours Postgrade en Informatique Technique 1992,  
Réseaux de neurones biologiques et artificiels)

Juin 1993

### **Zusammenfassung**

Diese Arbeit beschreibt ein System, das basierend auf Geopotentialfeldern eines numerischen Modells die Abschätzung der Bewölkung über dem Wallis erlaubt. Die Methode besteht aus der Aneinanderreihung zweier neuronaler Netzwerke und eines einfachen Expertensystems.

Das erste Netzwerk ist eine Kohonen-Karte, die eine Klassifikation der Strömungsmuster auf 500 hPa vornimmt. Das zweite ist ein Perceptron, das eine Diagnose der auf der Kohonen-Karte definierten Kategorien durchführt. Schliesslich werden die Aktivitäten der Neuronen am Ausgang des Perceptrons in logische Einheiten übersetzt, die in Prolog formuliert sind und in der dynamischen Datenbasis des Expertensystems stehen.

Einige Gedanken über die Art und Weise der Verbindung eines neuronalen Konzeptes mit einem regelbasierten System schliessen die Arbeit ab.

### **Résumé**

Le présent travail décrit un dispositif de diagnostic de la nébulosité sur le Valais opérant à partir des champs géopotentiels livrés par un modèle numérique en mettant en oeuvre successivement deux réseaux de neurones et un petit système expert.

Le premier des réseaux est une carte de Kohonen réalisant une classification des flux à 500 hPa. Le second est un Perceptron portant un diagnostic sur les catégories définies sur le Kohonen. Enfin, un interface traduit les activités des neurones de sortie du perceptron en faits logiques, rédigés selon le formalisme de Prolog, et inscrits dans la base de faits dynamiques du système expert.

Quelques réflexions portant sur les modalités de connexion d'un dispositif neuronal à un système à règle sont présentées en guise de conclusion.

## **Riassunto**

Si presenta una descrizione di un dispositivo di diagnosi della nuvolosità sul Vallese usando i campi di geopotenziali provvisti da un modello numerico e operando successivamente su due reti di neuroni ed un sistema esperto.

La prima rete è una carta di Kohonen che seleziona in diverse categorie i flussi a 500 hPa. La seconda rete è un "perceptron,, che produce una diagnosi sulle categorie definite dalla rete Kohonen. Al termine una traduzione dei risultati dati dal "perceptron,, in fatti logici sono trascritti secondo il formalismo del Prolog e iscritti nella base di dati dinamici del sistema esperto.

In conclusione, sono presentate qualche riflessione sulle modalità delle connessioni tra un dispositivo neuronale ed un sistema esperto con regole.

## **Summary**

This work describes a device for diagnosing cloudiness over the region of Valais based on geopotential fields produced by a numerical model. These data provide the input to a sequence of two neural networks followed by a small expert system.

The first of these networks is a Kohonen map producing classifications of the 500 hPa flow fields. The second one is a perceptron performing a diagnosis of the categories defined by the Kohonen. Finally, activities of the output neurons of the perceptron are translated into logical facts written under Prolog specifications and registered in the dynamic data base of the expert system.

A short reflection on the way of connecting a neural device to a rule-based system concludes the paper.

# **Plan**

## **1. Introduction**

## **2. Les données météorologiques de base**

## **3. Le réseau de Kohonen**

### **3.1 L'architecture**

### **3.2 L'auto-organisation**

## **4. Le perceptron**

### **3.1 L'architecture**

### **3.2 L'apprentissage**

## **5. Le couplage à un système expert**

### **5.1. La structure du petit système expert**

### **5.2 Conversion de l'activité d'un neurone en la plausibilité d'un prédicat**

## **6. Le problème général du couplage**

## **7. Conclusion**

## **8. Bibliographie**

Ce travail a été effectué dans le cadre du Cours Postgrade d'Informatique de l'Ecole Polytechnique Fédérale de Lausanne, consacré en 1992 aux «Réseaux de neurones biologiques et artificiels». Il a fait l'objet d'un examen de défense à la suite duquel une attestation de troisième cycle a été remise à l'auteur.

## 1. Introduction

La prévision météorologique n'est plus concevable, en cette fin de siècle, sans l'intervention des modèles numériques. Ceux-ci autorisent des prévisions à grande échelle portant sur 5 à 7 jours. Diverses techniques sont par conséquent engagées afin de réduire l'échelle spatiale. L'une d'entre elles consiste à imbriquer un modèle numérique à fine maille (Péridot de Météo-France, Mesoscale Britannique, Mésomod Germano-Suisse) dans un modèle à grande échelle dont la maille est plus grossière. Une variante nouvelle de cette approche, encore en développement mais promise à un bel avenir, revient à construire un modèle spectral auquel est associée une grille de maille variable, fine sur la région où la prévision est utilisée, plus grossière ailleurs (Arpège de Météo-France).

La méthode envisagée dans le cadre du projet présenté ici est au contraire de nature *diagnostique*.

Elle consiste à établir une relation de nature heuristique, ou statistique, entre la circulation générale telle qu'elle est prévue par un modèle numérique à grande échelle, et le déroulement local d'un phénomène météorologique, ici l'évolution de la nébulosité sur le Valais.

Les champs en prévision élaborés par le modèle numérique, outre qu'ils sont globaux, ne sont entachés d'aucune probabilité, le modèle étant déterministe, et les algorithmes diagnostiques à envisager devront donc traduire des prévisions *globales et déterministes* en prévisions *locales*, pouvant être affectées de *facteurs de plausibilité*.

Le projet est organisé selon la figure 1:

Un champ météorologique à large échelle, le géopotential 500 hPa sur l'Europe, est tout d'abord présenté à une carte de Kohonen grâce à laquelle est réalisée une classification : à chaque lieu de la carte doit correspondre un type de circulation météorologique (cyclonique zonale, dépression sur le sud des Alpes, anticyclonique du nord-ouest, ...).

La carte de Kohonen est ensuite présentée à un perceptron disposant de quatre neurones dans sa couche de sortie, correspondant respectivement à l'évaluation des quatre éléments suivants :

- . Le caractère cyclonique ou anticyclonique de la circulation présentée.
- . La présence d'un thalweg sur le flanc ouest de la Chaîne des Alpes.
- . La présence d'un jet au nord-ouest de la Chaîne des Alpes, orienté du nord-ouest au sud-est.
- . Une évaluation de la nébulosité couvrant le Valais.

Les activités des trois premiers neurones sont enfin exploitées pour rédiger trois faits en langage Prolog, introduits dans la base de faits dynamiques d'un petit système expert, représenté au bas de la figure 1, et qui, en tenant compte également de la force et de la direction du vent soufflant sur la Suisse à 500 hPa, élabore un diagnostic de nébulosité-enseuillement sur le Valais.

Les buts que je me suis fixés sont :

- . Ecrire un réseau de Kohonen, en maîtriser l'auto-organisation.
- . Ecrire des perceptrons à deux et trois couches. Maîtriser les processus d'apprentissage.
- . Elaborer un dispositif traduisant les données émanant d'un réseau de neurones en prédicats Prolog inscrits dans la base de faits dynamiques d'un petit système expert.

A titre personnel enfin, une connaissance fine des algorithmes neuronaux, de leurs limites, a primé l'objectif de réalisation d'un système devant absolument fonctionner de manière opérationnelle. J'ai donc souvent, dans ma démarche, privilégié les essais et les voies de traverse.

Donnée d'entrée :

Champ géopotential 500 hPa

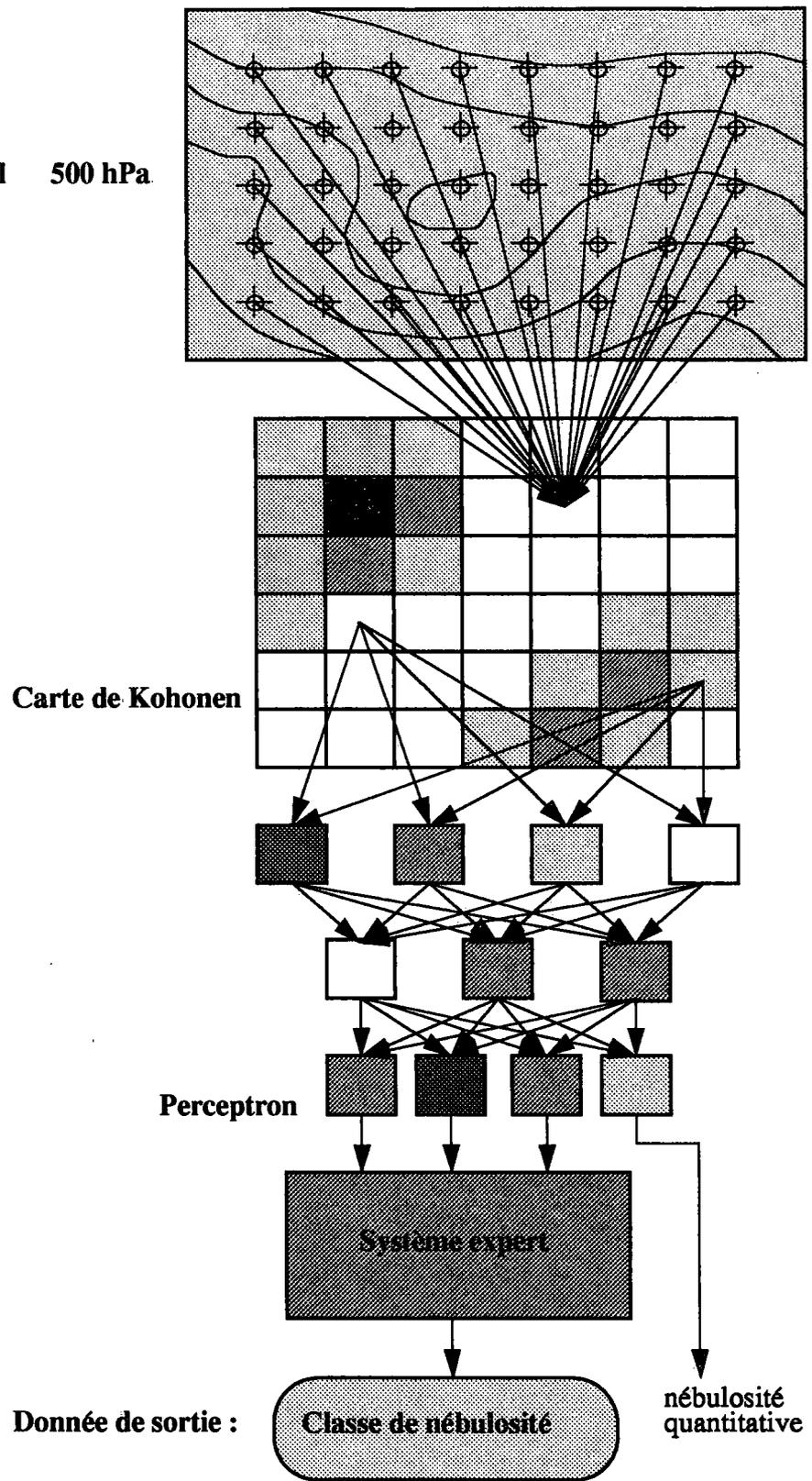


Figure 1 : Organisation générale du projet

## 2. Les données météorologiques de base

La donnée entrant dans le dispositif est le *champ géopotential 500 hPa*, défini comme suit :

$$h(x, y) = \frac{R}{g} \int_{p_0}^{p_1} T(x, y, p) d \ln(p)$$

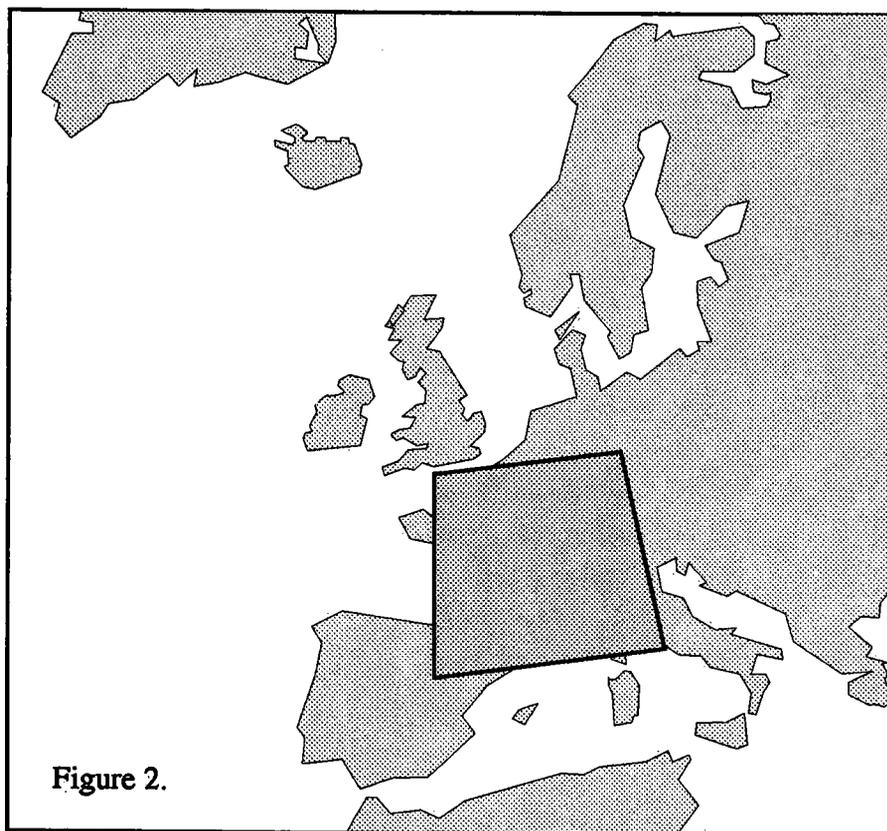
où  $p_0 = 500$  hPa,  $p_1$  est la pression à la surface,  $g$  l'accélération de la pesanteur,  $R$  la constante du gaz parfait et  $T(x, y, p)$  la température à chaque niveau de pression  $p$  au point de coordonnées géographiques  $(x, y)$ . Le géopotential a une dimension de longueur et représente en fait l'altitude à laquelle est situé, en un endroit donné, le niveau de pression 500 hPa.

Un exemple de ce champ, fondamental pour le météorologue, est représenté par ses isohypses sur la carte de la figure 4.

350 champs, donnés chaque jour à 0 h. GMT, du 2 mars 1990 au 14 février 1991, ont été utilisés. Ils sont spécifiés en  $10 \times 10$  points régulièrement distribués sur une grille latitude-longitude recouvrant le carré curviligne dessiné sur la carte de la figure 2, avec les conventions suivantes :

Latitude min =  $42^\circ$  N,  
Longitude min =  $1,5^\circ$  W  
delta latitude =  $1,0^\circ$

Latitude max =  $51^\circ$  N  
Longitude max =  $12,0^\circ$  E  
delta longitude =  $1,5^\circ$ .



Toutes les cartes présentées ci-après correspondent exactement à ce domaine.

A chaque champ 500 hPa, et pour chaque jour, sont affectées trois caractéristiques qualitatives et une donnée quantitative. Ces quatre paramètres déterminent essentiellement la nébulosité sur le Valais, ce sont :

1. La vorticité, notée  $\xi$ , cyclonique (figure 3.1 gauche), ou respectivement anticyclonique :

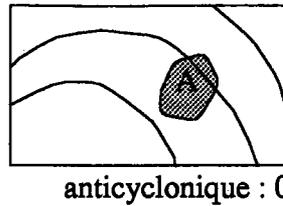
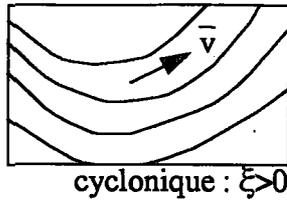


figure 3.1.

Formellement, la vorticité  $\xi$  est définie par :

$$\xi = \lim_{|A| \rightarrow 0} \oint_{\partial A} \bar{V} \wedge \bar{V} \cdot \vec{dl}$$

où  $\bar{V}$  désigne le vent dans le champ de géopotential, et A un élément de surface que l'on fait tendre vers zéro. On peut démontrer que la vorticité est approximée par le Laplacien du géopotential. Dans la pratique, un météorologue expérimenté distingue «à l'oeil» et avec une grande assurance si, en un endroit donné, la circulation est cyclonique (vorticité positive) ou si elle ne l'est pas, (vorticité négative).

2. La présence d'un thalweg à l'ouest de la chaîne des Alpes, (Figure 3.2) :

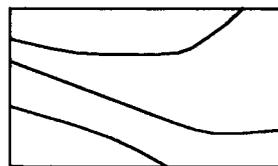
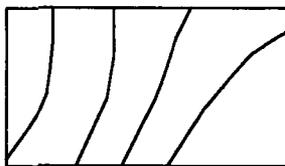


figure 3.2.

3. La présence au nord-ouest des Alpes d'un jet-stream orienté du nord-ouest au sud-est, (figure 3.3).

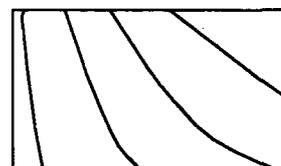
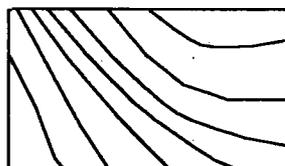


figure 3.3

Deux autres caractéristiques objectives, quantitatives, sont introduites :

4. Le vent soufflant à 500 hPa à 0 h. GMT à Payerne (direction et force).  
A nouveau, le vent peut être approximé par une forme différentielle:

$$\bar{V} = \frac{g}{2\Omega \sin \lambda} (\bar{k} \wedge \nabla h)$$

dans laquelle g est l'accélération de la pesanteur,  $\Omega$  la vitesse angulaire de la planète,  $\lambda$  la latitude,  $\bar{V}h$  le gradient du géopotential et  $\bar{k}$  le vecteur unitaire normal à la surface terrestre.

5. La nébulosité moyenne observée sur le Valais, évaluée à partir des données émanant des stations météorologiques de Sion, Zermatt, Montana et, avec une légère extension, Le Sepey.

Ces cinq caractéristiques, dont les trois premières ont été évaluées à l'oeil et à la main par deux météorologues expérimentés en portant un jugement subjectif sur les cartes synoptiques, sont stockées et font partie du fichier d'apprentissage.

Le choix de ces paramètres n'est pas fortuit. Les quatre premiers forment l'ensemble des faits pris en



### 3. Le réseau de Kohonen

La fonction du réseau de Kohonen est d'établir une classification des situations météorologiques présentées.

L'ensemble des algorithmes décrits ci-après ont été programmés en Pascal sur une machine HP-Vectra dotée d'un processeur 386. Les structures ont été implémentées en construisant des pointeurs dynamiques en mémoire.

Tous les graphiques ont été réalisés à l'aide d'une procédure de dessin d'isolignes écrite en Pascal et procédant par interpolation bilinéaire.

#### 3.1. L'architecture du réseau

Le champ géopotential, stimulus d'entrée du réseau de Kohonen, est représenté par la matrice  $X_{kl}$ , où  $k = 0 \dots k_{\max}$  et  $l = 0 \dots l_{\max}$ , avec  $k_{\max} = l_{\max} = 9$ .

Chaque neurone est distingué sur la carte par un couple  $(i,j)$  d'indices, où  $i = 0 \dots i_{\max}$  et  $j = 0 \dots j_{\max}$ .

A chaque neurone  $(i,j)$  est associée sa matrice de poids synaptiques  $\underline{W}^{ij} = W_{kl}^{ij}$ .

L'activité de chaque neurone est notée  $n_{ij}$ .

L'indice  $t = 0 \dots T_x$  compte les pas d'apprentissage.

La figure 5 illustre ces conventions :

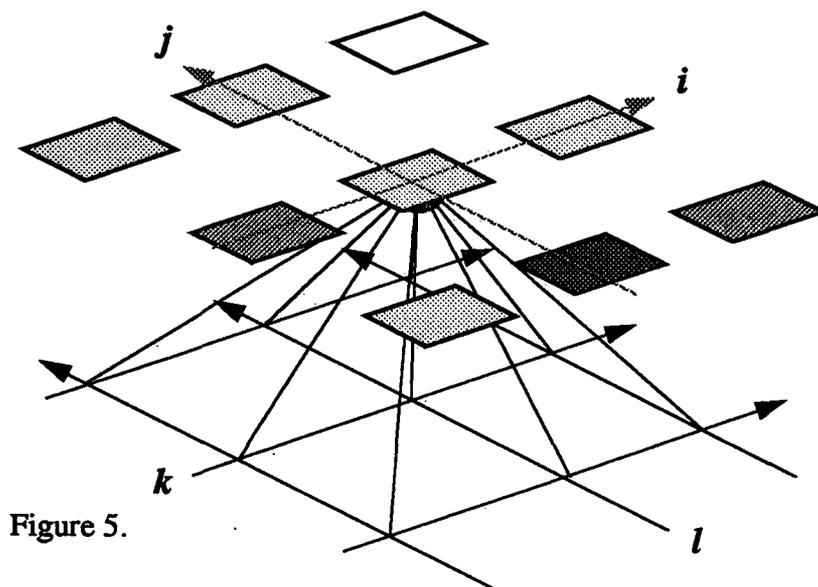


Figure 5.

Deux distances sont définies :

L'une sur la carte de Kohonen, entre tout couple de neurones  $(i,j)$  et  $(i',j')$  :

$$\delta_{((i,j), (i',j'))} = \sqrt{(i-i')^2 + (j-j')^2}$$

L'autre sur l'espace des stimuli :

$$d^2_{(X, \underline{W}^j)} = \sum_{k,l} (X_{kl} - W_{kl}^{ij})^2$$

Deux fonctions d'apprentissage sont données, dépendant du pas t d'apprentissage :

L'une spécifiant la taille de la bulle d'activité dans le réseau:

$$\sigma_{(t)} = \frac{\sigma_0}{1 + e^{\left[\frac{2t}{Tx} - 1\right]}} \quad \sigma_0 = \frac{\sqrt{i_{max}^2 + j_{max}^2}}{4}$$

$\sigma_0$  est choisi égal au quart de la longueur de la diagonale de la carte de Kohonen.

L'autre le taux d'apprentissage :

$$\tau_{(t)} = \tau_0 \left[ \frac{1}{1 + \nu d^2_{(X, \underline{W}^j)}} \right] \left[ \frac{1}{1 + e^{\left[\frac{2t}{Tx} - 1\right]}} - \frac{1}{1 + e} \right] \quad \tau_0 = 1 \quad \nu = \frac{1}{10}$$

Deux termes interviennent dans le calcul du taux d'apprentissage :

Le second tout d'abord, correspondant au deuxième crochet, diminue le taux à chaque pas t et tend vers 0 lorsque t tend vers Tx. Le premier enfin, correspondant au premier crochet, inhibe légèrement l'apprentissage lorsque la distance entre le champ présenté  $\underline{X}$  et la matrice des poids synaptiques du neurone élu  $\underline{W}^j$  est grande.  $\tau_{(t=0)}$  est de l'ordre de 0.3.

L'algorithme d'apprentissage est programmé comme suit :

**Pas initial t = 0 :**

Pour tout neurone (i,j), affecter aux poids synaptiques  $W_{kl}^{ij}$  de petites valeurs aléatoires.

**Pas général t :**

L'algorithme d'apprentissage détermine tout d'abord un neurone élu (I,J), satisfaisant :

$$\min_{ij} [d^2_{(X, \underline{W}^j)}] = d^2_{(X, \underline{W}^{IJ})}$$

Puis, pour tout neurone (i,j) tel que :

$$\delta_{((i,j), (I,J))} \leq \sigma_{(t)}$$

Effectuer :

$$\underline{W}^{ij}_{(t+1)} = \underline{W}^{ij}_{(t)} + \tau_{(t)} [X - \underline{W}^{ij}_{(t)}]$$

Normaliser :

$$\underline{W}^{ij}_{(t+1)} \leftarrow \frac{\underline{W}^{ij}_{(t+1)}}{\|\underline{W}^{ij}_{(t+1)}\|}$$

**Fin de l'algorithme.**

### 3.2. L'auto-organisation du réseau

La figure 6 illustre l'état du réseau au début de l'auto-organisation. Le champ géopotentiel présenté au réseau est dessiné en bleu clair à l'arrière plan. L'état de la matrice des poids synaptiques du neurone élu  $I=13$ ,  $J=15$ , représentée par les lignes isohypses, est dessiné en blanc. L'activité des neurones est figurée sur la grille.

Les figures 7 à 10 exhibent quelques réponses de la carte de Kohonen après l'auto-organisation. Les isolignes du champ météorologique sont en rouge, celles qui sont associées à la matrice des poids synaptiques en noir. Les cartes de Kohonen sont représentées en bas à droite des figures.

- La situation du 2 mars 1990, caractérisée par un vigoureux flux cyclonique d'ouest à 500 hPa, est présentée au réseau sur la figure 7. La bulle d'activité est dessinée en blanc. La matrice synaptique du neurone élu (0,0) reproduit relativement fidèlement le champ présenté et détecte la cyclonalité.

- Un flux anticyclonique est présenté au réseau sur la figure 8. La bulle d'activité, de dimension plus considérable, est grossièrement située au milieu du réseau. A nouveau, la matrice synaptique du neurone élu (8,5) reproduit le flux présenté en en détectant parfaitement le caractère anticyclonique.

- Une situation de dépression centrée au sud-est des Alpes illustre la figure 9. La coïncidence entre le flux météorologique et les poids synaptiques du neurone élu est moins bonne, bien que le neurone élu caractérise une situation de même type.

- La figure 10 enfin, caractérise une situation de col, moins fréquente. Le neurone élu reproduit cependant lui aussi un col.

Deux remarques :

. A l'examen de nombreuses cartes de Kohonen, on observe que les bulles d'activité sont de petit diamètre lorsque les situations météorologiques sont actives, notamment lorsque le gradient du géopotentiel est élevé, alors que les bulles sont de diamètre plus grand dans les situations peu actives, atypiques, essentiellement anticycloniques.

. Sur un plan plus général, le stimulus d'entrée  $\underline{X}$  dans le Kohonen et la matrice des poids synaptiques  $\underline{W}^{ij}$  sont tous deux éléments du même espace  $\mathbb{R}^N$ , muni de la distance  $d^2$  définie au paragraphe 3.1, et tel que  $N = (k_{\max}+1) \times (l_{\max}+1)$ .

Ainsi donc, s'il est possible, comme on le fait en météorologie, de représenter et interpréter un stimulus  $\underline{X}$ , soit pour nous un champ géopotentiel, il doit être nécessairement possible d'appliquer le même canevas de représentation et d'interprétation à la matrice  $\underline{W}$  des poids synaptiques de chaque neurone du réseau de Kohonen, après auto-organisation.

Cette propriété apparaît naturellement sur les figures 7 à 10, où l'on constate que l'on peut discerner une circulation cyclonique ou anticyclonique, un col ou la présence d'une dépression, *aussi bien sur le champ généré par les poids synaptiques que sur le champ géopotentiel.*

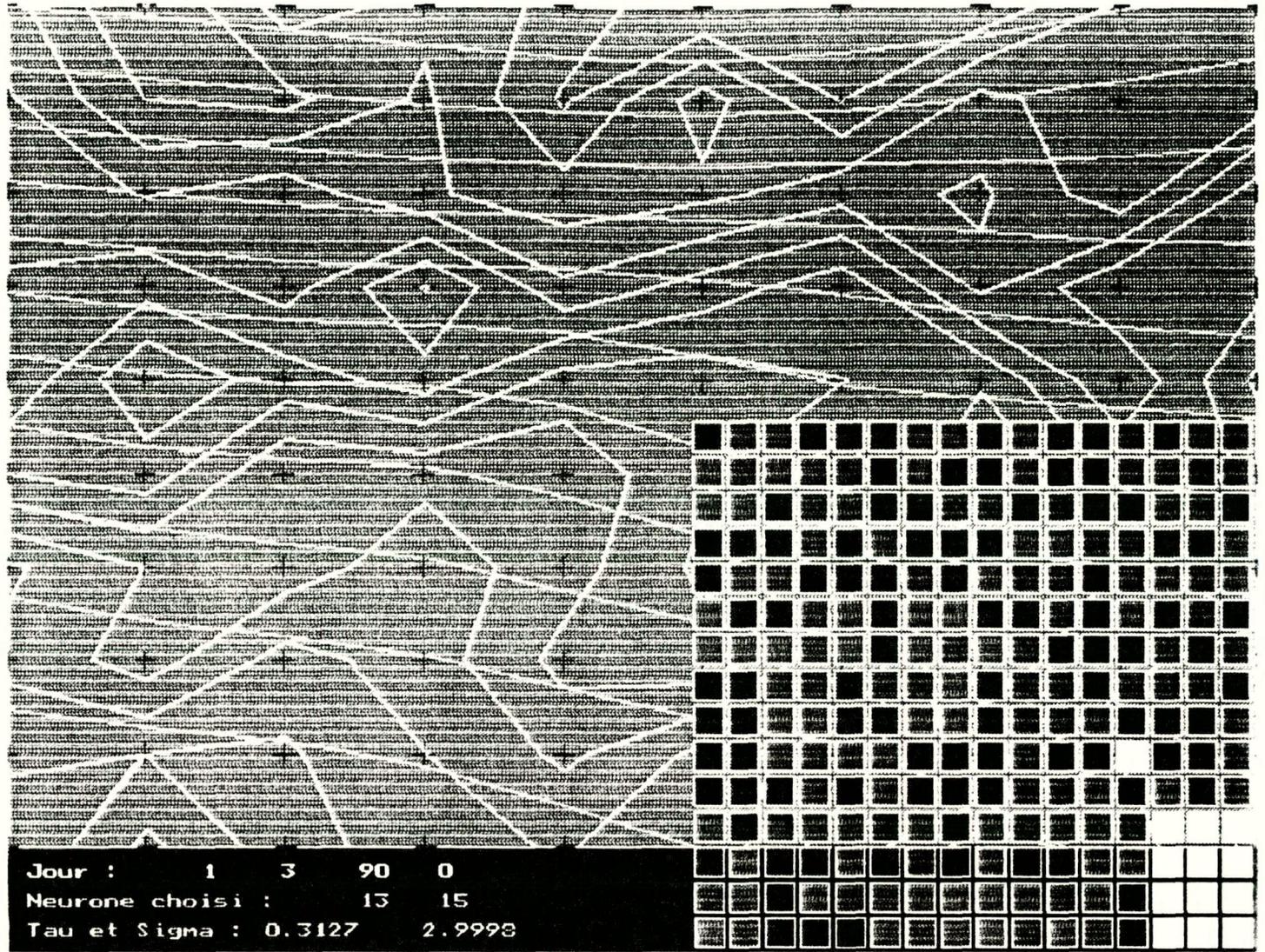


FIGURE 6.

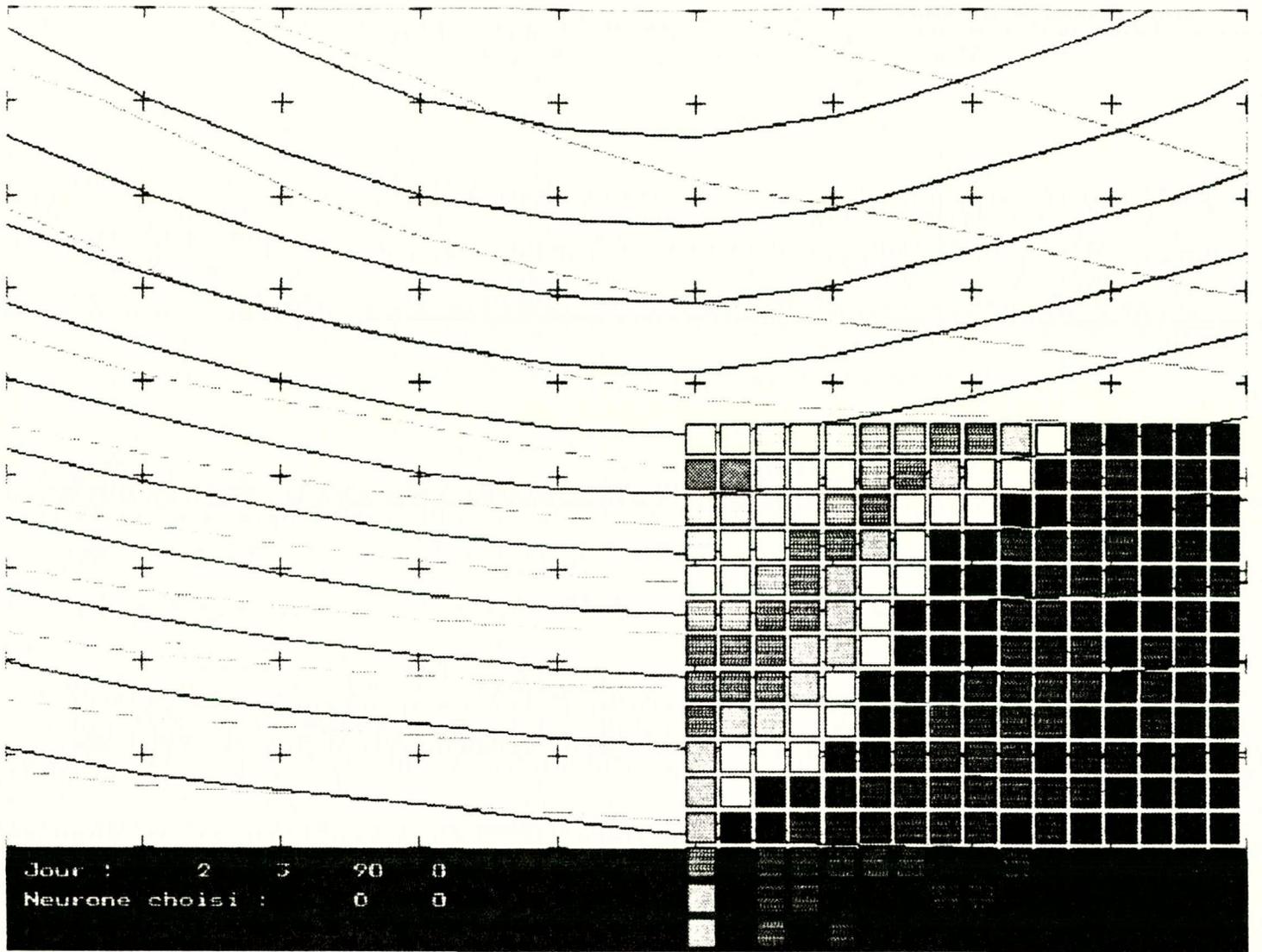


FIGURE 7.

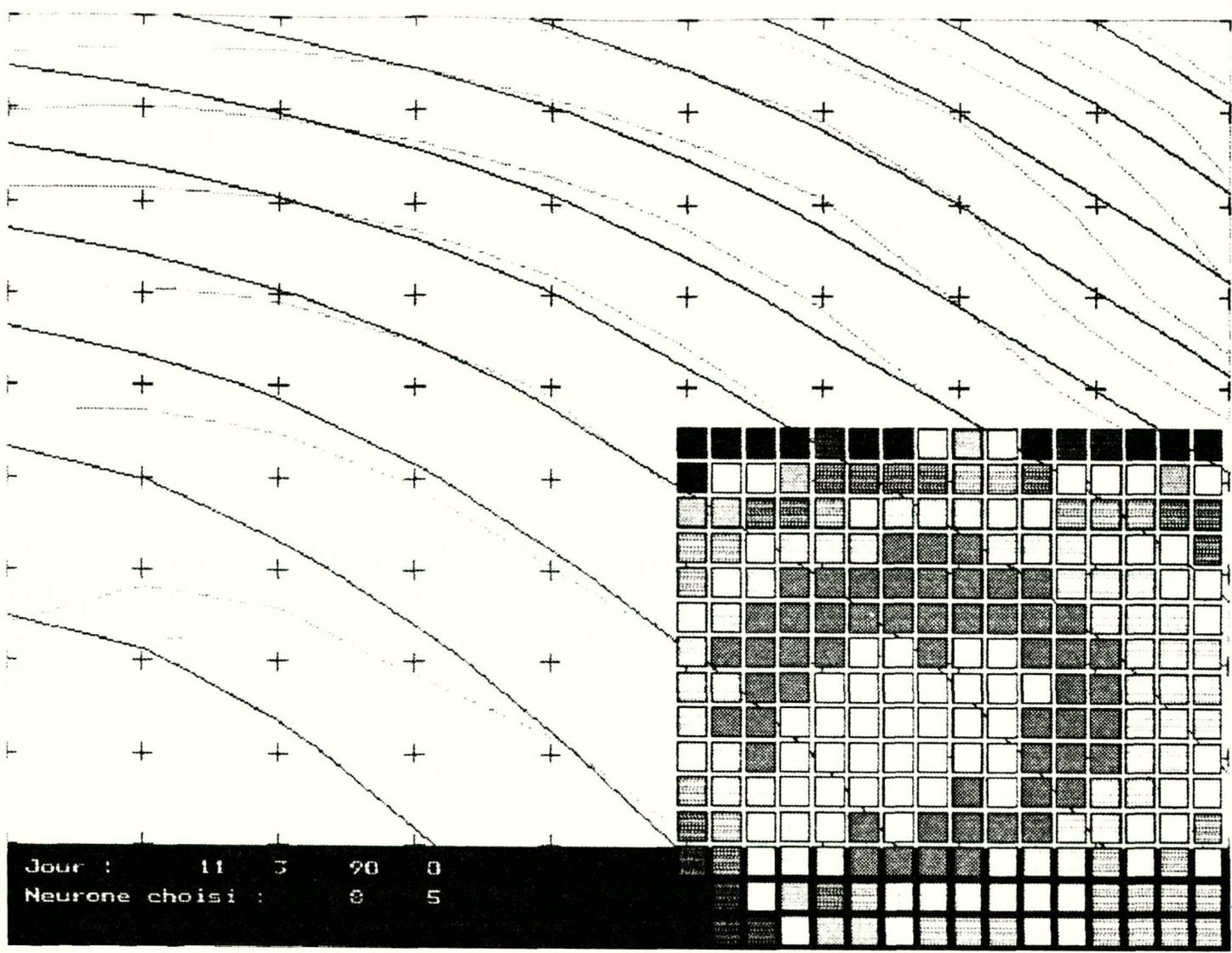


FIGURE 8.

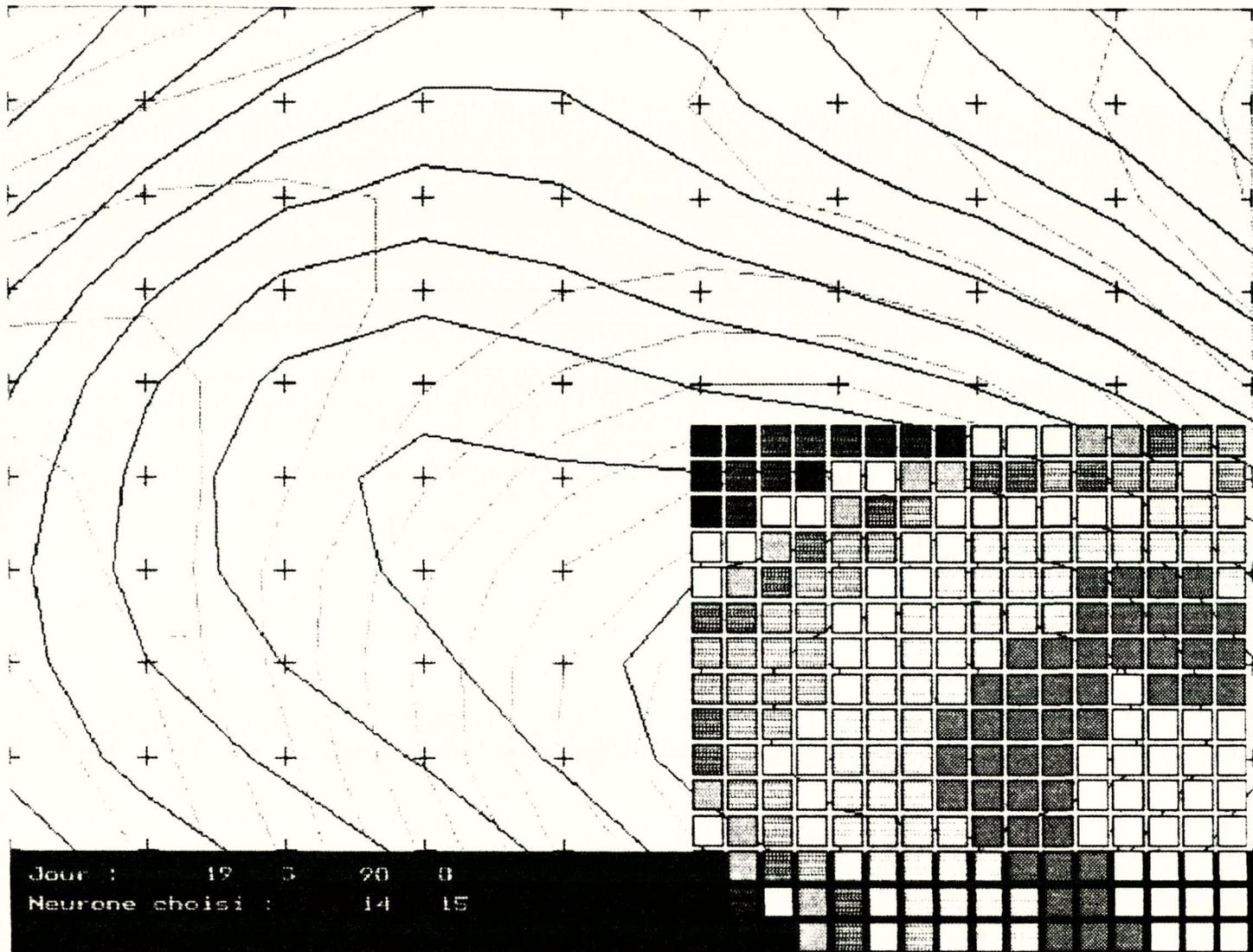


FIGURE 9

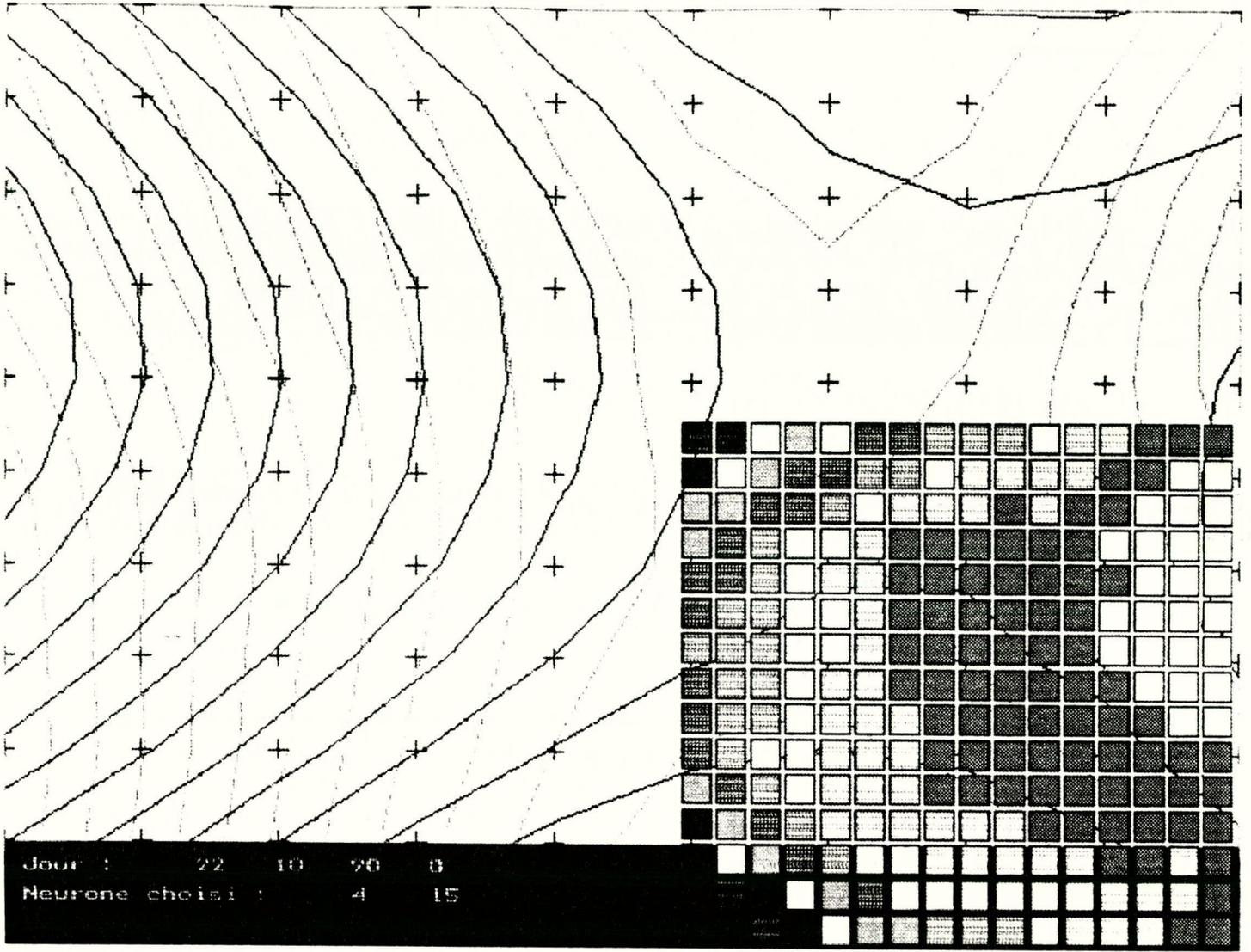


FIGURE 10

## 4. Le perceptron

La fonction du perceptron est de diagnostiquer les quatre paramètres de vorticit , de pr sence de thalweg, de jet, et la n bulosit .

Ici  galement, le langage Pascal a  t  utilis  sur la machine Vectra/386, et les structures sont g n r es sous forme de pointeurs dynamiques.

### 4.1 L'architecture

Deux perceptrons ont  t  consid r s, l'un   trois couches, l'autre   deux couches.

Le premier, comportant trois couches, dont une cach e, est form  de  $n_1=16$ ,  $n_2=12$  et  $n_3=4$  neurones. Il est pr sent    la figure 11. Le perceptron y est dessin    gauche de la carte de Kohonen, avec, de haut en bas, les quatre neurones de sortie: vorticit , thalweg, jet et n bulosit . Les isohypses 500 hPa apparaissant en rouge en arri re plan sont celles de la situation du 15 mars 1990, qualifi e d'anticyclonique par l'expert, sans qu'il y ait de jet ni de thalweg. A 500 hpa   Payerne, le vent ce jour-l  souffle d'est - sud-est   5 noeuds et le ciel est clair sur le valais avec une nebulosit   gale   0/8.

Les quatre carreaux jaunes et verts, d crivant l' tat de l'apprentissage, sont expliqu s   la fin du paragraphe 4.2. consacr    l'apprentissage.

N'ayant pas converg  de mani re satisfaisante, il a  t  transform  en un perceptron   deux couches de 12, respectivement 4 neurones (figure 12).

La matrice des poids synaptiques de la  $k^{\text{ me}}$  couche du perceptron est not e  $\underline{W}^k$

La fonction sigmo de utilis e, et sa d riv e, sont :

$$\sigma_{\kappa}(x) = \frac{1}{1 + e^{-\kappa x}} \quad \sigma'_{\kappa}(x) = \frac{\kappa e^{-\kappa x}}{(1 + e^{-\kappa x})^2}$$

.  $\kappa$  d termine la r ponse de la sigmo de, et est  quivalent   l'inverse d'une temp rature.

### 4.2 L'apprentissage

La t che T assign e est d finie par :  $T = \{\{\underline{n}^1 \dots \underline{n}^{\mu} \dots \underline{n}^m\}, \{\zeta^1 \dots \zeta^{\mu} \dots \zeta^m\}\}$

o  : .  $\mu$  est l'indice comptant les jours dans le fichier d'apprentissage conform ment   l'ordonnement sp cifi    la fin du paragraphe 2.

.  $\underline{n}^{\mu}$  est le vecteur form  des activit s  manant de la carte de Kohonen, calcul es le jour  $\mu$ .

.  $\zeta^{\mu}$  est un vecteur d termin  par les param tres d'apprentissage, form  des quatre quantit s suivantes :

$\zeta_1^{\mu} = 0.85$  si la circulation est cyclonique,  $\zeta_1^{\mu} = 0.15$  si elle ne l'est pas.

$\zeta_2^{\mu} = 0.85$  si un thalweg est pr sent,  $\zeta_2^{\mu} = 0.15$  s'il n'y a pas de thalweg.

$\zeta_3^{\mu} = 0.85$  si un jet est pr sent,  $\zeta_3^{\mu} = 0.15$  s'il n'y a pas de jet.

$\zeta_4^{\mu} = \text{neb}/8$ , o  neb est la quantit  de n bulosit  donn e dans les param tres d'apprentissage.

Les deux fonctions d'apprentissage sont simplement linéaires décroissantes :

$$\eta_{(t)} = \eta_0 + \left(\frac{t}{T_x}\right) (\eta_1 - \eta_0) \quad \eta_0 = 0,6 \quad \eta_1 = 0,3$$

$$\alpha_{(t)} = \alpha_0 + \left(\frac{t}{T_x}\right) (\alpha_1 - \alpha_0) \quad \alpha_0 = 0,4 \quad \alpha_1 = 0,2$$

L'algorithme d'apprentissage procède à une rétropropagation classique, avec momentum :

**Pas initial t = 0 :**

A tous les poids synaptiques du perceptron sont affectées de petites valeurs aléatoires.

**Pas général :**

Répéter :

1. Calcul direct des activités dans les trois couches :

$$a_k = \sigma_{\kappa_1} \left( \sum_{i,j} W_{ij}^k n_{ij} \right) \quad \tilde{a}_k = \sigma'_{\kappa_1} \left( \sum_{i,j} W_{ij}^k n_{ij} \right) \quad k = 1 \dots n_1$$

$$a_l = \sigma_{\kappa_2} \left( \sum_k W^l_k a_k \right) \quad \tilde{a}_l = \sigma'_{\kappa_2} \left( \sum_k W^l_k a_k \right) \quad l = 1 \dots n_2$$

$$a_m = \sigma_{\kappa_2} \left( \sum_l W^m_l a_l \right) \quad \tilde{a}_m = \sigma'_{\kappa_2} \left( \sum_l W^m_l a_l \right) \quad m = 1 \dots n_3$$

2. Calcul rétrograde des deltas dans les trois couches :

$$\delta_m = \tilde{a}_m [c_m^\mu - a_m] \quad m = 1 \dots n_3$$

$$\delta_l = \tilde{a}_l \sum_m W^m_l \delta_m \quad l = 1 \dots n_2$$

$$\delta_k = \tilde{a}_k \sum_l W^l_k \delta_l \quad k = 1 \dots n_1$$

3. Adaptation des poids synaptiques (avec momentum) :

$$\Delta W^k_{ij} = \delta_k n_{ij} \quad W^k_{ij} = W^k_{ij} + \eta_{(t)} \Delta W^k_{ij} + \alpha_{(t)} \Delta W^k_{ij \text{vieux}} \quad \Delta W^k_{ij \text{vieux}} = \Delta W^k_{ij}$$

$$\Delta W^l_k = \delta_l a_k \quad W^l_k = W^l_k + \eta_{(t)} \Delta W^l_k + \alpha_{(t)} \Delta W^l_{k \text{vieux}} \quad \Delta W^l_{k \text{vieux}} = \Delta W^l_k$$

$$\Delta W^m_l = \delta_m a_l \quad W^m_l = W^m_l + \eta_{(t)} \Delta W^m_l + \alpha_{(t)} \Delta W^m_{l \text{vieux}} \quad \Delta W^m_{l \text{vieux}} = \Delta W^m_l$$

**Fin de l'algorithme.**

Le choix des deux bornes  $\zeta = 0.85$  et  $\zeta = 0.15$ , n'est pas fortuit. Requérir en effet que les valeurs de  $\zeta$  s'approchent de 0 et de 1 conduirait à forcer les réponses des neurones de sortie du perceptron sur les plateaux asymptotiques des sigmoïdes. Dans une telle circonstance, les dérivées des sigmoïdes seraient voisines de zéro et tout le processus d'apprentissage serait arrêté, attendu que ces dérivées interviennent de manière multiplicative dans le calcul rétrograde des  $\delta$ .

Après de nombreux essais pour le moins laborieux, portant notamment sur le nombre de neurones formant la couche d'entrée et la couche cachée, et sans qu'aucune convergence irréfutable ne soit observée, je me suis résigné à transformer le perceptron à trois couches en un perceptron à deux couches. Dès lors, une faible convergence est apparue et, en jouant sur les valeurs de  $\kappa$  dans l'une et l'autre des deux couches, il s'est avéré que la meilleure convergence était obtenue avec des valeurs variables de  $\kappa$  d'une couche à l'autre.

A cet égard, des outils théoriques objectifs permettant de déterminer les dimensions optimales des couches, de même que les valeurs optimales des coefficients  $k$ , se seraient avérés utiles.

Finalement, le meilleur apprentissage obtenu, conduit selon l'algorithme spécifié ci-dessus sur le perceptron à deux couches présenté à la figure 12 avec  $\kappa_1 = 0.5$  et  $\kappa_2 = 2.0$ , a abouti aux fonctions d'erreur présentées ci-dessous

**Table 1:**

pas	cyclonalité	thalweg	jet	nébulosité
1	6.92	7.58	5.15	11.23
2	7.84	8.09	6.00	11.19
3	6.78	6.87	4.64	9.07
4	7.13	6.38	4.16	9.78
5	5.42	4.99	3.80	8.19
6	4.95	4.62	3.87	7.89
7	4.69	4.36	3.11	7.43
8	4.18	4.34	3.09	7.06
9	4.04	3.91	2.77	6.51
10	3.97	3.84	2.76	5.79
11	3.78	3.59	2.66	5.43
12	3.58	3.46	2.54	5.11
13	3.43	3.32	2.47	4.88
14	3.38	3.28	2.49	4.75
15	3.49	3.39	2.62	4.72

Les figures 11 et 12 illustrent le processus d'apprentissage du perceptron :

A la figure 12, la situation du 25 mars 1990 est présentée au réseau de Kohonen déjà organisé, qui prend une configuration, livrée en entrée au perceptron :

La circulation ce jour-là est cyclonique,  $\zeta_1 = 0.85$  (dans le premier casier vert au bas de la figure), et le neurone «cyclonalité» répond avec une activité  $a_1 = 0.98$ .

Un thalweg, léger, est repéré à l'ouest des Alpes:  $\zeta_2 = 0.85$ , le neurone «thalweg» répond avec une activité  $a_2 = 0.62$ .

Il n'y a pas de jet,  $\zeta_3 = 0.15$ , la réponse du neurone correspondant est  $a_3 = 0.15$ .

Enfin, la nébulosité moyenne estimée sur le Valais est ce jour là de 6/8,  $\zeta_4 = 0.75$ , et le neurone «nébulosité», avec une réponse  $a_4 = 0$ , est totalement faux.

Le code de couleur vert jaune et rouge indique simplement si le neurone est proche du stimulus (vert), s'il en est peu ou très éloigné (jaune, respectivement rouge).

Trois raisons au moins peuvent être envisagées pour expliquer ce faible résultat :

- La première résulte de la nature même du fichier des paramètres d'apprentissage décrit à la fin du paragraphe 2. Pour de nombreuses situations météorologiques en effet, il apparaît que la vorticité, la présence d'un thalweg ou d'un jet sont ambiguës. Les situations sont en effet souvent peu cycloniques, ou faiblement anticycloniques et les jets sont parfois mal marqués, de même que les thalwegs peuvent être peu accentués, légèrement décalés, ...

Dans toutes ces situations intermédiaires, mal déterminées, les deux experts météorologues se sont astreints à trancher de manière dichotomique, alors qu'une graduation eût été nécessaire.

- En deuxième lieu, la rétropropagation du gradient est, comme son nom l'indique, une méthode du premier ordre. Des méthodes de deuxième ordre, faisant usage de l'inverse des matrices Hessiennes, dérivées secondes des fonctions d'erreur par rapport aux poids synaptiques, auraient pu conduire à de meilleures convergences.

- Enfin, d'autres types de couplage entre le réseau de Kohonen et le perceptrons, discutés au paragraphe 6, sont à envisager.

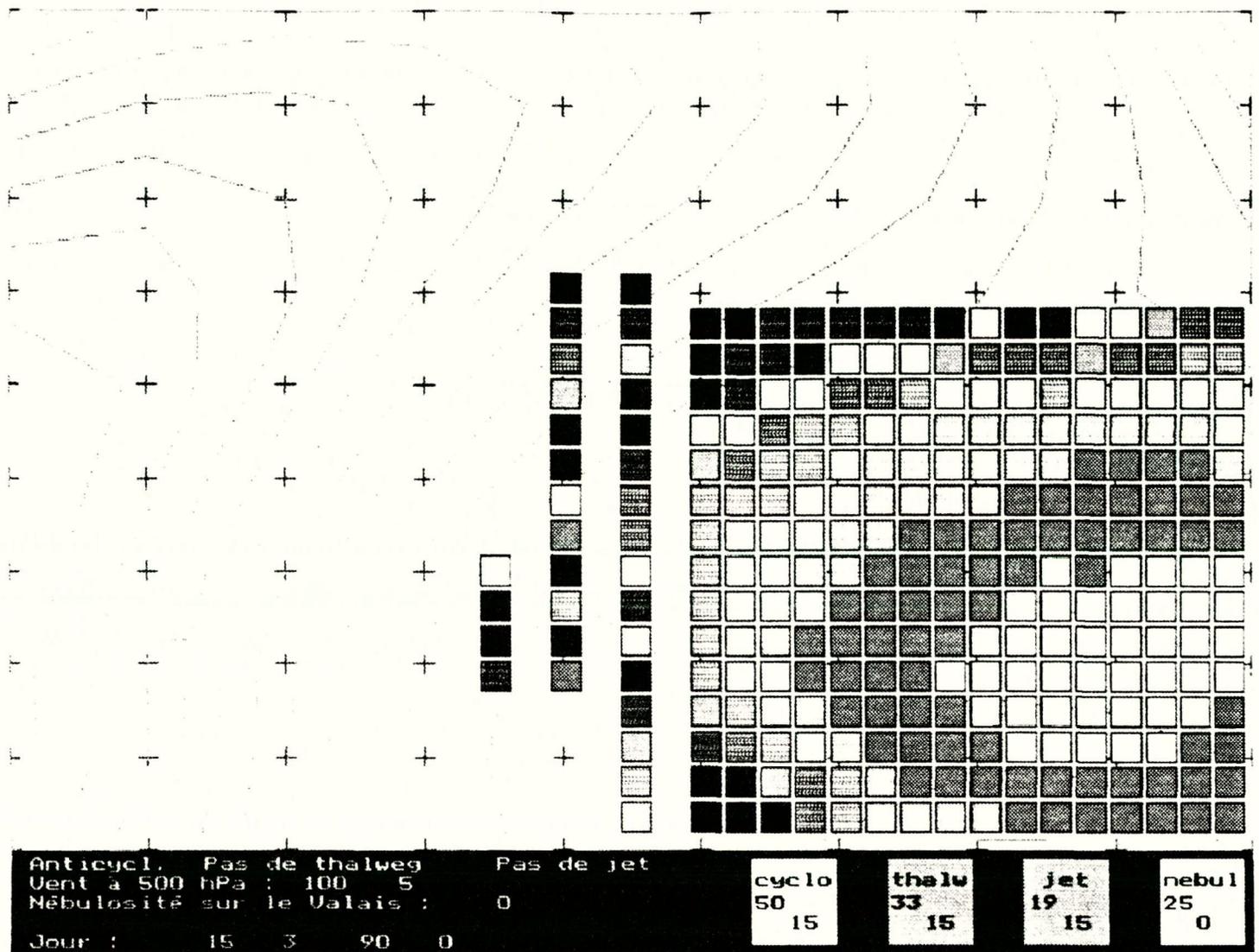
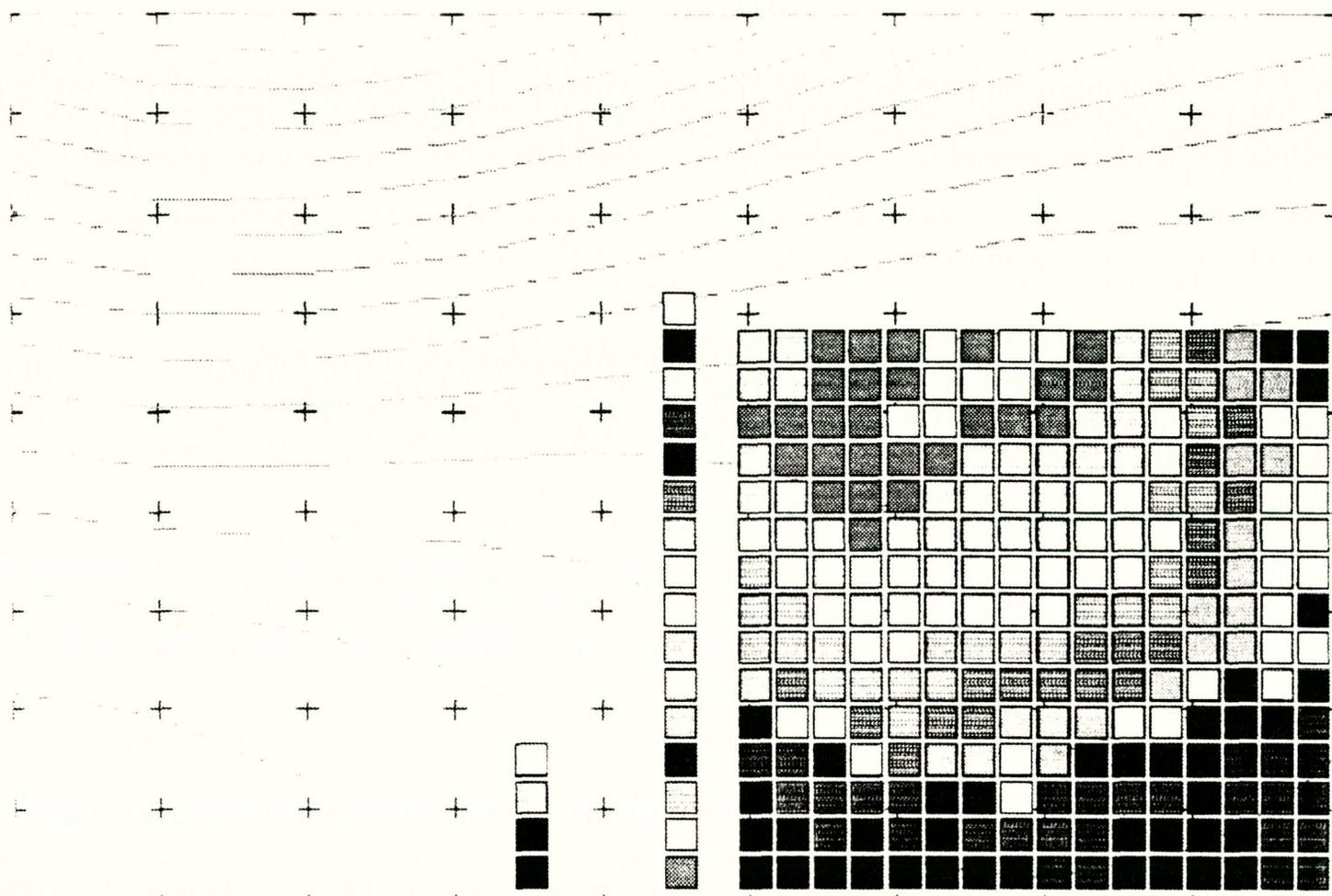


FIGURE 11



Cyclonique Thalweg W Alpes	Pas de jet	<b>cyclo</b>	<b>thalu</b>	<b>jet</b>	<b>nebul</b>
Vent à 500 hPa : 260 20		98	62	15	0
Nébulosité sur le Valais : 6		85	85	15	75
Jour : 25 3 90 0					

FIGURE 12

## 5. Le couplage à un système expert

Disposant d'un mini système expert écrit en Prolog, opérant en chaînage arrière, et dont la fonction est d'estimer la nébulosité moyenne sur le Valais, les trois sorties des neurones «cyclonalité», «thalweg» et «jet», ainsi que le vent à 500 hPa, ont été utilisées pour élaborer directement les prédicats entrant dans la base de faits dynamiques du système expert.

### 5.1. La structure du petit système expert

Les quatres prédicats entrant dans la base de faits dynamiques sont :

<i>cyclo_dba(cyclo,p<sub>1</sub>)</i>	<i>cyclo:</i>	<i>{faible,forte}</i>
<i>jet_dba(occurence,p<sub>2</sub>)</i>	<i>occurence:</i>	<i>{oui,non}</i>
<i>thalweg_dba(localité,p<sub>3</sub>)</i>	<i>localité:</i>	<i>{in0_10est,ext0_10est}</i>
<i>vent_dba((force,direction),p<sub>4</sub>)</i>	<i>force:</i>	<i>{ms_20Kt,20_29Kt,30_39Kt,40_49Kt, ps50Kt}</i>
	<i>direction:</i>	<i>{nnw_ne,e_s_w_nw},</i>

Les paramètres  $p_1, \dots, p_4$ , compris entre 0 et 1, mesurent la plausibilité attribuée à chaque prédicat. Ils sont déterminés à partir des activités des neurones de sortie du perceptron, selon une règle expliquée ci-après. Les suffixes *dba* spécifient simplement la «database», soit la base de faits dynamiques. Les cinq classes de nébulosité suivantes sont distinguées par le système expert :

- classe 0 : Couvert, nébulosité abondante.
- classe 1 : Le plus souvent très nuageux, peu de soleil.
- classe 2 : Passages nuageux importants, partiellement ensoleillé.
- classe 3 : Peu nuageux, assez ensoleillé.
- classe 4 : En général ensoleillé.
- classe 5 : Ensoleillé.

La base de connaissance, reflétant quelques usages des prévisionnistes du Centre Météorologique de Genève, très simple dans ce prototype, comprend quatorze règles de la forme :

```

Valais(classe_1,X):-
    cyclonalité(faible,X1),
    axe_thalweg(ext0_10est,X2),
    jet(oui,X3),
    seq_liste(X,[X1,X2,X3]),
    !.

```

Le prédicat *seq\_liste(X,[X1,X2,X3])* calcule la plausibilité X, fonction des plausibilités X1,X2,X3, associée à la classe de nébulosité lorsque la règle est satisfaite. Cette estimation, pouvant être effectuée selon plusieurs modalités différentes, est réalisée ici en calculant la moyenne géométrique des plausibilités X1,X2,X3.

Il reste à construire l'opérateur associant, à l'activité de chaque neurone de sortie du perceptron, un prédicat et sa plausibilité.

## 5.2 Conversion de l'activité d'un neurone en plausibilité d'un prédicat

Quelques conventions sont établies : Etant données les deux bornes d'apprentissage fixées dans la tâche T au paragraphe 4.2 :  $\zeta_{sup} = 0.85$  si une clause est réalisée,  $\zeta_{inf} = 0.15$  si elle ne l'est pas, il faut définir une séparatrice entre ces deux valeurs, telle que le prédicat reçoive la valeur «vrai» si l'activité du neurone est supérieure ou égale à cette séparatrice, la valeur «faux» sinon. Dans le cadre de ce travail, la valeur de la séparatrice est arbitrairement fixée à 0.6.

Ainsi donc, si par exemple le neurone cyclonalité, à la fin d'une chaîne de calcul, a une activité égale ou supérieure à 0.6, le prédicat inscrit dans la base de faits dynamiques est *cyclo\_dba(forte,P<sub>1</sub>)*. A contrario, si l'activité est inférieure à 0.6, *cyclo\_dba(faible,P<sub>2</sub>)* est inscrit dans la base de faits.

Les plausibilités P1 et P2 sont déterminées à l'aide de la fonction représentée à la figure 13 ci-dessous.

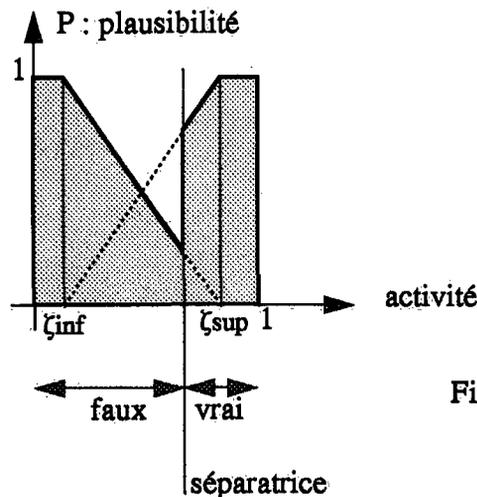


Figure 13 .

La plausibilité est égale à 1 lorsqu'un neurone répond avec une activité supérieure ou égale à  $\zeta_{sup}$ , ou inférieure ou égale à  $\zeta_{inf}$ , et le prédicat reçoit la valeur «vrai», respectivement «faux». La plausibilité est inférieure à 1 lorsque l'activité est comprise entre  $\zeta_{inf}$  et  $\zeta_{sup}$ .

Dans le cas non symétrique présenté ici, où la séparatrice est située à la valeur 0.6, la plausibilité d'une réponse peu supérieure à la séparatrice est bien plus grande que la plausibilité d'une réponse peu inférieure à la séparatrice. Si par exemple un neurone répond avec une activité de 0.61, le prédicat est codé «vrai» avec une bonne plausibilité, alors qu'il est codé «faux» sans grande assurance si l'activité est de 0.59.

A titre d'exemple, les prédicats livrés au petit système expert, pour le 3 mars 1990, sont:

```
date_dba(3,3,90)
vent_dba(wind(35,135),1.0)
cyclo_dba(forte,0.80)
athal_dba(ext0_10est,0.54)
jetnw_dba(non,0.67)
```

et le système répond, pour le Valais :

Passages nuageux importants, partiellement ensoleillé.

Plausibilité : 0.73.

## 6. Le problème général du couplage

Le perceptron s'avérant être le maillon le plus faible de la chaîne développée ci-dessus, deux modifications du couplage entre la carte de Kohonen et le perceptron, susceptibles d'améliorer la performance de l'apprentissage, ainsi qu'un couplage direct Kohonen -> système expert, ont été envisagées.

1. La première modification conduit à ne livrer au perceptron que les coordonnées normalisées du neurone élu  $IJ$  sur la carte de Kohonen, soit en fait le vecteur formé des deux composantes :

$$\left\{ \frac{I}{i_{max}}, \frac{J}{j_{max}} \right\}$$

Cette simplification, pour le moins drastique, n'a fait que dégrader la convergence de l'apprentissage.

2. La deuxième modification considérée a consisté à présenter en entrée du perceptron la matrice des poids synaptiques du neurone élu le jour  $\mu$ , notée  $(W^{IJ})^\mu$ .

Ainsi donc, au lieu d'assigner la tâche  $T = \{\{n^1 \dots n^\mu \dots n^m\}, \{\zeta^1 \dots \zeta^\mu \dots \zeta^m\}\}$  à l'apprentissage, il a été substitué la nouvelle tâche  $T' = \{\{(W^{IJ})^1 \dots (W^{IJ})^\mu \dots (W^{IJ})^m\}, \{\zeta^1 \dots \zeta^\mu \dots \zeta^m\}\}$ .

Et là les choses sont devenues inextricables pour le modeste ordinateur HP-Vectra utilisé ... A chaque pas d'apprentissage, il s'est agi de :

- . Générer une carte de Kohonen en pointeurs dynamiques
- . Charger ses poids synaptiques depuis le disque dur
- . Calculer une activité et identifier le neurone élu à partir d'un champ géopotential présenté
- . Mémoriser en mémoire vive la matrice synaptique de ce neurone élu
- . Détruire la carte de Kohonen (release heap)
- . Générer un perceptron en pointeurs dynamiques
- . Charger les poids synaptiques du perceptron depuis le disque dur
- . Effectuer un cycle complet de rétropropagation avec momentum sur les données émanant de la matrice synaptique du neurone élu
- . Sauver les nouveaux poids synaptiques du perceptron sur le disque dur
- . Détruire le perceptron
- . Et tout recommencer ...

Bien que ce dispositif ait effectivement été programmé, il n'a pas été possible de le faire fonctionner à un rythme suffisamment rapide pour pouvoir réaliser un apprentissage complet en moins de 24 heures de calcul.

Cette deuxième perspective est cependant intéressante. De par la nature même de l'auto-organisation du Kohonen et suivant la deuxième remarque de la fin du paragraphe 3.2, *la matrice des poids synaptiques de chaque neurone de ce réseau représente une situation météorologique typique*. On le vérifie sur les figures 7 à 10: les lignes de flux dessinées en bleu sur chaque figure sont en fait des isohypses du champ formé par les poids synaptiques du neurone élu.

L'apprentissage du perceptron ne porte alors plus sur un échantillon de situations météorologiques réelles, mais sur un *catalogue fini de cas typiques*, tels qu'ils ont été identifiés par le Kohonen.

3. Enfin, du point de vue du couplage du réseau de neurones au système expert, une génération directe des prédicats, réalisée depuis le Kohonen est à envisager. Elle est également fondée sur la remarque du paragraphe 3.2 :

Une approche nouvelle de l'apprentissage supervisé émergerait alors. Elle consisterait à requérir de l'expert, du superviseur, d'examiner le champ météorologique associé à la matrice synaptique de chaque neurone du réseau de Kohonen. Pour chaque neurone, et donc pour chaque cas typique du catalogue, l'expert devrait décider quelles sont, des caractéristiques météorologiques du champ représenté

par la matrice des poids synaptiques, celles qui doivent être annoncées au système expert. Le générateur de prédicat recevrait alors en entrée la coordonnée du neurone élu et livrerait sous forme de prédicats les propriétés associées au champ de sa matrice des poids synaptiques, telles qu'elles auraient été définies par l'expert.

Cette approche permettrait de s'affranchir totalement du perceptron, et des lourds problèmes associés à la convergence.

Enfin, des extensions pourraient être envisagées, où par exemple tous les neurones situés dans la bulle d'activité livreraient leurs prédicats associés.

## 7. Conclusion

Un météorologue examinant une carte semblable à celle qui est présentée à la figure 4 parvient en peu de temps à discerner les motifs qui lui paraissent importants, à les nommer, en leur associant des symboles, puis à effectuer des raisonnements portant sur ces symboles.

Le but du présent travail était de construire un système réalisant cette succession d'opérations, et le bilan doit être qualifié de mitigé. En effet, si l'auto-organisation du réseau de Kohonen s'est avérée acceptable, la convergence de l'apprentissage des perceptrons est restée relativement médiocre.

L'une des difficultés, intrinsèque au problème, a résulté de la difficulté d'affectation des qualités météorologiques aux situations de l'échantillon, telle qu'elle a été mentionnée à la fin du paragraphe 4.

D'autre part, le manque d'instruments théoriques, permettant notamment de mesurer le degré d'auto-organisation du réseau de Kohonen, de déterminer objectivement la dimension des couches des perceptrons, ainsi que les valeurs des coefficients  $\kappa$  intervenant dans les sigmoïdes, s'est avéré pénalisant en entraînant de multiples tâtonnements.

La perspective à mon sens la plus favorable de réalisation satisfaisante d'un tel système est mentionnée au point 3 du paragraphe 6: c'est l'interprétation directe des poids synaptiques des neurones du Kohonen par l'expert, conduisant à l'élaboration automatique de prédicats explicitement définis. Cette approche éviterait simplement l'usage du perceptron.

De manière générale, je pense qu'il faut, en matière de diagnostic, distinguer deux classes de problèmes.

La première classe est formée de diagnostics portant sur des séquences d'événements relativement équiprobables et pour lesquels les apprentissages supervisés - telle la rétropropagation - sont adéquats. L'on pourrait appeler *implicites* de tels modes d'apprentissage.

Le diagnostic d'événements rares, apparaissant aléatoirement dans les séquences d'événements équiprobables, tels par exemple l'apparition de lettres grecques dans un texte écrit en français, ou l'occurrence de pluies congelantes au cours d'un hiver, appartient en revanche à la seconde classe. Les diagnostics de ce type requièrent des formes *explicites* d'apprentissage basées sur l'exploitation d'une connaissance formelle des événements rares. L'usage de réseaux de Kohonen associés à des systèmes experts est alors préférable.

Lausanne, le 3 décembre 1992

## 8. Bibliographie

- Amaldi Edoardo      *Apprentissage supervisé dans les réseaux en couches*  
XIIIe Cours Postgrade 1992
- Ambühl Jacques      *Systèmes experts d'aide à la prévision*  
Arbeitsberichte der Schweizerischen Meteorologischen Anstalt  
No 161. Mars 1991
- Blayo François      *Réseaux de neurones formels à apprentissage non supervisé*  
XIIIe Cours Postgrade 1992
- Elsner J.B. & Tsonis A.A. *Nonlinear Prediction, Chaos and Noise*  
Bulletin of American Meteorological Society. Vol. 73 No 1  
Janvier 1992
- Herz J. Krogh A. Palmer R. G. *Introduction to the theory of neural computation*  
Santa Fe Institute in the Sciences of Complexity 1991
- Mac Cann Donald      *A Neural Network Short-Term Forecast of Significant Thunderstorms*  
Weather and Forecasting. Vol. 7, Septembre 1992
- Pour la Science - Scientific American      *Le cerveau et la pensée*  
Numéro spécial - Ouvrage collectif. No 181, Novembre 1992

### Adresse de l'auteur:

Jacques Ambühl  
Centre météorologique régional et  
de l'aéroport de Genève  
Case postale 312  
1215 Genève 15

