Neural, Parallel, and Scientific Computations 18 (2010) 343--356

ON THE FORECASTING ABILITIES OF A TIME VARYING AUTO-ADAPTING ALGORITHM

ANTONIO BALLARIN AND SIMONA GERVASI
Stefano Bacchetti, Umberto Capponi, Stefano Costi, Kristian A. Gervasi Vidal, Paul B. Moore, Carlo Nardone, Giovanni Passali, Francesco Sagone, Morena Signori, Fabrizio Vollera
Research Department, Quantum Forecasting LLC, 50 Broad Street - Suite 1911, New York, NY 10004.

Abstract. Physical motivations, theoretical aspects, and practical applications of a time-varying, auto-adaptive algorithm are described, as well as the results obtained through its application in some practical examples; these results were reached during a time span of over ten years from its first presentation. The intrinsic non-ergodicity of the physical phenomena leads us to hypothesize the existence of a *characteristic time parameter*, specific for each single physical phenomenon, uniquely valid in the temporal interval during which the same phenomenon is observed, in such a way as to transform the ergodic hypothesis into a locally valid ergodic approximation. The theoretical approach for determining the form of this time parameter springs from learning processes that take place without total memory loss. The algorithm's application to time series forecasts of any nature shows an extreme ease of utilization and an elevated forecasting capability, which vastly overcomes expected performances of forecasts obtainable through the use of tools derived from classical statistical methods.

1. INTRODUCTION

In Statistical Mechanics a macroscopic system made up of *N* microscopic components can be represented, at time *t*, by a *6N* dimension vector $\mathbf{x}(t) \equiv (\mathbf{q}_1(t),\ldots,\mathbf{q}_N(t),\mathbf{p}_1(t),\ldots,\mathbf{p}_N(t))$ for which $\mathbf{q}_i$ and $\mathbf{p}_i$ represent, respectively, the position vectors and the impulse vectors. The observables of the system, defined in the 6*N* dimensional space in the Γ phases, are represented by the functions $A(\mathbf{x}(t))$. Such a system is, generally, dynamic, or rather, for this system there exists a law of deterministic evolution, in Γ, of the type $\mathbf{x}(0) \to \mathbf{x}(t) = U^t\mathbf{x}(0)$ and an invariant measure $d\rho(\mathbf{x})$ under the evolution given by U^t. The system thus characterized is defined as ergodic if, for every integrable function $A(\mathbf{x}(t))$ and for nearly all the initial conditions $\mathbf{x}(t_0)$, we have:

$$\overline{A} \equiv \lim_{T\to\infty}\frac{1}{T}\int_{t_0}^{t_0+T} A(\mathbf{x}(t))dt = \int_{\Gamma} A(\mathbf{x})d\rho(\mathbf{x}) \equiv \langle A\rangle \tag{1}$$

Equation (1) describes the so-called *ergodic hypothesis*, as formulated by Boltzmann and Gibbs (Boltzmann, 1964; Gibbs, 1981), according to which the fraction of time, during an interval of time *T, sufficiently long*, within which can be found a determined system $\mathbf{x}(t)$ in any region of the phase space, is equal to the fraction of all the spaces of the *statistical ensemble* which represent it and which are found in the same region of the phase space. In other words, point **x** in Γ, representative of the macroscopic state of the system, evolves along the entire space, spending in every sub-volume $\Omega \subset \Gamma$, a time proportional, on average, to Ω itself. Thus, if the system is observed in a certain instance of time, chosen randomly and in a very long temporal interval (long enough to allow the trajectory to explore the entire volume of the phase space of interest), the probability of finding the system in a generic set coincides with its volume Ω. This characteristic is known as the *equiprobability* of microscopic states, and all of Statistical Mechanics, in each of its formulations, is based on this principle.

Received April 25, 2010

Another way of interpreting the ergodic hypothesis is the following: the probability of finding some physical system considered in a particular state, in a certain instance of time, is equal to the probability of selecting a system chosen randomly from among the corresponding systems of the statistical ensemble in that state (or rather, within the considered region of the phase space). In this sense, the temporal mean of any physical entity $A(\mathbf{x}(t))$, pertaining to the particular observed system of origin, will be equal to the corresponding mean carried out for that entity among the systems of the corresponding statistical ensemble. The importance of this vision resides in the fact that, through the identity that it is possible to establish between the two kinds of means, it is possible to acquire a better understanding of some notions of the system by observing not necessarily all of the possible and single states into which it could evolve during its lifetime, but only a part. In fact, (1) represents an arrival point according to which the average obtained on a series of consecutive trials in the same experiment will be in agreement with the average on the ensembles, allowing, in this way, for every particular physical system observed to exhibit a behavior very different than its mean, in time.

Even if the ergodic hypothesis and the consequent ergodic theory have taken their origins in considerations on thermodynamic systems, in physical reality, very often, systems which are not necessarily thermodynamic, are approximated as such. In other words, there exist certain conditions within which one can assert that, *in substance, on average, in summary*, etc., there isn't a great difference between the specific system being considered and a thermodynamic system, as long as it is endowed with certain characteristics that render it similar to the former. There exist many examples for which such similitude is called into play, such as for the neurons of the cerebral cortex in the human brain; or, for the stars of a galaxy; or, yet, in the sequences of rolling dice or flipping a coin. At the same time, one can plainly see how these similarities are often completely out of place, due to the simple fact that the conditions within which (1) remains valid must never be ignored.

In (1), the average carried out on infinity exists and *the limit* does not depend on the choice of the starting point of the determined trajectory considered (Birkhoff, 1931; von Neumann, 1932); still, the complete independence *of the system* from its starting point is not guaranteed. Moreover, a condition that is necessary and sufficient, and within which the ergodic hypothesis is valid, involves the fact that the phase space is not subdivided into two parts, each in a positive measure, in which the invariance with respect to the dynamic U^t continues to hold true; but in general it is not possible to decide *a priori* whether a particular system can satisfy this condition or not. Finally, the main problem regarding the usability of (1) involves the effective size of T, or rather the physical approximation of the concept of infinity: how large must T be in order for the temporal mean to result in being equal to the mean of the ensemble?

In practice (1) is always approximated in the following form:

$$\overline{A} \cong \frac{1}{\Delta T} \int_{t_0}^{t_0+\Delta T} A(\mathbf{x}(t))dt \cong \int_{\Gamma} A(\mathbf{x})d\rho(\mathbf{x}) \equiv \langle A \rangle \tag{2}$$

for which, at once it is made clear that: "ΔT must be sufficiently large". What significance must be attributed to the concept underlined by the phrase *sufficiently large*? The same original formulation of the ergodic hypothesis raises this question, imposing very long times for the observation of the considered systems for its validity.

In practice, leaving behind the initial states and even in conditions very far from equilibrium, the ergodic principle hypothesizes that the underlying dynamic *pushes* the system to a certain volume in the phase space; once this state is reached and occupied in a *dense* manner, the system will spend the vast majority of its lifetime there, except for small, possible-but-rare fluctuations. In *this* state the equiprobability principle is valid, and the transient represents, uniquely, the process of closing in on equilibrium. If, then, any measurement should be carried out which is far from a state of equilibrium, the result would be strongly influenced by the system's initial conditions. Now, in the real world, it is very difficult to have any *a priori* knowledge of the intimate nature of a phenomenon being observed, and furthermore, for this reason, it is impossible to understand if that which is being observed is in a state of equilibrium or not. One can postulate some presumptions such as, for example, the necessity of considering the system *rigorously separated* from the rest of the universe (that is, the external environment), so that external forces cannot intervene on it. So, hypothesizing carrying out a measurement in a state in which the system in transient is not at all out of place; indeed, given the general ignorance regarding the phenomenon, it is all the more appropriate and prudent. In any case, the effective observation of such a phenomenon and its associated system takes place, by its nature, in a physically limited time. These measurements are usually conducted in a time that is anything but infinite, since they are characterized by an initial instance of observation (in which the system is in a very well defined state) and a present instance, after the first survey, in an interval during which the system has typically evolved, having been strongly influenced by the conditions observed at the time of the initial measurements.

In synthesis, the ergodic hypothesis is quite far from being demonstrated in real systems (see for example, Gallavotti, 1982) and non-ergodicity as a condition is much more probable, in nature, than ergodicity itself (Buonomano, 1987). After all, reason brings us to assert that, in flipping a coin a billion times, the averages of heads and tails obtainable could be, with a very good approximation, very close to the result that could be obtained in flipping a billion similar coins, once each; however, the same reason imposes upon us doubts if such equivalence could remains true even when the number of flips is greatly diminished. So then, what is the inferior limit that defines the ergodicity of a system?

The factual statement of the impossibility of knowing which value of ΔT correctly approximates correctly ergodic conjecture, forms the base of the hypothesis described in the present work, in which there is postulated the existence of a certain *time characteristic parameter* (or, more correctly *characteristic parameter functional*), peculiar for every single physical phenomenon, uniquely valid in the interval of observation. In practical terms, if the observation time interval of the phenomenon is not *sufficiently long*, if the observation goes beyond that *time characteristic parameter* for the considered phenomenon, then the ergodic hypothesis will not be true, either. In other words, we hypothesize the local non-validity of the ergodic hypothesis and we rewrite the left side of (1) in the following form:

$$\overline{A} = \int_{t-T(t)}^{t} A(\mathbf{x}(t))dt \tag{3}$$

where the *true* unknown quantity of the problem is the term $T(t)$ which appears at the inferior limit of the integral. Let us attribute to $T(t)$ the role of element capable of supplying a description of the phenomenon being observed, in the considered temporal window. Equation (3) thus formulated, describes a completely natural behavior, in the sense that it expresses the observation of a physical phenomenon not in an interval that spans from $-\infty$ to $+\infty$, but rather in a temporal window that is necessarily limited, and, in this interval, calls into question the existence and the nature of a *characteristic time*, or rather, an observable capable of *describing the system* and such for which (1) is true.

2. THEORETICAL CONSIDERATIONS

The search for an explicit form for $T(t)$ has been the leitmotif of the investigations carried out in these past years, being that we are aware that its determination would have characterized a better comprehension in the physical observation of a system. In fact, the existence of a *control parameter* able to govern the dynamics phase transaction processes (such as, for example, in chaotic systems, regarding order/disorder transients) is well noted in the literature (see, for example, Arecchi, 1990). Let us focus our attention, then, on a typical system governed by a similar parameter, namely a system of interacting neurons. Such systems are undoubtedly more complex, and the results obtained in this sphere can easily be transferred to fields in which such complexity is less. The process for determining a possible explicit form for $T(t)$ is long and arduous, and so, in the present work, only the key points of these processes will be enounced; the demonstrations of the propositions to be encountered in this article can be found in other articles.

As already described (Ballarin *et al.*, 1995), every learning mechanism with the goal of being biologically plausible, must satisfy two important properties which are very general and able to be found in nature, and that can be summarized, in extreme synthesis, in the following statements:

1. The connections – in this case, the synapses – must be correlated to the activities of the sites (neurons) that the synapses themselves connect.
2. The overloading in the memorization processes must not imply a complete memory loss in a system; in these situations, it is preferable to substitute the old memories with the new ones (palimpsest property).

A simultaneous response to both the properties, is given by a type model like the following (Shinomoto, 1987):

$$\Delta W_{ij} = -\gamma W_{ij} + \left\langle s_i s_j \right\rangle_W \qquad \forall \; i, j \tag{4}$$

in which W_{ij} represents the connection between the sites i and j and the W index indicates the average is calculated with a constant, predefined value of the synaptic matrix. The model synthesized in (4a), postulates two basic assumptions: first of all, the time scale of the synaptic modifications is hypothesized as sufficiently ample with respect to that of the neuronal modifications; secondly, for the synaptic connections one presumes a certain form of limitation by means of a rule of a physiological nature. The extension of such a model, in line with the considerations expressed in the introduction (Basti *et al.*, 1991; Ballarin *et al.*, 1995), transforms (4a) as:

$$\Delta W_{ij} = -\gamma W_{ij} + \left\langle s_i s_j \right\rangle_{T(t)} \qquad \forall \; i, j \tag{5}$$

that is, passing to the derivative:

$$\frac{dW_{ij}(t)}{dt} = -\gamma W_{ij}(t) + \int_{t-T(t)}^{t} s_i(x) s_j(x) dx = -\gamma W_{ij}(t) + Q_{ij}(t) \qquad \forall \; i, j \tag{5b}$$

where the activation dynamic assumes the standard form

$$\frac{ds_i(t)}{dt} = -s_i(t) + g\left\{ \sum_{j=1}^{N} W_{ij}(t) s_j(t) \right\} \qquad \forall \; i = 1, 2, \ldots \tag{5c}$$

and $g(\bullet)$ represents a sigmoid function. The model hypothesized here represents an entirely connected matrix $N \times N$ in which $W_{ij}(t)$ is the variable connection in time between sites i and j, every site can emit an informative unit, $s_i(t) = +1$, or else it can be in a silent state, $s_i(t) = -1$; in this way, we obtain that $s_i(t) s_j(t) = \pm 1$ for each pair i, j, and this implies

$$\left| Q_{ij}(t) \right| = \left| \int_{t-T(t)}^{t} s_i(x) s_j(x) dx \right| \le T(t) \qquad \forall \; i = 1, 2, \ldots \tag{6}$$

In the model here described, we hypothesize that at a certain time, t_0^m, the system is stimulated with an input, i.e. the system reaches the m-th pattern $\bar{\xi}^m = \left(\xi_1^m, \xi_2^m, \ldots, \xi_N^m \right)$. In this time we see that $\bar{s}\left(t_0^m\right) = \bar{\xi}^m$. The matrix of the synaptic connections $\boldsymbol{W} = \left\lfloor W_{ij} \right\rfloor$ is supposed, in general, to be non-symmetric: $W_{ij}\left(t_0^m\right) - W_{ji}\left(t_0^m\right) \ne 0$. The simulations show that, if $T(t)$ satisfies certain conditions described further on, the system given by the combination of (5b) and (5c), when facing a stimulus after the pattern presentation, evolves in a chaotic manner in the beginning, then, after a certain time period, becomes stable. Let us call t_1^m the time around which the system enters into a stable state. The behaviour to which the system is subject, after the presentation of an input, plays a fundamental role in the theory described here; basically, depending on the mutual relationship established in the second member of the learning dynamic (5b), different behaviours are observed within the system itself. It results as being stable if:

$$Q_{ij}(t) >> \left|\gamma W_{ij}(t)\right| \qquad \forall \; i, j \text{ and } t \geq t_1^m \tag{7}$$

while it is chaotic, or noisy, in the case for which

$$Q_{ij}(t) << \left|\gamma W_{ij}(t)\right| \qquad \forall \; i, j \text{ and } t_0^m \leq t \leq t_1^m \tag{8}$$

In the transition phase, or rather for values of the correlation integral of the type

$$Q_{ij}(t) \approx \left|\gamma W_{ij}(t)\right| \qquad \forall \; i, j \text{ and } t \approx t_1^m \tag{9}$$

the system is chaotic. These behaviours can be verified through the determination of Lyapunov exponents. From (5b) and (5c) it is not difficult to construct a model for which, at time t_0^m of the presentation of the input pattern of the presentation of the input, chaos condition described in (8) will be satisfied, and which, with the flowing of time, can pass over, with *continuity*, to a state in which the condition of stability described by (7) is valid. In substance, the stress caused by a new stimulus in the system described here, leads to the recording of a decidedly confused activity in its initial phase, and, with the flowing of time, it reaches an equilibrium phase. From the physical point of view, such a system reflects a behaviour synthesized in (7)-(9), and phases such as these are characterized by their following one another in a continuous manner; that is, no discontinuity is observed in the system's transition from one state to another. During the stable phase, that is, when (7) is valid, we have:

$$\frac{ds_i(t)}{dt} = 0 \qquad \forall \; i = 1, 2, \ldots \text{ and } t \geq t_1^m \tag{10}$$

while, in the chaotic state, the following is valid:

$$\frac{ds_i(t)}{dt} \neq 0 \qquad \forall \; i = 1, 2, \ldots \text{ and } t \approx t_1^m \tag{11}$$

Equations (7) through (11), along with the assumption of a continuous transition of the system itself from a state of instability to a state of stability, impose certain conditions for the assumed value of the control parameter $T(t)$ which governs the described dynamic. The system will be in one condition or the other as a function of the value assumed by $T(t)$. First of all, it is possible to demonstrate that, to each oscillation of the system's limit cycle, of the type $s_i(t+\vartheta) = s_i(t)$, with ϑ being the period of the cycle, there corresponds a periodic behaviour of $T(t)$, that is, with the same period

$$T(t+\vartheta) = T(t) \tag{12}$$

The general solution of the learning dynamic (5b) is of the following type:

$$W_{ij}(t) = W_{ij}(t_0^m)e^{-\gamma t} + e^{-\gamma t}\int_{t_0^m}^{t} e^{\gamma x}\int_{x-T(x)}^{x} s_i(u)s_j(u)\,du\,dx = W_{ij}(t_0^m)e^{-\gamma t} + e^{-\gamma t}\int_{t_0^m}^{t} e^{\gamma x}Q_{ij}(x)\,dx \qquad \forall \; i, j \text{ and}$$
$$t_0^m \leq t \leq t_1^m \tag{13}$$

still, in the transition phase, when $t \approx t_1^m$, the mutual relationships between the learning dynamic correlation integral and synaptic energy, are defined by equation (9), thus, in the hypothesis that $-\gamma W_{ij}(t) > 0$, we can also write that:

$$Q_{ij}(t) \approx -\gamma e^{-\gamma t}\left[W_{ij}(t_0^m) + \int_{t_0^m}^{t} e^{\gamma x}Q_{ij}(x)\,dx\right] \qquad \forall \; i, j \text{ and } t \approx t_1^m \tag{14}$$

and this equation is a particular form of the first species equation of Volterra, integrable by reducing it into a differential equation which, for $t_0^m \le t$ furnishes the information that

$$Q_{ij}(t) \approx -\gamma e^{-2\gamma t} W_{ij}(t_0^m) \qquad \forall\ i, j \text{ and } t \approx t_1^m \tag{15}$$

In a totally analogous manner, let us consider the system in the phase in which a stable behaviour is found. During phase such as this, the relationship found in (7) holds true, thus, maintaining the algebraic sign and reconsidering the reasoning that brought about the determination of (15), we conclude that the system's state of stability implies that:

$$-\gamma e^{-2\gamma t} W_{ij}(t_0^m) << Q_{ij}(t) \qquad \forall\ i, j \text{ and } t \ge t_1^m \tag{16}$$

and thus, due to (6), we obtain another general condition that must be satisfied by $T(t)$ so that the system can find itself in a state of stability, and that is:

$$-\gamma e^{-2\gamma t} W_{ij}(t_0^m) << T(t) \qquad \forall\ i, j \text{ and } t \ge t_1^m \tag{17}$$

Obviously, the algebraic sign in (16) inverts, in systems that don't satisfy the conditions for stability imposed by the model, meaning in the time interval in which (8) is satisfied. Let us suppose that the second member of (17) is positive: (15) and (16) suggest that we seek out an integral of correlation (and thus, a function $T(t)$) of the following type:

$$Q_{ij}(t) \approx -\gamma W_{ij}(t_0^m) e^{-\gamma t} f(t) \tag{18}$$

for which $f(t)$ is a function which satisfies the condition, $f(t) > e^{-\gamma t}$ when $t \ge t_1^m$. Equation (18) is, indeed, a general form for $Q_{ij}(t)$ (and therefore, for $T(t)$). At this point we observe that, if the system is in a stable state, (7) will be valid, and so the learning dynamic (5b) can be rewritten as:

$$\frac{dW_{ij}(t)}{dt} = \int_{t-T(t)}^{t} s_i(x) s_j(x) dx = Q_{ij}(t) \qquad \forall\ i, j \text{ and } t \ge t_1^m$$

whose immediate solution is:

$$W_{ij}(t) = W_{ij}(t_1^m) + \int_{t_1^m}^{t} \int_{x-T(x)}^{x} s_i(p) s_j(p) dp dx = W_{ij}(t_1^m) + \int_{t_1^m}^{t} Q_{ij}(x) dx \qquad \forall\ i, j \text{ and } t \ge t_1^m \tag{19}$$

but the general solution to the learning dynamic is given by (13). Therefore, for the solution sought for $Q_{ij}(t)$ and, thus, $T(t)$ to be congruent with the two expressions (19) and (13), which need to be valid simultaneously for $t \ge t_1^m$, due to the continuity with which the system passes from one phase to the next, the following condition must be satisfied:

$$W_{ij}(t_0^m) e^{-\gamma t} + e^{-\gamma t} \int_{t_0^m}^{t} e^{\gamma x} Q_{ij}(x) dx = W_{ij}(t_1^m) + \int_{t_1^m}^{t} Q_{ij}(x) dx \qquad \forall\ i, j \text{ and } t \ge t_1^m \tag{20}$$

Equation (20) represents the true key for the problem, has been the most studied object during the period of formation of the theory, and it can be transformed, through appropriate steps, into a form that is compatible with Riccati's equation. It provides a functional relationship for $Q_{ij}(t)$, that is, the learning dynamic correlation integral. The trivial solution is obtained for $Q_{ij}(t) = 0$ assuming, at the same time, $W_{ij}(t_1^m) = W_{ij}(t_0^m) e^{-\gamma t}$, which is reasonable, since it derives from the solution for the synaptic connection during the chaotic phase, that is, for $t_0^m \le t \le t_1^m$. A more general, analytic solution can be obtained using (18) in (20), and by means of suitable manipulations we obtain an explicit form for $Q_{ij}(t)$, of the following type:

$$Q_{ij}(t) = \frac{-\gamma W_{ij}(t_1^m) e^{\frac{\gamma(t_1^m - t)}{2}} \left[\sqrt{3}\cos\left(\frac{\sqrt{3}}{2}\gamma t\right) + \sin\left(\frac{\sqrt{3}}{2}\gamma t\right) \right]}{\sqrt{3}\cos\left(\frac{\sqrt{3}}{2}\gamma t_1^m\right) - \sin\left(\frac{\sqrt{3}}{2}\gamma t_1^m\right)} \qquad \forall\ i, j \text{ and } t \geq t_1^m \tag{21}$$

and thus, finally, in the diagonal case,

$$T(t) = \frac{-\gamma W_{ii}(t_1^m) e^{\frac{\gamma(t_1^m - t)}{2}} \left[\sqrt{3}\cos\left(\frac{\sqrt{3}}{2}\gamma t\right) + \sin\left(\frac{\sqrt{3}}{2}\gamma t\right) \right]}{\sqrt{3}\cos\left(\frac{\sqrt{3}}{2}\gamma t_1^m\right) - \sin\left(\frac{\sqrt{3}}{2}\gamma t_1^m\right)} \qquad \text{for } t \geq t_1^m \tag{22}$$

From (22) we can note that the solution found depends on the diagonal element of the matrix of the correlations; thus, in general, there exist *N* different *T*(*t*), which diversify among themselves by the amplitude. This consideration does not seem to limit the validity of the sought solution since, with the goal of an effective functioning of the proposed model, the size of the *N* different elements of correlation $W_{ii}(t_1^m)$ is substantially uniform, so that, as a reference value of *T*(*t*), it is reasonable to consider that which, as a constant multiplicative factor, the average carried out on the diagonal elements, that is, $\frac{1}{N}\sum_{i=1}^{N} W_{ii}(t_1^m)$. The solution found for *T*(*t*) is undoubtedly an approximation of a complex process; in any case, it appears evident that a reasonable solution for *T*(*t*) can be of the following type:

$$T(t) \approx C_0 \sum_j \left[a_j e^{\varphi_j - \gamma t}\cos(\gamma t + \varphi_j) + b_j e^{\psi_j - \gamma t}\sin(\gamma t + \psi_j) \right] \qquad \text{for } t \geq t_1^m \tag{23}$$

where C_0 is a constant and $a_j, b_j, \varphi_j, \psi_j$ are parameters of the solution that can be determined by substituting (23) into (20).

There are at least three noteworthy considerations to be made, regarding the developed theory:

A. Considering time as a parameter of the model, not all time values are allowed for the stabilization of the dynamic of the model, and, in this sense, we can introduce the concept of *time quantization*.
B. Analysis of the model's dynamic stability clearly shows the Hebb's model is a *static approximation* of a learning dynamic deducible as a consequence of the present theory.
C. The *noise* which the system is capable of tolerating and which, at the same time, does not interfere with its capacity in terms of stabilization and recognition, is equal to approximately 50%.

Proposition A derives from the fact that, from the simulations, we observe that the system passes from a chaotic behaviour for $t_0^m \leq t \leq t_1^m$ to a stable one for $t \geq t_1^m$, while the transition phase takes place around t_1^m. It is precisely for this time value that the relationship between the two quantities of the second member of the learning dynamic is governed by (9), from which, if we let $i = j$, we obtain $T(t_1^m) \approx -\gamma W_{ii}(t_1^m)$ and writing (22) for t_1^m (which is possible if arriving at the limit value t_1^m from the right) we can deduce that such a condition will be satisfied if and only if:

$$t_1^m = \frac{2}{\sqrt{3}} \frac{l\pi}{\gamma} \qquad l = 0, 1, 2, \ldots \tag{24}$$

In (24) if $l = 0$, then $t_1^m = 0$, that is, the system is stable at the origin of times, which coincides with the time in which the input is presented to the system. But the notable consequence of (24) consists of the fact that only *some* time values are allowed for the stabilization of the system. In substance, the passage from the noisy-chaotic state to the stable state happens for:

$$\delta = t_1^m - t_0^m \approx \frac{K}{\gamma} \tag{25}$$

that is, after a finite time interval (K is a constant). This expression induces an interesting observation: one would expect the stabilization time to vary along with the variations of the considered pair i, j. For this, this time should be the maximum from among as many values as there are possible pairs of i, j for which (7) can be verified. Instead, due to the particular form singled out for $T(t)$, we observe that the stabilization time is independent of the specific i, j pair considered. Moreover, in varying the presented pattern, one would expect a $T(t)$ dependent on the particular pattern presented. Instead, (22) varies only in the amplitude upon variation of the input pattern, and not in the form of the wave. For this reason, neither (25) varies upon the variation of the input pattern, and δ assumes an invariant form *for every* pair i, j and *for every* input pattern presented to the system. A more detailed analysis shows that the system's stability can be had for the instances of time belonging to the set Θ defined as:

$$\Theta : \left\{ t \mid t_1^m \le t < \frac{10\sqrt{3}}{9}\frac{\pi}{\gamma}; \frac{4\sqrt{3}}{3}\left(l - \frac{2}{3}\right)\frac{\pi}{\gamma} < t < \frac{4\sqrt{3}}{3}\left(l - \frac{1}{6}\right)\frac{\pi}{\gamma}, l = 1,2,\ldots \right\}$$

Proposition B derives from an analysis on dynamic stability (Amit, 1989) defined as $h_i(t)s_i(t) > 0$ in which $h_i(t) = \sum_{j=1}^{N} W_{ij}(t)s_j(t)$, applied to the present model. With suitable steps, we arrive at the conclusion that, in a stable state in the sense described by (7) and (10), it follows that:

$$W_{ij}(t_0^m) \approx C_{ij}\xi_i^m \xi_j^m \tag{26}$$

that is, Hebb's rule, which maintains its validity only in approximations in which the system is not subject to a dynamic evolution.

Finally, proposition C derives from the fact that, in a stable phase (10) must be satisfied and thus the activation dynamic (5c) assumes the form

$$s_i(t) = g\left\{ \sum_{j=1}^{N} W_{ij}(t)s_j(t) \right\} \quad \forall \; i = 1, 2, \ldots, N \text{ and } t \ge t_1^m$$

and it is shown that this last equation can be rewritten as

$$s_i(t) = g\left\{ \sum_{j=1}^{N} r(t)C_{ij}\xi_i^m \xi_j^m s_j(t) \right\} \qquad \forall \; i = 1, 2, \ldots, N \text{ and } t \ge t_1^m \tag{27}$$

where $r(t)$ is a damped oscillator. Introducing noise into a system means forcing some components of input vector $\bar{\xi}^m$ to assume values that are different than the original ones; or rather, it is the equivalent of carrying out $\xi_i^m \to -\xi_i^m$ type operations for some values of i. This brings about, generally, an inversion of the sign in $W_{ij}(t_0^m)$, as can be seen in (26), or, in other words, the operation $W_{ij}(t_0^m) \to -W_{ij}(t_0^m)$, is carried out q times (where $q \le N$). It is evident, from the analysis of (27) for ferromagnetic paramagnetic, and mixed cases, that the value assumed by $s_i(t)$ will, at stability, be conserved, even if $0.5N - 1$ terms in the sum of the argument of the sigmoid function should undergo an operation of inversion following the introduction of noise. And therefore, the system is capable of recognizing the presented pattern, even if the noisy components number a small amount more than 50%. Definitively, even if that percentage of input components were corrupted, the system would be capable of bring itself back to a stable condition.

3. TIME PARAMETER AND TIME SERIES

A time series of the type $\left(x(t_1), x(t_2),\ldots,x(t_k),\ldots\right)$ is a particular form of in input vector for the model exposed in this paper. For these inputs t_k represents a discrete instance of time. Just as $T(t)$ is able to describe the properties of a model, defined by (5b) and (5c), with the goal of arriving at a state of equilibrium, with all that this achieves, the application of $T(t)$ to time series has, in recent years, shown itself to be extremely efficient as well. Here follows a description of the utilization of the time characteristic parameter with reference to the forecasting of the trends in these series.

The algorithm, in its basic structure, is extremely simple, and can be articulated in four reiterated steps:

1. Reduction to the series of first differences for which the k-th element is given by $y_k = x(t_k) - x(t_{k-1})$.
2. Determination of a value, V, for which:
   a. if $y_k < -V \Rightarrow z_k = c_1$
   b. if $-V \le y_k \le V \Rightarrow z_k = c_2$
   c. if $y_k > V \Rightarrow z_k = c_3$

   The clusterization into three groups, of the series of the differences, is purely indicative in the goal of comprehending the model; in fact, a larger number of clusters allows for a more accurate subdivision of the analysed series. The operation described in this passage is carried out on a temporal window Δt, for which the three groups, initially, result in being equally distributed; this, to maintain the initial condition of equiprobability. This implies a tight correlation between V and Δt, which need to be determined opportunely. The series $\boldsymbol{z} = \left(z_2, z_3,\ldots,z_k\right)$ resulting from the present passage, is a classic step function series: the verification of a particular cluster in the quantized time, defines a new step in the graphic.
3. Association for each cluster h calculated in the previous passage (in which h =3) of a time parameter T_h as defined by (23). The coupling element between cluster h and the respective T_h is given by the value of time, t, which appears in the arguments of the exponentials and trigonometric functions.
4. Determination of a set of values $a_j, b_j, \varphi_j, \psi_j$ in (23), unique for all the h clusters, so that it supplies values T_h which, once put in order, represent the probability with which the clusters at the next time k + 1 of the considered series $\boldsymbol{z}$ will be verified. In other words, for every cluster, there is calculated a characteristic, specific time parameter for that cluster. The single value T_h thus obtained is related to the probability that cluster h has of representing itself at the system's next time k + 1.

As is quite obvious, the key to determining time parameters able to define the probability of verification of the different clusters at the next times, is contained in the capacity of termination of the coefficients of (23). In order to obtain satisfactory results, it is necessary to reiterate steps 1 through 4 of the described algorithm, on portions of the series in which the result to be produced, that is, the cluster of arrival, and that is the order of the T_h (or of the associated function), ia *a priori* known. Successively, it is sufficient to use the same parameters to evaluate the probability of future clusters. Observing the use of time parameters in this study, a certain similarity can be noted with the concept of probability associated with the wave function $\psi(q, t)$ in Quantum Mechanics, in which the squared module represents the probability of finding a determined particle in a spatial interval $(q, q + dq)$. In the case considered here, the single cluster takes the place of the particle and instead of considering the squared module of T_h, its inverse is considered; from this, we derive the probability of finding, in the next time interval, the determined cluster. Regarding this, we observe that the inverse of the time parameter has the same nature as a frequency, that is, a local frequency valid for the observation interval Δt, as if the probability p_h derived from the frequency of the observations of the clusters, were adequately represented by a function of the following type:

$$\phi_h = \frac{1}{T_h} \approx p_h = \frac{v_h}{N(\Delta t)}$$

with v_h being the number of occurrences for cluster h, and $N(\Delta t)$ the number of observations carried out in the considered window Δt, defined at step 2 as a portion of the analysed series.

4. FORECASTING ABILITIES OF THE TIME PARAMETER: SOME APPLICATIONS

The use of the time parameter as expressed in (23) and the developed algorithm described in the previous section, has been continuing for over ten years. Equation (23) and the forecasting algorithm were used to analyze hundreds of time series, of extremely diverse kinds, with the goal of confirming the surprising predictive capabilities and the synthesized results in the present work. The basic reason for such a long waiting period before divulging the results is found in the fact that the work group has always maintained a great prudence in ascertaining the forecasting performances, so as not to commit the error of attributing properties to the utilized algorithm that eventually might turn out untrue or unbelievable. Now, though, these performances, after years of study and analysis, as well as certification by external certification firms that analysed the time series in the financial world, give us convinced confidence with regard to the strength and forecasting capabilities of the developed model.

In the present work, by way of example, we present the results of three practical forecasting applications, concerning:

A. The futures contract of the Euro vs. USD, sampled every ten minutes from 2nd June 2006 to 30th March 2010 (figures 1 - 3).
B. The concentration of PM10 (*particulate matter*, with a diameter of ≤ 10 μm), air pollution recorded in Rome and sampled hourly from 3rd July 1999 to 14 July 2000 (figures 4 and 5).
C. The extractions at the roulette table number 4 in the Wiesbaden, Germany, casino, recorded from 30th November 2009 to 31st December 2009 (figure 6).

For all three of the above examples, the algorithm described in step 1 – 4 of the preceding paragraph was applied exactly as already described. According to the logical scheme, the time series analysed is the series of the first differences, (more precisely the *backward differences*), of the type $\nabla_j[x](t) = x(t) - x(t-j)$ where usually, but not necessarily j = 1, and the forecast is carried out for step t + 1. In the approach here adopted, the values of series $\boldsymbol{z}$ are effectively regrouped into three clusters c_h with h = 1, 2 or 3. The re-transformation to the initial series of the forecast carried out on the clusters, implies the subdivision of future space into three possible regions: higher, if $x(t+1) > x(t) + V$; lower, if $x(t+1) < x(t) - V$ and central, if $x(t) - V \leq x(t+1) \leq x(t) + V$. In these expressions V is the parameter calculated in Δt following the logic expressed in step 2. The model assigns the probability that the forecasted value should fall in a subset (two) of these regions. In substance, it is like *playing* with a three-faced die; the probability, *a priori*, that each side of the die has of winning in the next throw is equal to 1/3; thus, *betting* on two sides for each throw the total probability of winning is equal to 2/3 and, after about a hundred throws, one would expect to *guess*, correctly, more or less 66 times total. The use of the algorithm described here, instead, shows that the number of times the correct region is guessed correctly is around, or higher than, 80%.

In example A the total values of the series and the relative forecasts carried out are 123,524; of these forecasts, which were carried out according to the modalities described above, 109,185, or 88.39%, were correct. This kind of correctness (more or less 88%) were observed distributed uniformly for the year considered in the test. In example B, the data set is made up of

3456 cases and there were 2926 correct predictions, that is, 84.66%. It should be noted that even on the basis of the work to carry out the present model, it was possible to create a system for forecasting atmospheric pollution for the city of Rome. In example C, the data set is composed of 9871 cases and the correct predictions were 7880, or 79.8%.

These results were obtained without the application of any kind of filter for the manipulation of the data in the series. Little manipulations of original data bring about values which are usually higher for the correct forecasts, such as, for example, when applying the forecasting model not to the series of values but instead to the derivative of the series. In this latter hypothesis, the percentage of correctly predicted cases for example C was 88.29%.

The examples given here are not special cases, nor were they chosen in such a way as to showcase the good forecasting capabilities of the algorithm; they are, rather, completely typical and representative of our experiences carried out in these years.

5. CONCLUSIONS

The search for an explicit form for $T(t)$, valid for a specific observed physical phenomenon and after a specific time interval (limited by definition), has been the object of studies and research for over ten years; in this time, we have come to known that its determination, for the system under observation, is better able to describe the system itself. Until $T(t)$ has been determined, the hypothesis of *a priori* equiprobability cannot be valid either. The results achieved have allowed us to verify whether the form for $T(t)$, defined by partial observations, was effectively able to better forecast the analysed phenomena. As a result we have verified that the use of such a time parameter $T(t)$ as an element able to forecast, allow us to determine trend evolutions reaching higher levels of success with respect to the levels that classical statistics is able to furnish. In other words, given a system that evolves in time, observing that system in an adequate temporal window, calculating the characteristic time parameter for that phenomenon in that temporal window, grouping the variations of the analysed data into three clusters, we are able to predict the future time evolution of the same phenomenon with a precision that is correct, on average, around 80 times out of 100, contrasted with the 66 times established by classical theory. Another notable consequence of this theory is that it assumes time to be an intrinsic variable of the system: a variable which is not banally a parametric element, but implicitly written into the observed system, in such a way that the same ergodic hypothesis is, logically, not respected. It is as if the observed system were not able to free itself from time, with its representation in the phase space. But if this ergodic hypothesis is no longer valid, then neither will the hypothesis at the base of the ergodic hypothesis be valid, that is, the *a priori* equiprobability for the results of an experiment. In this way, it is reasonable to think that in a time series, a generic point considered at instance k will influence the value at the following instance, $k + 1$, in some way, and that it was in turn influenced by the preceding value, $k - 1$, in such a way that not all the possible values are permitted *a priori*, but only some, in a determined range of possibilities. In this sense, the non-assumption of the principle of equiprobability has made possible the realization of tools better able to predict the evolutions of a system.

Undoubtedly, to assume the equiprobability principle as marginal for describing the observed physical phenomenon, within limited timeframes, opens the way for profound reflections on the meaning of randomness itself, but it would be more correct to say, at this point, that it opens considerations on the meaning of order. These reflections could be readily left to philosophical and theological speculation.

6. REFERENCES

1. Boltzmann, L. (1964): Lectures on the Gas Theory, University of California, Berkeley.
2. Gibbs, J (1981): Elementary Principles in Statistical Mechanics, Ox Bow Press, (reprint).
3. Birkhoff, G. D. (1931): Proof of the Ergodic Theorem. Proceedings of the National Academic Society, U.S.A. v 17.
4. von Neumann, J. (1932): Proof of the Quasi-Ergodic Hypothesis. Proceedings of the National Academic Society, U.S.A. v 18.
5. Gallavotti, G. (1982): Aspetti della teoria ergodica qualitativa e statistica del moto. Pitagora, Bologna.
6. Buonomano, V. (1987): The Non-Ergodic Interpretation of Quantum Mechanics. In van der Merwe, A., Selleri, F., Tarozzi, G. (1987): Microphysical Reality and Quantum Formalism - volume 2. Kluwer Academic Publisher, Dordrecht.
7. Arecchi, T. (1990): I simboli e la realtà. In Selleri, F. (1990): Che cos'è la realtà. Jaca Book, Milano.
8. Ballarin, A., Gervasi, S., Cannatà, V. (1995): A Classification Neural Algorithm Using a Time Varying Learning Rule. Proceedings of the First International Conference on Neural Parallel and Scientific Computation, Atlanta, 1995.
9. Shinomoto, S. (1987): Memory Maintenance in Neural Networks. Journal of Physics A, v. 20.
10. Basti, G., Perrone, A., Ballarin, A., Cimagalli, V., Morgavi, G. (1991): Second-Order Statistics in a Non-Stationary Neural Network. Proceedings of the Fourth Italian Workshop on Parallel Architectures and Neural Networks. World Publishing, Singapore.
11. Amit, D. J. (1989): Modeling Brain Function. Cambridge University Press, Cambridge.

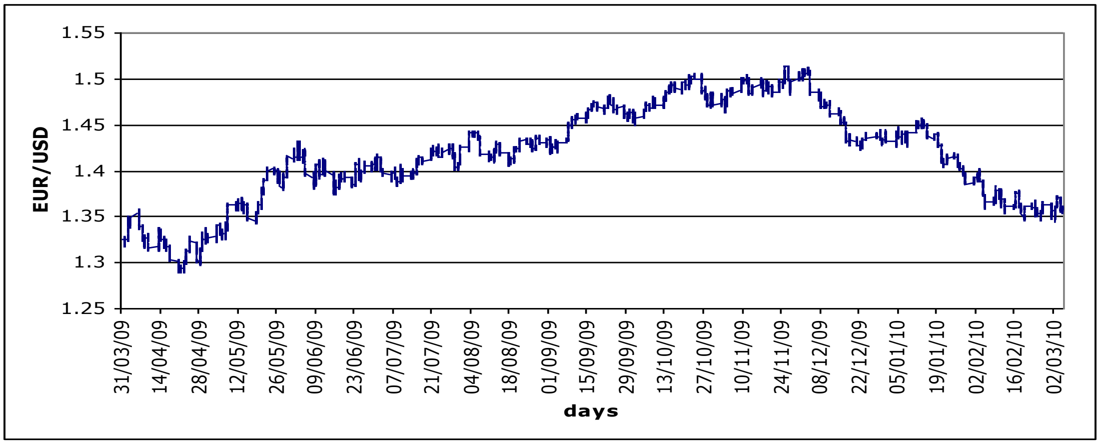

Figure 1. Sub-time series from example A representing the behaviour of the Euro vs. USD futures contract, sampled every ten minutes from March 3rd, 2009 to March 30th, 2010.

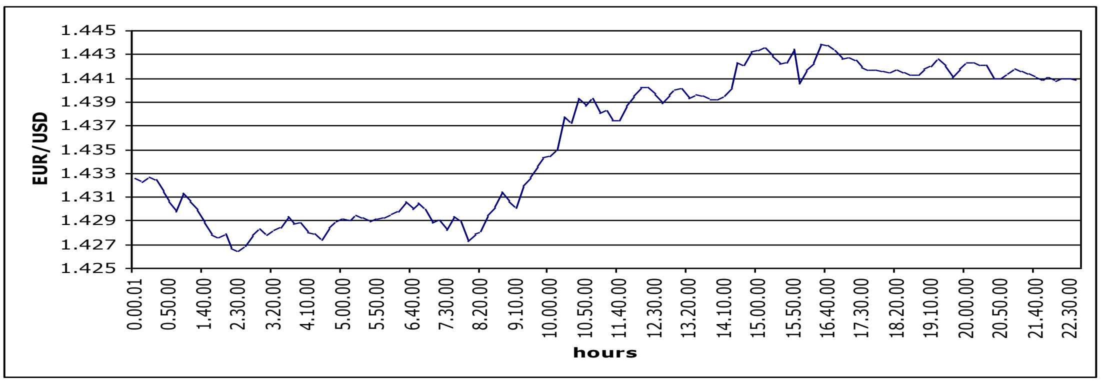

Figure 2. Sub-time series from example A representing the futures contract of the Euro vs. USD sampled every ten minutes on January 4th, 2010.

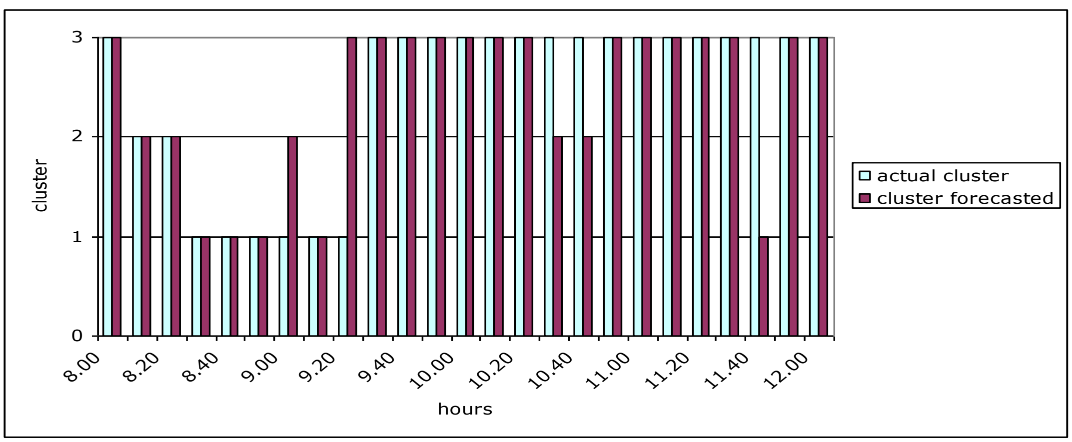

Figure 3. Values of the time series clustered as described in step 2 of the algorithm. The picture represents an example of the sequence of actual clusters in which the values are grouped as recorded from 8 am to 12 am just for one day, January 4th, 2010. The forecast developed by the algorithm is reported beside each cluster.

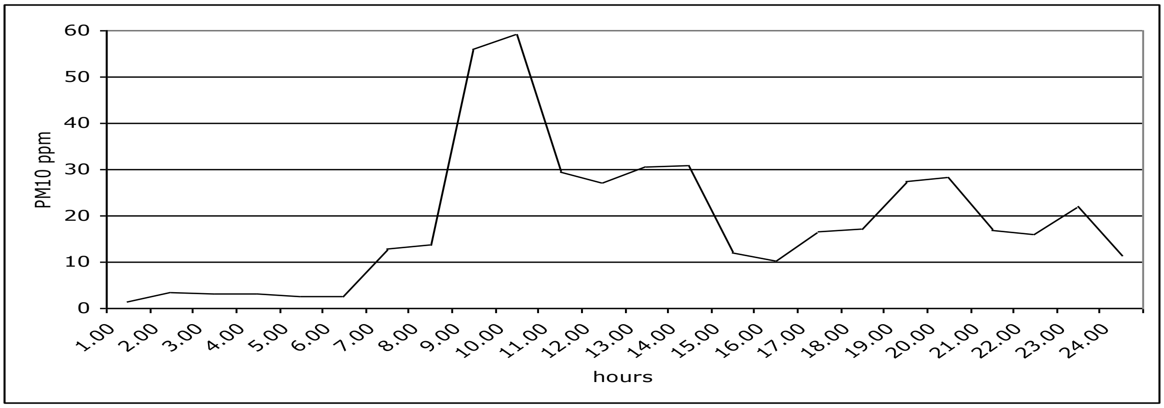

Figure 4. Sub-time series from example B describing the concentration of PM10 sampled every hour on July 13th, 2000.

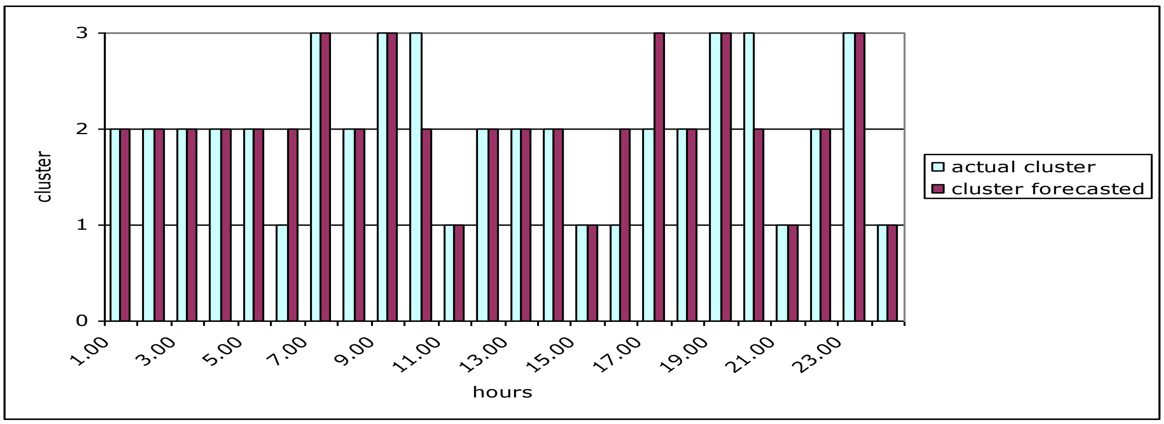

Figure 5. Example of the sequence of clusters in which the values of experiment B are grouped: this represents the sequence of actual clusters sampled every hour on July 13th, 2000 and, for each hourly cluster, the associated forecast.

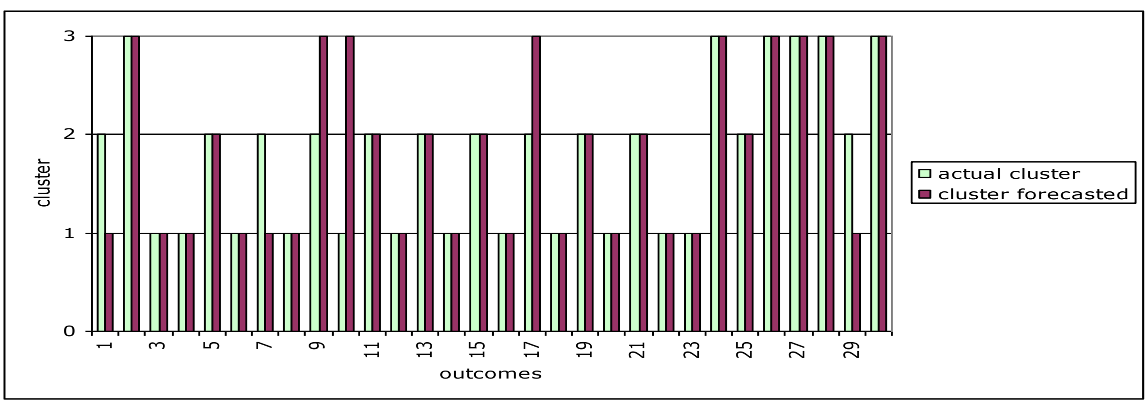

Figure 6. Representation of the last 30 roulette launches from example C. Outcomes are grouped into three sets, corresponding to the three vertical columns of the roulette table (for simplicity, 0 is considered to belong to column 2).