

# Methodology for sensor data forecast.

Michail Zak

Senior Research Scientist (Emeritus)

Jet Propulsion Laboratory California Institute of Technology  
Pasadena, CA 91109, USA.

## PART I. THEORY and ALGORITHMS.

### Section 1. Background and Overview.

One of the fundamental objectives of mathematical modeling is to **interpret** past and present, and, based upon this interpretation, to **predict** future. The use at time  $t$  of available observations from a time series to forecast its value at some future time  $t+l$  can provide basis for 1) model reconstruction, 2) model verification, 3) anomaly detection, 4) data monitoring, 5) adjustment of the underlying physical process. Forecast is usually needed over a period known as the lead time that is problem specific. For instance, the lead time can be associated with the period during which training data are available. The accuracy of the forecast may be expressed by calculating probability limits on either side of each forecast. These limits may be calculated for any convenient set of probabilities, for example, 50% and 90%. They are such that the realized value of the time series, when it eventually occurs, will be included within these limits with the stated probability, Fig.1.

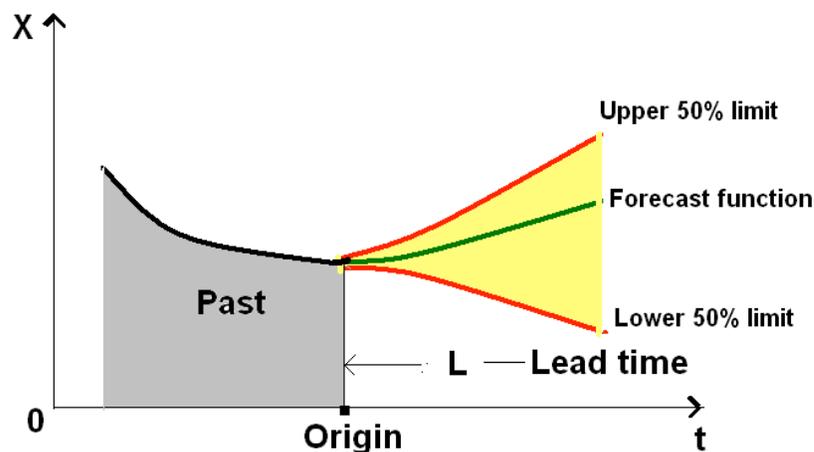


Figure 1. Forecast and probability limits.

When past and present are given in the form of observed or measured time series, both interpretation and prediction can be performed efficiently if one can **reconstruct** an underlying dynamical system that may **reproduce** the given time series. Indeed, in that case, all the abnormalities in the system behavior will be reflected in the corresponding changes in the parameters of the underlying dynamical system, while the prediction of future behavior can be obtained by running the dynamical system ahead or actual time.

In the simplest case when the time series are smooth, i.e. differentiable as many times as necessary, the analysis of abnormalities is trivial, while the prediction of future is reduced to an extrapolation, based upon the Taylor expansion, Fig.2,A. However, in case of **complex** structures, such as spacecrafts or engines, the situation is much more sophisticated. There are two important properties that prevent one to exploit the Taylor expansion and make the reconstruction of the underlying dynamical system more difficult. Firstly, these time series are not differentiable: even the first derivatives are not well defined, (Fig. 2, B). Secondly, they are non-deterministic: each observation represents a sample of an underlying stochastic process, and it can be reproduced only to accuracy of the probability density, (Fig.2, B).

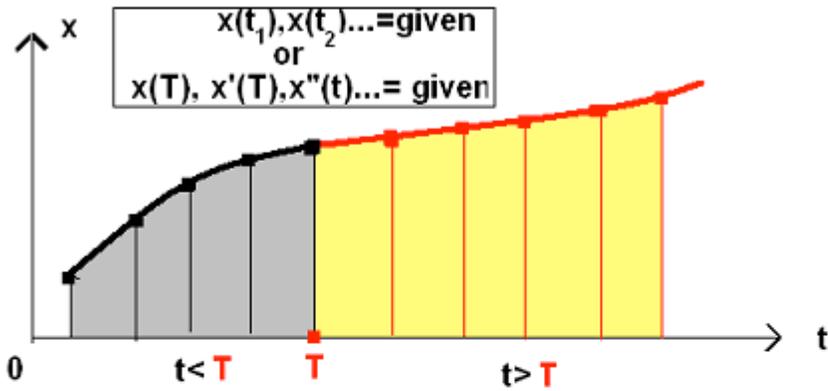
Development of the methodology for forecasting sensor data is the main **objective** of this report. The basic idea of the proposed methodology is the following: in order to reconstruct the model of a stochastic process, one has to have an ensemble of samples, unless the stochastic process is stationary, (see Fig.3). Since the ensembles are never available, and the stochastic processes of interest, in general, are not stationary, the approach should be based upon the decomposition of the given sample into stationary and non-stationary components. The latter can be divided into trends and oscillations. The trends are detected and removed by simple differencing of the sample: the first order differencing detects a linear trend, the second order – a parabolic trend, etc. The oscillations are detected by so called seasonal differencing.

The criterion of stationarity is associated with the area enveloped by the autocorrelation function: the smaller the area the closer the process to a stationary one. After the stationary component of the sought stochastic process is found, one can reconstruct its mathematical model. The model for the stationary component consists of the linear part (in the form of autoregressive Yule-Walker equations), the nonlinear part (in the form of feed-forward neural nets), and the error (i.e. the difference between the analytically reproduced data and the original data) that represents a stationary non-correlated stochastic process.

As soon as the model for the stationary component is found, one can reconstruct the model for the whole stochastic process by applying to this model the summation operators that are inverse to the differencing operators being applied earlier

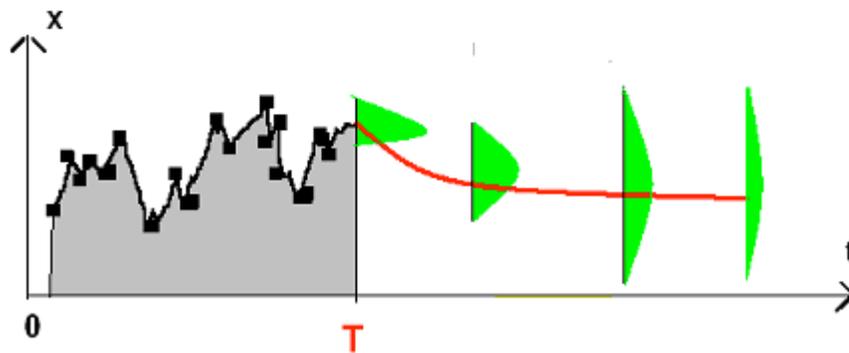
## Time series forecast

A. Deterministic case.

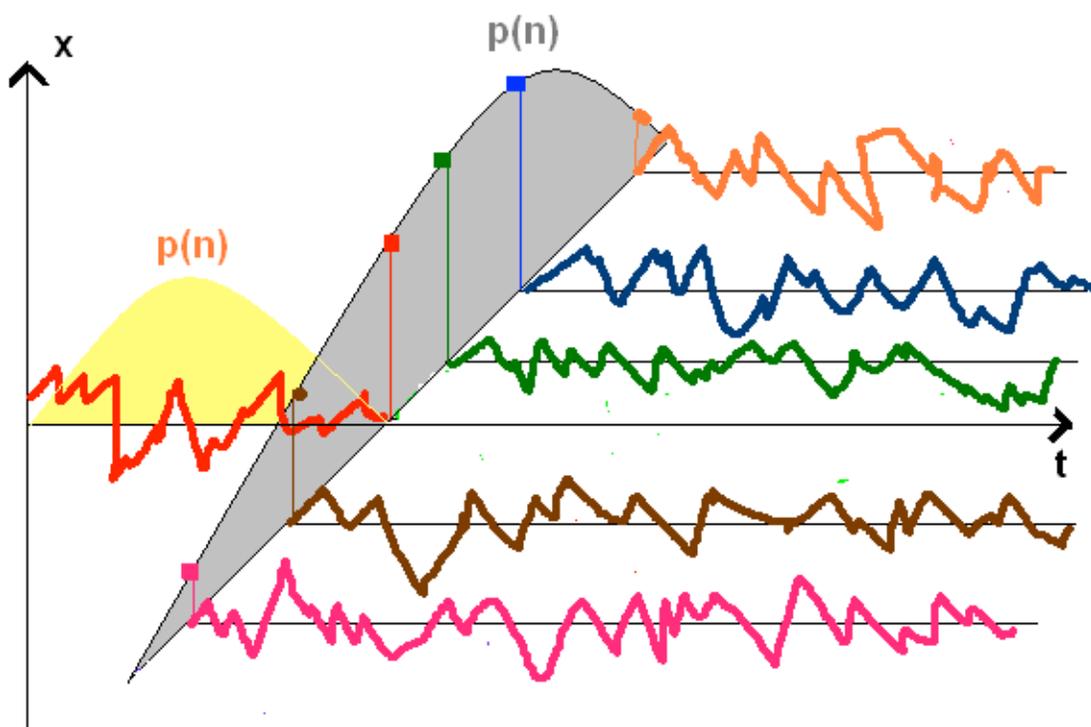


B. Stochastic case .

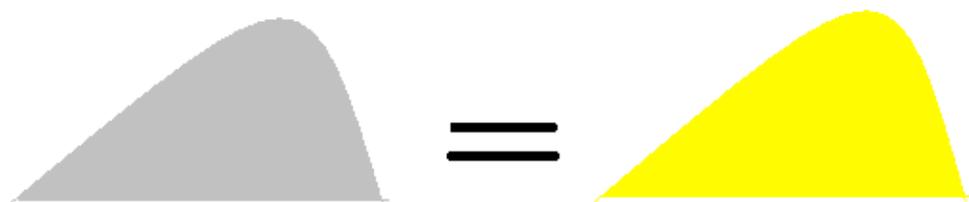
given:  $x(t_1), \dots, x(t_n)$ ; find:  $\langle x \rangle, \langle x \rangle, \dots$  for  $t > T$



**Figure 2. Background.**



For **STATIONARY** data the ergodic hypothesis holds

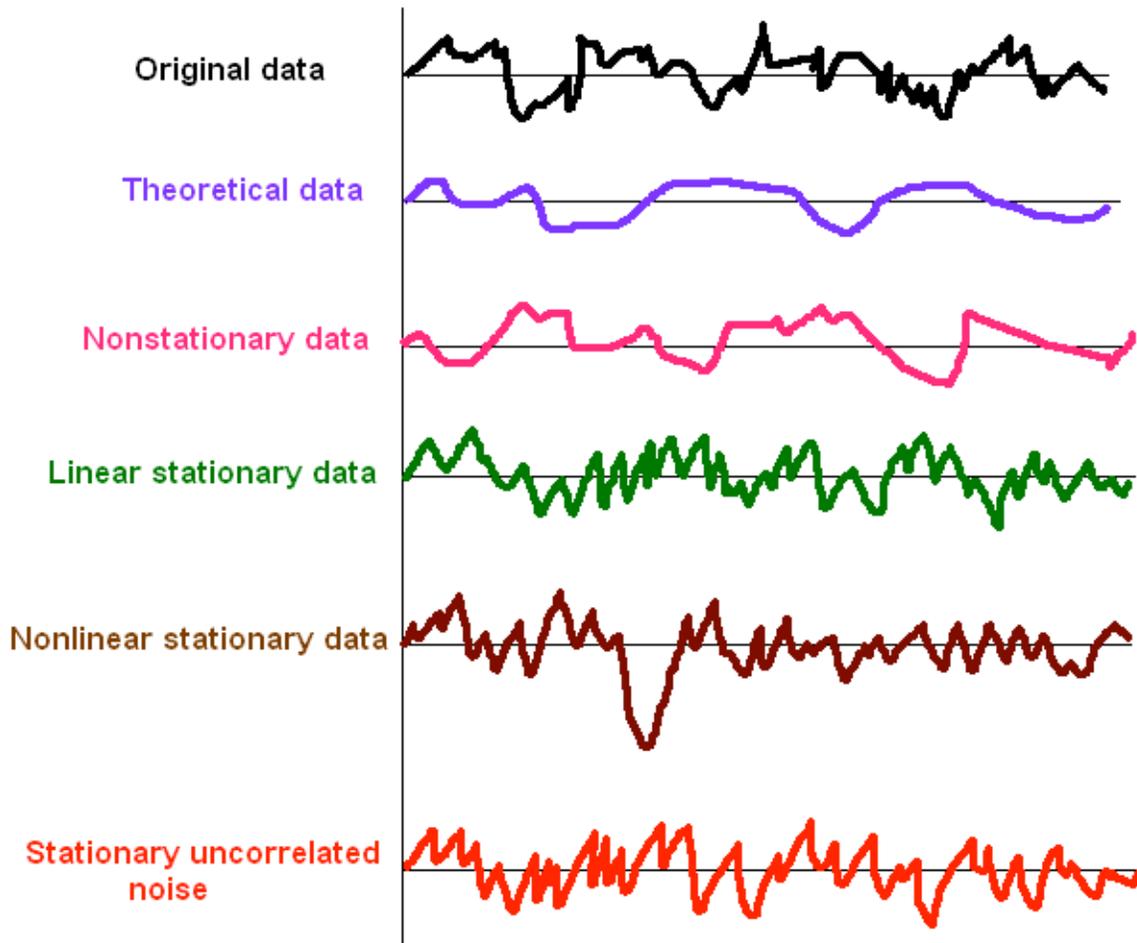


For **NONstationary** data



**Figure 3. Ergodicity**

The decomposition procedure is illustrated in Fig.4



$$O = T + NS + LS + NLS + N$$

**Figure 4. Data decomposition**

The model in a compressed form can be written as

$$X_{t+1} = F(X_t, \dots, X_{t-m}) + R_{t+1}, \quad (1)$$

Here  $X$  is the observed variable at times  $t, t-1, \dots, t-m$ ;  $F$  is the function that includes a linear component in the form of the autoregressive process, nonlinear component in the form of the feed-forward neural net; it also includes the result of the summation operators that incorporate the non-stationary components;  $R_t$  is an error of the representation of the original time series by the analytical expression in Eq. (1) for the time  $t$ .

## Section 2.

### 1. Origin of non-stationary stochastic processes.

In this section, for the purpose of illustration, we will discuss the types of stochastic processes (SP) that can be generated by **gravity field** since gravity forces represent the most important component of external forces in space, especially when the data are associated with the **spacecraft trajectory**. But first we recall some definitions. SP is stationary (SSP) if its probability density does not depend upon time, Fig.5,A). Otherwise it is non-stationary (NSP), Fig.5, B).

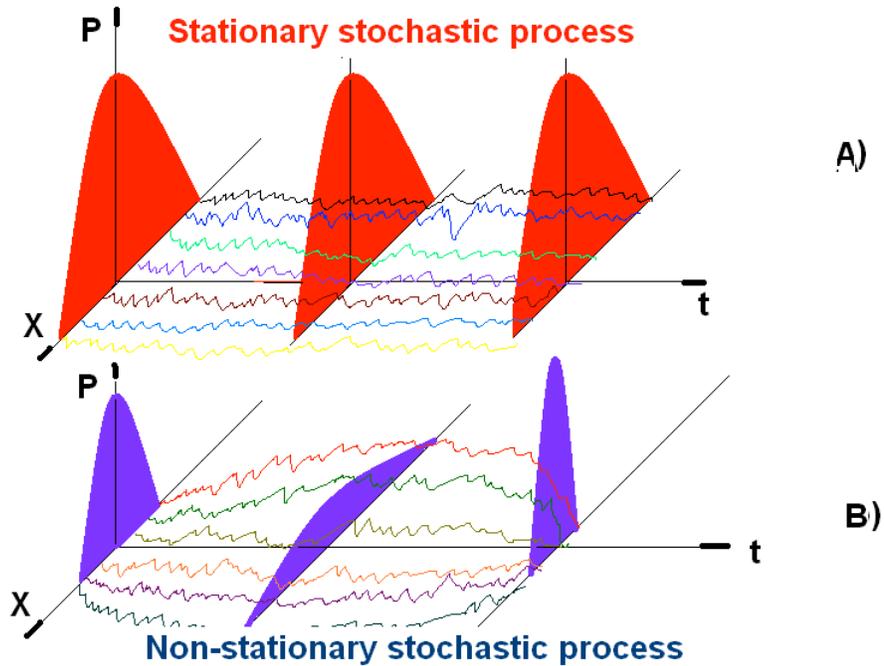


Figure 5.

However, this criterion can be useful only if the whole ensemble of samples is available. Based only upon one sample (as usually is the case), a simpler criterion can be applied. This criterion is based upon the behavior of the autocorrelation function. This function describes the strength of dependence between the values of the series at different times. It can be **estimated** using the following formula:

$$\rho(m) = \frac{1}{N-m} \sum_{n=1}^{N-m} X_n X_{n+m} \quad (2)$$

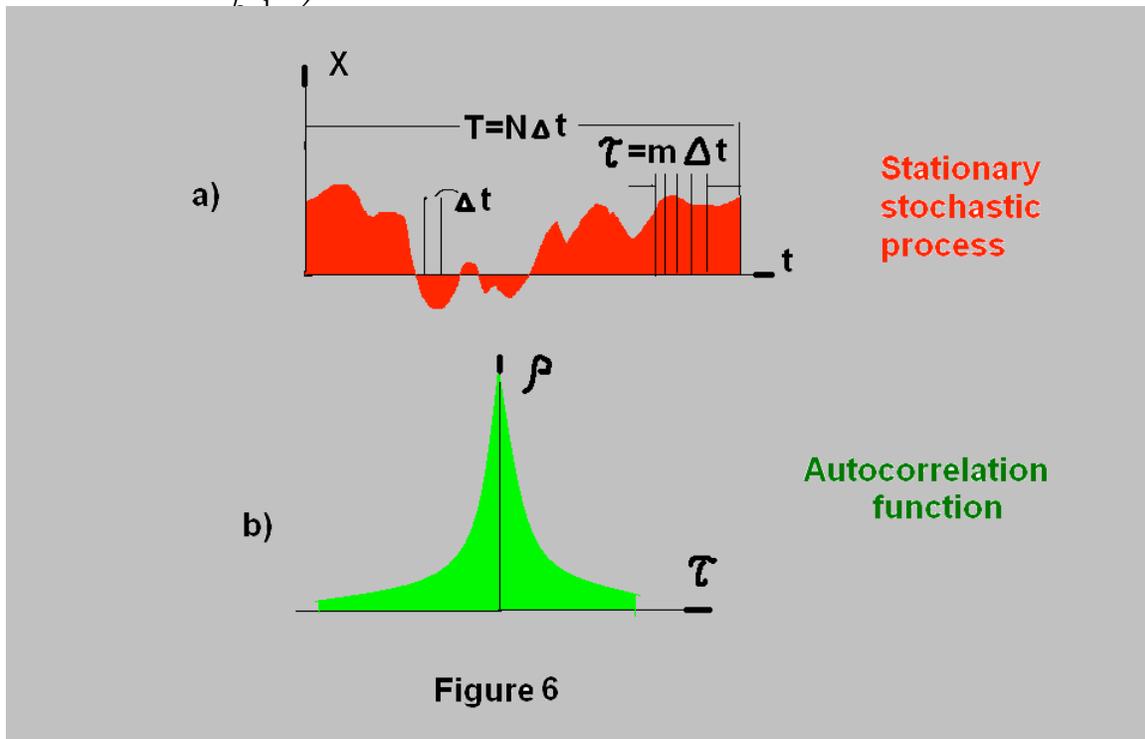
where  $N$  is the number of selected data separated by equal time intervals  $\Delta t = T/N$ , and  $T$  is the corresponding total time interval, Fig.6, A).

It has been proven that for stationary stochastic processes that the function (1) decays exponentially with increase of  $m$ , Fig.6,B). For instance, a typical form of this function for stationary processes is the following

$$\rho(\tau) = De^{-\alpha|\tau|}, \quad \tau = m\Delta t, \quad D, \alpha \text{ are const.}, \quad (3)$$

For non-stationary processes the decay is much slower than exponential, or no decay at all. For instance, a deterministic function decomposable in Fourier series has a non-decaying autocorrelation function

$$\rho(\tau) = A_0^2 + \sum_{k=1}^{\infty} \frac{A_k^2}{2} \cdot \cos \omega_k \tau \quad (4)$$



In general, it can be shown that most of typical non-stationarities in stochastic processes can be presented as a weighted combination of polynomial trends and oscillations.

The property of the autocorrelation function described above will be exploited for identification of stationarity of the stochastic processes as well as for separation of non-stationary components.

## 2. Contribution of gravitational fields.

In this section we will discuss specific contributions of the gravity field into positions of a satellite. The gravitational potential  $U$  must satisfy the Poisson equation

$$\nabla^2 U = -4\pi\rho(x, y, z, t), \quad (5)$$

where  $\rho$  is the mass density that depends upon the mass location (with respect to the satellite), and time  $t$  (if the masses are moving with respect to the satellite). The solution to this equation for the gravitational potential of the Geoid is

$$U = -\frac{GM}{r} + \frac{GMa^2 J_2}{2r^3} (3 \sin^2 \Phi - 1) \quad (6)$$

in which  $G$  is the universal gravitational constant,  $M$  the mass of the Earth,  $r$  the distance from the satellite to the Earth center,  $\Phi$  the latitude of the satellite,  $a$  the Earth radius, and  $J_2$  the difference between the biggest and the smallest of the moments of inertia of the Earth.

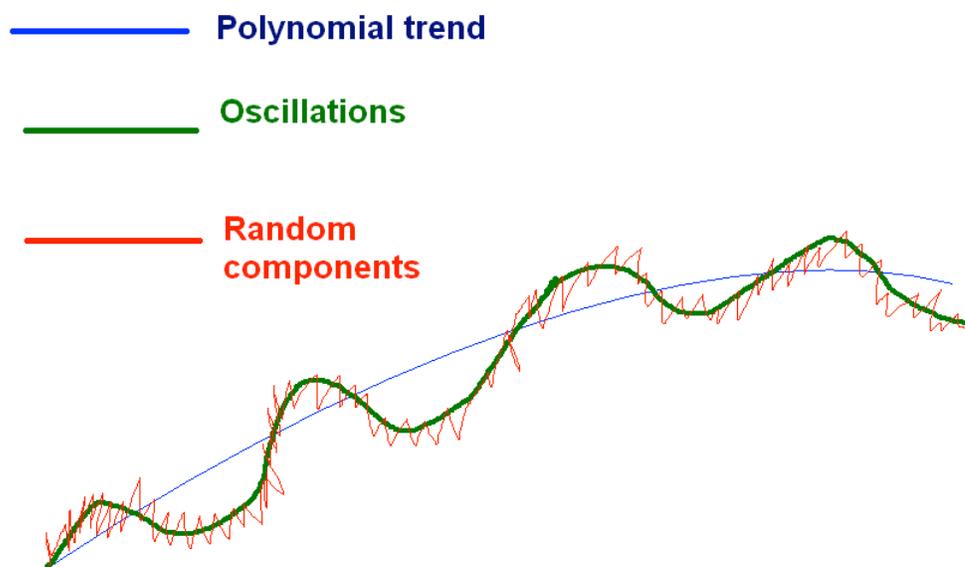
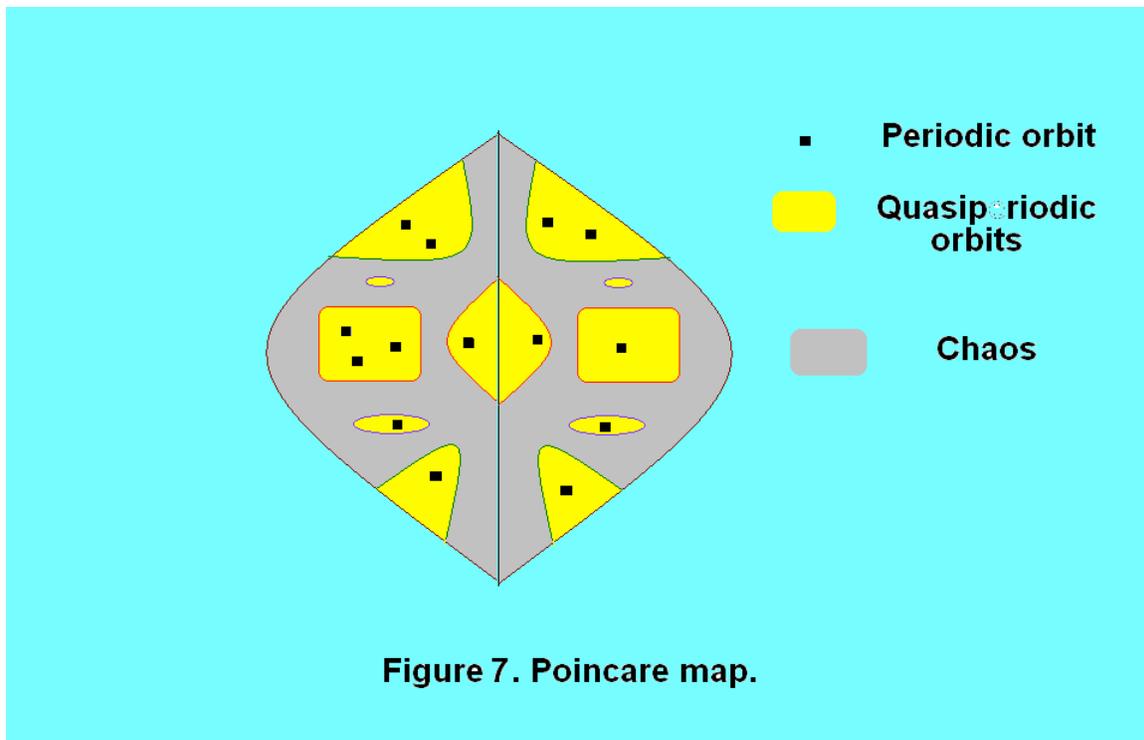
The force  $F$  generated by the potential (6) and applied to the satellite is

$$\bar{F} = \nabla U \quad (7)$$

As follows from Eqs. (6) and (7), both  $r$  and  $\Phi$  become functions of time when the satellite is moving along its orbit, and therefore, the gravitational force attains polynomial trends as well as periodic components. Similar components are incorporated into the gravitational force by other masses like sun, moon, asteroids, etc.

So far we have discussed the fully deterministic contributions of the gravity field into the sensor data describing the position or velocity of the satellite. Its random contributions are coming from a more sophisticated source. It has been proven that the  $n$ -bodies interacting via gravity attraction are **unstable** if  $n > 2$ . This instability has the same origin as those in turbulence that is supersensitivity to small changes in initial conditions. However, unlike a viscous fluid which is a dissipative system, and therefore, it has a chaotic attractors, the interacting bodies in a gravity field form a conservative system, and instead of chaotic attractors, they can have only a conservative chaos where deterministic motions jump into random area back and force. This picture is illustrated by a Poincare map in Fig.7. This map describes the Earth motion as the Earth cuts through the Jupiter-Sun plane in the rotating system. The map display a dot once each time the Earth passes Jupiter as both orbit the Sun. If the Earth comes back to the same relative position each time it crosses the Jupiter-Sun axis, it will always come out at the same point on the map. If the Earth comes back every second, or third, or  $n$ -th time around, several different points appear on the map. When  $n$  becomes large, these points confluent into curves, but the motion is still deterministic. Finally, there are fuzzy regions filled up densely by dots. Trajectories in these regions are sensitively dependent on their initial conditions, and the motion becomes chaotic. These chaotic motions generate random gravity forces that in turn, cause the stochastic components of the satellite motion.

As a result the satellite motion can be decomposed into the following three components: Deterministic polynomial trend, deterministic multiperiodic oscillations, and a stationary stochastic process, Fig.8.



### Section 3. Data decomposition

#### 1. Building a dynamical model.

Our approach to analysis of time series is based upon reconstruction an underlying dynamical model that can reproduce the whole stochastic process including the given time series as a sample of this process. This approach is based upon progress in three independent fields: nonlinear dynamics, theory of stochastic processes, and artificial neural networks.

From the field of nonlinear dynamics, based upon the Takens theorem, any dynamical system that converges to an attractor of a lower (than the original) dimensionality, can be simulated (with a prescribed accuracy) by a time-delay equation

$$x(t) = F[x(t - \tau), x(t - 2\tau), \dots, x(t - m\tau)] \quad (8)$$

in which  $x(t)$  represents a given time series, and  $\tau = \text{constant}$  is the time delay.

It was proven that the solution to Eq. (8) subject to appropriate initial conditions converges to the original time series

$$x(t) = x(t_1), x(t_2), \dots \text{etc} \quad (9)$$

when  $m$  in (8) is sufficiently large.

However, the function  $F$ , as well as the constants  $\tau$  and  $m$ , are not specified by this theorem. But the most serious limitation of the model (8) is that the original time series must be **stationary** (since it represents an attractor): this means that for non-stationary time series, the solution to (8) may not converge to (9) at all. Actually this limitation has deeper roots: it is linked to the problem of stability of the model (8), i.e. to sensitivity to errors in initial conditions, and that will be discussed below. In view of our **particular** problem when the origin of randomness is associated with chaotic instability of  $n$ -body problem, the Takens theorem is of special importance.

Much earlier than the Takens theorem has been introduced, mathematical statisticians have developed a different approach to the same problem: they approximated a **stationary** stochastic process by a linear autoregressive model

$$x(t) = \alpha_1 x(t-1) + \alpha_2 x(t-2) + \dots + \alpha_n x(t-n) + N, \quad n \rightarrow \infty \quad (10)$$

where  $\alpha_i$  are constants, and  $N$  represent the contribution of noise.

As shown by Wold, any zero-mean purely non-deterministic **stationary** process  $x(t)$  possesses a linear representation as in (10) with

$$\sum_{j=0}^{\infty} \alpha_j^2 < \infty \quad (11)$$

(the inequality (11) is equivalent to the condition of stationarity). These conditions are equivalent to the conditions of stability of solutions to Eq.(10), and they can be expressed via the characteristic roots  $G_i$ , i.e. the process is non-stationary if

$$|G_i| \geq 1 \quad (12)$$

The case  $|G_i| < 1$  is usually excluded from considerations since it corresponds to an exponential instability that is unrealistic in physical systems under observation. However,

the case  $|G_i| = 1$  is realistic. Real and conjugates  $G_i$  incorporate trend and oscillations, respectively, into the time series (2), Fig.9.

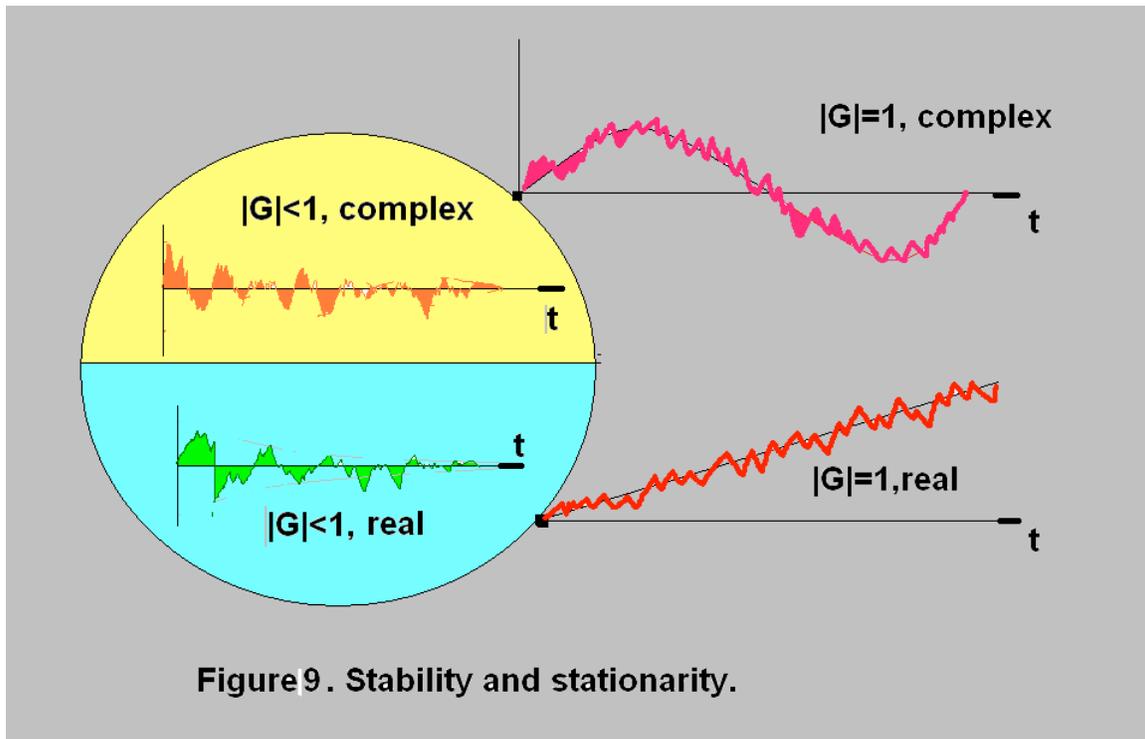
## 2. Separation of trends.

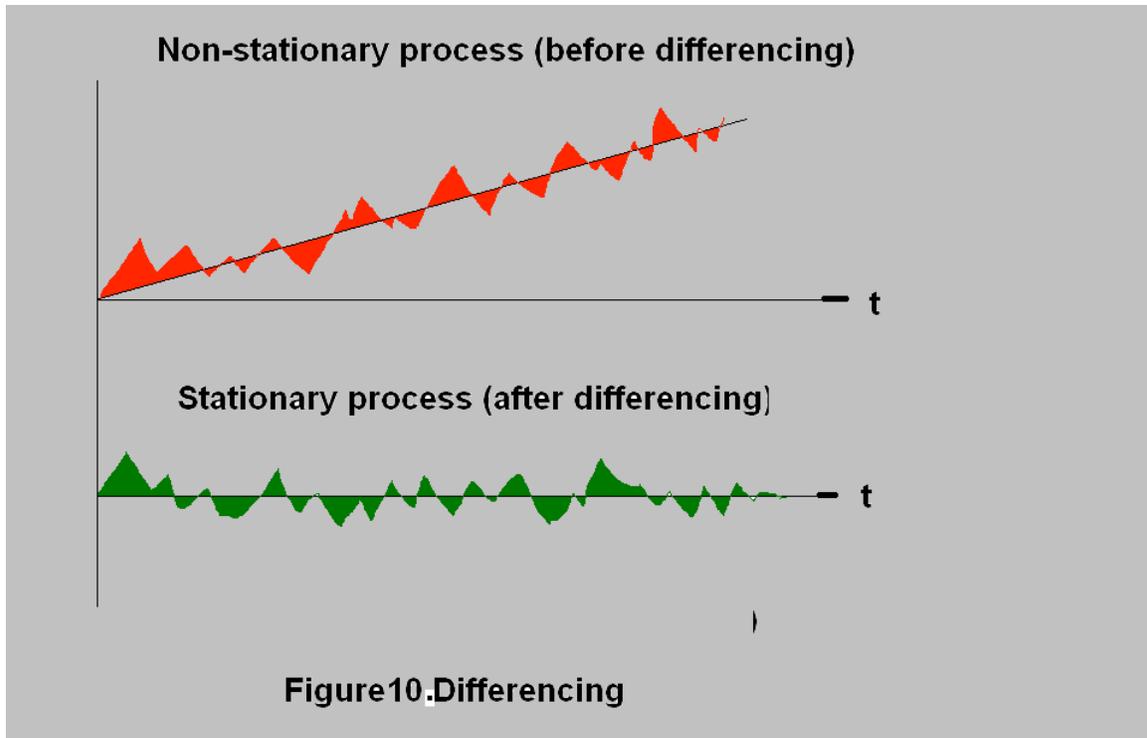
As mentioned above, the non-stationary components of time series are composed of polynomial trends and oscillations. The decomposition of time series in stationary and non-stationary components starts with removal of trends. The idea is very simple; to apply a difference operator to the original time series

$$\nabla x_t = x_t - x_{t-1} = (1 - B)x_t \quad (13)$$

in which B is the backward shift operator.

Obviously, the difference operator is a discrete analog of differentiation, and being applied once, it removes a linear trend; for removal polynomial trend of m-th order, the operator should be applied m times sequentially, Fig.10.





**Figure10.Differencing**

It should be noticed that the process plotted in Fig. 10 is highly idealized (for the purpose of better illustration); actually even a linear trend does not have to be deterministic: it can include random components, and then it cannot be so easily detected by just looking at the data. For the same reason, it is not known in advance how many times the operator (13) should be applied to the original time series. The most reliable approach to this problem is to check autocorrelation function after each differencing. As noticed in the report 2, the simplest condition of stationarity is an exponential decay of the autocorrelation function (see Fig 10, and Eqs. (2), (3)). Theoretically this property follows from the stability analysis described above. As a result, the time series

$$S = \sum_{l=1}^{\infty} \rho_l < \infty \quad (14)$$

must converge.

Here  $\rho_l$  is the autocorrelation between  $x_t$  and  $x_{t+l}$  :

$$\rho_l = \frac{1}{C_0 N} \sum_{t=1}^N (x_t - \mu)(x_{t+l} - \mu) \quad (15)$$

in which

$$\mu = \frac{1}{N} \sum_{t=1}^N x_t, \quad C_0 = \frac{1}{N} \sum_{t=1}^N (x_t - \mu)^2 \quad (16)$$

To enforce the stationarity of  $x(t)$ , one has to find the optimal number of differencing  $k$  that provides a minimum value of the sum (14), or the minimum of the area enveloped by the autocorrelation function, Fig.11.

As shown in Fig.11, the optimal number of differencing is  $k^*$ .

### 3. Separation of oscillations.

The removal of non-stationary components in the form of oscillations is based upon similar idea of differencing, but this differencing must be of special type: it is so called seasonal differencing implemented by the following operator

$$\nabla_s x_t = (1 - B^s)x_t = x_t - x_{t-s}, \quad (17)$$

in which  $s$  is the number of points corresponding to the period of oscillations.

Although the idea of the approach is simple, the problem is that the period  $s$  is not known in advance. However, in our particular case, since we are dealing with a physical system, these oscillations are deterministic, and they can be found from the Fourier expansion, while the sharp values in the spectra correspond to frequencies of the deterministic oscillations with the periods  $s', s'', s'''$ .

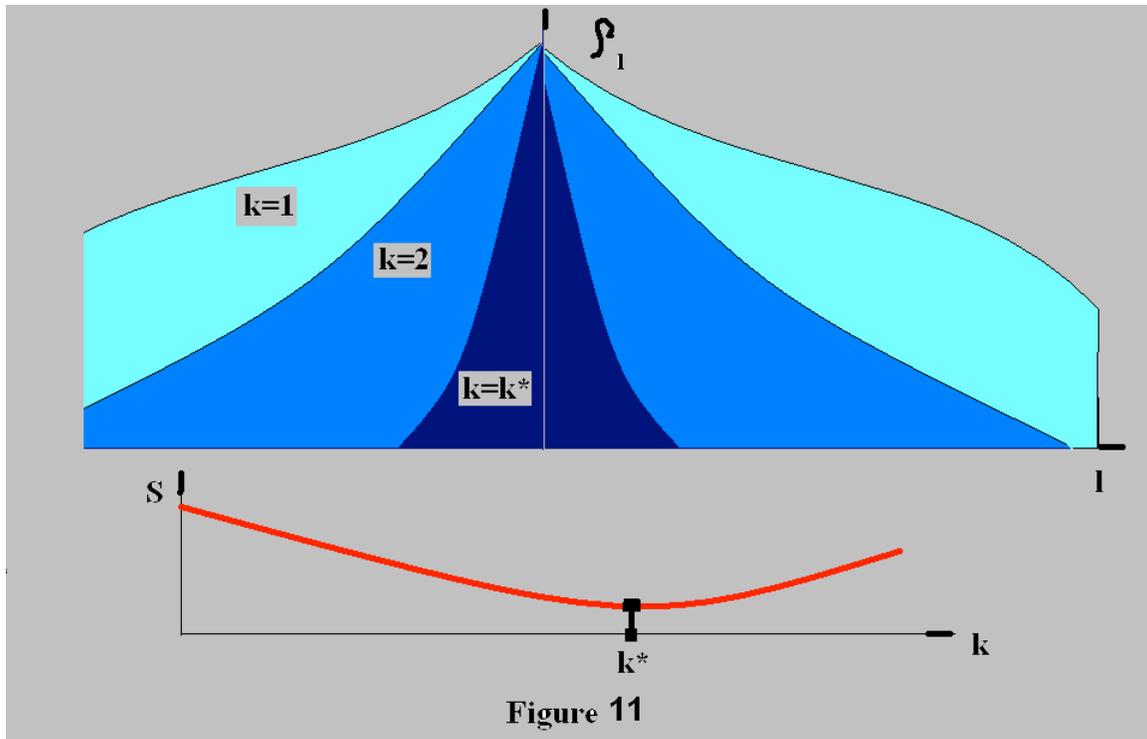


Figure 11

## Section 4. Model fitting

### 1. Linear approximation.

A first step in the reconstruction of the dynamical model after the stationary data  $x_i$  are extracted from the original data (see Section 3) is to derive a **linear** approximation in the following form

$$x'(t) = \alpha_1 x'(t-1) + \alpha_2 x'(t-2) + \dots + \alpha_n x'(t-n) + N, \quad n \rightarrow \infty \quad (18)$$

where the constant coefficients  $a_i$  are found from the Yule-Walker estimates

$$a = P_n^{-1} \rho_n, \quad a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, \quad \rho_n = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_n \end{bmatrix}, \quad P_n = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{n-1} \\ \rho_1 & 1 & \rho_2 & \dots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{n-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{n-1} & \rho_{n-2} & \dots & \rho_{n-3} & 1 \end{bmatrix} \quad (19)$$

Here  $\rho_1$  is the autocorrelation between  $x'_t$  and  $x'_{t+1}$ :

$$\rho_l = \frac{1}{C_0 N} \sum_{t=1}^N (x'_t - \mu)(x'_{t+l} - \mu) \quad (20)$$

in which

$$\mu = \frac{1}{N} \sum_{t=1}^N x'_t, \quad C_0 = \frac{1}{N} \sum_{t=1}^N (x'_t - \mu)^2 \quad (21)$$

The physical meaning of this representation is the following: if equation (18) is solved as a difference equation with respect to  $x'$ , then the solution will reproduce the stationary data extracted from the original data to accuracy of the last term  $N$  that represents uncorrelated noise, Fig 12.

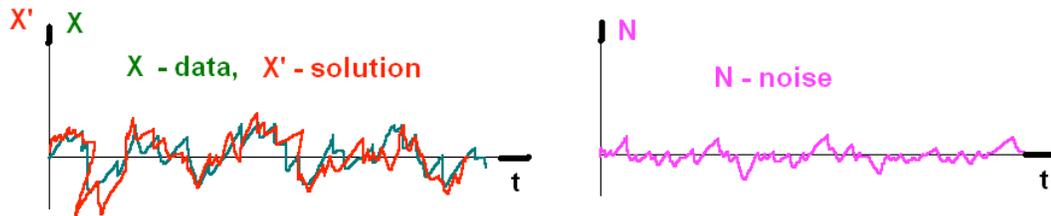


Figure12

## 2. Neural-net-based nonlinear approximation.

Our experience has shown that in some cases,  $N$  is not fully uncorrelated and is not sufficiently small. This means that the original model is **nonlinear**, and therefore, it cannot be approximated by the linear difference equation (18) with acceptable

accuracy. In this case, we will consider Eq.(18) as a first approximation, and introduce the following nonlinear residual data

$$x'' = x - x' \quad (22)$$

Here  $x$  is the stationary data extracted from the original data, and  $x'$  is the linear approximation to this data. Obviously,  $x''$  must be sought as a nonlinear function of the same arguments as those in Eq. (18), i.e. in the form

$$x''(t) = F[x''(t-1), x''(t-2), \dots, x''(t-m)] \quad (23)$$

We will parameterize this equation using the formalism of feed-forward neural networks

$$x''(t+1) = \sigma \left\{ \sum_j W_{ij} \sigma \left[ \sum_k w_{jk} x''(t-k) \right] \right\}, \quad k = 1, 2, \dots, p \quad (24)$$

where  $W_{ij}$  and  $w_{jk}$  are constant weight to be found, and

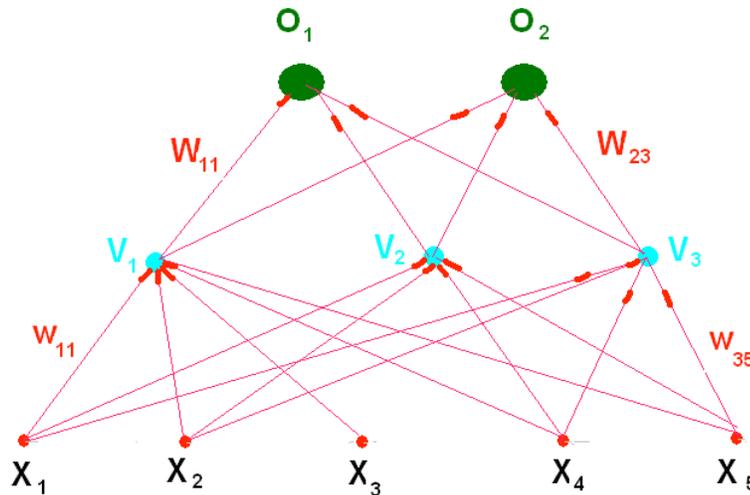
$$\sigma(x) = \tanh \beta x \quad (25)$$

is the sigmoid function providing universal properties of feed-forward neural nets to approximate a broad class of nonlinear functions, Fig.(13). The neural net has two layers. In general, it can have many layers

$$x''(t+1) = \sigma \left\{ \sum_{j'} W'_{i'j'} \sigma \left\{ \sum_{j''} W''_{i''j''} \dots \right. \right\} \quad (26)$$

Our experience suggests that two layers are sufficient for our purposes.

Let us now turn to Eq. (24) and Fig.13. The input units denoted there by  $X_i$  represent the known values of the data  $x''(t-k)$ , where  $k = 1, 2, \dots, N$ , and  $N$  is the number of inputs.



**Figure13. Feed-forward neural nets.**

The output denoted by  $O_1$  represents the computed values of  $x''(t)$ . The so called hidden units  $V_j$  do not represent any known values and they served for expanded connection between the inputs and the output. . The best-fit approach includes several different

patterns. For instance, the first pattern  $\mu = 1$  includes: the inputs  $X_1^1 = x''(t-1), \dots, X_5^1 = x''(t-5)$  and the output  $O_1 = x''(t)$ ; the second pattern  $\mu=2$  includes: the inputs  $X_2^2 = x''(t-2), \dots, X_6^2 = x''(t-6)$  and the output  $O_1 = x''(t-1)$ , etc.

The best-fit procedure can be described as following.

Given pattern  $\mu$ , hidden unit  $j$  receives a net input

$$h_j^\mu = \sum_k w_{jk} X_k^\mu \quad (27)$$

and produces the output

$$V_j^\mu = \sigma(h_j^\mu) = \sigma\left(\sum_k w_{jk} X_k^\mu\right) \quad (28)$$

Output unit thus receives

$$h_1^\mu = \sum_j W_{1j} V_j^\mu = \sum_j W_{1j} \sigma\left(\sum_k w_{jk} X_k^\mu\right) \quad (29)$$

and produces the final output

$$O_1^\mu = \sigma(h_1^\mu) = \sigma\left(\sum_j W_{1j} \sigma\left(\sum_k w_{jk} X_k^\mu\right)\right) \quad (30)$$

The error measure or cost function

$$E = \frac{1}{2} \sum_\mu (\zeta_1^\mu - O_1^\mu)^2 \quad (31)$$

compares the computed output  $O_1^\mu = x''(t)$  and its actual value  $\zeta_1^\mu$  taken from the data. With the reference to Eq. (28), one obtains

$$E = \frac{1}{2} \sum_\mu [\zeta_1^\mu - \sigma\left(\sum_j W_{1j} \sigma\left(\sum_k w_{jk} X_k^\mu\right)\right)]^2 \quad (32)$$

This is a continuous differentiable function of every weight, and these weights should be found from the condition that they minimize the error (32). The minimization procedure can be performed by the gradient descend method. For the hidden-to-output connections the gradient descend rule gives

$$\Delta W_{1j} = -\eta \frac{\partial E}{\partial W_{1j}} = \eta \sum_\mu \delta_1^\mu V_j^\mu \quad (33)$$

where

$$\delta_1^\mu = \sigma'(h_1^\mu) [\zeta_1^\mu - O_1^\mu], \quad \text{and} \quad \sigma'(h) = \beta(1 - \sigma^2) \quad (34)$$

For the input-to-hidden connections

$$\Delta w_{jk} = -\eta \frac{\partial E}{\partial w_{jk}} = \eta \sum_{\mu} \delta_j^{\mu} X_k^{\mu} \quad (35)$$

where

$$\delta_j^{\mu} = \sigma'(h_j^{\mu}) W_{1j} \delta_1^{\mu} \quad (36)$$

Then

$$w_{jk}^{new} = w_{jk}^{old} + \Delta w_{jk} \quad (37)$$

The process continues until

$$\Delta w_{jk} = 0 \quad (38)$$

Since Eqs(33) and (35) represent a gradient system for which E plays the role of a Lyapunov function, the solution to these equations eventually converge to a minimum. But since this system is nonlinear, this minimum may be local, Fig. (14 A). Hence, in order to find the global minimum, one should start the algorithm with many different initial conditions for the weights w and W, while each solution can be run independently. Then, the best-fit solution will correspond to the lowest (global) minimum.

In neural network theory this algorithm is called back-propagation, and it has many modifications that are problem-specific.

### 3. Alternative approach to model fitting.

Back-propagation algorithm applied to detection of the best-fit parameters in the model representation

$$x(t+1) = \sigma\left\{\sum_{ij} W_{ij} \sigma\left[\sum_k w_{jk} x(t-k)\right]\right\}, \quad k = 1,2,\dots,p \quad t = 1,2,\dots,q \quad (39)$$

is represented by the equation that is nonlinear with respect to the state variables x as well as with respect to the unknown parameters w and W. Because of the nonlinearity with respect to the parameters w and W TO BE DETECTED, the search for these parameters (implemented by the back-propagation), is implemented by the gradient descent algorithm that is slow (it has to be repeated several times to guarantee that the minimum is global), see Fig. 14 A..

Modify Eq. (1) to the following form

$$x(t+1) = \sum_k W_k \sigma[x(t-k)] \quad k, = 1,2,\dots,p. \quad t = 1,2,\dots,q \quad (40)$$

This equation is still nonlinear with respect to the state variables x (and therefore, it captures all the nonlinear effects), but it is linear with respect to the SOUGHT parameters  $W_k$ . The solution to this equation is trivial: it is reduced to the inversion (or pseudo-inversion) of the matrix

$$W = A * Y \quad (41)$$

where  $A^*$  is the pseudo-inverse of the matrix  $A$

$$A^* = (A^T A)^{-1} A^T \quad (42)$$

while

$$A_{kt} = \{\sigma[x(t-k)]\}, \quad Y_t = x(t+1), \quad W = \begin{Bmatrix} W_1 \\ \dots \\ W_p \end{Bmatrix} \quad (43)$$

Usually  $q > p$ , i.e. the number of equations is bigger than the number of unknowns, and the system is overdetermined. In that case, the pseudo-inverse of the rectangular matrix provides the best solution in terms of the minimal-norm of the residual vector.

The capacity of the model (40) is equivalent to the capacity of the model (39) *without* the hidden layer. However, the model (40) can be extended to include the second-order products of the sigmoid functions

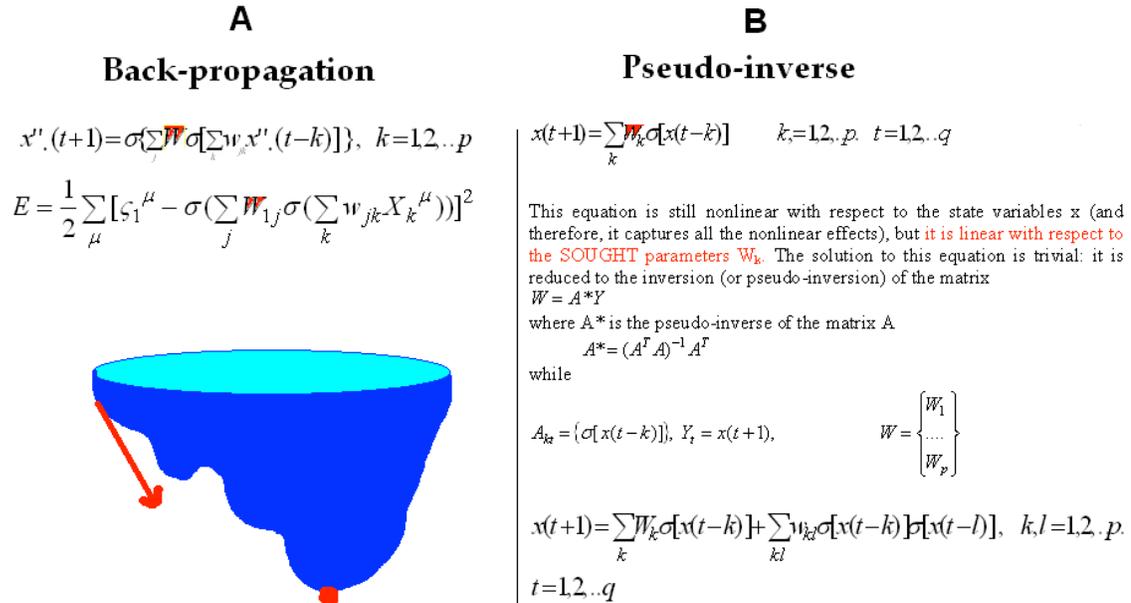
$$x(t+1) = \sum_k W_k \sigma[x(t-k)] + \sum_{kl} w_{kl} \sigma[x(t-k)] \sigma[x(t-l)], \quad k, l = 1, 2, \dots, p. \quad (44)$$

$$t = 1, 2, \dots, q$$

so that its capacity becomes equivalent to the model (39) with the hidden layer. At the same time, the solution to this system is still trivial, and it is given by the similar pseudo-inverse. Obviously, extension of Eq. (44) by including triple-products of sigmoid functions is equivalent to the second hidden layer in the model (39), etc, see Fig.14 B.

It should be emphasized that solution to the problem does not include dynamical convergence, local minima etc, and therefore, it can be easier generalized to multi-dimensional case.

## Model fitting



**Figure 14**

### Section 5. Criteria of abnormalities

#### 1. Criterion of structural abnormalities.

There are two causes of abnormal behavior of the solution to the governing equation of the dynamical model: changes in external forces or initial conditions, and changes in the parameters  $W_{ij}$ ,  $w_{jk}$ , i.e. changes in the structure of the dynamical system. (These changes can be associated with structural abnormalities, and can be linked to the theory of catastrophe).

We will introduce the following criterion of the structural abnormality:

$$\xi = \sum [(W_{ij} - W_{ij}^0)^2 + (w_{jk} - w_{jk}^0)^2] \quad (45)$$

where the parameters  $W$  and  $w$  are introduced in Eqs.(24). Their nominal values (i.e. the values corresponding to “healthy” state of the dynamical system) is denoted by the super index “0”.

Obviously, if

$$\xi = 0, \quad \text{or} \quad |\xi| < \varepsilon \quad (46)$$

where  $\varepsilon$  is sufficiently small, then there is no structural abnormalities. The advantage of this criterion is in its simplicity: it can be periodically update, and therefore, the structural “health” of the process can be easily monitored. The only limitation of this criterion is that it does not specify a particular cause of n abnormal behavior. Obviously, this limitation can be removed by monitoring each parameter of  $W_{ij}$ ,  $w_{jk}$  separately.

#### 2. Reconstruction of trends and oscillations.

After fitting the model

$$x = x' + x'' \quad (47)$$

where

$$x'(t) = \alpha_1 x'(t-1) + \alpha_2 x'(t-2) + \dots + \alpha_n x'(t-n) + N \quad (1)$$

$$x''_*(t+1) = \sigma \left\{ \sum_j W_{ij} \sigma \left[ \sum_k w_{jk} x''_*(t-k) \right] \right\}, \quad k = 1, 2, \dots, p$$

we have obtained an analytical formulation for the stationary component of the model. Now we can obtain an analytical formulation for the original model that will include both the trends and oscillations.

In order to reconstruct trends, one should exploit the summation operator

$$(1-B)^{-1} z_t = Sz_t = \sum_{j=0}^{\infty} z_{t-j} = z_t + z_{t-1} + z_{t-2} + \dots \quad (48)$$

that is inverse to the difference operator being used for the extraction of the stationary component from the original data. For instance, if the stationarization procedure has been performed by the operator

$$\nabla x_t = x_t - x_{t-1} = (1-B)x_t \quad (49)$$

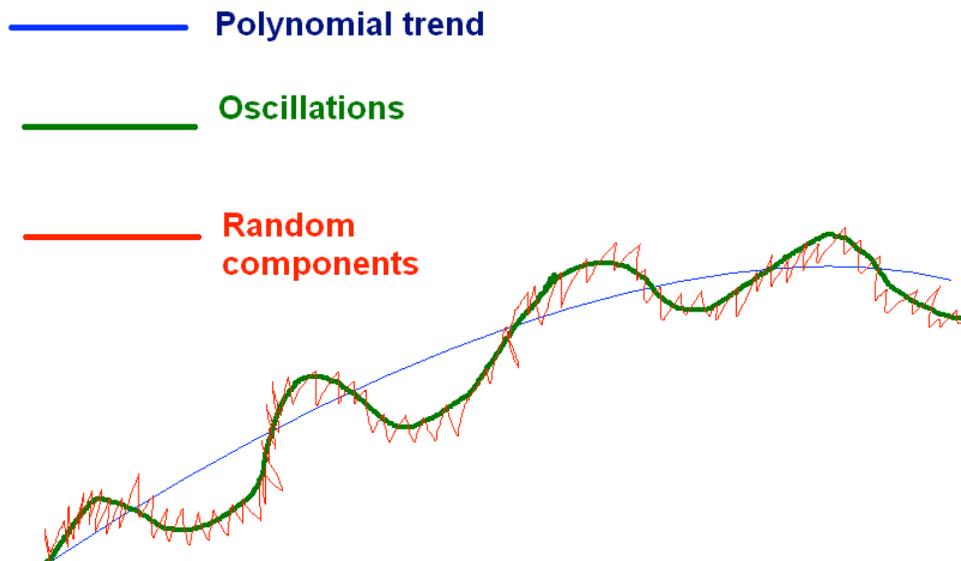
applied to the original time series  $y(t)$ , then the analytical formulation of the underlying dynamical model is

$$Y_t = x_{t-1} + x' + x'' \quad (50)$$

In case of several differencing of the original data, the summation operator should be applied the same number of times. Obviously, the additional terms appearing after summations are to be found from the original data.

Similarly, the oscillations that have been filter out by seasonal differencing by applying the differencing operator  $(1-B^S)$  are to be obtained by using the inverse (summation) operator  $(1-B^S)^{-1}$ .

The reconstruction is illustrated in Fig.15.



**Figure 15** Reconstruction of trend and oscillations

**Remark.**

When the original data are non-stationary, and trend or/and oscillations are stochastic, some of the characteristic roots of Eq. (10) are located at the unit circle (Figure (9)), and the solutions to Eq.(10) are neutrally stable. That may cause the divergence of the summation operator (48) for infinite number of terms in the sum

(48). Instead one can consider the finite summation operator

$$S = 1 + B + B^2 + \dots + B^{m+1} = \frac{1 - B^{m+1}}{1 - B}$$

For that purpose, the whole interval of summation should be divided into several sub-intervals of lengths  $m$ , and the end value datum of the previous interval should serve as the initial condition for the next interval. Obviously, selection of  $m$  is data-specific.

When the original data are non-stationary, but trends and oscillations are deterministic, the originally stable data will remain stable since deterministic non-stationarities do not affect the characteristic roots. In this case, the replacement (48a) are not necessary.

### 3.Trend and oscillation invariants.

The difference between the original time series  $y(t)$  and the stationary components  $x(t)$  that is obtained as a result of an appropriate number of differencing represent the non-stationary component

$$U(t) = x(t) - y(t) \quad (51)$$

Let us decompose this component into a polynomial trend

$$U_1(t) = \sum_{k=0}^m \gamma_k t^k \quad (52)$$

and the periodic components

$$U_2(t) = \sum_{j=1}^{[S/2]} [\beta_{1j} \text{Cos}(\frac{2\pi j}{S} t) + \beta_{2j} \text{Sin}(\frac{2\pi j}{S} t)] \quad (53)$$

Here  $m$  is the number of differencing in the data processing,  $S$  is the power of the seasonal differencing operator  $(1-B^S)$ , while

$$[S/2] = \begin{cases} 0.5S & \text{if } S \text{ is even} \\ 0.5(S-1) & \text{if } S \text{ is odd} \end{cases} \quad (54) \quad \text{In}$$

general, both trends and periodic components are not deterministic, and therefore, the trend coefficients  $\gamma$  and the amplitudes  $\beta$  can be represented by their least square estimates. In order to implement these representations, let us turn to the original time series  $Y(t)$ . Denote the result of seasonal differencing (with the periods) as  $y_s(t)$ . Then

$$U_2(t) = Y(t) - Y_s(t) \quad (55)$$

$$U_1(t) = U(t) - U_2(t) \quad (56)$$

Now the least square estimate for the coefficients  $\beta$  are

$$\beta_{1i} = \frac{2}{3} \sum_{t=1}^S U_2(t) \text{Cos} \frac{2\pi j}{S} t \quad (57)$$

$$\beta_{2j} = \frac{2}{3} \sum_{t=1}^S U_2(t) \sin \frac{2\pi j}{S} t \quad (58)$$

The least square estimates for the trend coefficients  $\gamma$  are found from the system

$$U_1(t_i) = \sum_{k=0}^m \gamma_k t_i^k, \quad i = 1, 2, \dots, N \quad (59)$$

where  $N$  is the number of points in the time series

$$\bar{\gamma} = A^* \bar{U}_1 \quad (60)$$

Here

$$\bar{\gamma} = \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \dots \\ \gamma_m \end{bmatrix}, \quad \bar{U}_1 = \begin{bmatrix} U_1(t_1) \\ U_1(t_2) \\ \dots \\ U_1(t_m) \end{bmatrix}, \quad A = \begin{bmatrix} 1t_1t_1^2 \dots t_1^m \\ 1t_2t_2^2 \dots t_2^m \\ \dots \\ 1t_Nt_N^2 \dots t_N^m \end{bmatrix} \quad (61)$$

and  $A^*$  is the pseudo-inverse of the matrix  $A$

$$A^* = (A^T A)^{-1} A^T \quad (62)$$

where  $A^T$  is the transpose of  $A$ .

#### 4. Noise invariants.

Noise represents the difference between the original data  $y(t)$  and the data reproduced by the reconstructed model  $Y(t)$

$$R = y(t) - Y(t) \quad (63)$$

As follows from the data processing discussed above, noise is not correlated and it is stationary since all the correlated and non-stationary components have been extracted. Therefore, one can apply the ergodic hypothesis and find all the statistical invariants through the summation over time (instead the summation over ensemble), namely, the moments

$$R_1 = \frac{1}{N} \sum_{k=0}^N R(t_k) \quad (64)$$

$$R_2 = \frac{1}{N} \sum_{k=0}^n R^2(t_k) \quad (65)$$

$$R_q = \frac{1}{N} \sum_{k=0}^N R^q(t_k) \quad (66)$$

It should be noticed that the noise invariants (64)-(66) make important contribution into the identity of the underlying model.

### Section 6. Nominal confidence intervals.

In the previous section, a state of the underlying dynamical system generating the observed sensor data was defined by the dynamical invariants  $a_j$  that include the coefficients of the trend polynomials  $\gamma_j$ , amplitudes and frequencies of the dominating harmonics  $\beta_j$ , autoregressive coefficients  $\alpha_j$  and neural net weights  $w_{ij}, W_{ij}$  representing the stationary component of the stochastic process, and the moments of noise distribution  $R_j$ :

$$Y(t) = S(\gamma_j, \beta_j) \{ \alpha_1 Y(t-1) + \alpha_2 Y(t-2) + \dots + \alpha_n Y(t-n) + N + \sigma [ \sum_j W_{ij} \sigma [ \sum_k w_{jk} Y(t-k) ] ] \}, \quad k = 1, 2, \dots \quad (67)$$

Here  $S$  is the summation operator that is the inverse to the set of the difference operators being applied for the detection of the stationary component of the stochastic process.

The invariants  $a_j$  were calculated during a selected training period that included  $N$  values of the original time series. This period can be associated with the short-term training. Obviously, during this period all the invariants  $a_j$  constants.

In order to introduce a long-term training period, let us return to the original data

$$x = x(t_i), \quad i = 0, 1, \dots, etc \quad (68)$$

and consider them within the interval shifted forward by  $q$  points:

$$x_1 = x_1(t_i), \quad i = q, q+1, \dots, etc$$

$$x_2 = x_2(t_i), \quad i = 2q, 2q+1, \dots, etc$$

.....

$$x_p = x_p(t_i), \quad i = pq, pq+1, \dots, etc \quad (69)$$

where  $p$  is the number of  $q$ -shifts.

For each time series of data (69), one can compute the same invariants  $a_i$  by applying the same sequence of algorithms as those applied to the original data (2). In general, even in case of absence of any abnormalities, the values for  $a_i$  for different  $p$  are different because of measurement and computational errors, so that  $a_i$  will occur as series of  $p$ :

$$a_i = a_i(p), \quad p = 1, 2, \dots, etc \quad (70)$$

Since  $p$  is proportional to time

$$p \propto q\Delta t \quad (71)$$

where  $\Delta t$  is the sampling interval, Eq. (70) actually represents another time series, and therefore, it can be treated in the same way as the original time series (68). However, such a treatment applied to each invariant  $a_i$  is very costly, and for most of the practical cases, it is unnecessary. Indeed, since all the significant non-stationarities (in the form of the least-square polynomial trends and dominating harmonics) were already identified, it is unlikely that the time series (70) contains additional non-stationarities. Besides of that,

since these invariants associated with certain physical properties of the underlying dynamical system, their deviations from the original constant values can be interpreted as a result of errors in measurements and computations. That is why a simple statistical analysis of the time series (70) is sufficient for defining the nominal confidence intervals.

In order to perform the statistical analysis for the time series (70), one could generate a histogram.

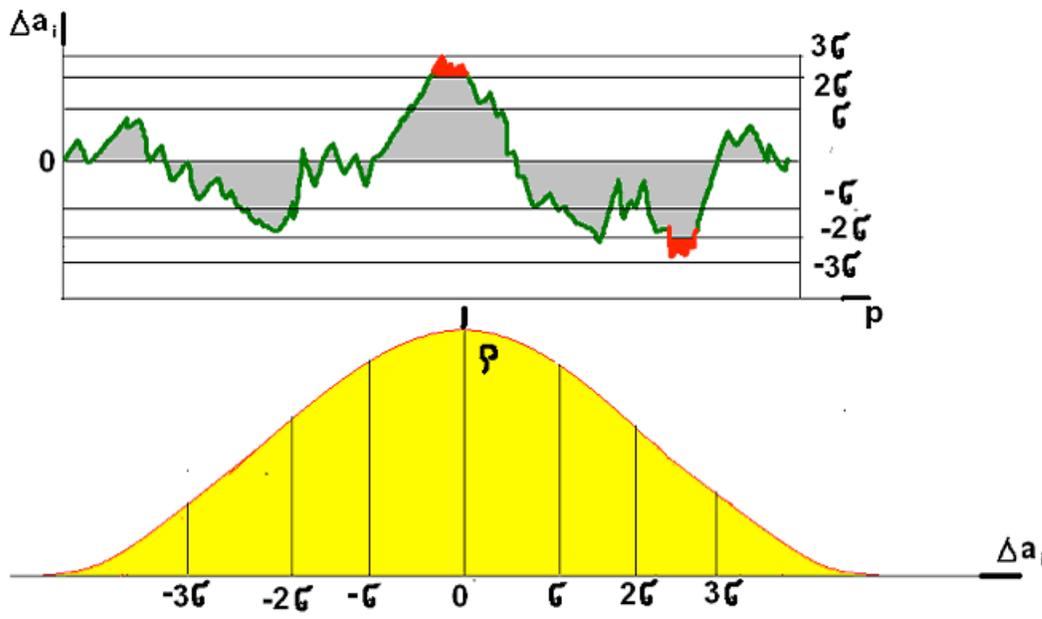


Figure16 . Nominal confidence intervals

However, in many practical cases, it can happen that such a histogram exhibit non-Gaussian distribution that makes it harder to define the confidence intervals. That is why it is more convenient to apply a bootstrap approach as a preprocessing procedure in the following way:

1. Choose randomly P samples of data from (70) with the same sample size  $n \sim P/2$ .

$$a^{(1)}_p = a^{(1)}_{p1}, a^{(1)}_{p2}, \dots, a^{(1)}_{pn},$$

.....

$$a^{(p)}_p = a^{(p)}_{p1}, a^{(p)}_{p2}, \dots, a^{(p)}_{pn}$$

2. Find the sample means:

(72)

$$\tilde{a}^1_p = \frac{1}{n} \sum_{i=1}^n a^{(1)}_{pi} \quad (73)$$

As shown in statistics, the distribution of the means  $\tilde{a}_p$  will be Gaussian (even if the original distribution is non-Gaussian) (see Fig. 1). Hence, the nominal confidence intervals that include 68%, 95% and 100% of all the sample means are, respectively:

$$\mu_a - \frac{\sigma_a}{\sqrt{n}} < \tilde{a}_p^{(i)} < \mu_a + \frac{\sigma_a}{\sqrt{n}} \quad (74)$$

$$\mu_a - 2 \frac{\sigma_a}{\sqrt{n}} < \tilde{a}_p^{(i)} < \mu_a + 2 \frac{\sigma_a}{\sqrt{n}} \quad (75)$$

$$\mu_a - 3 \frac{\sigma_a}{\sqrt{n}} < \tilde{a}_p^{(i)} < \mu_a + 3 \frac{\sigma_a}{\sqrt{n}} \quad (76)$$

where  $\mu_a$  and  $\sigma_a$  are the mean and the standard deviation of the sample's mean distribution.

## Section 7. Proposed forecast methodology.

We will start with analysis of Eq.(77) representing the reconstructed model in a compressed form

$$X_{t+1} = F(X_t, \dots, X_{t-m}) + R_{t+1}, \quad (77)$$

In principle, this equation allows one to predict the next values  $X_{t+1}$  given the previous values  $X_t, X_{t-1}, \dots, X_{t-m}$  and the *next* value of the error  $N_{t+1}$ . But the latter has not been measured yet, and it can be only evaluated based upon statistical considerations. That is why a deterministic forecast of future data, in general, cannot be achieved. Turning to the probabilistic approach, one should recall that the time series  $N_t$  are represented by a sequence of independent random values whose probability density does not depend upon time. Indeed, the time dependence of statistical properties of this sequence was eliminated in the course of stationarization while the correlated components were captured by Eq. (77). Hence, the only statistical invariants of the process  $R_t$  are the invariants of its probability density, i.e. the mean, the variance, and the higher moments. This means that all the values  $R_t, R_{t+1}, \dots$  can be drawn randomly from the time series  $R_0, R_1, \dots, R_{t-1}$ . Obviously, each sample

$$R_t^i, R_{t+1}^i, \dots, etc, \quad i = 1, 2, \dots, m, \quad (78)$$

will lead to different predicted values

$$X_t^i, X_{t+1}^i, \dots, etc, \quad i = 1, 2, \dots, m, \quad (79)$$

However, since all the samples in (79) are statistically identical, all the predicted values will represent different realizations of the same stochastic process forming an ensemble shown in Fig.2. For each  $t_j$  one can find the probability function

$$f(X_j^{(i)}) = P(x = X_j^{(i)}), \quad j = 1, 2, \dots, n; \quad i = 1, 2, \dots, n. \quad (80)$$

by plotting a histogram for the function

$$X_j^{(1)}, X_j^{(2)}, \dots, X_j^{(m)}, \quad j = 1, 2, m. \quad (81)$$

Since the original time series, in general, are non-stationary, the probability functions (81) will depend upon time. Therefore, for each time  $t_j$  one can compute the statistical invariants: the mean

$$\mu_j = \sum_{i=1}^m X_j^{(i)} f_{j,j}(X_j^{(i)}) \quad (82)$$

The standard deviation

$$\sigma_j = \left[ \sum_{i=1}^m (X_j^{(i)} - \mu_j)^2 f_j(X_j^{(i)}) \right]^{1/2} \quad (83)$$

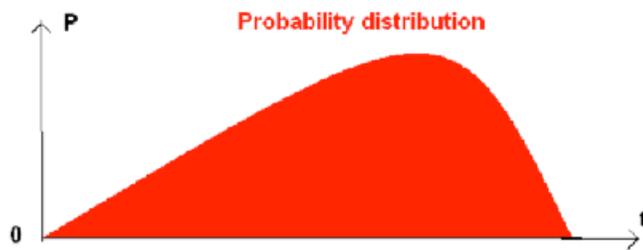
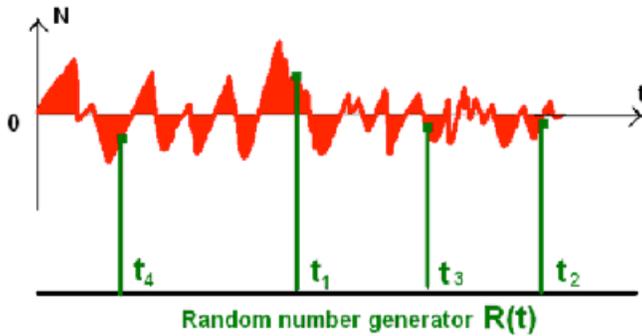
as well as higher moments

$$M_j^q = \sum_{i=1}^m (X_j^{(i)} - \mu_j)^q f_j(X_j^{(i)}) \quad (84)$$

The proposed forecast methodology can be implemented in the following steps.

1. Generate the error- time- series (78) in the time interval  $[0, T]$  as a difference between the original data and the data produced by the analytical model, (see Eq.(77)).
2. Apply a random number generator to produce random, equally probable, numbers  $m_j$  in the interval  $[0, 1]$ .
3. Multiply these numbers by the time-period  $T$  and find the corresponding random times  $t_j = Tm_j$ , Fig.17.
4. Find the corresponding value of the error by substituting  $t_j$  into the series (78).
5. Using these error series as an input  $R_{t+1}$  to the analytical data, produce the calculated data (77).
6. Repeat this procedure many ( $\sim 100$ ) times to produce an ensemble representing the sought stochastic process, Fig.18.
7. For each time  $t_j$ , find the probability distribution (80) as well as its invariants (82)-(84), Fig.18. In case the probability distributions have several maxima, find the global maxima that would correspond to the maximum likelihood for the corresponding data to appear. (This value may not coincide with the mathematical expectation (82)).
8. The Figure 19 illustrates future tasks based upon the proposed forecast.

$$x_t = L(x_{t-1}, \dots, x_{t-n}) + NL(x_{t-1}, \dots, x_{t-n}) + R_t$$



**Figure 17. Proposed methodology**

An alternative approach that is statistically equivalent, but *computationally* may be more effective is the following.

1. Generate the error- time- series (78) in the time interval  $[0, T]$  as a difference between the original data and the data produced by the analytical model, (see Eq.(77)).
2. Divide the time-interval  $[0, T]$  on  $n$  equal time-steps.
3. Shift the time series one time-step forward (replacing the datum at  $t=0$  by those at  $t=T$ ).
4. Adding the shifted error time series to the calculated data (see Eq. (77), produce the data that statistically equivalent to the original data.
5. Continue this process until the shift forward is equal to  $T$ .
6. For each time  $t_j$ , find the probability distribution (80) as well as its invariants (82)-(84), Fig.17a. .

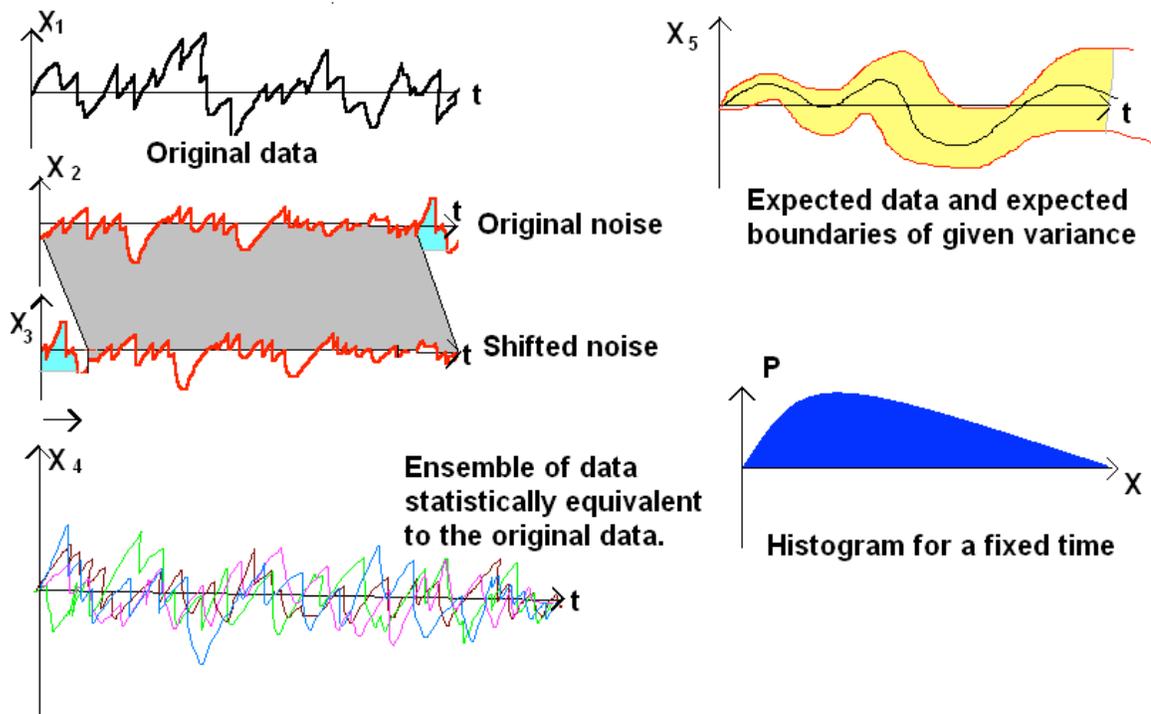


Figure 17 a. Alternative approach.

It should be noticed that although the alternative approach is simpler, it may be less reliable in case when the stationarization procedure is not perfect, and therefore, noise may be correlated. In the first version of the proposed forecast methodology, such correlation will be eliminated due to random picks of noise values; in the second version, the residual correlation will remain.

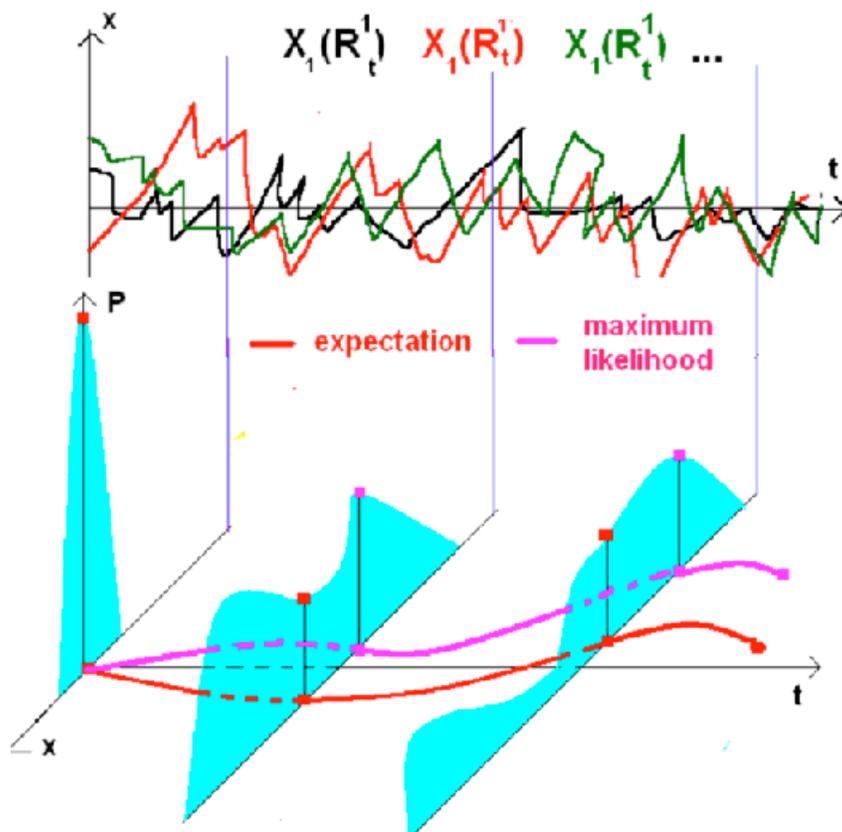


Figure 18. Representation of data forecast.

## 1. Forecast interpretation:

a) **Diagnosis**

b) **Prognosis**

## 2. Forecast-based **reactive** control.

a) **Monitoring**

b) **Adjustment**

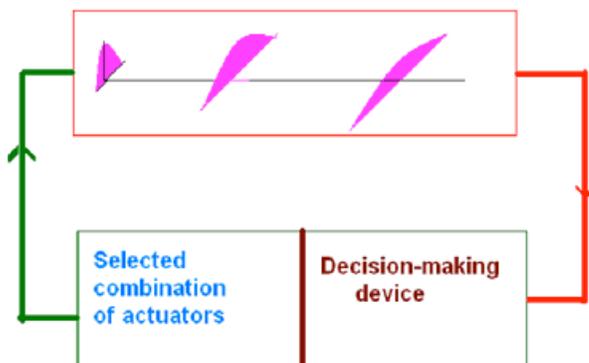


Figure 19. Forecast- based following tasks.

## Section 7. Risk assessment.

In this section we will briefly describe the role of data forecast in risk assessment. For that purpose, let us turn to Eq.(77) that includes the parameters  $a_i$  found as a result of the model fitting

$$X_{t+1} = F(X_t, \dots, X_{t-m}, a_1, a_2, \dots, a_n) + R_{t+1}, \quad (77a)$$

This equation can be considered as a discrete version of  $m$  coupled first order differential equations of the Langevin type (due to presence of random components  $R_{t+i}$ ), and errors in the parameters  $a_i$  may lead to such a catastrophe phenomena as loss of stability, auto-oscillations, linear or nonlinear resonance, stochastic resonance, etc, (see Figure 20). If Eq. (77a) is linear, or the deviations of the state variables  $X_{t+i}$  are small, and therefore, this equation is linearizable with respect to their stationary values, the safe boundaries shown in Figure 20 can be found analytically. However, in general case, these boundaries depend not only upon the parameters  $a_i$ , but upon the current values of the state variables as well, and the problem becomes analytically intractable. Here we are proposing a naïve, but reliable numerical approach: turn to the ensemble of data shown in Figure 17a, and calculate what percent of curves deviate away of the safe interval; this percent can be taken as a risk assessment.

### Parametric sensitivity analysis

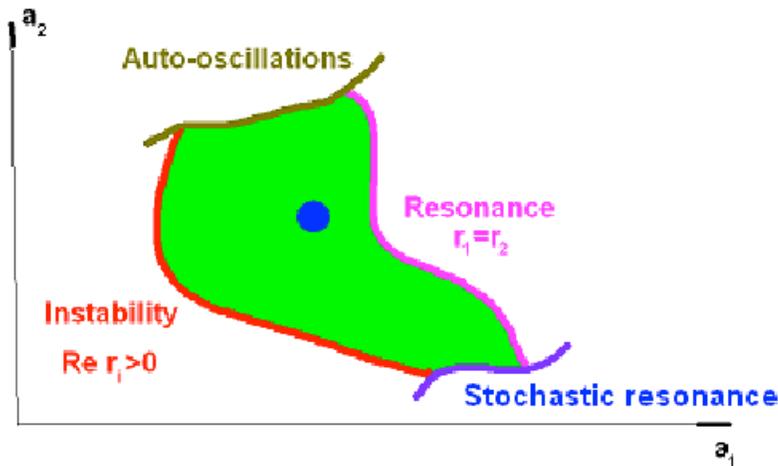


Figure 21. Safe boundaries.

## PART II

### EXAMPLE: GAS TURBINE POWER PLANT

#### Section 1

#### Description of the plant.

As an example, we demonstrate the data forecast methodology on a simulation of the shaft rotation speed of gas turbine with time delay feedback actuator. A schematic of the system is shown in Fig. 21.

Ignoring the thermal inertia of the combustion chamber, one can write the following dynamic equation for the angular velocity,  $\omega$ , of the shaft as:

$$T_s \frac{d^2 \omega}{dt^2} + \left(1 + \frac{T_s}{T_1}\right) \frac{d\omega}{dt} + \frac{\omega}{T_1} + \frac{\omega(t-\tau)}{\delta' T_2} = 0, \quad (85)$$

where  $T_s$  is the time constant for servo-motor (actuator);  $T_1$ ,  $T_2$  are the time constants for turbo-compressor;  $\delta'$  is the time delay characteristic of controller; and  $\tau$  is the total time delay from (1) to (2) in Fig. 20. The primary effects that are modeled by Eq. (85) are the rotational dynamics of shaft, the dynamics of actuator, and the time-delay of feedback loop.

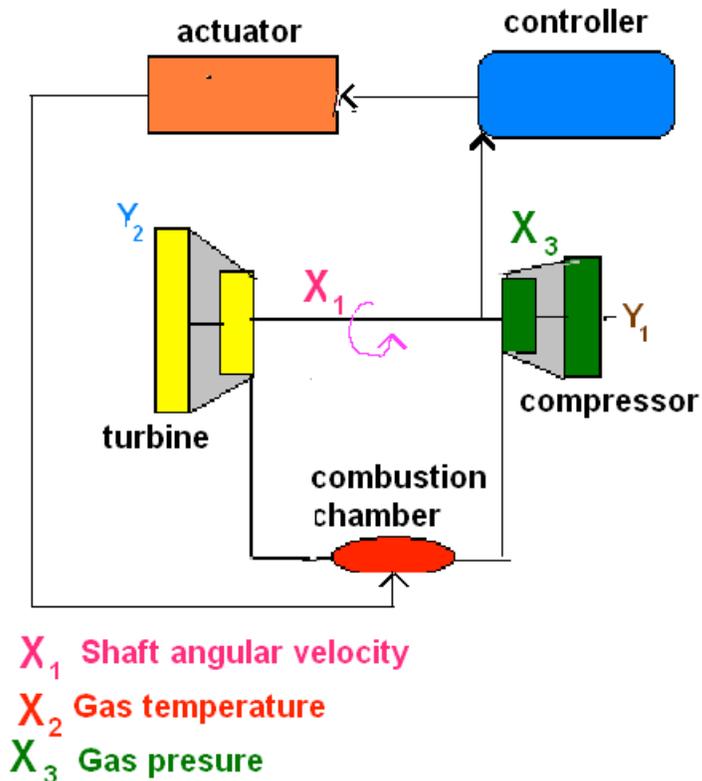
In a real system, however, it is expected that other effects, such as the torsional bending of shaft, the dynamics of combustor, blade flutter, etc. will also govern the speed of the gas turbine. In addition, there are stochastic effects such as friction and thermal effects that will also affect the speed of the gas turbine. In order to simulate these effects, two forcing terms are added to the left side of the discretized form of Eq. (85). The first term is an autoregressive linear process,  $z(k)$ :

$$z(k) = 0.75z(k-1) - 0.5z(k-2) + a(k), \quad (86)$$

where  $k$  is the  $k$ 'th sample, and  $a(k)$  is uncorrelated white noise. The second forcing term is a nonlinear logistic map,  $\Gamma_d(k)$ :

$$\Gamma_d(k) = 0.5\Gamma_d(k-1) + (1 - \Gamma_d(k-1)). \quad (87)$$

## Gas turbine power plant



*Fig. 21. Schematic of gas turbine with time delay feedback actuator.*

## Section 2.

### Sensor data processing and detection of abnormalities.

The result of the simulation of Eq. (85) with the two forcing terms is shown in Fig. 4(a). The parameters used for this simulation are:  $T_s = 0.0356$  sec,  $T_1 = 1$  sec,  $T_2 = 1$  sec,  $\delta' = 0.08$  sec, and  $\tau = 0.100$  sec. The neutral stability point for the system is  $\tau = 0.103$  sec. Because it is operating so close to its neutral stability point, the system oscillates when driven by the two linear and nonlinear forcing terms.

Figure 22(c) shows the residual generated when data is filtered using the deterministic component (white-box) as shown in Fig. 22(b). The filtering is accomplished by taking

the difference between the next time step projection using the deterministic model, i.e., (85), and the actual simulation data.

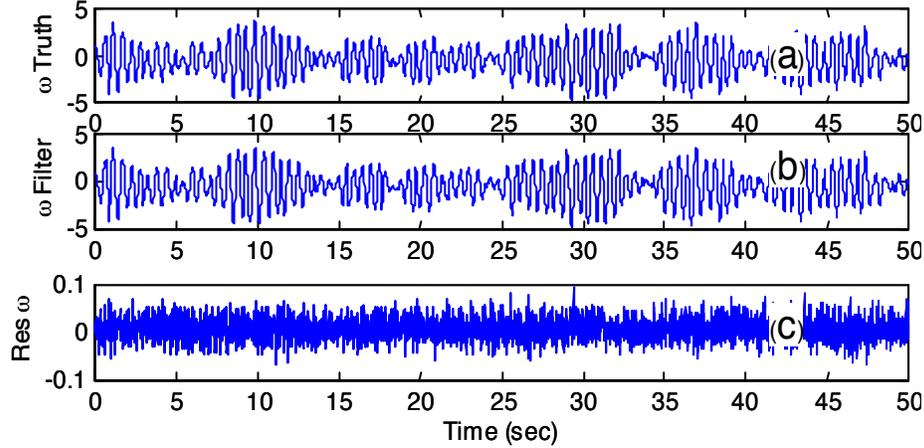


Fig. 22. *Simulation of gas turbine with time delay feedback actuator. The top plot (a) is the rotational speed. The middle plot (b) is the filter, and the bottom plot (c) is the residual.*

Once the filtering is accomplished, the residual (Fig. 23(a)) can be separated into its linear, nonlinear, and uncorrelated noise components as shown in Fig. 23(b) – (d).

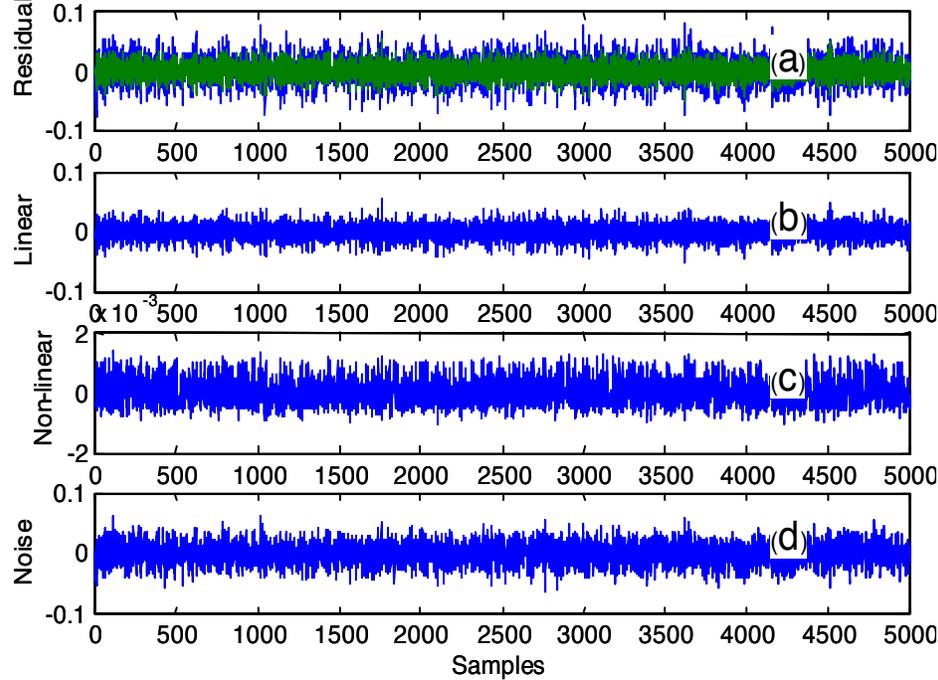


Fig. 23. *Separated components of the residual. The top plot (a) is the residual. Plots (b), (c), and (d) are the linear, nonlinear, and noise components, respectively. The light color trace in plot (a) is the reconstructed residual using the linear and non-linear components.*

The linear components were extracted by fitting a fourth order auto-regressive process as outlined in Eq. (88).

$$\begin{aligned}
 x(t) &= a_1x(t-1) + a_2x(t-2) + \dots \\
 &+ a_n(t-n) + z(t) \text{ as } n \rightarrow \infty
 \end{aligned}
 \tag{88}$$

The nonlinear components were extracted by first subtracting the linear component from the residual and then fitting a time-delay feed-forward network with five weights and biases as given by Eq. (89).

$$z(t) = \sigma \left\{ \sum_{j=1} W_{1j} \sigma \left[ \sum_{k=1}^m w_{jk} z(t - k\tau) \right] \right\} \quad (89)$$

Finally, the noise component is the remaining component.

For fault detection, the separated components of the residual are monitored by building a stochastic model. By examining the model parameters, the performance of the system can be monitored. An example, Fig. 24 shows the model coefficients for the fourth order linear autoregressive process. A window size of 500 samples were used to compute the parameters for each interval. It demonstrates that the stochastic parameters can be computed at regular intervals and be monitored for any anomalies. The dashed lines are the 99% confidence intervals for each of the parameters. If any parameters fall outside the confidence interval, the system would be classified as “anomalous,”

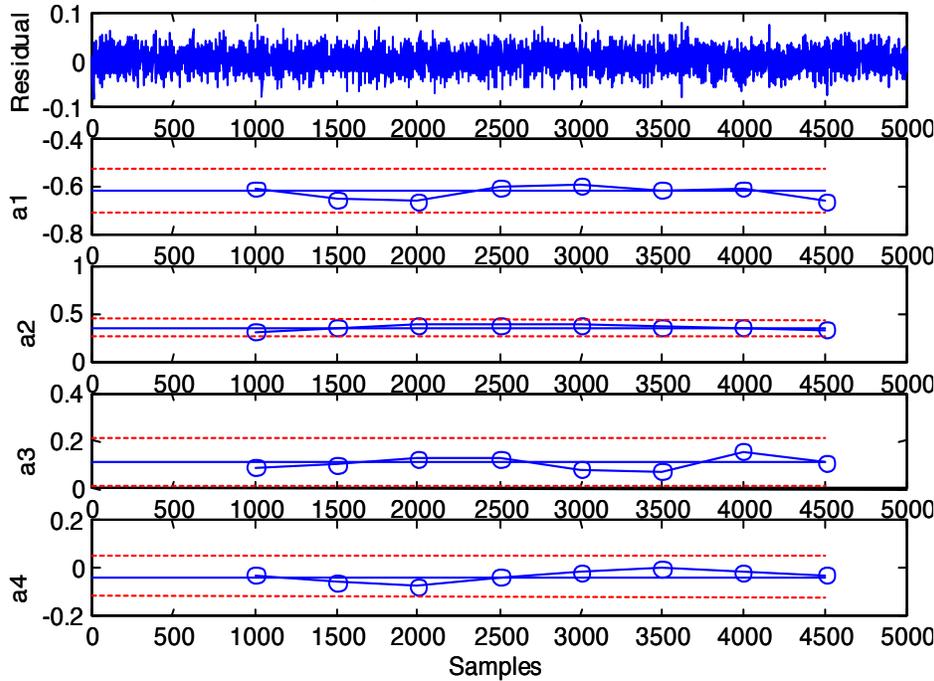


Fig. 24. Coefficients of the fourth order linear auto-regressive process.

### Section 3. Sensor data forecast and risk assessment.

#### Sensor data forecast.

In order to illuminate the specific features of the sensor data forecast, we will rearrange the data introduced in Section 1 as follows. The sensor data characterizing the gas turbine power plant can be subdivided into three components: the solutions to Eqs.(85), (86), and (87), respectively. The first component represents smooth theoretical data that are differentiable as many times as necessary; therefore, their forecast is trivial since it can be

based upon Taylor-series extrapolation. For that reason, they will be excluded from our consideration. The second component

$$y(k) = 0.75y(k - 1) - 0.5y(k - 2) \quad (86a)$$

represents deterministic, but irregular data. They can be forecasted by running the solution to Eq.(86), (see Fig. 25)

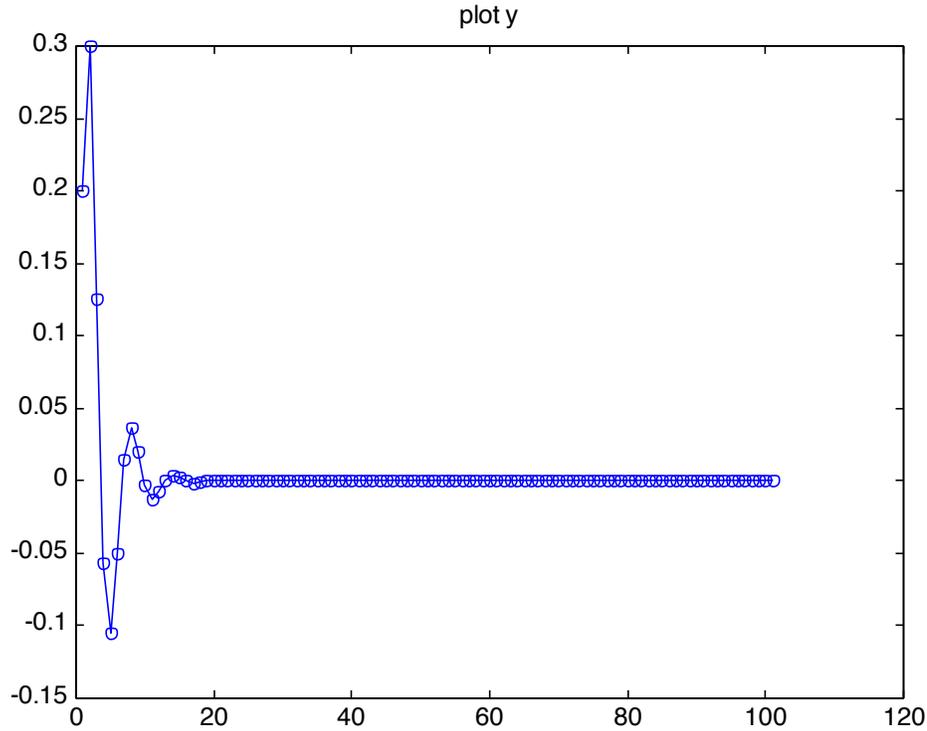


Figure 25. Deterministic data forecast.

As follows from Figure 25, the deterministic process described by Eq.(86) settles down to a stationary state after  $t > 20$ , and that is typical for a stable plant.

The third component represents uncorrelated noise. For better illustration, we will compose it from the logistic map similar to those in Eq.(87)

$$x(m + 1) = 4x(m)[1 - x(m)] \quad (87a)$$

The solution to this equation is indistinguishable from white noise (see Fig. 26).

Applying the proposed forecast methodology (see Part I, Section 7), Eq.(86a) should include noise (87a) in the following way

$$y(i, n + 2) = 0.75y(i, n + 1) - 0.5y(i, n) + x(n + i - 1) - \text{mean}(x) \quad (88)$$

Here  $I = 1, 2, \dots, 1000$  stands for the number of a sample chosen from the ensemble of 1000 samples. The difference between samples is due to the different noise contributions. The last term provides zero mean of the total noise contribution.

In this example we are using the second version of our methodology for incorporation of noise into the model (illustrated in Fig. 17a.)

Fig. (27) demonstrates samples of the solution to Eq. (88) for different numbers of  $i$ .

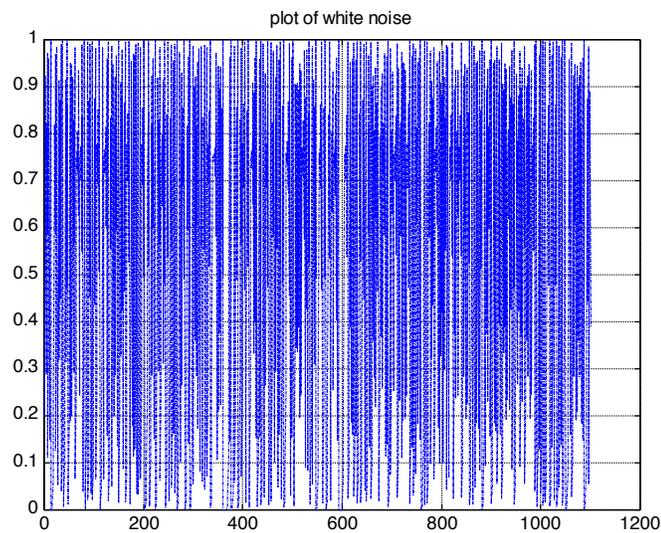
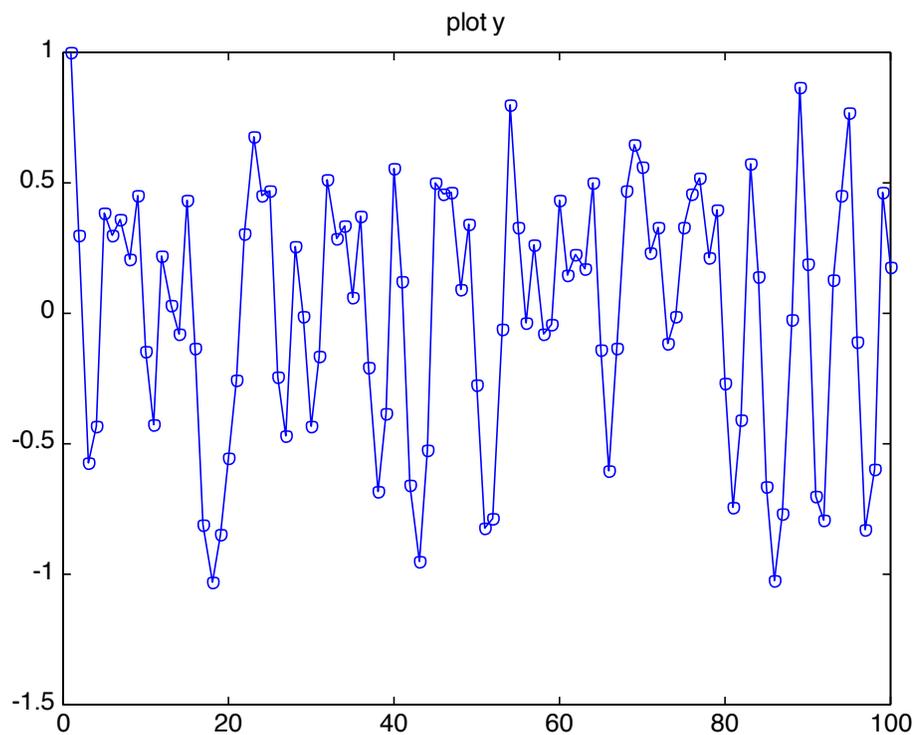


Figure 26. Logistic map representation of white noise.



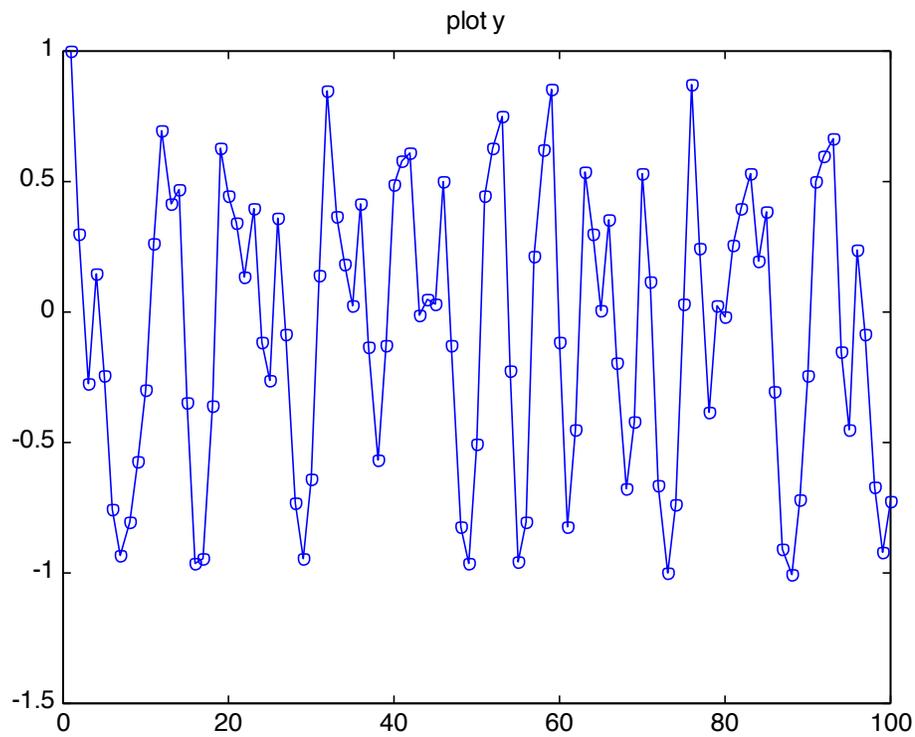
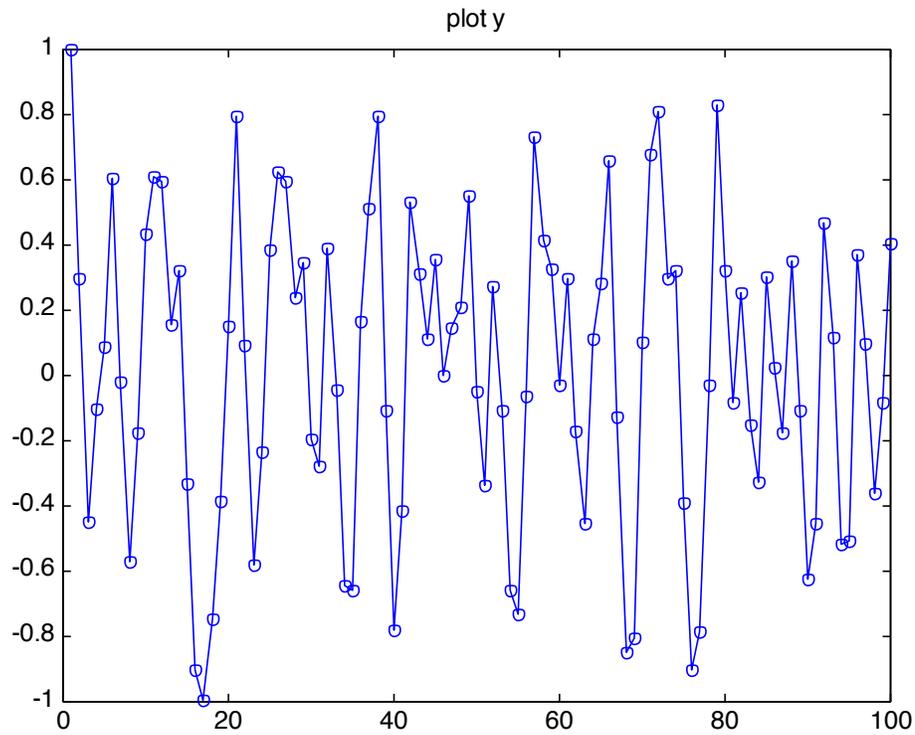


Figure 28. Samples of solution for  $i=10, 20,$  and  $30$ .

The most important characteristics of a forecast is given by the mean calculated for the whole 1000 samples of the ensemble since it expresses the *expected* data. For our example, they are plotted in figure 29

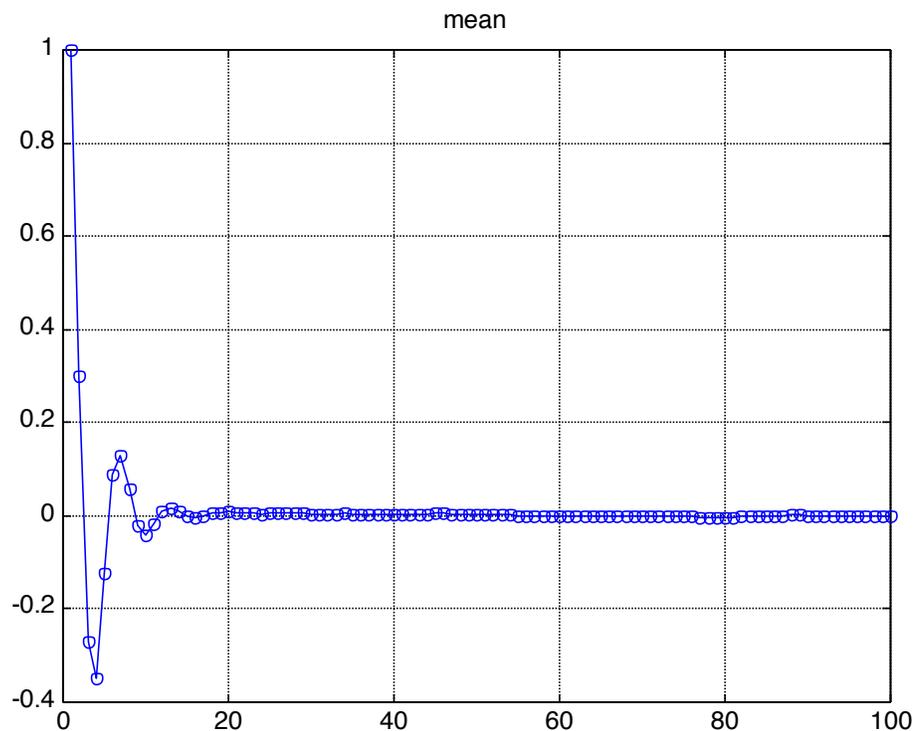


Figure 29. Expected data.

As follows from the figure 29, the mean depends upon time for  $t < 20$ ; after that it does not change, and that means that the corresponding stochastic process becomes stationary.

The second important characteristic of the forecast is the standard deviation that shows an expected error of the mean-based forecast. The standard deviation is plotted in Fig. 30. As follows from Fig. 30, the standard deviation grows sharply from  $t = 0$  to  $t = 20$ , approaching the value close to 0.5, and then it remains constant, and that expresses the second law of thermodynamics: the standard deviation is proportional to the entropy of the process, and that entropy can only increase until the process becomes stationary. It is impotent to emphasize that the speed of stationarization as well as the level of the stationary error, and therefore, accuracy of prediction, depend upon the degree of stability of the spacecraft dynamics, and in particular, on the degree of stability of Eq.(88). Indeed, let us replace Eq. (88) by a less stable equation

$$y(i, n + 2) = -0.75y(i, n + 1) - 0.95y(i, n) + x(n + i - 1) - \text{mean}(x) \quad (89)$$

The evolution of the standard deviation for that equation is plotted in Fig. 31. As shown there, the standard deviation now approaches the value 1.2, while the process becomes stationary only at  $t > 100$ . The histograms for Eqs (88) and (89) are plotted in Fig. 32. They demonstrate that an underlying dynamics with lower stability has a flatter histogram.

Figure 33 demonstrate dependence of data forecast upon the stability of the underlying dynamical system.

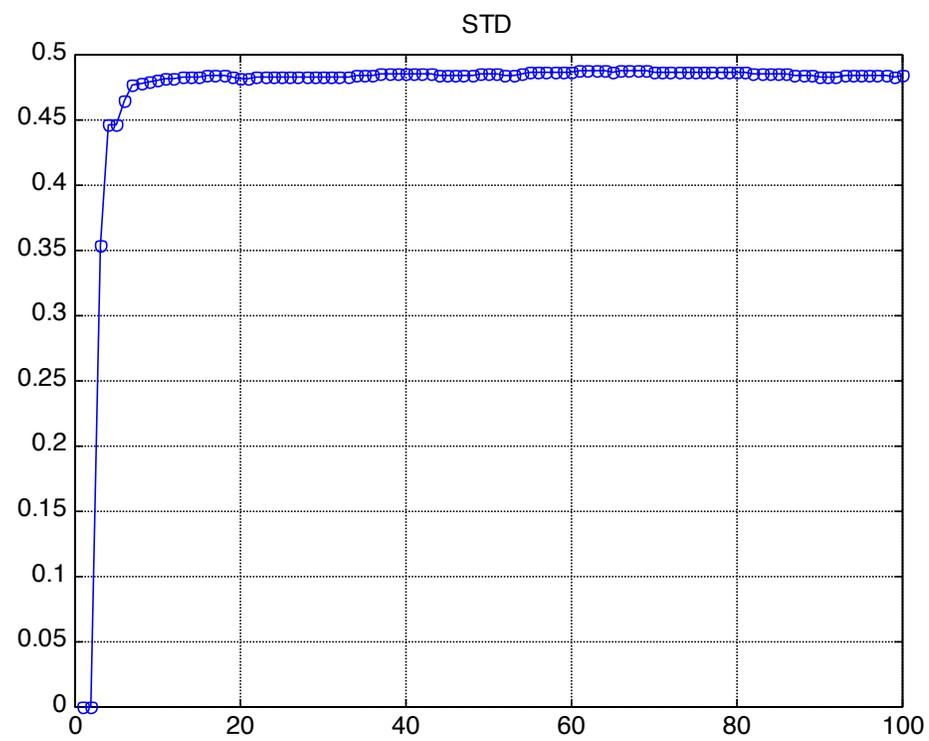


Figure 30. Standard deviation.

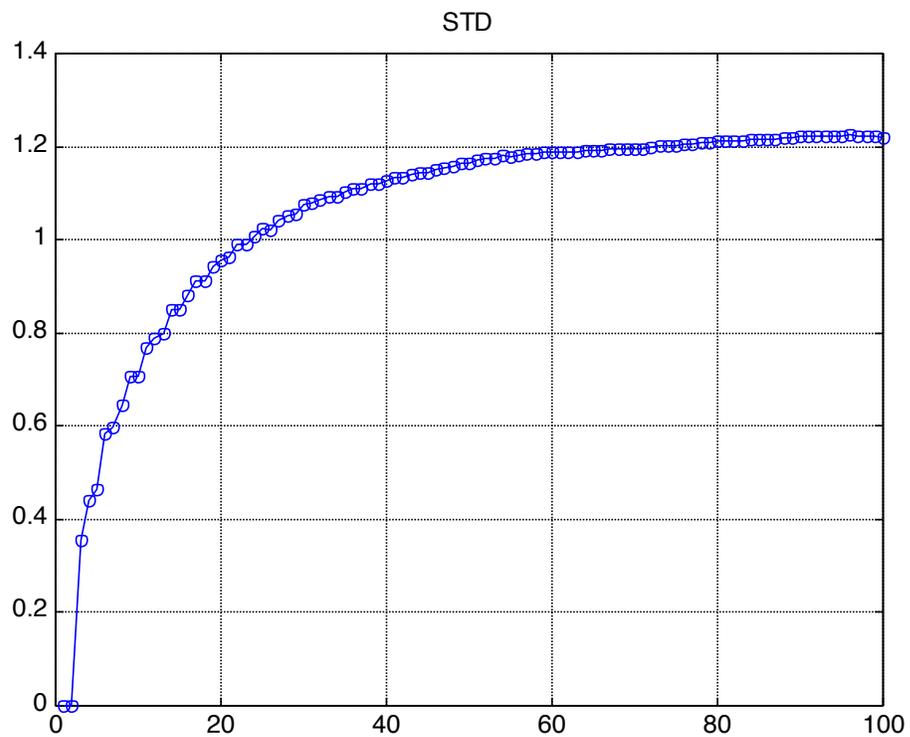


Figure 31. Standard deviation for a less stable process.

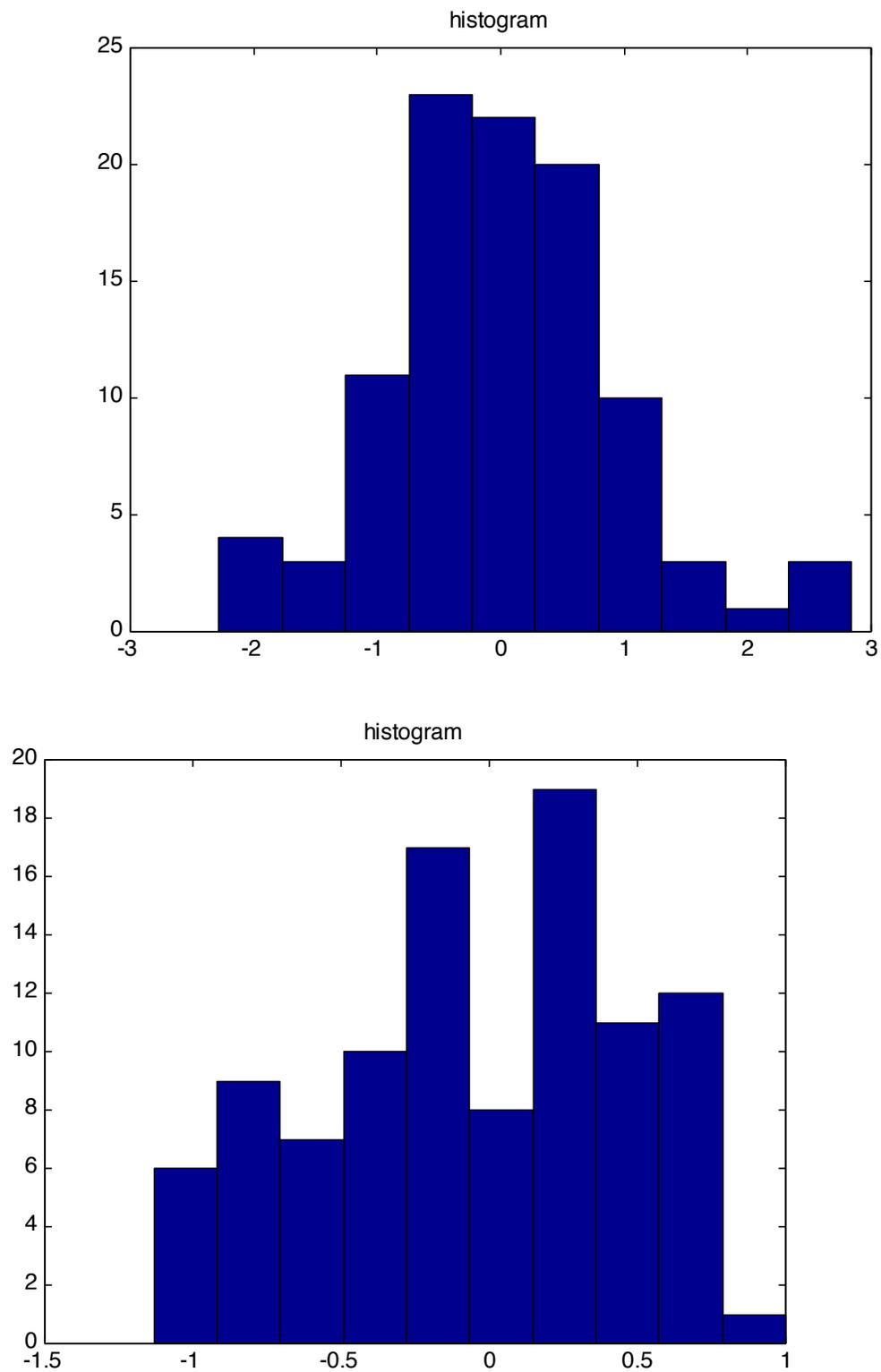


Figure 32. Typical histograms for Eq. (88) and (89), respectively.

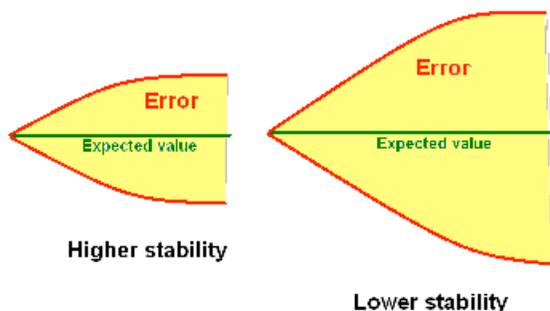
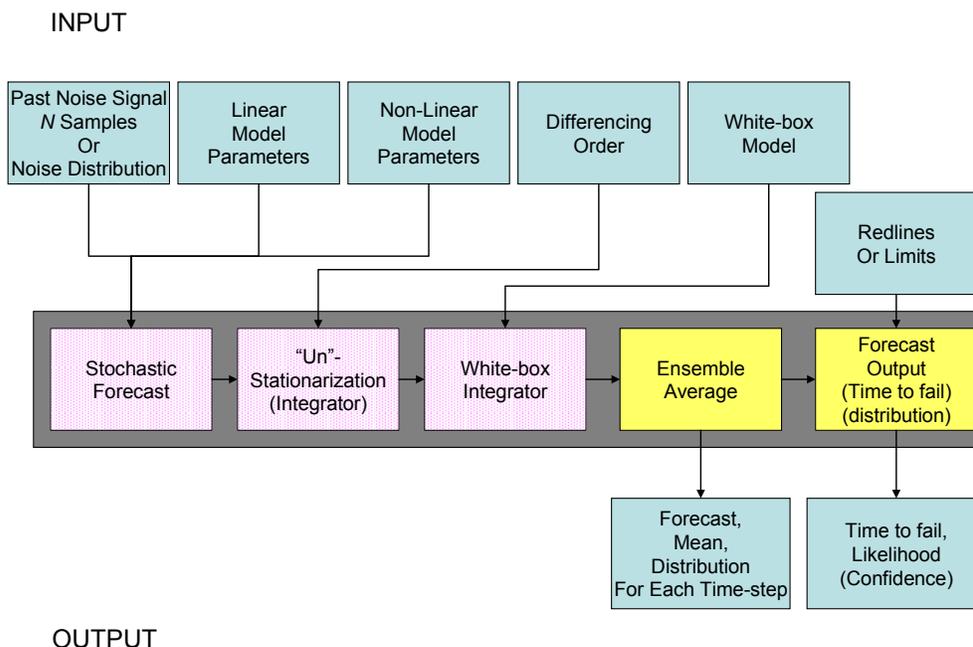
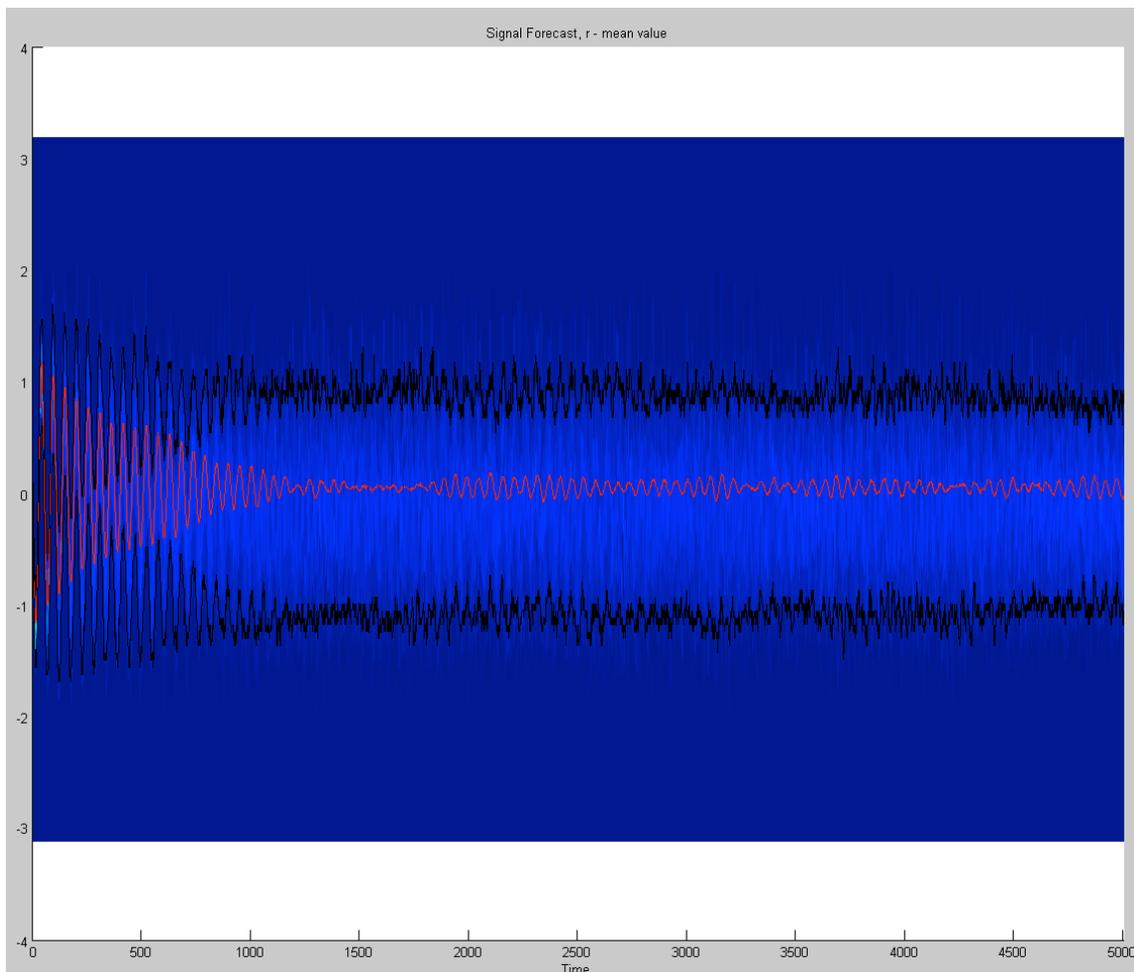


Figure 33. Data forecast in terms of expected values and standard deviation.

In order to verify the *first* version of the proposed methodology of incorporation noise into the model (illustrated in Fig. 17) and to confirm the validity of both versions, we have performed new computational runs. In addition to that, this time we produced noise not synthetically, but as the difference between the original (raw) data and their analytical approximation, (see Eq. (86)). Functional flow-diagram of the input and output of the forecast system illustrates the logical structure of the algorithm.



Plot of forecast ensemble mean (red, 100 trials) and error bounds (black, 99% confidence level) are presented below. (See Fig. 34). Comparison of the ensemble mean in figures 29 and 34 shows qualitative similarity. Slight quantitative difference is caused by the difference in the number of trials that is 1000 and 100, respectively. The error behavior is also qualitatively the same: the error starts from zero and asymptotically approaches a stationary value. The quantitative difference between errors behavior in Figures 30 and 34 is due to the fact that in Figure 34 error is represented by the standard deviation, while in Figure 34 – by error bounds.



**Figure 34.** Plot of forecast ensemble mean and error bounds

## 2. Risk assessment.

There are two types of parameters in the reconstructed dynamical model described by Eq. (77): the structural parameters  $w$ ,  $W$ , (see Eqs. (26) or (39)) and the operational parameters  $\gamma, \beta$  (see Eqs. (52), and (53), respectively). In order to understand the difference between them, start with the following trivial example. Consider a mass  $M$  suspended on a spring of elasticity  $C$ . The ratio  $C/M$  represents a *structural* parameter (the squared frequency). Changes of this parameter manifest an anomaly: the spring is broken or the mass is lost. However, change of the amplitude can occur without changes of  $M/C$  (it depends upon initial position and initial velocity). The conclusion about abnormality can be made only after observing whether the motion is within safe interval or not. Hence, the amplitude is not a structural, but rather an *operational* parameter.

This analogy can be applied to the BEAM methodology. As a result of BEAM, we have an analytical model (77) that includes the parameters  $w$ ,  $W$ ,  $\gamma$ , and  $\beta$ . The parameters  $w$ , and  $W$  are *structural*: they are obtained by fitting stationarized data into the model. (compare with  $C/M$ ). The difference between the structural parameters of training and actual data manifests anomaly *regardless* of the forecast. The parameters  $\gamma$ , and  $\beta$  are operational. They are obtained by fitting the data complemented by trends and

oscillations into the model obtained by integration as a reverse to differencing.(see Eq48)). The arbitrary constants of integration are found from initial position and initial velocities that were “lost” in the course of differencing, (compare with the amplitude of the mass-spring example). Hence, a mismatch between operational parameters of training and actual data does not YET manifest anomaly: that can be a result of obtaining the training and actual data at different starting point, or at different external forces. ONLY data forecast will show the significance of this mismatch: if it leads to violation of the prescribed safety interval, it is an anomaly, otherwise it is OK.

The purpose of this section is to assess risk based upon data forecast when only operational anomalies have been detected. In particular, we will be interested in probability of data running off a prescribed safe interval during a prescribed lead time.

Consider the data following from Eq. (88), and plotted in Figures 28, 29, and 30, and the histogram in Figure 32, and assign the safe interval for  $y$ :  $|y| < 1$ . Calculating the area off the safety interval and taking into account that the total area of the histogram is equal to 1, one obtains that for leading time  $t > 20$ , the risk that the data will approach the safe boundary is 16%. If the safe interval is narrowed to  $|y| < 0.5$ , the risk increased from 16% to 23%.

It should be noticed that the same risk will stay for  $t > 20$  since after  $t = 20$ , the process becomes stationary.

Turning to the less stable data following Eq. (89), (see Figure 31), one obtains, for the leading time  $t = 20$  and the safe interval  $|y| < 0.5$ , the risk 35%.

### 3. Comparison with theory of stochastic processes.

From the viewpoint of theory of stochastic processes, Eqs. (88) and (89) represent a discretized version of the Langevin equation that, along with Newtonian forces, includes forces on thermodynamical scale (such as noise). The plots in Figure 28 illustrate the solutions to the corresponding Langevin equation for different values of noise. The sequence of histograms in Figure 32 illustrate the solution to the corresponding Fokker-Planck equation.

The general form of the corresponding one-dimensional Langevin equation is

$$\dot{x} + \gamma x = q\Gamma(t) \quad (90)$$

Here  $\gamma x$  represents a deterministic (Newtonian) force, and  $\Gamma(t)$  stands for stationary uncorrelated noise where

$$\langle \Gamma(t) \rangle = 0, \quad \langle \Gamma(t)\Gamma(t') \rangle = 2\delta(t - t') \quad (91)$$

and  $q = \text{const}$  is the strength of noise.

The closed form solution to this equation subject to the initial condition  $x = x_0$  at  $t = 0$  is

$$x(t) = x_0 e^{-\gamma t} + \int_0^t e^{-\gamma(t-t')} \Gamma(t') dt' \quad (92)$$

The solution is random due to the random noise contribution  $\Gamma(t)$ . That is why it is more convenient to deal with the corresponding Fokker-Planck equation describing the evolution of the probability distribution of the error due to noise. This equation reads

$$\frac{\partial \rho}{\partial t} = \gamma \frac{\partial}{\partial X} (X\rho) + q^2 \frac{\partial^2}{\partial X^2} \rho \quad (93)$$

Here  $\rho(X,t)$  is the error probability distribution.

The closed form solution to this equation subject to the sharp initial condition is

$$\rho = \sqrt{\frac{\gamma}{2\pi q^2 (1 - e^{-2\gamma(t-t')})}} \exp\left[-\frac{\gamma(X - e^{-\gamma(t-t')} X')^2}{2q^2 (1 - e^{-2\gamma(t-t')})}\right] \quad (94)$$

As follows from this equation, for large time

$$\gamma(t - t') \gg 1 \quad (95)$$

the solution approaches a stationary distribution

$$\rho_{\infty} = q\sqrt{\lambda/2\pi} \exp[-\gamma X^2 / (2q^2)] \quad (96)$$

while the larger the coefficient  $\gamma$ , i.e. the higher the degree of stability of the underlying dynamics, the sooner the process approaches its stationary distribution, and the lower the stationary value of the standard deviation, and that is in qualitative agreement with the results discussed above, (see Figures 30 and 32).

## Discussion and Conclusion

There are two basic problems in Engineering Science: Direct Problem and Inverse Problem.

A Direct problem is posed as following: given the mathematical model, find a solution. Usually the model is formulated in the form of differential equations derived from the First Principles of physics. From the viewpoint of information processing, Direct Problems are associated with data *decompression*: a compact formulation of the Newton law or of the Maxwell equation give rise to billions of data describing behavior of planets, hurricanes, electro-magnetic fields etc. In the past, Direct Problem was dominating in the Engineering Science, and it inspired development of several branches of mathematics associated with existence, uniqueness and stability of solutions.

An Inverse problem is posed as following: given experimental data, *discover* an underlying dynamical model. Usually most of experimental data are coming from sensors in the form of time series. Since recent years are characterized by large massifs of data obtained from NASA missions, Inverse problems are becoming more and more important. From the viewpoint of information processing, inverse problems are associated with data *compression*: billions of data are ‘hidden’ in the form of *discovered* dynamical models that are presented in a parameterized form, Fig. 35.

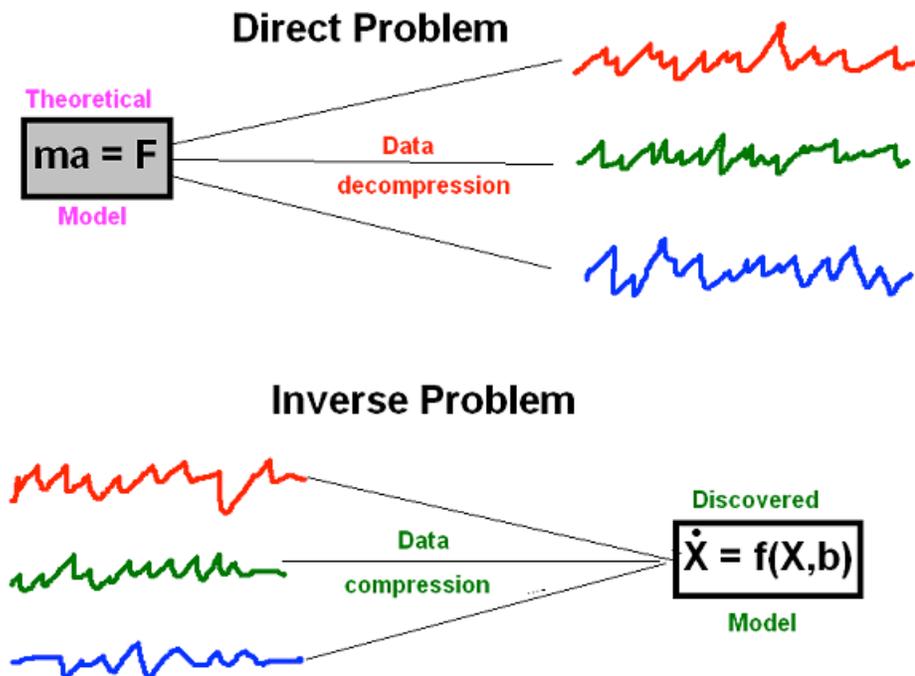


Figure 35.

The approach to sensor data forecast developed in this work is based upon inverse problem, and in particular, upon *discovery* of dynamical model that reproduces the original data to accuracy of statistical invariants. As soon as such model is ready, one can perform the following tasks:

1. Detect abnormal behavior of the object by observing changes in the model parameters, (See the parameters  $\mathbf{b}$  in Fig. 34).
2. Predict future object behavior by running the model ahead of actual time.
3. Perform risk assessment by evaluating uncertainty in data forecast along with the safety boundaries.
4. Analyze a hypothetical (never observed) object behavior by appropriate changes of the model parameters.

The methodology for all these tasks were developed and illustrated in these report.

The following tasks are ready for a start in future:

1. Expansion of the developed methodology to correlated sensors.
2. Development optimal strategy for reactive recovery by manipulating the model actuators.
3. Performance of preemptive control based upon data forecast.

The most challenging and still unsolved problems are:

1. Discrimination between abnormalities in the object and in its *sensors*.
2. Extension the methodology to include *human factor*.