Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement des Innern EDI
Bundesamt für Meteorologie und Klimatologie MeteoSchweiz

Veröffentlichung MeteoSchweiz Nr. 85

Neural interpretation of ECMWF ensemble predictions

Jacques Ambühl, Daniel Cattani, Pierre Eckert

Herausgeber
Bundesamt für Meteorologie und Klimatologie, MeteoSchweiz, © 2010

MeteoSchweiz
Krähbühlstrasse 58
CH-8044 Zürich
T +41 44 256 91 11
www.meteoschweiz.ch

Weitere Standorte
CH-8058 Zürich-Flughafen
CH-6605 Locarno Monti
CH-1211 Genève 2
CH-1530 Payerne

Veröffentlichung MeteoSchweiz Nr. 85

ISSN: 1422-1381

Neural interpretation of ECMWF ensemble predictions

Jacques Ambühl, Daniel Cattani, Pierre Eckert

Bitte zitieren Sie diese Veröffentlichung folgendermassen
Ambühl, J; Cattani, D; Eckert, P: 2010, Neural interpretation of ECMWF ensemble predictions, *Veröffentlichungen der MeteoSchweiz*, **85**, 47 pp.

Herausgeber
Bundesamt für Meteorologie und Klimatologie, MeteoSchweiz, © 2010

MeteoSchweiz
Krähbühlstrasse 58
CH-8044 Zürich
T +41 44 256 91 11
www.meteoschweiz.ch

Weitere Standorte
CH-8058 Zürich-Flughafen
CH-6605 Locarno Monti
CH-1211 Genève 2
CH-1530 Payerne

Abstract

This study presents a neural system operated since the end of the last century at MeteoSwiss. The neural network firstly performs an automatic classification of weather patterns, secondly apply this classification in order to cluster members provided by the ECMWF Ensemble Forecasting System. Weather elements, quantitative as well as probabilistic, are then derived from the classification and used as basis for the elaboration of middle range forecasts at MeteoSwiss.

.

Contents

.

1 Introduction

Classifications of large scale atmospheric features always aroused fascination among meteorologists. Valuable information, that would otherwise end up concealed in people's memories and dusty archives, remains by that way perennial and available.

Accurately typecast weather patterns enable the definition of decision schemes reaching beyond the scope of strict meteorology. Allowing for example the description of weather configurations inducing environmental impediments, such classifying schemes bear societal impact.

An application of the mathematical chaos theory, the Ensemble Prediction Technique is one of the major achievements that occurred in meteorology during the last few decades [8]. Acknowledging the chaotic behaviour of atmospheric flows as a genuine propriety of Nature, the Ensemble Technique generates probabilistic forecasts inducing alternative weather scenarios having to be classified too.

A vast array of classification methods have burgeoned since the second world war. First of them consisted in sets of formal rules checked by bench forecasters whose decisions were aimed at their best judgement. The Perret-Bresovski classification [1] belongs to that category. Later, the emergence of numerical weather forecasting methods triggered the development of point to point algorithms implemented in suitable geometric spaces. Adequate metrics were accordingly identified. Factor- and Principal Component Analysis, both methods belonging to the broader category of Canonical Analysis, were introduced then.

The classification technique presented here is based on neural computating. Unreservedly conscious that biological neurons and organs are tremendously more sophisticated structures than the networks hereafter presented, we dare, however, to imitate Nature in trying to emulate her solutions into our algorithms. Instead of lying in the bare simulation of neurons, the challenge consists here in establishing learning processes operating upon neural assemblies. Such a mindset, aimed at reproducing biological structures in technology, is called Bionics. It frames many currently ongoing technological endeavours.

2 Scope and aim

This report presents work that were pursued during the nineties but remained almost concealed, sparsely documented and seldom presented [2, 3]

Our almost unstuffy project was aimed at developing an algorithm firstly realizing an automated classification of meteorological patterns, subsequently identifying and classifying newly presented patterns in accordance with that classification. Last but not least, gripped as we were by potential applications of then highly fashionable neural algorithms, their implementation downstream of the Ensemble Forecasting System installed at ECMWF was not our weakest motivation.

Formally, our objectives consisted in:

1. Replace old intuitive methods by formal systems.
2. Develop algorithms able to cope with the then emerging Ensemble Technique.
3. Implement automatic classification systems in production suites.

Most of the objectives have been reached. The neural system is still operated at the Forecasting Division of MeteoSwiss. Insights in subtle connexions between neural computation, biological considerations, ensemble technique, information theory, quantum aspects and, of course, meteorology, were gained.

After all these years, our desire consist now in making them available to a broader community. Making use of graphical systems that simply did not exist fifteen years ago, we allow us to present old results in quite bombastic graphics.

A final remark is related to the EU-Cost Action 733 Harmonisation and Applications of Weather Type Classifications for European Regions[1]. Our project was achieved long before this Cost Action would be started and was not considered within this Action.

[1] www.geo.uni-augsburg.de/lehrstuehle/phygeo/forschung/klima/COST733/; www.cost733.org/about cost733.htm

Acknowledgment

My first thanks are addressed to the organizers of a Post Doc seminar devoted to the "Réseaux de neurones biologiques et artificiels", held in 1992 at the Ecole Polytechnique Fédérale de Lausanne. Christophe Voisard, Andreas Asch and Rudolph Füchslin deserve special thanks for their generous help, enthusiasm and encouragements during the redaction and the reviewing of this report.

3 Meteorological Data

The learning process being based on actual meteorological patterns, geopotential 500 hPa and Temperature 850 hPa fields were chosen, expressed in meters for the geopotential, in tenth of Kelvin for the temperature. The former provided valuable information on the flow in the middle of the troposphere, the latter clues about the possible presence of frontiers between air masses. Broad enough to capture large scale patterns encompassing the whole alpine domain, the geographical area was defined by a grid of 10 x 10 points located in the area comprised between latitudes 36^o N and 54^o N, and longitudes 9^o W and 18^o E with intervals of 2, respectively 3 degrees. Figure 1 exhibits two such fields, as well as the underlaying geographical area.

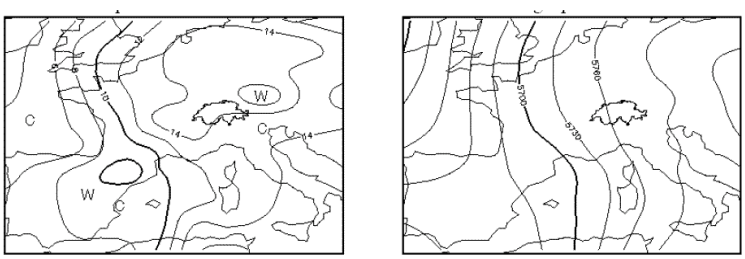

Figure 1: *Left panel: Temperature Field on the Surface 850 hPa, [degrees Celsius]. Right panel: 500 hPa geopotential field, [geopotential metres].*

The database onto which our algorithms were developed and tested comprised 9000 such situations, retrieved from the ECMWF archive at verifying times of 00Z and 12Z.

4 Neural learning

Having in mind a classification system able to gather similar weather patterns provided by the Ensemble Forecasting System into suitable clusters, our approach consisted in designing an artificial neural network able to distinguish and classify such patterns, and consecutively to gather them into clusters.

Neural networks are broadly cast in two main categories: those who learn under supervision, those who do not. Members of the first category operate following so called Supervised Learning algorithms. Members of the second one follow Unsupervised Competitive Learning algorithms, where the term competitive is explained in the sequel. Both classes of algorithms proceed by iterative adjustment of internal parameters in the networks.

Supervised learning requires at each step of the learning process a valuation in the sense that a comparison is realized between the outcome delivered by the network, when provided with an example pattern on its input, and a prerequisite answer corresponding to that pattern, expected by the environment[2]. A gradient, derived from this comparison, is then used to define corrections to be applied onto the internal parameters of the network (indeed its synaptic weights). The whole cycle is repeated with new examples until a required score is attained.

Contrarily, unsupervised competitive learning operates without any definition of predefined correct answers. The only valuation considered by the algorithm is the correlation occurring between patterns newly presented to the network, and patterns already mastered and registered in its circuits (its synaptic weights). Competition occurs within the network, among its neurons, where the most successful of them, realizing the highest correlation with the incoming pattern, takes the hand, adapts its synaptic weights to the incoming pattern and forces its direct neighbours to copy it either. It furthermore squeezes more distant neurons in the network[3]. Learning thus occurs as an emergent process, driven without master[4].

4.1 Unsupervised competitive learning

Mapping features into categories is a crucial mental faculty. Teuvo Kohonen proposed in 1982 an unsupervised learning algorithm strictly based on neurological observation [4]. He was aimed at clustering, or categorizing data: similar features, presented on the input of the network, were to be classified in corresponding categories. Those were to be found by the network itself, from the correlation of the input data with the already mastered and identified patterns. In the Kohonen network, the requirement for learning and increasing knowledge was satisfied in a topological sense. The network was framed as a two dimensional area on which specific items of knowledge were to be stored at specific locations. Such mappings are common in the brain. As for example, two closely located spots on the body are mapped onto two neighboring areas of the somato-sensorial cortex.

The architecture of our network is presented in the following sections. Geometrical considerations, playing a key role in unsupervised learning, are introduced first, in sections 4.1.1 to 4.1.3. Factors acting on a temporal basis

[2]Most frequently life, or a teacher.

[3]Unsurprisingly, this technique is dubbed "the winner takes all" in financial circles.

[4]Biologists consider that supervised learning emerged first in the course of the evolution, in processes driven by primordial feelings like anger, fear, aggressiveness, affection, compassion. Such networks are now mostly located in our limbic brains. In contrast, unsupervised competitive learning is a distinctive trait of the modern cortex.

during the learning process are introduced is section 4.1.4. The learning algorithm itself is presented in section 4.2.

4.1.1 Planar neural network, basic settings and definitions

In our setting, the neural network is framed as a square of $\{1, \cdots, M_{max}\} \times \{1, \cdots, N_{max}\}$ neurons. Each neuron is identified by its indices $\{n, m\}$ and is fully specified by its synaptic weights, given by matrices $W^{m\,n} = w^{m\,n}{}_{i\,j}$. Therefore, $\{m, n\} \in \{1, \cdots, M_{max}\} \times \{1, \cdots, N_{max}\}$ spans the network and $\{i, j\} \in \{1, \cdots, I_{max}\} \times \{1, \cdots, J_{max}\}$ spans the vertexes on the meteorological map, where geopotential height and temperature are considered, as given in section 3. With $M_{max} = N_{max} = 12$ and $I_{max} = J_{max} = 10$, $M_{max} \times N_{max} \times I_{max} \times J_{max} \times 2 = 28'800$ connections are simultaneously established from the meteorological map onto the neural network. This structure[5] is sketched in following Figure 2.

For the sake of typographical compactness, any neuron[6] will be referred in the sequel by its coordinates, e.g. $\{m, n\}$, and the whole neural network as $\mathcal{N}$.

[5]It is worth stressing that all the structures hereafter presented were never physically realized, but only virtually implemented in computer memories.

[6]At this point, an important precision has to be brought: computer scientists speak of "neurons". However, neurological structures actually corresponding to those "computer neurons" are in reality highly complex assemblages of 40'000 to 80'000 biological neurons, disposed in the cortex in parallel arrays of vertical columns (each 0.3-0.5 m in diameter). Reference: http://bmi.epfl.ch/page61216.html.

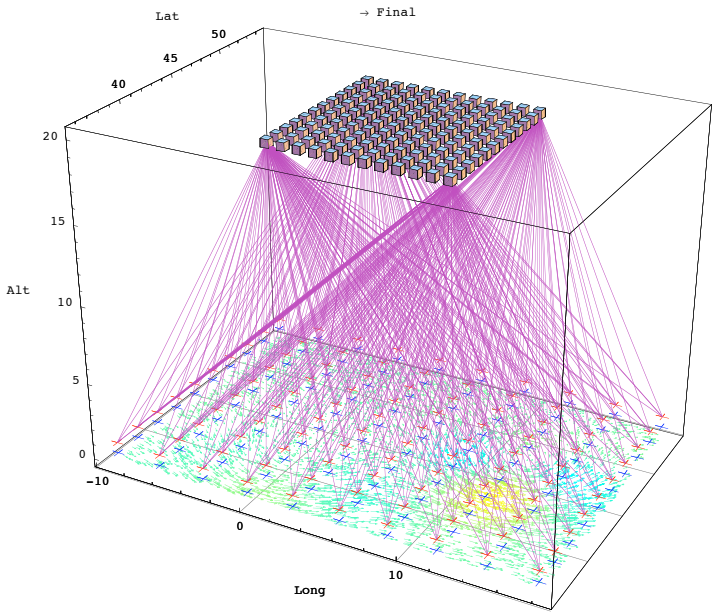

Figure 2: *Planar neural network and artistic representation of the connection established between the neural network, sketched "in the sky", and the weather map, drawn on the floor. Instead of the 28'800 connections implemented in the real network, only 400 of them are drawn here.*

4.1.2 Metrics and topologies

Unsupervised learning requires the definition of a geometric distance among neurons. The neurons being disposed on a planar quadratic network, as sketched in Figure 2, the distance between two of them, $\{m, n\}$ and $\{m', n'\}$, is simply provided by the euclidian metric:

$$\Delta_{\mathcal{E}\{\{m,n\},\{m',n'\}\}} = \sqrt{(m-m')^2 + (n-n')^2}.$$

Experience, however, unveils that unsupervised competitive learning is hampered by solitary neurons, as those lying at the corners of a quadratic network, or along its edges. A more sophisticated topology being sought in order to circumvent this foible, the obvious answer consists in gluing together opposite edges of the network, thus transforming it into a torus. The network is morphed into a closed surface that is exempt of boundary. Rewritten in accordance with the toroidal topology, the metric reads now:

$$\begin{aligned} \Delta^2_{\mathcal{T}\{\{m,n\},\{m',n'\}\}} &= \min\left\{|m + M_{max} - m'|, |m - m'|, |m - M_{max} - m'|\right\}^2 \\ &+ \min\left\{|n + N_{max} - n'|, |n - n'|, |n - N_{max} - n'|\right\}^2 \end{aligned} \quad (1)$$

Both metrics, $\Delta_{\mathcal{E}}$ and $\Delta_{\mathcal{T}}$, evaluated on $\mathcal{N}$, are presented in Figure 3.

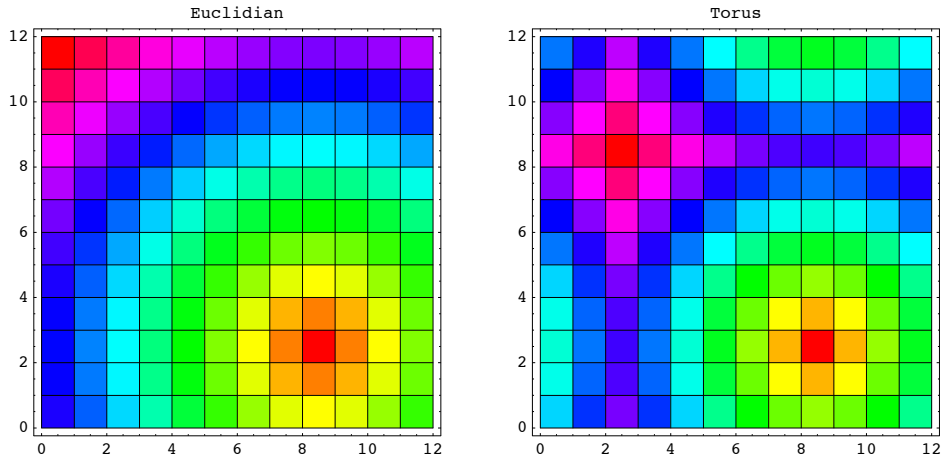

Figure 3: *Euclidian and Toroidal metrics computed between neuron* $\{9, 3\}$ *and all other neurons on each network.*

Distance monotonically increases from the location (neuron) $\{8, 3\}$ on the left chart, it exhibits a dual cyclic structure on the right one. In this setting, the most distant neuron from the lower right red one is located at the center of the purple cross. They are antipodal locations on the torus[7].

[7] In the sense that New Zealand and Switzerland are antipodal locations on the terrestrial sphere.

Figure 4 depicts the transformation of the square network into a toroidal one, with the corresponding transport of the metric. In that figure, distances are computed from the orange spot located near the upper edge of the square, and the antipodal area is pictured in purple.

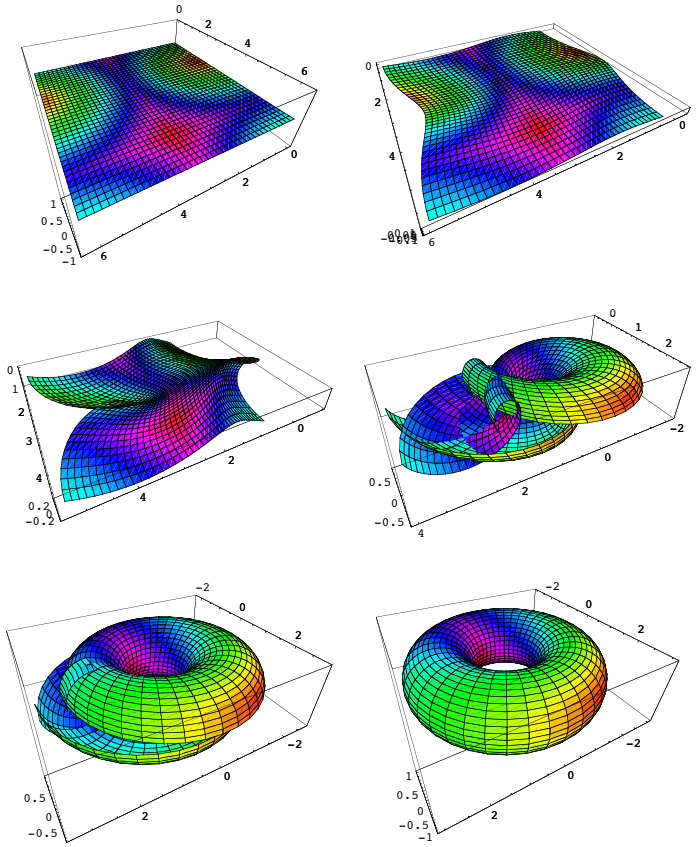

Figure 4: *Transformation of a square network into a toroidal one, with the corresponding transport of the metric.*

4.1.3 Input space and its corresponding metric

The signal delivered by any neuron $\{m, n\}$ expresses the distance $\mathcal{D}_{\{m,n\}(X)}$ between its synoptic weights and a meteorological signal X presented to the network, computed according to a metric having the general form:

$$\mathcal{D}^2_{\{m,n\}(X)} = [X - W^{m\,n}]^T \; \mathcal{V}^{-1} \; [X - W^{m\,n}]. \tag{2}$$

As specified in Section 3, X is a meteorological field provided on the map described in that section on the grid $\{i, j\} \in \{1, I_{max}\} \times \{1, J_{max}\}$.

The Matrix $\mathcal{V}$ is the variance-covariance matrix computed on the database mentioned in Section 3. Its virtue consists in increasing the homogeneity of the task presented to the network during the learning process[8]. However, taking into consideration the fact that all norms are equivalent in a (metric) topological space of finite dimension, $\mathcal{V}$ has been neglected in the setting chosen here and replaced by the unity matrix. Accordingly simplified, the metric, expressed as the euclidian distance, is given by:

$$\mathcal{D}_{\{m,n\}(X)} = \sqrt{\sum_{i,j=1}^{I_{max},J_{max}} [x_{i\,j} - w_{i\,j}^{m\,n}]^2}$$

A further refinement has to be introduced, due of the fact that the network - and its neurons - apprehends both geopotential and temperature fields simultaneously. The notation is therefore extended in the following way: the meteorological field X receives two components and is written $\mathcal{X} = X^k$ with $k = \{\Psi, T\}$ for the geopotential, respectively the temperature field. The same occurs for the synaptic matrix of the neuron, that is modified in $W^{k\,m\,n}$ or, in components, in $w_{i\,j}^{k\,m\,n}$. $\{m, n\}$ designates a neuron, $\{i, j\}$ refers to one of its synapes and $k = \{\Psi, T\}$, as previously specified. A mixing parameter α, specifying the relative importance of both geopotential and temperature fields, is arbitrarily set at $\alpha = 0.1$. Eqquiped in this way, the metric finally reads:

$$\mathcal{D}_{\{m,n\}(\mathcal{X})} = \sqrt{\sum_{i,j=1}^{I_{max},J_{max}} [\alpha\,[x_{i\,j}^{\Psi} - w_{i\,j}^{\Psi\,m\,n}]^2 + (1-\alpha)\,[x_{i\,j}^{T} - w_{i\,j}^{T\,m\,n}]^2]} \quad (3)$$

A generalization to more than two meteorological fields would be obvious.

The $M_{max} \times N_{max} \times I_{max} \times J_{max} \times 2 = 28'800$ connections established from the meteorological map onto the neural network correspond to the synaptic weights incapsulated in the $W^{k\,m\,n}$ matrices. Altogether, they actually shape the "memory" of the network. The learning process to be implemented will be given the task of determining a suitable value for all these weights, adequately representing not only one weather pattern, but a array of weather situations.

Expression (3) shows that each neuron $\{m, n\}$ computes the euclidian distance existing between the set of its synaptic weights, provided by its $W^{k\,m\,n}$ matrix, and meteorological $\mathcal{X}$ fields. It must be stressed that this

[8]The ubiquitousness of such metrics in numerical wheater forecasting is a striking feature in modern meteorology. They appear as penalty functions in assimilation schemes for 3D- and 4D-Var algorithms, as well as in Ensemble Kalman assimilation schemes.

distance is not measured in the usual geometrical or geographical space. It is instead considered in what will be called in the sequel the input space, or even the "cognitive space" of the network.

4.1.4 Activity bubbles and their temporal evolution

As mentioned in section 4.1, the learning process proceeds through simulation of neurons closely located to a distinguished, elected, neuron, and inhibition of neurons lying at larger distance from that elected neuron. This procedure requires the construction of an operator acting on the network and propagating the so called "activity bubble" defined as:

$$\mathcal{B}_{\{m',n'\}_{(p)}} = \left\{ \{m,n\} \in \mathcal{N} \mid \Delta_{\mathcal{T}\{\{m,n\},\{m',n'\}\}} \leq \sigma_{(p)} \right\} \tag{4}$$

Following the convention introduced in section 4.4.1, the elected neuron is designed by $\{m',n'\}_{(p)}$, where $p = 1, \cdots, P_{max}$ represents the p^{th} learning step and P_{max} the maximum number of such steps.

Two functions are to be provided, specifying the size and the intensity of the activity bubble. The former, $\sigma_{(p)}$, expresses the extensiveness of the bubble at learning step p, the latter, $\tau_{(p)}$, its intensity at step p:

$$\sigma_{(p)} = \frac{\sigma_0}{1 + \frac{p}{P_{max}}};\ \tau_{(p)} = \frac{\tau_0}{1 + \frac{p}{P_{max}}}$$

The parameters σ_0, τ_0 and P_{max} are pragmatically fetched with, in our setup, $\sigma_0 = \sqrt{M_{max}^2 + N_{max}^2}$, $\tau_0 = \frac{1}{4}$, $P_{max} = S_{data} \cdot C_{max}$, where $C_{max} = 100$ is the number of learning cycles, $S_{data} = 8000$ the size of the sample of meteorological cases considered, defined in Section 3.

Being both monotonically decreasing functions of p, they reduce the extension as well as the intensity of the activity bubble in the course of the learning process. Through their action, the neural network, initially subjected to a broad and intense stimulating activity, eventually freezes and reaches by the end of the process its learned state[9].

4.2 The learning process

All elements required to define the learning algorithm, but one, are now at hand. Any seasonal bias has yet to be removed from the task being submitted to the learning process.

[9] Acting onto the whole network, $\sigma_{(p)}$ and $\tau_{(p)}$, might indeed be seen as kinds of simplified neurotransmitters.

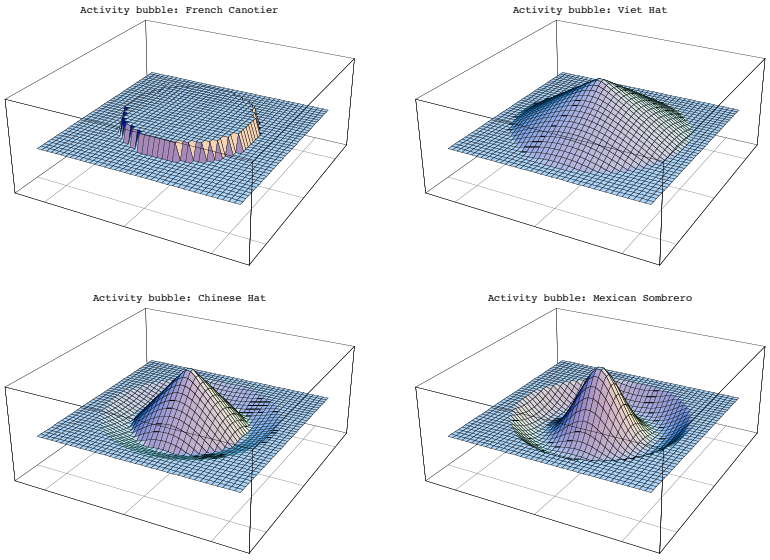

Figure 5: *Examples of activity bubbles. Although all of them have been tested, the cone, dubbed here "viet hat", has been chosen for the development of neural networks operationnaly run at MeteoSwiss. In that case the* $\tau_{(p)}$ *function determines the height of the summit, the* $\sigma_{(p)}$ *function the radius of the base of the cone. Both decrease as the learning process proceeds.*

4.2.1 Debiased learning task

Formally, the task is the array of geopotential and temperature fields (Ψ at 500, respectively T at 850 hPa, see section 3) presented to the network during the learning process. It is shaped as: $\overline{\mathcal{X}} = \left[\overline{\mathcal{X}}_d\right]_{d=1}^{Dmax}$ with d defining a day $d \in S_{data}$ and $D_{max} = |S_{data}|$.

Debiasing factors $\overline{\Psi}_{i,j}^{d}$ for geopotential and $\overline{T}_{i,j}^{d}$ for temperature are defined at location $\{i, j\}$ for the day d of the task. They are computed following the expressions:

$$\overline{\Psi}_{i,j}^{d} = p_{i,j}^{\Psi}\ sin(\frac{2\pi\ d^*}{365} + q_{i,j}^{\Psi}) + r_{i,j}^{\Psi}$$
$$\overline{T}_{i,j}^{d} = p_{i,j}^{T}\ sin(\frac{2\pi\ d^*}{365} + q_{i,j}^{T}) + r_{i,j}^{T}$$

Those sinusoidal functions are smoothing operators acting on daily values. Based on daily averages established on a 20 years basis, they are computed at each grid point for both fields. The day d^* of the year is retrieved from the d of the task, the matrixes $p_{i,j}^{k}$, $q_{i,j}^{k}$ and $r_{i,j}^{k}$ express amplitudes, phases and translations at location $\{i, j\}$ for both fields $k = \{\Psi, T\}$.

Finally, an element of the debiased task reads: $\overline{\mathcal{X}}_d =$

$$\begin{bmatrix} \left\{\Psi_{1,1}^{d} - \overline{\Psi}_{1,1}^{d}, T_{1,1}^{d} - \overline{T}_{1,1}^{d}\right\} & \cdots & \left\{\Psi_{1,10}^{d} - \overline{\Psi}_{1,10}^{d}, T_{1,10}^{d} - \overline{T}_{1,10}^{d}\right\} \\ \vdots & \ddots & \vdots \\ \left\{\Psi_{10,1}^{d} - \overline{\Psi}_{10,1}^{d}, T_{10,1}^{d} - \overline{T}_{10,1}^{d}\right\} & \cdots & \left\{\Psi_{10,10}^{d} - \overline{\Psi}_{10,10}^{d}, T_{10,10}^{d} - \overline{T}_{10,10}^{d}\right\} \end{bmatrix}$$

4.2.2 Unsupervised competitive learning algorithm

Having now all required elements at hand, the learning process itself, hereafter presented, happens to be of evangelical simplicity[10]. Furthermore, beside of being operated without supervision, it deeply relies on stochasticity.

The process is initialized by setting all the synaptic weights of all the neurons to small random values. Then, at each learning step $p \geq 1$, a day d is randomly determined and the corresponding input field $\overline{\mathcal{X}}_d$ is fetched from the task $\overline{\mathcal{X}}$. Relabeled $\overline{\mathcal{X}}^p$, it is then presented to the network. At each presentation p, a elected neuron $\{m', n'\}_{(p)}$ is determined, with the propriety of having, between all the neurons of the network, the smallest $\Delta_{\mathcal{T}}$ distance to the presently presented input field $\overline{\mathcal{X}}^p$. An activity bubble $\mathcal{B}_{\{m',n'\}_{(p)}}$ is

[10] Wether this simplicity could, amid other factors, account for the rapid development of cortical structures in evolutionary zoology, remains an open issue among scientists.

then determined around this elected neuron.

The neurons located in the bubble are forced to move their synaptic weights towards the input field $\overline{\mathcal{X}}^p$ currently presented, while the others are weakened, thus creating the effect of cooperation with close neighbours, and competition against distant neurons. The size of the bubble, given by the $\sigma_{(p)}$ function, decreases as the learning process goes on and, at the end, only local interactions may still occur. The function $\tau_{(p)}$ specifies through the height of the bubble the intensity of the learning process. It decreases during the learning process too, and contributes to freeze the network in its organized state.

Now rigorously formulated, the **unsupervised competitive learning algorithm** reads:

0: Initialize -
assign small random value to all synaptic weights:
$\forall \{m, n, i, j\} \in \mathcal{N} : w_{i\,j}^{m\,n} \longmapsto \mathit{Random}$;
initialize the step counter: $p \longmapsto 0$;

1: Randomly choose a pattern in the task -
increment the step counter: $p \longmapsto p + 1$;
randomly[11] choose a day index: $\tilde{d} \in \{1 \cdots D_{max}\}$;
fetch the corresponding pattern: $\overline{\mathcal{X}}_{\tilde{d}} \longmapsto \overline{\mathcal{X}}^p$;

2: Determine the elected neuron -
find the neuron $\{m', n'\} \in \mathcal{N}$ such that:
$\{m', n'\}_{(p)} = \min_{\{m,n\} \in \mathcal{N}} [\mathcal{D}_{\{m,n\}(\overline{\mathcal{X}}^p)}]$

3: Compute the activity bubble -
$\mathcal{B}_{\{m',n'\}_{(p)}} = \left\{ \{m, n\} \in \mathcal{N} \mid \Delta_{\mathcal{T}\{\{m,n\},\{m',n'\}\}} \leq \sigma_{(p)} \right\}$

4: Adapt synaptic weights -
$\forall \{m, n\} \in \mathcal{B}_{\{m',n'\}_{(p)}}$:
$W^{m\,n} \longmapsto W^{m\,n} + \tau_{(p)} \cdot [\overline{\mathcal{X}}^p - W^{m\,n}]$

5: Iterate -
if $p = P_{max}$ then exit, else return to 1.

4.3 Discussion

The random choice of the days, and therefore of the patterns in the task, is easily justified: initially, the learning process appears to be rapid. Hence,

[11] The random generator operates on the basis of an uniform distribution. This hypothesis will be later with so called "favoured learning" partially relaxed.

if the chronological order of the task elements were preserved, the network would organize itself, in the first learning steps, so as to reflect the climatological features present at the beginning of the season. Destroying the chronological order prevents the network performing such a seasonal overfitting.

Psychologists and neurologists agree to explain that learning occurs when people sleep. Significant events that occurred in the awaken life are then presented from the short time memory (mostly located in the hyppocampus) to long term memories (predominantly located in cortical areas) in a way that is similar, however infinitely more complex than the scheme described in this report. Furthermore, significant or even traumatic events seem to be recognised in the limbic system and trigger much intense and broader activity bubbles than uneventful stimuli. A more sensitive memorisation of significant events is accomplished in this way.

This property of biological learning could be considered to develop neural systems aimed at detecting potentially threatening weather patterns. To this purpose, our learning algorithm should be be improved in two ways. The first one would consist in repeating in the learning task significant patterns, for example those days where adverse weather occurred, the second one in associating an intensification factor to such patterns.

By the first technique, events that are deemed to be significant would be presented more frequently than ordinary patterns. The second technique would consist in associating an intensification factor characterising each element of the task. It would be set at one for ordinary events, and at a higher value for significant events. This intensification factor would then magnify the parameters of the activity bubble defined in section 4.1.4 at each step of the learning process[12]. This issue will be further discussed in section 5.1.

Conclusively, a surprising connection can be established between neural learning and numerical weather forecasting, at least at theoretical level. Firstly, both 4D-Var assimilation schemes and supervised learning algorithms proceed through the retro-propagation of a gradient (with the adjoint model playing the role of the retro-propagator in the case of 4D-Var). Secondly, Nudging assimilation schemes and unsupervised competitive learning algorithms are closely related too. Both guide the dynamical behaviour of a process in imposing external signals onto its components. Accordingly, neural scientists are inclined to consider assimilation schemes implemented in numerical meteorology as learning processes operated in real time in fore-

[12] A mathematical technique based on "Voronoï Tesselations" enables the visualization of the induced effect on in the input space.

casting suites. Ironically, the most sophisitcated assimilation scheme implemented in meteorology happens to be related to the oldest learning process invented by Nature, and vice versa.

Section 6 will be devoted to the assessment of both the performance of the algorithm and the quality of its outcome. Let us first, in the following section, marvel at the produced patterns.

5 Synoptic weather patterns

Expression (3) in section 4.1.3 shows that the distance $\mathcal{D}_{\{m,n\}(\mathcal{X})}$ is measured between a couple of geopotential and temperature fields, and a matrix of synaptic weights. Indeed, as suggested at the end of section 4.1.3, for the neural network, the $\{\Psi, T\}$ fields as well as the $W^{m\,n}$ synaptic matrices lie in the same input space. As a consequence, any interpretation relevant for one of the quantities has to be relevant for the other one as well. Therefore, once the network is organized, each neuron $\{m, n\} \in \mathcal{N}$, corresponds to a specific meteorological situation that is coded in the synaptic weights of its $W^{m\,n}$ matrix. Let us exhibit this fact in the two following figures 6 and 7.

5.1 Synaptic matrices as weather fields

Neuron $\{2, 5\}$ is presented in figure 6, neuron $\{0, 9\}$ in figure 7. Synaptic equivalent of the geopotential field at 500 hPa (black isolines) and the temperature field at 850 hPa (coloured features) are drawn on the upper panels of both figures, were, as defined in section 3, the geographical area is comprised between latitudes 36^o North and 54^o North, and longitudes 9^o West and 18^o East. The bottom panels present in both figures the corresponding geostrophic wind fields coded in temperature, with the altitude of the wind arrows representing the the geopotential height.

Figure 6 depicts a northeast stream over Switzerland, a "Bisenlage". It is powered by a low over Sicily, on the bottom right corner of the area, and a high over the British Isles, on the upper right corner. Figure 7 is characterized by a low off the Portugal coast and an area of flat pressure distribution over Italy. Although the southeast stream is weak over the Alps, it suffices to induce a strong Föhn on the northern slopes of the alpine chain, well reproduced by the temperature gradient over Switzerland.

It is worth stressing that these fields are synaptic and not meteorological fields. They emerged in the network by the virtue of the unsupervised competitive learning. Furthermore, a glance at figures 6 and 7 shows that meteorological properties can be ascribed to them. This characterisation

can be realized at least in the different ways, that are now presented.

The first one simply consists in describing the field verbally, as a meteorologist is used to doing it. By that way, a portrayal description is attributed to each neuron, as for example in qualifying neuron $\{2, 5\}$ as "Bisenlage" and neuron $\{2, 5\}$ as a "Föhnlage". More specific features could even be described. As for example, seasoned swiss forecasters know that the weak southeast flow depicted in figure 7 over the Alps is likely to trigger the quite pugnacious "Guggiföhn", experienced in such situations in the Berner Oberland.

The second way consist in elaborating a climatology of each neuron, describing its characteristics in terms of precipitation, gales, ... for specific areas. It enables the elaboration of customer oriented products and services, as for example the statistics of fine particle distribution over cities. The following tables are extracts of the climatological information related to the presented neurons, both for spring. The stations are: OTL Osservatorio Locarno Monti, GVE Geneva, SMA MeteoSwiss Zürich (previously Schweizerische Meteorologische Anstalt), SIO Sion, PAY Payerne. Probabilities of occurrence of precipitation, relative sunshine and gusts are presented in the first three columns. PM10, concentration of particulate matter (PM) or fine particles, top of the stratus over the swiss midlands in meters above mean sea level and expected forecasting scores (OPKO: Objektive Prognose KOntrolle:) are given in the last three columns.

The third way would consist in identifying specific neurons that could be ascribed to potentially devastating events in specific regions. As an example, neurons describing weather patterns by which floods are to be expected in alpine valleys would be highlighted. Intensification parameters discussed in section 4.3 could then be applied to such neurons, or they could be repeated in the learning task, as explained in section 4.3. With such a setting, the neural network would then be trained to identify those critical weather patterns emphasized by the intensification parameters. This information could be taken into account to trigger warning systems.

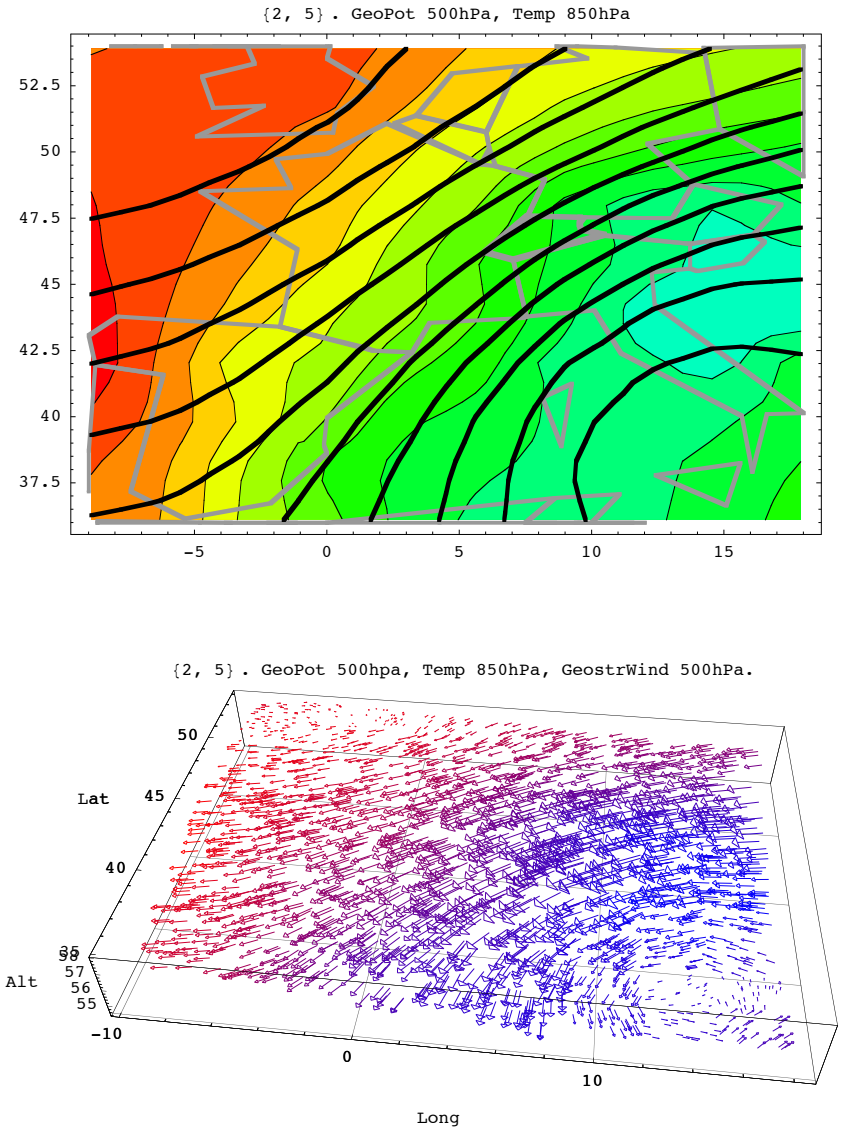

Figure 6: *Example of a synaptic field: Neuron* $\{2, 5\}$*, depicting a Bisenlage over the Alps. Latitude* 36^o *N -* 54^o *N, longitude* 9^o *W -* 18^o *E. Top panel: synaptic equivalent of geopotential field at 500 hPa (black isolines) and temperature field at 850 hPa (coloured features). Bottom panel: corresponding geostrophic wind field coded in temperature, with the altitude of the wind arrows representing the geopotential height.*

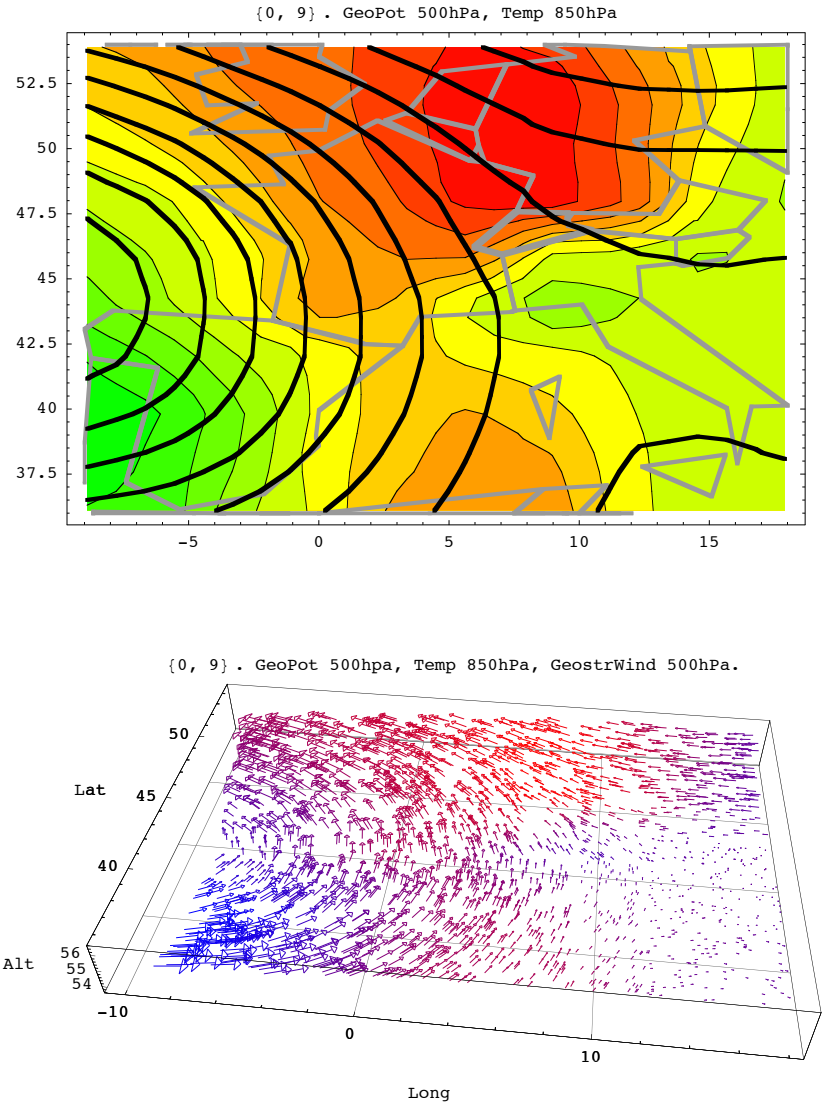

Figure 7: *Example of a synaptic field: Neuron* $\{0, 9\}$*., depicting a Föhnlage over the Alps. Latitude* 36^o *N -* 54^o *N, longitude* 9^o *W -* 18^o *E. Top Panel: synaptic equivalent of geopotential field at 500 hPa (black isolines) and temperature field at 850 hPa (coloured features). Bottom panel: corresponding geostrophic wind field coded in temperature, with the altitude of the wind arrows representing the geopotential height.*

	Prob prec gt than						Relative sunshine			Prob gusts gt than				PM10	Top stratus	OPKO
Station	0,1mm	1mm	5mm	10mm	20mm	50mm	70-100 %	25-70 %	0-25%	45km/h	60km/h	75km/h	100km/h	Mean value µg/m^3	Mean value m	Mean value %
OTL	0	0	0	0	0	0	75	17	8	20	10	0	0	--	--	--
GVE	25	25	0	0	0	0	40	33	27	40	10	0	0	19	--	82
SMA	50	50	0	0	0	0	25	35	40	0	0	0	0	--	--	--
SIO	25	0	0	0	0	0	75	14	11	30	0	0	0	--	--	82
PAY	25	25	25	0	0	0	50	19	31	20	0	0	0	--	1200	82

Climatological probability of exceeding a threshold, in %

Unit (2,5)
Season : spring

Figure 8: *Sample of the climatology related to the neuron* $\{2, 5\}$ *for spring. The stations are: OTL Osservatorio Locarno Monti, GVE Geneva, SMA MeteoSwiss Zürich, SIO Sion, PAY Payerne. First three columns: Probabilities of occurrence of precipitation, relative sunshine and gusts. PM10: concentration of particulate matter (PM) or fine particles. Top Stratus over the swiss midlands in meters above mean sea level. OPKO: Objektive Prognose KOntrolle: Expected forecasting score.*

	Prob prec gt than						Relative sunshine			Prob gusts gt than				PM10	Top stratus	OPKO
Station	0,1mm	1mm	5mm	10mm	20mm	50mm	70-100 %	25-70 %	0-25%	45km/h	60km/h	75km/h	100km/h	Mean value µg/m^3	Mean value m	Mean value %
OTL	90	70	40	30	10	0	7	21	72	0	0	0	0	--	--	--
GVE	50	20	10	10	0	0	7	26	67	0	0	0	0	28	--	83
SMA	20	10	10	10	0	0	12	36	52	7	7	0	0	--	--	--
SIO	20	0	0	0	0	0	36	33	31	40	0	0	0	--	--	83
PAY	50	40	20	10	0	0	5	33	62	7	0	0	0	--	850	83

Climatological probability of exceeding a threshold, in %

Unit (0,9)
Season : spring

Figure 9: *Same as figure 8: sample of the climatology related to the neuron* $\{0, 9\}$ *for spring.*

5.2 Topological gathering

Let us consider another key feature of the Kohonen network, its ability to gather similar patterns on a topological basis.

Figures 10 and 11 exhibit this propriety. The area on the network encompassing neurons $\{3, 3\}$ to $\{5, 5\}$ is presented in figure 8, where cyclonic weather patterns are gathered. The area encompassing neurons $\{7, 3\}$ to $\{9, 5\}$ is presented in figure 9, for anticyclonic patterns. The topological mapping makes sense as closely related weather patterns appear to be allocated to close-lying neurons on the network. Cyclonic and anticyclonic circulations are gathered, as well as weak and strong gradients and flow direction change smoothly from one neuron to its neigbours.

A further unmistakable trait in these panels is the refinement of the 850 hPa temperature fields. Exhibiting structures that are far more subtle than those barely provided by geopotential fields, they underline the fact that a multi-layers approach should be favored in the business of weather classification.

Enabling a definition of the coherence of signals produced by the network, topological gathering will happen to be the cornerstone of our ensemble interpretation technique. This coherence measure, interpreted as predictability of a forecast and based on entropy, will be presented in section 7. The ensemble interpretation will be described in section 8. The next section is devoted to the classification decision, automatically taken by the network.

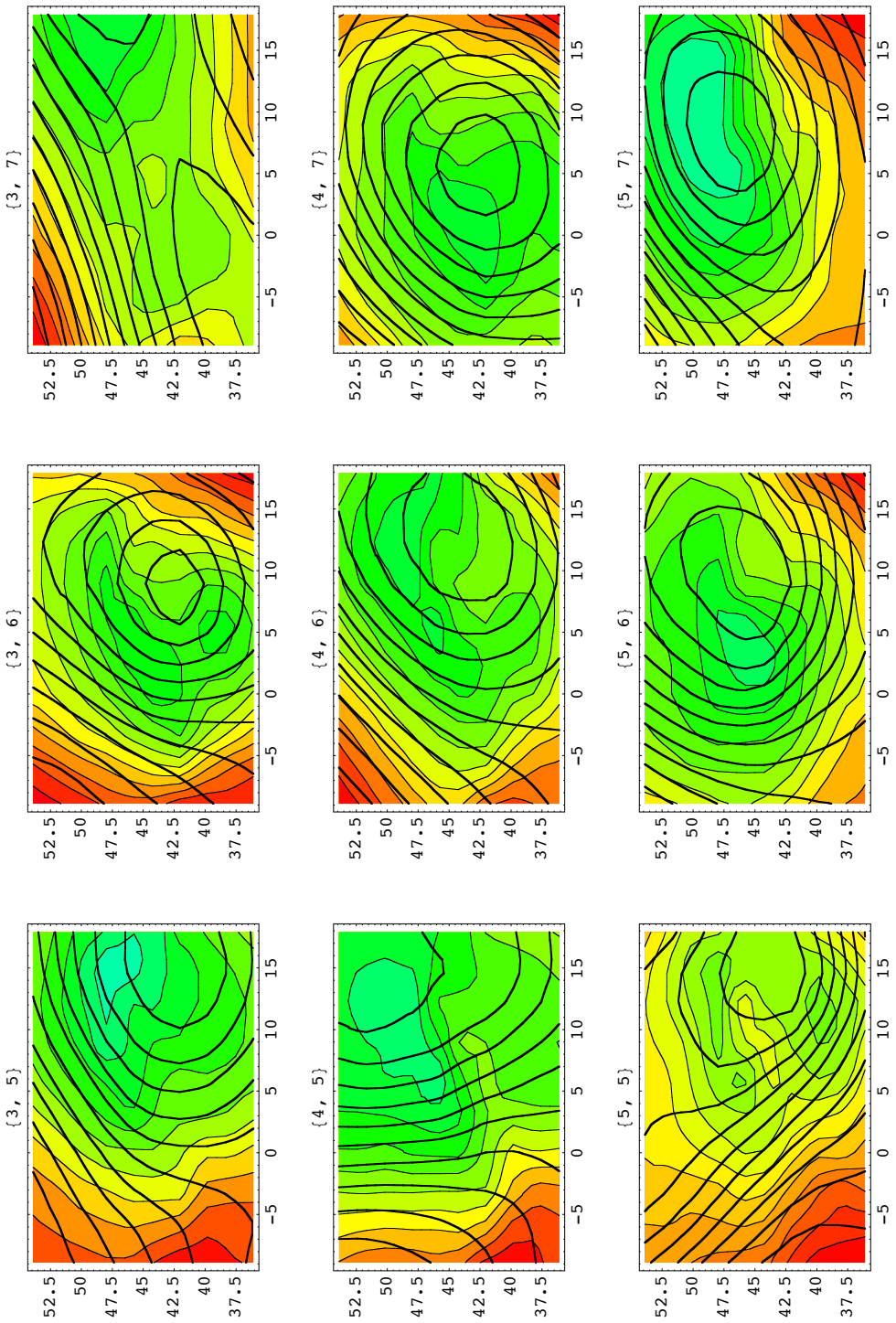

Figure 10: *Gathering of patterns on the network: cyclonic sector*

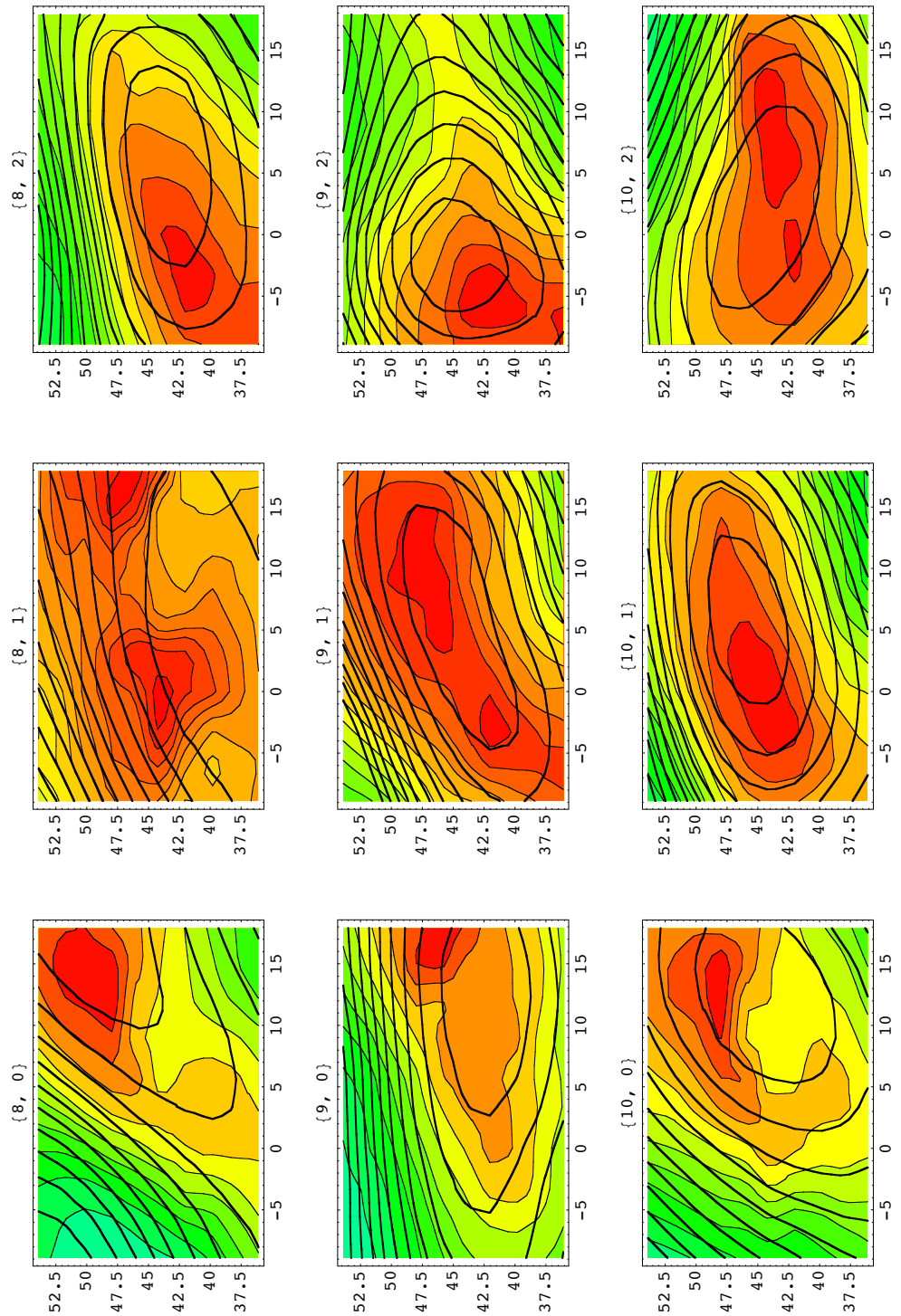

Figure 11: *Gathering of patterns on the network: anticyclonic sector*

6 Classification decision

Once organized, frozen and implemented in operational duty, the elementary action of the network consists in allocating one neuron to any newly presented meteorological field $\mathcal{X}$.

This action is performed through determination of the closest neuron $\{m', n'\}$ to $\mathcal{X}$, that is once again called the elected neuron. The evaluation of the $\mathcal{D}_{\{m,n\}(\mathcal{X})}$. distances thus settles the classification decision. The task given at the beginning of section 4.1, mapping features into categories, is completed. Categories are defined here by the synaptic matrices of the neurons, together with the definition of the $\mathcal{D}$-metric. This fundamental propriety of the Kohonen Algorithm provides the aggregating tool needed for the classification problem of ensemble forecasts.

Attempts to exhibit this decision process are presented in figures 12, 13 and 14. In each figure, the weather field that is submitted to the neural network is sketched on the floor of the cube, with the network itself hovering in the sky. The synaptic sensors are described as small yellow dots on the flow, with straight lines emanating from them and converging towards the elected neuron, in this case $\{10, 10\}$. The response of the network is reported by the colour of its neurons, with brown colours for those neurons whose synaptic matrices are distant from the submitted field, and in green colours for those who are close.

Another field, with the corresponding answer of the network is given in figure 11 with elected neuron $\{4, 7\}$. Figure 12 is the exact correspondent of figure 11, with elected neuron $\{4, 7\}$. However, the transformation from the square to the torus, given in figure 4, has been reverted, thus allowing a better visualization of the answer of the whole network.

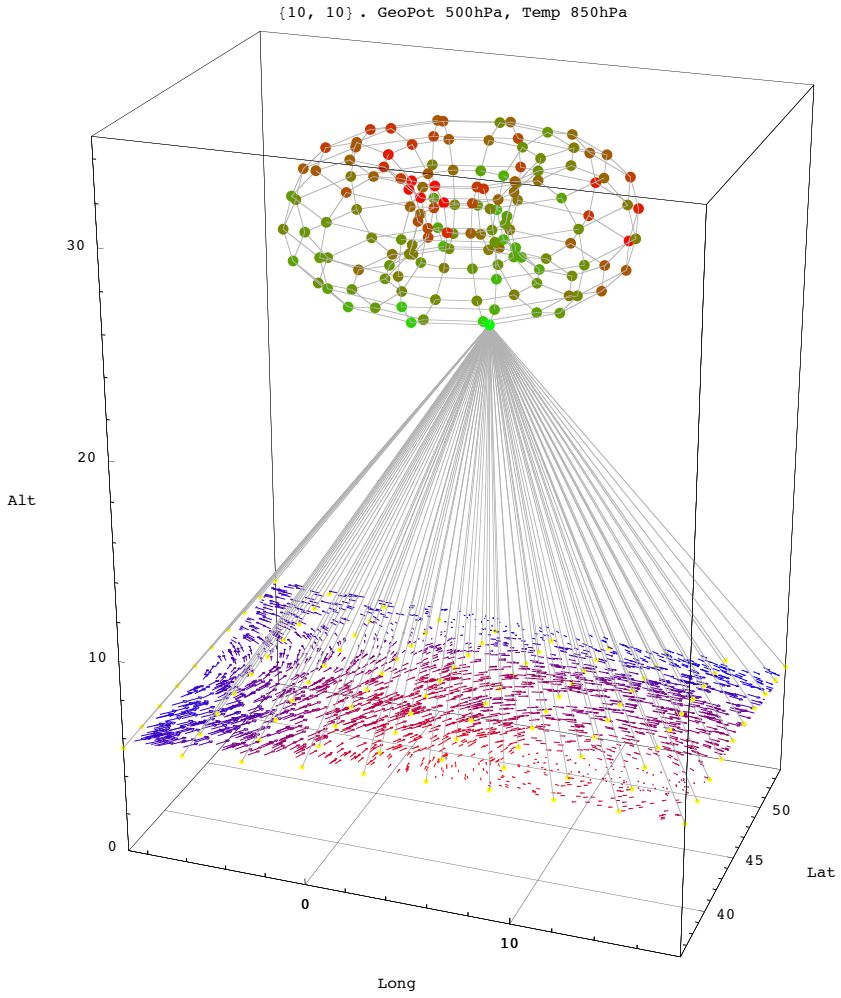

Figure 12: *Identified meteorological field: neuron* $\{10, 10\}$.

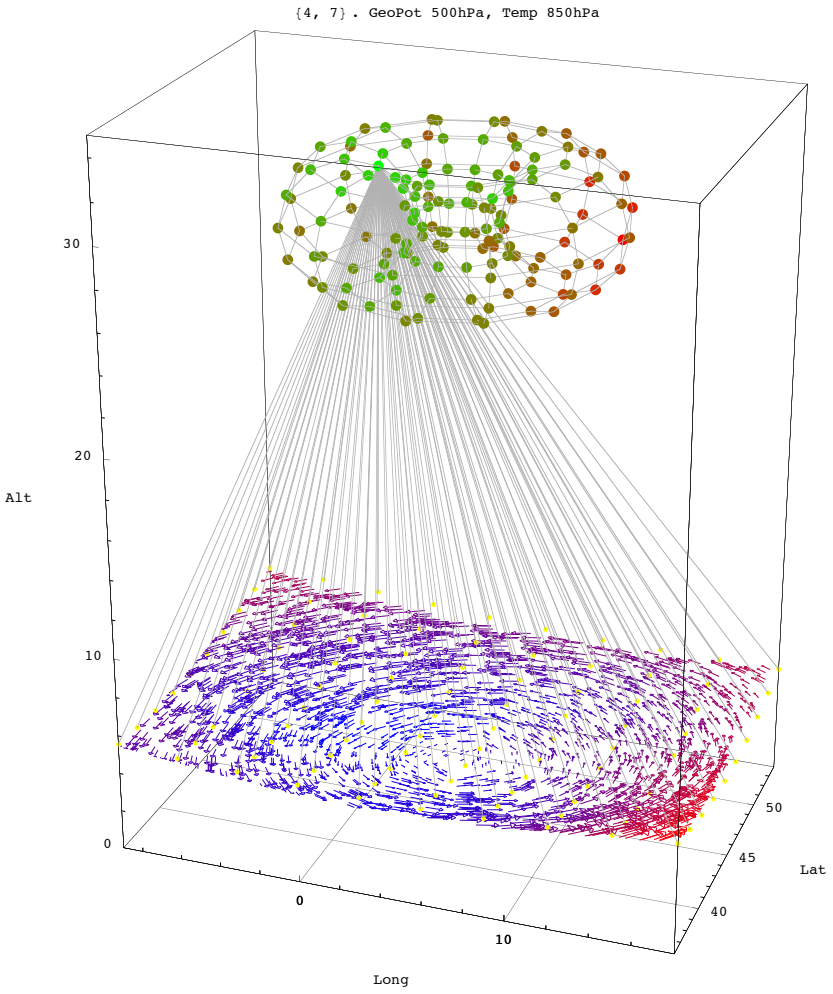

Figure 13: *Identified meteorological field: neuron* $\{4, 7\}$.

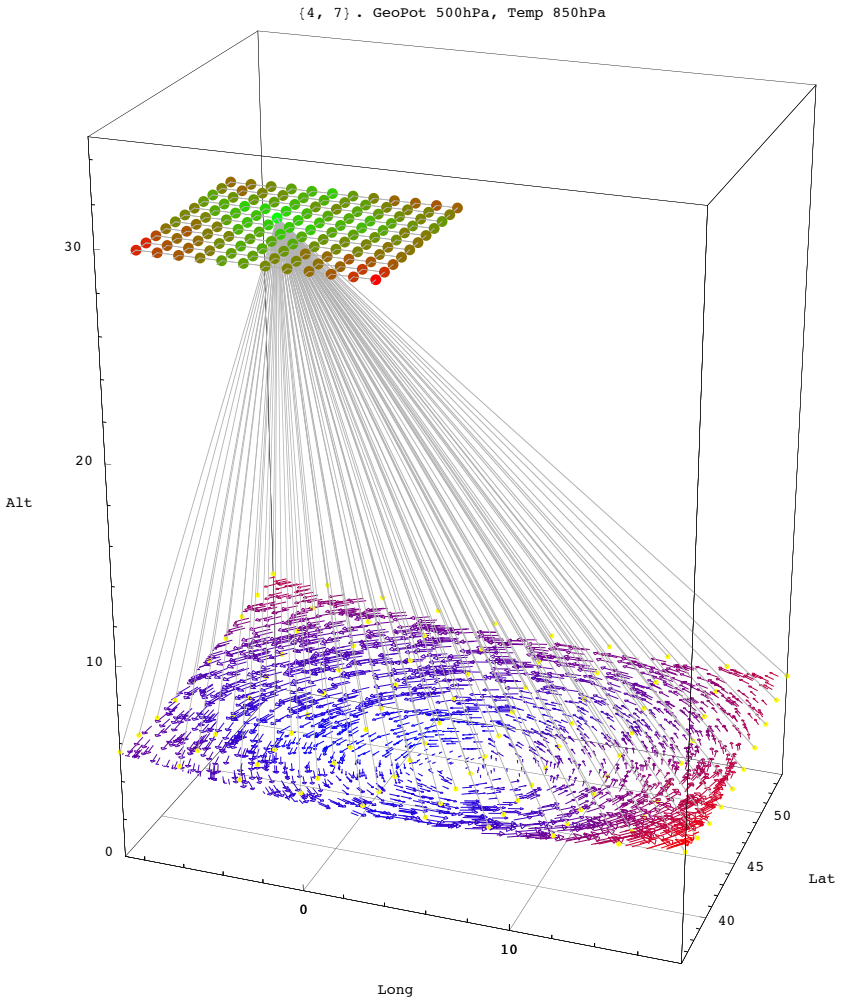

Figure 14: *Identified meteorological field: neuron* $\{4,7\}$. *As in figure 12, network unfolded to a square.*

7 Performance assessment

Various techniques have been implemented in order to asses the performance of the learning process and the quality of its outcome. Three of them, developed within the frame of our project, are hereafter presented. All of them provide a dynamical measure of the quality of the learning process as it goes on, until the network freezes in its learned state.

7.1 Geometric assessment

The most intuitive technique consists in measuring at each step p of the learning process (section 4.2.2) the distance $\mathcal{D}_{\{m',n'\}(\overline{\mathcal{X}}^p)}$ between the elected neuron $\{m', n'\}$ and $\overline{\mathcal{X}}^p$, the presented input pattern. At the beginning of the process, only few of the neurons having been submitted to the forcing of their synaptic weights, these distances are large. As learning proceeds, synaptic weights move in the input space towards the presented patterns. By the end of the process, when the network starts to be frozen in its learned state, for most of the presented patterns, a suitable neuron can be elected with the propriety of being, in the synaptic space, close to that pattern. Accordingly, the average Root Mean Square Distance converges towards smaller becoming values. This evolution, computed at the end of each cycle c, is presented in the upper panel in figure 15.

7.2 Entropic assessment

Entropy measures are used for two different purposes in the present work. They are firstly implemented as a measure of the quality of the learning process, secondly as a measure of the predictability of a meteorological ensemble forecast. Let us consider now the first application. Predictability measures will be discussed in section 8.1.

7.2.1 Standard information entropy

The learning process is divided in learning cycles[13] in order to evaluate, at the end of each cycle c, the frequency $P^c_{m\,n}$ at which the neuron $\{m, n\}$ has been chosen as the elected neuron during that cycle. Then the information entropy is evaluated following its classical definition:

$$\mathcal{S}_{(c)} = \sum_{m,n=1}^{M_{max},N_{max}} P^c_{m\,n}\; ln[\frac{1}{P^c_{m\,n}}] \tag{5}$$

[13] defined is section 4.1.4 with C_{max} being the number of such cycle and S_{data} the size of such a cycle, indeed the size of the sample of meteorological cases considered, also defined in Section 3

Two bounds on the entropy are easily defined. Firstly, if a same neuron $\{\tilde{m}, \tilde{n}\}$ has - pathologically- always been elected in a cycle c, then $P^c_{\tilde{m}\,\tilde{n}} = 1$ and $P^c_{m\,n} = 0$ for all other neurons, thus, accordingly, $\mathcal{S}_{(c)min} = 0$. Alternatively, if all neurons have been elected at the same frequency, then $\forall\,\{m,n\} \in \mathcal{N} :\ P^c_{m\,n} = \frac{S_{data}}{M_{max}\,N_{max}}$ and $\mathcal{S}_{(c)max} = ln[M_{max}\ N_{max}]$[14]. The evolution of the information entropy during the learning process, computed at the end of each cycle c, is presented in the middle panel in figure 15.

7.2.2 Geometric entropy

This extended definition of entropy is introduced in order to take in account the dispersion on the neural network of the signal produced either during the learning phase, or in operation, when the dispersion of ensemble prediction fields is assessed. The aforementioned standard entropy is a mapping of the network onto the real numbers: $\mathcal{N} \longmapsto \Re$. The geometric entropy is a similar mapping, this time established from the cartesian product of the network with itself, onto the real numbers: $\mathcal{N} \times \mathcal{N} \longmapsto \Re$.

In this new set up, the $\mathcal{D}_{(c)\{m,n\}\,(W^{o\,p})}$ metric is no longer computed between an input field $\mathcal{X}$ and the synaptic matrix of the neuron $\{m,n\}$, as presented in the definition of the learning algorithm, section 4.2.2. It is computed between the synaptic matrices of two neurons, in this case $\{m,n\}$ and $\{o,p\}$. For the stake of clarity, this intra-network distance will be written as $\mathcal{M}_{(c)(W^{m\,n},W^{o\,p})} = \mathcal{D}_{(c)\{m,n\}\,(W^{o\,p})}$. The evolution of the information entropy during the learning process, computed at the end of each cycle c, maps the cartesian product of the network with himself, onto the real numbers:

$$\begin{aligned} \mathcal{N} \times \mathcal{N} &\longmapsto \Re \\ \{\{m,n\}\ ,\ \{o,p\}\}_{(c)} &\longmapsto \mathcal{S}_{G(c)}, \end{aligned}$$

where the geometrical entropy is defined as:

$$\mathcal{S}_{G(c)} = \sum_{m,n}\sum_{o,p} \mathcal{Q}^{\{m,n\}\{o,p\}}_{(c)}\ ln[\frac{1}{\mathcal{Q}^{\{m,n\}\{o,p\}}_{(c)}}] \tag{6}$$

with:

$$\begin{aligned} \mathcal{Q}^{\{m,n\}\{o,p\}}_{(c)} &= P^c_{m\,n}\,P^c_{o\,p}\,\mathcal{M}_{(c)(W^{m\,n},W^{o\,p})}\,F_{(c)} \\ F_{(c)} &= \frac{M_{max}\,N_{max}\,(M_{max}\,N_{max}-1)}{\Delta_{(c)}\,S^2_{data}} \\ \Delta_{(c)} &= \sum_{m,n}\sum_{o,p}\mathcal{M}_{(c)(W^{m\,n},W^{o\,p})}. \end{aligned}$$

[14] Being originally an extensive thermodynamical quantity, the entropy is proportional to the surface of the neural network. Interestingly, the entropy of a cosmological black hole is, as in our case, proportional to the surface of its event horizon.

$P^c_{m\,n}$ and $P^c_{o\,p}$ are defined in section 7.2.1. $\Delta_{(c)}$ is the sum over $\mathcal{N}$ of all distances between the neurons at the end of the cycle c, where the distance of a neuron with itself is zero: $\mathcal{M}_{(c)(W^{m\,n},W^{m\,n})} = 0$[15]. $F_{(c)}$ is a normalization factor taking into account the size of the network, as well as $\Delta_{(c)}$, the average distance among all neurons at the end of the cycle c. The evolution of the geometrical information entropy during the learning process, computed at the end of each cycle c, is presented in the lower panel in figure 15.

7.3 Discussion

The three panels presented in figure 13 exhibit optically convincing convergences. However, a closer analysis yields interesting clues. On the top panel, where averages of RMS distances between elected neurons and presented fields are reported, one notices that the horizontal asymptote is not laid at 0 but leaves a residual distance. This means that none of the $S_{data} = 8000$ meteorological fields is exactly represented by a given neuron. Indeed, the typical weather patterns memorized in the network are averages of the initial fields.

Furthermore comparing the two lower panels, related to the entropies, one notices that the geometrical entropy converges slower towards its asymptotic plateau than the classical entropy. This demonstrates the fact that the former delivers a more sensitive measure of the quality of the learning process than the latter.

[15] As $\lim_{x\to 0}\ x\ ln(x^{-1}) = 0$, no $\{m,n\} \neq \{o,p\}$ restriction has to be introduced.

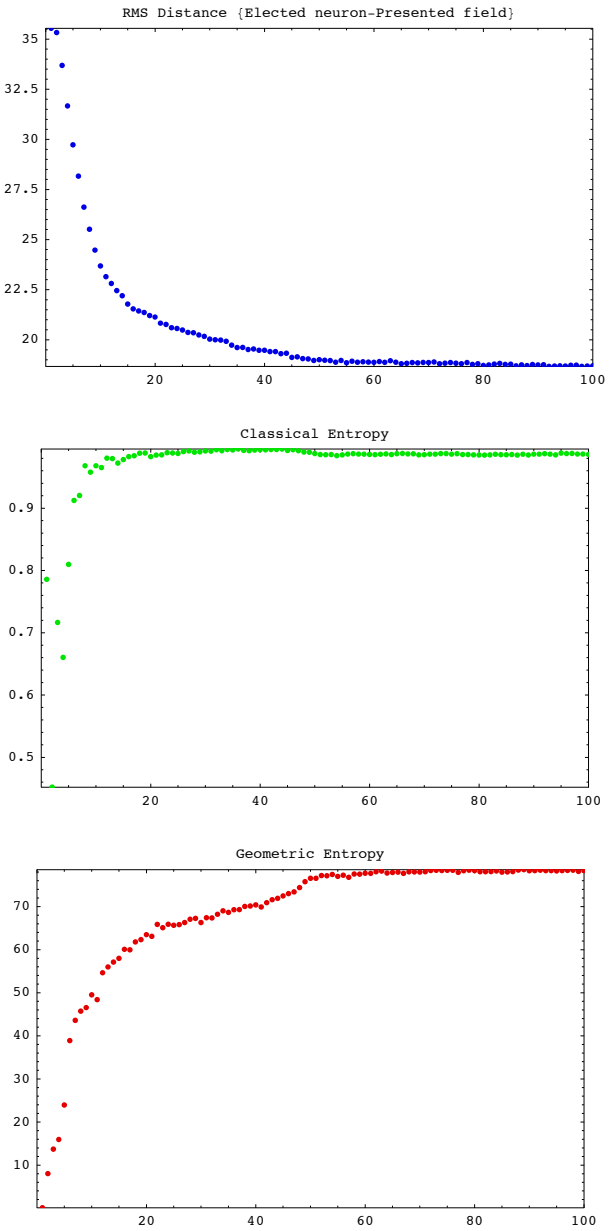

Figure 15: *Learning scores. Higher Panel: average of RMS distances between elected neurons and presented fields during the corresponding cycle. Middle panel: normalized information entropy of the neural chart at the end of the corresponding cycle. Bottom panel: geometrical entropy of the neural chart at the end of the corresponding cycle. On each panel the abscissa represents the learning cycles, as defined in sections 4.1.4 and 4.2.2.*

8 Ensemble Prediction System

The Ensemble Prediction System (EPS) was introduced about 15 years ago by ECMWF and has been constantly refined since then [8, 10]. Taking into account the chaotic behavior of the atmosphere, it computes alternative evolutions resulting in the production of differentiated weather patterns.

Technically, EPS is based on the whole set of observations available on the planet at a given time, that is fused into an initial condition established twice a day with a 4D-Var assimilation scheme. Beside of the forecast delivered by the deterministic operational model at maximal resolution (T799), an unperturbed "control" forecast is computed with a coarser (T399) resolution. Then, after having determined the most unbalanced vector states of the initial condition provided to the control forecast, called singular vectors [9], EPS generates 50 alternative -perturbed- forecasts[16] based on these singular vectors, computed on the same coarser grid as the control. Together with the deterministic and control forecasts, they provide 52 differentiated scenarii of the weather evolution by time steps of 12 hours at 00 and 12 UTC up to forecast day 15, with a (T399) resolution from day 0 to day 10, and with lower (T255) resolution from day 10 to day 15. In the vertical, all integrations have 62 levels, with the model top at 5 hPa.[17].

Those patterns, representing a huge amount of information, are finally to be interpreted on a local frame either by weather forecasters, or by any automated system. The methodology implemented at MeteoSwiss to achieve this interpretation is presented in the following section. It is based on the example provided by the operational ensemble prediction dispatched on 26 November 2009 at 12 UTC by ECMWF.

8.1 Neural interpretation of forecasts provided by the Ensemble Prediction System

Figure 16 describes the answer of the network for that ensemble forecast (26 November 2009 at 12 UTC) for verifying times elapsing from Monday November 30. 2009 12Z (+96 hours, upper left chessboard looking like square) to Sunday 6. December 2009 12 Z (+240 hours, bottom right chessboard looking like square).

The corresponding answers of the neural network, exhibited on the chessboards, are computed in the following way: for each verifying time, all the 52 members are submitted to the network. Following the scheme explained in section 6, it answers in taking 52 classification decisions, whereas each

[16] State of the matter at the time of writing.

[17] These alternative scenarii might be interpreted as being "clones" of the unperturbed control forecast.

member of the ensemble forecast is allocated to one neuron, indeed to one pattern of the meteorological classification. Taking advantage of the topological gathering explained in section 5.2, similar members of the ensemble are allocated either to a same neuron, or to neurons that are closely gathered on the network. Dissimilar members are allocated to distant neurons.

The frequency at which each neuron has been elected is provided by the colour code sketched on the bottom of figure 16. As for example, on Wednesday 2. December, neuron $\{5, 10\}$ (ordinate 5, abscissa 10) captured 8 members, as well as the control run. Neuron $\{4, 4\}$ only one member. The toroidal structure of the network can once again be noticed, for example on the Tuesday, where a bimodal distribution with only two clusters is sketched. The second one, quite loose, is in the centre of the chessboard, around neurons $\{5, 5\}$ and $\{5, 6\}$. The first one, much tighter and encompassing neurons $\{5, 11\}$, $\{5, 0\}$ and $\{4, 11\}$, lays on the vertical borders.

The increasing spread of the signal provided by the network, as the verifying time increases from +96 to +240 hours, can be used as a measure of the reliability of the ensemble forecast. Is the spread low, as for the Monday (+96 hours verifying time), then the reliability of the forecast is high. Is it large, then the reliability is accordingly low, as for the Sunday (+240 hours). Thus the geometric entropy, defined in section 7.2.2, expression (6), and providing a measure of the spread, is interpreted as a measure of the forecast predictability. It is computed and pictured as the continuous blue line in the bottom right panel in figure 16, with the dotted purple line representing a climatology of that predictability. Once normalized, this predictability is reported as a confidence index, evaluated between 0 and 10, and sketched as a red figure at bottom right corner of each box. This methodology is thoroughly presented and discussed in [6].

Let us now concentrate our attention on the Thursday, (+168 hours verifying time). It corresponds to the forecasting day 7 in the entropic diagram in figure 16 and will be analyzed in the following section and in figure 17.

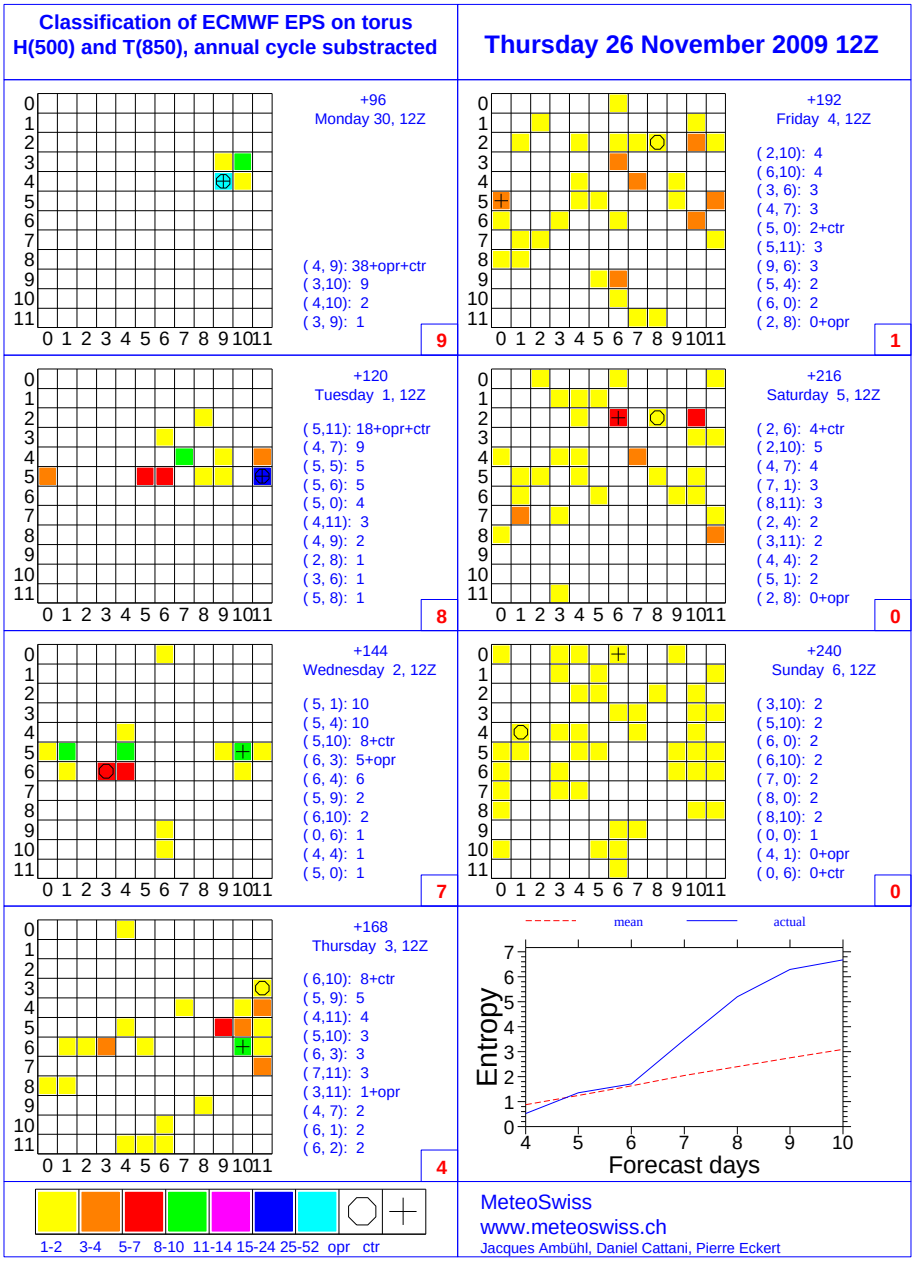

Figure 16: *Interpretation of the ensemble forecast from the 26 november 2009. Explanation provided in section 8.1.*

8.2 Neural interpretation of alternative scenarii

Figure 17, now discussed, presents the information concealed in the bottom left box of previous figure 16 for Thursday, December 3. 2009, (+168 hours verifying time).

The top panel represents the 500 hPa field averaged on the 52 members of the ensemble at each grid point of the ensemble grid, at T399 resolution. Not related to the neural computation, it provides an overall grasp of the general circulation forecasted by the ensemble over North Atlantic and Europe. Coloured patterns represent standard deviations of the flow, computed at each grid point of the ensemble grid. They are coded in the small rectangles provided on the top of the figure, from 2-4 decameters and 4-6 decameters in beiges, up to 12-20 decameters in lugubrious brown. The confidence laid in this forecast appears to be seriously hampered over the North Atlantic, as well as over Northeast France. The confidence index already presented in figure 16 is repeated in the small box laid at the bottom right corner of the panel: $CI = 4$ (in a scale running from 1 to 10). It should be stressed that the confidence index is not evaluated on the whole area of figure 17, but on the smaller domain onto which the neural network operates, defined in section 3, sketched in figures 6 and 7, and repeated in figure 18.

Let us now concentrate on the four small panels on the bottom of the figure. They correspond to neurons selected for the Thursday on preceding figure 16 and are labelled as units $\{6, 10\}$, $\{5, 9\}$, $\{4, 11\}$ and $\{6, 3\}$ on each panel. Not the synaptic map of the elected neuron, but the closest ensemble member to that synaptic map, as having been chosen by the classification decision, is presented on each of the four panels. Precisely expressed, the top left panel corresponds to the neuron $\{6, 10\}$, for which the closest field of the ensemble forecast, labelled run No 8, is pictured. The top right panel corresponds to the neuron $\{5, 9\}$, for which the closest field of the ensemble forecast, labelled run No 10, is pictured, and so on. Geopotential 500 hPa and temperature 850 hPa are shown, as usual. Neurons $\{6, 10\}$ and $\{5, 9\}$ being quite close neigbours on the network, the meteorological fields presented on the corresponding panels are closely related. More distant neurons, as for example $\{4, 11\}$ and $\{6, 3\}$, exhibit, as expected from the competitive non supervised learning algorithm, weakly related meteorological features.

As for the evaluation of the confidence index, the classification decision is not taken on the whole domain of the maps under discussion, but on the smaller domain onto which the neural network operates, defined in section 3 and sketched in figure 18, as well as in figures 6 and 7. The decision is taken on a local, European basis. Its consequences encompass, however, a

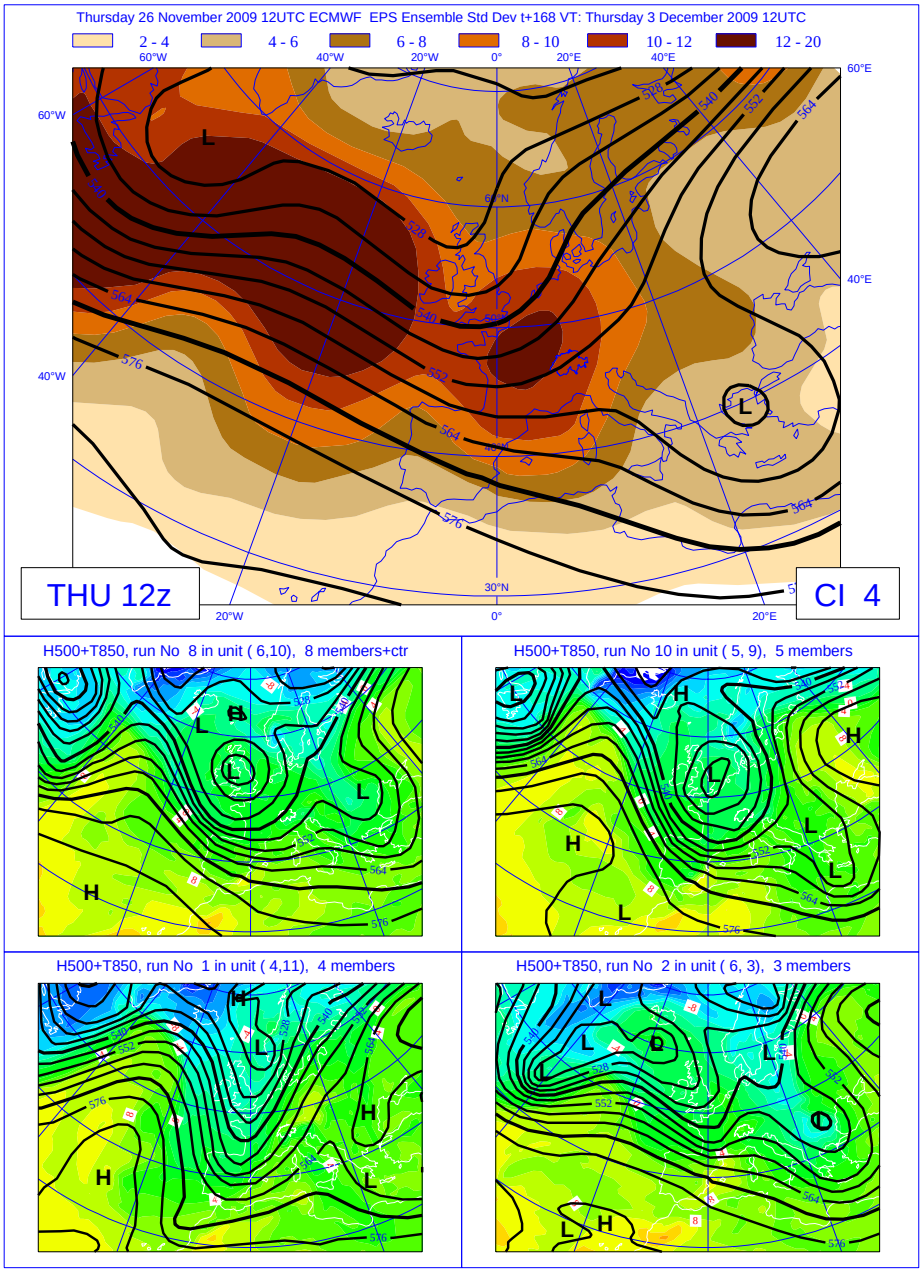

Figure 17: *Interpretation of the ensemble forecast from the 26 november 2009. Explanation provided in section 8.2.*

broader realm including the North Atlantic and wast areas of the occidental Mediterranean basin.

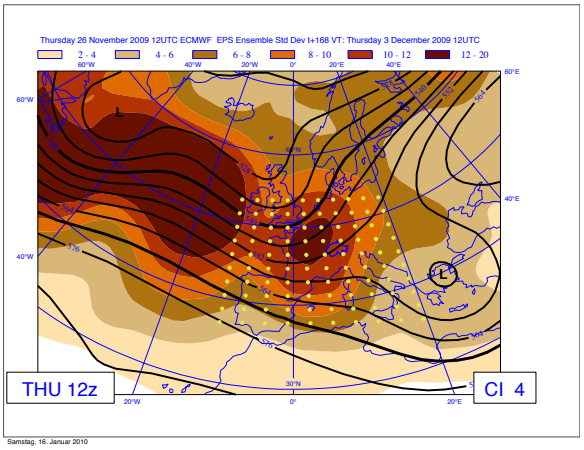

Figure 18: *Location of the synaptic contacts of the neural network.*

Retrospectively summing the number of actions taken by the network on a 10 days ensemble forecast, one notices that it has compared 10×52 ensemble members (alternative forecasts or, even, clones) with 144 neurons. It has taken 520 classification decisions after having considered $144 \times 520 = 74880$ alternatives. This system is operated twice a day at the forecasting division of MeteoSwiss.

8.3 Quantitative elements

Figure 19, now discussed, outlines the quantitative information extracted from the four runs. Once again does the top left panel correspond to the neuron $\{6, 10\}$, for which the run No 8, is pictured. The top right panel corresponds to the neuron $\{5, 9\}$ with run No 10, the bottom left panel to the neuron $\{4, 11\}$ with run No 1, the bottom right panel to the neuron $\{6, 3\}$, with run No 2. Following elements of information are provided in the compartments: first box: coordinates of the neuron and selected run. Second box: Geopotential heights 500 and 850 hPa over Payerne, expressed in decameters, corresponding temperatures in degrees Celsius, wind direction and velocity in knots. Altitude of the snowfall limit expressed in meters for western, eastern and southern Switzerland. Temperature difference between 850 and 500 hPa for those regions. Amounts of precipitation in 24 hours, expressed in mm, are provided for Geneva (GVE), Sion (SIO), Neuchâtel (NEU), Adelboden (ABO), Zürich Airport (KLO), Samedan (SAE), Lugano (LUG) and Domodossola (DOM). The confidence index $CI = 4$ is once again reported in the upper left, most populated neuron.

Of course ought bench forecasters to choose the evolution they consider the most likely to occur. The two first proposals, neurons $\{6, 10\}$ and $\{5, 9\}$, represent the most populated neurons (figure 14) and therefore express the most likely weather evolutions. Comparing the forecasted precipitation amounts for Geneva (GVE) and Adelboden (ABO) in the top panels of figure 16, it appears that tiny differences in the meteorological patterns induce substantial variations in precipitation amounts. Were, however, forecasters asked to consider exclusively likeliest evolutions, they would be required to operate with the most populated neuron, in that case $\{6, 10\}$. The meteorological decision would then *de facto* be automatically taken by the neural system. By the time of writing, forecasters at MeteoSwiss are still allowed to balance their decision between the few most populated neurons, thus maintaining a human component in the forecasting decision.

At this point, it is worth stressing that the temporal sequence of forecasts delivered by a deterministic model is physically consistent. On the contrary, a sequence of clusters provided by the algorithm, or chosen by a forecaster, may lack of deterministic connection. This foible is inherent to any clustering action exerted on the ensemble members. It is likely to be amplified by inadequate decisions.

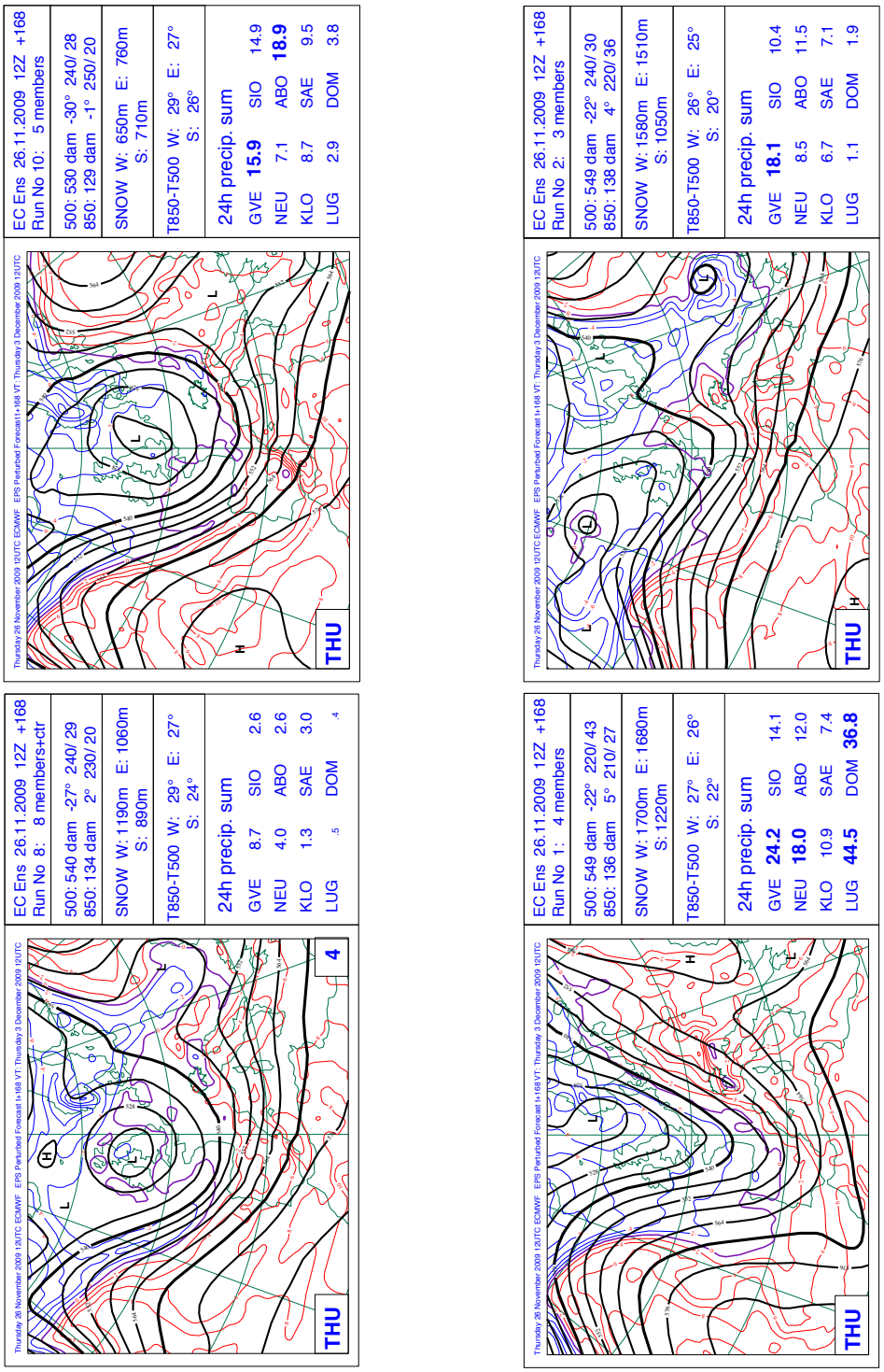

Figure 19: *Interpretation of the ensemble forecast from the 26 november 2009.*

Following figure 20 exhibits the ECMWF analysis for Thursday, December 3. 2009. Although the main trough over the British Isles and the North Sea is present, the two secondary throughs located over eastern Mediterranean and over Greece are presented in none of the four proposals provided in figures 17 and 18. This fact corroborates the low predictability ($CL = 4$) attributed to that 7 days forecast. Conclusively comparing the forecasted elements provided by neuron $\{6, 10\}$, figure 19, with the observations available for December 3., one notices that the weather elements were fairly forecasted for Switzerland with, however, an underestimation of the precipitation in the northeastern regions of the country.

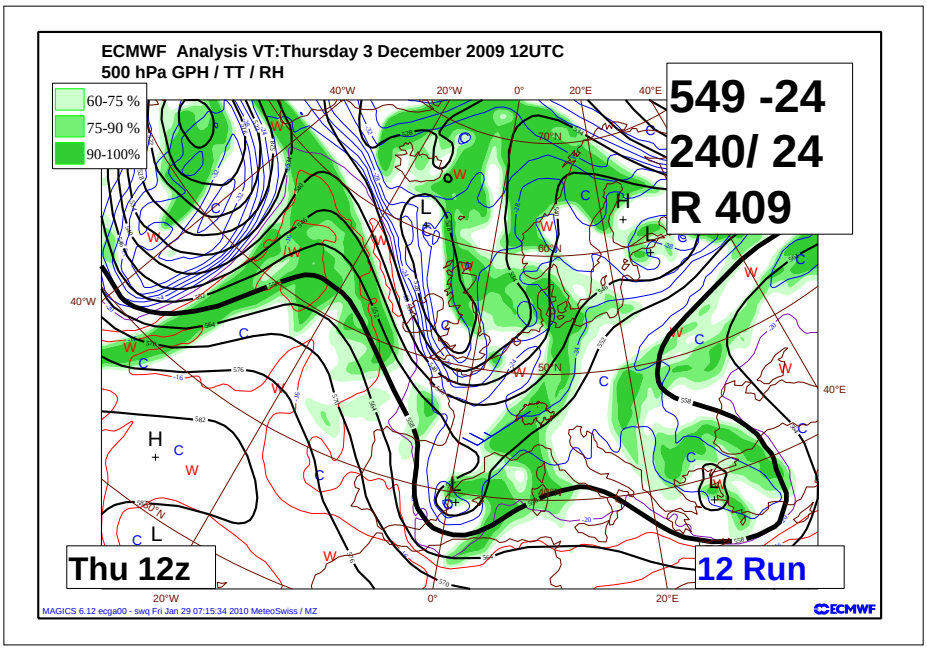

Figure 20: *Actual Analysis on Thursday 3 December 2009, 12 UTC. 500 hPa geopotential, temperature and relative humidity fields.*

9 Conclusion and outlook

Considering the informal structure of the project, it is fair to say that most of the objectives defined in section 2, Scope and aim, have been attained. The neural system has been implemented as the backbone of the decision methodology practiced at MeteoSwiss for middle range forecasts. It enables the formal and coherent interpretation of the information emanating from the Ensemble Prediction System. Furthermore, beside of fostering the unity of doctrine among forecasters, the neural interpretation ensures the quality and the coherence of products and services delivered by MeteoSwiss.

In a more philosophical stance, were the neural approach to be compared with the classical categorical approach established for databases in computing technology, it would become apparent that the neural network represents another, more holistic way of recording information. The emergence of patterns through unsupervised competitive learning provides the network with semantic capabilities that eventually transforms it into an intelligent meteorological knowledge basis.

Three possible avenues of improvement are hereafter outlined.

Geopotential 500 hPa and Temperature 850 hPa were chosen on a quite intuitive basis at an early stage of the project. Of course should other meteorological or geophysical parameters be considered, possibly addressing needs emanating from specific customers, as for example from hydrology or aviation [5, 7].

Within the frame of its Quality Assessment System, MeteoSwiss verifies elements of its middle range forecasts end products. The monitoring procedure takes into account progresses occurring along the whole forecasting chain, from assimilation systems over the models up to the interpretations schemes, and any specific impact of the neural interpretation remains concealed. A clear improvement over the years in the quality of middle range forecasts dispatched by MeteoSwiss is demonstrated (KOMIFRI[18]). However, a verification of the specific impact of the neural network on the forecasting suite could be emphasized.

Following the line of thought proposed in sections 4.3 and 5.1, neural networks could be conceived and specifically trained to identify potentially critical patterns, and consequently operated as warning systems. They would no longer be operated as all purpose classifiers, but would then be tailored to meet specific needs.

[18] Kontroll der MIttelFRIstigen Prognose

10 References

1. Perret R. 1983. Une classification des situations météorologiques à l'usage de la prévision, (Arbeitsbericht Nr. 46 der schweizerischen meteorologischen Anstalt)

2. Ambühl J. 1993. Réseaux de neurones en météorologie: une perspective d'application, (Arbeitsbericht Nr. 123 der schweizerischen meteorologischen Anstalt)

3. Eckert P, Cattani D, Ambühl J. 1996. Classification of ensemble forecasts by means of an artificial neural network (Meteorological Applications 3, pp. 169-178)

4. Kohonen T. 1990. The Self-Organizing Map. (Proceedings of the IEEE, vol 78, no 9, pp. 1464, 1480)

5. Michaelides S.C, Liassidou F, Schizas C.N. 2007. Synoptic Classification and Establishment of Analogues with Artificial Neural Networks. (Pure and appl. Geophys. 164(2007), pp. 1347-1364)

6. Scherrer C, Appenzeller Ch, Eckert P, Cattani D. 2003. Analysis of the Spread-Skill Relations Using the ECMWF Ensemble Prediction System over Europe. (Weather and Forecasting. Vol 19. No. 3, pp 552-565)

7. Bremmes J.B, Michaelidis S. Ch. 2007. Probability visibility forecasting using neural networks. Short range forecasting methods of fog, visibility and low clouds. (Cost action 722, final report. Pp. 421-426)

8. Palmer T N, Molteni F, Mureau R, Buizza R, Chapelet P, Tribbia J. 1993. Ensemble Prediction. (Proceedings 1992 ECMWF Seminar: Validation of Modes over Europe, pp 21-66, Reading, UK)

9. Palmer T N, Gelaro R, Barkmeijer J, Buizza R. 1998 Singular vectors, metrics and adaptative observations. (J. Atmos. Sci,. 55,633-653)

10. Palmer T N and all. 2007. The Ensemble Prediction System - Recent and Ongoing Developments. (ECMWF Technical Memorandum no. 540)

Author's Address

ambuhl@bluewin.ch; jacques.ambuehl@meteoschweiz.ch

Veröffentlichungen der MeteoSchweiz

84 Ambühl, J: 2010, Customer oriented warning systems, 86 pp, CHF 78.-

83 Ceppi, P: 2010, Spatial characteristics of gridded Swiss temperature trends: local and large-scale influences, 82 pp, CHF 76.-

82 Blanc, P: 2009, Ensemble-based uncertainty prediction for deterministic 2 m temperature forecasts, 90pp, CHF 78.-

81 Erdin R: 2009, Combining rain gauge and radar measurements of a heavy precipitation event over Switzerland: Comparison of geostatistical methods and investigation of important influencing factors, 109pp, CHF 81.-

80 Buzzi M: 2008, Challenges in Operational Numerical Weather Prediction at High Resolution in Complex Terrain, 186pp, CHF 103.-

79 Nowak D: 2008, Radiation and clouds: observations and model calculations for Payerne BSRN site, 101 pp, CHF 80.-

78 Arpagaus M, Rotach M, Ambrosetti P, Ament F, Appenzeller C, Bauer H-S, Bouttier F, Buzzi A, Corazza M, Davolio S, Denhard M, Dorninger M, Fontannaz L, Frick J, Fundel F, Germann U, Gorgas T, Grossi G, Hegg C, Hering A, Jaun S, Keil C, Liniger M, Marsigli C, McTaggart-Cowan R, Montani A, Mylne K, Ranzi R, Richard E, Rossa A, Santos-Muñoz D, Schär C, Seity Y, Staudinger M, Stoll M, Vogt S, Volkert H, Walser A, Wang Y, Werhahn J, Wulfmeyer V, Wunram C and Zappa M: 2009, MAP D-PHASE: Demonstrating forecast capabilities for flood events in the Alpine region. Report of the WWRP Forecast Demonstration Project D-PHASE submitted to the WWRP Scientific Steering Committee, 65pp, CHF 73.-

77 Rossa AM: 2007, MAP-NWS – an Optional EUMETNET Programme in Support of an Optimal Research Programme, 67pp, CHF 73.-

76 Baggenstos D: 2007, Probabilistic verification of operational monthly temperature forecasts, 52pp, CHF 69.-

75 Fikke S, Ronsten G, Heimo A, Kunz S, Ostrozlik M, Persson PE, Sabata J, Wareing B, Wichura B, Chum J, Laakso T, Säntti K and Makkonen L: 2007, COST 727: Atmospheric Icing on Structures Measurements and data collection on icing: State of the Art, 110pp, CHF 83.-

74 Schmutz C, Müller P und Barodte B: 2006, Potenzialabklärung für Public Private Partnership (PPP) bei MeteoSchweiz und armasuisse Immobilien, 82pp, CHF 76.-

73 Scherrer SC: 2006, Interannual climate variability in the European and Alpine region, 132pp, CHF 86.-

72 Mathis H: 2005, Impact of Realistic Greenhouse Gas Forcing on Seasonal Forecast Performance, 80pp, CHF 75.-

71 Leuenberger D: 2005, High-Resolution Radar Rainfall Assimilation: Exploratory Studies with Latent Heat Nudging, 103pp, CHF 81.-

70 Müller G und Viatte P: 2005, The Swiss Contribution to the Global Atmosphere Watch Programme – Achievements of the First Decade and Future Prospects, 112pp, CHF 83.-

69 Müller WA: 2004, Analysis and Prediction of the European Winter Climate, 115pp, CHF 34.

68 Bader S: 2004, Das Schweizer Klima im Trend: Temperatur- und Niederschlagsentwicklung seit 1864, 48pp, CHF 18.-

67 Begert M, Seiz G, Schlegel T, Musa M, Baudraz G und Moesch M: 2003, Homogenisierung von Klimamessreihen der Schweiz und Bestimmung der Normwerte 1961-1990, Schluss-bericht des Projektes NORM90, 170pp, CHF 40.-

Frühere Veröffentlichungen und Arbeitsberichte finden sich unter
www.meteoschweiz.ch » Forschung » Publikationen

Arbeitsberichte der MeteoSchweiz

229 Philipona R, Levrat G, Romanens G, Jeannet P, Ruffieux D and Calpini B, 2009: Transition from VIZ / Sippicanto ROTRONIC - A new humidity sensor for the SWISS SRS 400 Radiosonde, 37pp, CHF 66.-

228 MeteoSchweiz: 2009, Klimabericht Kanton Graubünden, 40pp, nur als .pdf erhältlich

227 MeteoSchweiz, 2009, Basisanalysen ausgewählter klimatologischer Parameter am Standort KKWLeibstadt, 135pp, CHF 88.-

226 MeteoSchweiz, 2009, Basisanalysen ausgewählter klimatologischer Parameter am Standort KKW Mühleberg, 136pp, CHF 88.-

225 MeteoSchweiz, 2009, Basisanalysen ausgewählter klimatologischer Parameter am Standort KKW Gösgen, 136 pp, CHF 88.-

224 MeteoSchweiz, 2009, Basisanalysen ausgewählter klimatologischer Parameter am Standort KKW Beznau, 135pp, CHF 88.-

223 Dürr B: 2008, Automatisiertes Verfahren zur Bestimmung von Föhn in den Alpentälern, 22pp, CHF 62.-

222 Schmutz C, Arpagaus M, Clementi L, Frei C, Fukutome S, Germann U, Liniger M und Schacher F: 2008, Meteorologische Ereignisanalyse des Hochwassers 8. bis 9. August 2007, 29pp, CHF 64.-

221 Frei C, Germann U, Fukutome S und Liniger M: 2008, Möglichkeiten und Grenzen der Niederschlagsanalysen zum Hochwasser 2005, 19pp, CHF 62.-

220 Ambühl J: 2008, Optimization of Warning Systems based on Economic Criteria, 79pp, CHF 75.-

219 Ceppi P, Della-Marta PM and Appenzeller C: 2008, Extreme Value Analysis of Wind Observations over Switzerland, 43pp, CHF 67.-

218 MeteoSchweiz (Hrsg): 2008, Klimaszenarien für die Schweiz – Ein Statusbericht, 50pp, CHF 69.-

217 Begert M: 2008, Die Repräsentativität der Stationen im Swiss National Basic Climatological Network (Swiss NBCN), 40pp, CHF 66.-

216 Della-Marta PM, Mathis H, Frei C, Liniger MA and Appenzeller C: 2007, Extreme wind storms over Europe: Statistical Analyses of ERA-40, 80pp, CHF 75.-

215 Begert M, Seiz G, Foppa N, Schlegel T, Appenzeller C und Müller G: 2007, Die Überführung der klimatologischen Referenzstationen der Schweiz in das Swiss National Climatological Network (Swiss NBCN), 47pp, CHF 68.-

214 Schmucki D und Weigel A: 2006, Saisonale Vorhersage in Tradition und Moderne: Vergleich der "Sommerprognose" des Zürcher Bööggs mit einem dynamischen Klimamodell, 46pp, CHF 68.-

213 Frei C: 2006, Eine Länder übergreifende Niederschlagsanalyse zum August Hochwasser 2005. Ergänzung zu Arbeitsbericht 211, 10pp, CHF 59.-

212 Z'graggen, L: 2006, Die Maximaltemperaturen im Hitzesommer 2003 und Vergleich zu früheren Extremtemperaturen, 74pp, CHF 75.-

211 MeteoSchweiz: 2006, Starkniederschlagsereignis August 2005, 63pp, CHF 72.-

210 Buss S, Jäger E and Schmutz C: 2005: Evaluation of turbulence forecasts with the aLMo, 58pp, CHF 70.–

209 Schmutz C, Schmuki D, Duding O, Rohling S: 2004, Aeronautical Climatological Information Sion LSGS, 77pp, CHF 25.–

Frühere Veröffentlichungen und Arbeitsberichte finden sich unter
www.meteoschweiz.ch » Forschung » Publikationen