time, and in place of the correlation function its time estimate is used

$$A_y(\tau) = \frac{1}{T-\tau}\int_0^{T-\tau} \overset{\circ}{y}(t)\,\overset{\circ}{y}(t+\tau)dt, \tag{2.76}$$

where T is the length of realization.

There is one-to-one correspondence between the correlation function and the power spectrum of the process; specifically, the power spectrum is the Fourier transform of the correlation function.

$$S_y(w) = \frac{1}{2\pi}\int_{-t}^{+t} A_y(\tau)e^{-jwt}d\tau. \tag{2.77}$$

In turn, the correlation function is defined in terms of the inverse Fourier transform,

$$A_y(\tau) = \int_{-\infty}^{+\infty} S_y(w)e^{jwt}dw; \tag{2.78}$$

i.e., the form of the correlation function depends essentially on the frequency spectrum of the original signal. The higher the frequency of the harmonics contained in that signal, the faster the correlation function decreases; a narrow spectrum corresponds to a broad correlation function and vice versa. In the limiting case, the correlation function of white noise is a *delta*-function with its singular point at the coordinate origin. Thus, the correlation function is a measure of the smoothness of the process being analyzed, and it can serve as a measure of the accuracy of prediction of its future values.

A relay autocorrelation function is called the sign-changing function $A'_y(\tau)$;

$$A'_y(\tau) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{+T} \overset{\circ}{y}(t)A\ \text{sign}\,[\overset{\circ}{y}(t+\tau)]dt. \tag{2.79}$$

Analogously, a relay cross-correlation function is given below.

$$K'_{xy}(\tau) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{+T} \overset{\circ}{y}(t)A\ \text{sign}\,[\overset{\circ}{x}(t+\tau)]dt. \tag{2.80}$$

Relay autocorrelation functions reflect only the sign and not the magnitude of $x(t)$. They have properties analogous to those of ordinary correlation functions, and in particular they coincide with them in sign. The advantage of relay functions (auto- and cross-correlations) is in the simplicity of the apparatus used for obtaining them. When the phase of the function $y(t)$ changes by 180°, the sign of the correlation function reverses. This means that in extremal regulation systems the correlation functions (ordinary or relay) can be used for determining which side of an extremum the system is on.

In practical computations associated with the random processes, one frequently estimates the so-called correlation interval, which is the time TV, over which the statistical connection between sections of the process is kept—in that the correlation moment between these sections exceeds some given level; for example, $|A(\tau)| > 0.05$ (Figure 2.8a).

Sometimes the meaning of the correlation interval is taken as the rectangular height $A(0)$ with area equal to the area under the correlation function (Figure 2.8b).

$$\tau_c = \frac{1}{A(0)}\int_{-\infty}^{+\infty} A(\tau)d\tau. \tag{2.81}$$

This is a convenient definition in case of a nonnegative correlation function.

The correlation time or interval is also defined as half the base of a rectangle of unit height whose area is equal to the area under the absolute value of the correlation function (Figure 2.8c).

$$\tau_c = \frac{1}{2}\int_{-\infty}^{+\infty} | A(\tau) | \, d\tau. \tag{2.82}$$

Among these three definitions we shall use the first one because of its simplicity.

3.2 Correlation interval as a measure of predictability

Various types of mathematical details (language) of modeling can be used. The influence of the degree of detailedness (sharpness) of the modeling language on the modeling accuracy—or in case of prediction, the limits of predictability of the process—is of great interest. One of the simplest devices for changing the diffuseness of description of a time series is to change the intervals of averaging (smoothing) of the data (for example, mean monthly, mean seasonal, mean annual, mean 11 years, etc.). The spectrum of the process in question then narrows down to the original and its correlation function broadens; that is, the correlation interval increases. This in turn extends the scope of predicting the process.

The problem encountered now is how to estimate, at least approximately, the achievable prediction time. The maximum achievable prediction time T_{pmax} of a one-step forecast is determined by the correlation interval time called coherence time τ_c of the autocorrelation function A_y. This time is equal to the shift that reduces the autocorrelation function (or its envelope) to a value determined by the allowed prediction error *8%* following this level which it no longer exceeds.

The maximum allowed prediction time of a multiple (step-by-step) forecast is equal to the coherence time multiplied by the number of steps; i.e., $T_{pmax} = n\tau_c$. The prediction error increases with each integration step, which imposes a definite limit on the step-by-step forecast. We give here a brief view on the maximum capabilities of multiple step-by-step prediction, assuming that they are determined by the coherence time in the same way as those for one-step prediction.

Because of one-to-one dependence between the correlation and spectral characteristics of a random process, one can use some limiting correlation frequency as a measure of process predictability instead of correlation interval. The spectrum amplitude for the limiting correlation frequency is less than some threshold $S(w) < 0$. Obviously these measures of diffuseness of the modeling language are not universal and are suitable only for evaluating certain mathematical modeling languages—primarily languages differing as regards the interval of averaging of the variables.

Example 1. Let us look at the influence of the interval of averaging on the form of its correlation function, its interval, and hence on the limit of its predictability; the example given here is an analysis on outflow $q(t)$of a river over a period of one hundred years [44]. The autocorrelation functions for different averaging times are constructed.

$$A_q(\tau) = \frac{1}{N-\tau}\sum_{i=1}^{N-\tau} q_i q_{i-\tau}, \tag{2.83}$$

where q is the mean monthly outflow, N is the number of data points, and r is the step in computation of the correlation function. It shows that averaging of variables in time increases the coherence time, in the same way as averaging time interval of variables over the surface of the earth, as shown in Figure 2.10.

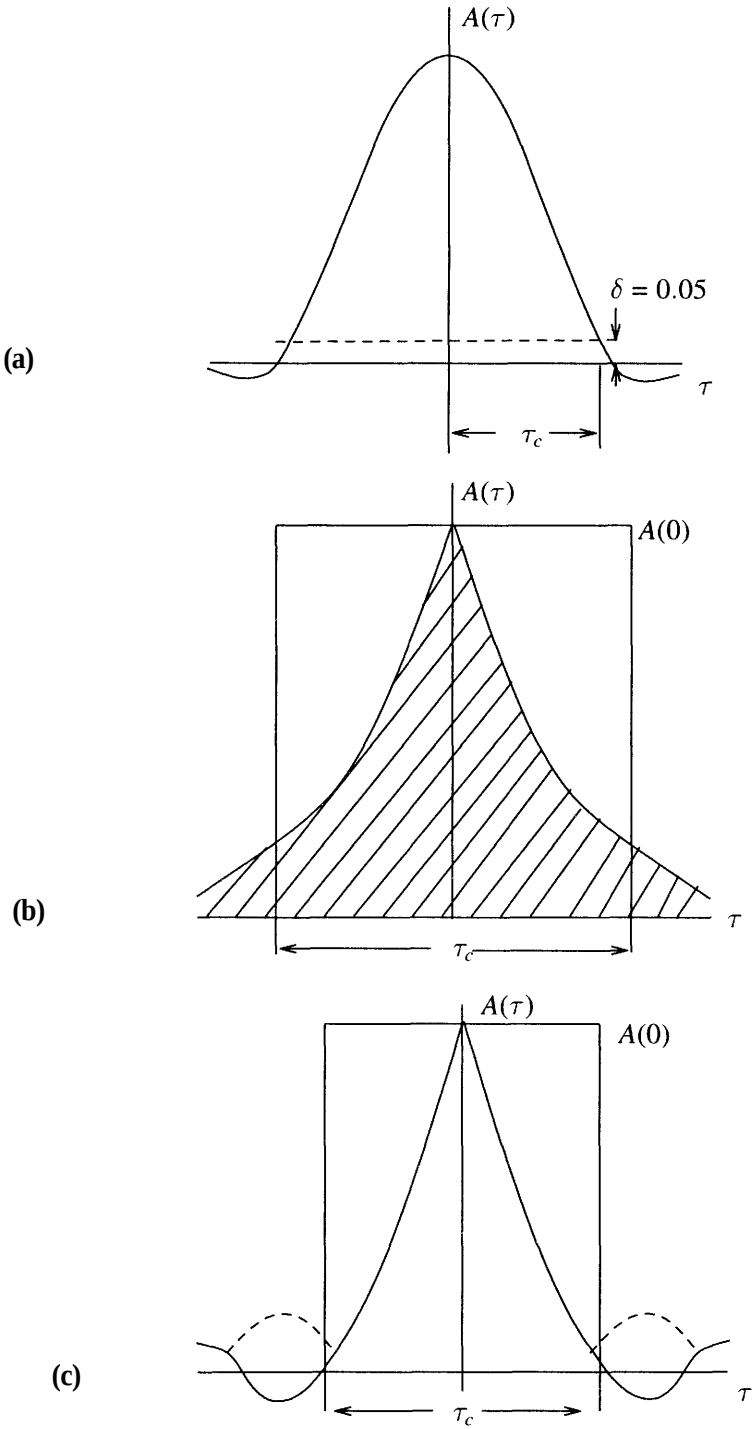

Figure 2.8. Three versions of defining the correlation interval

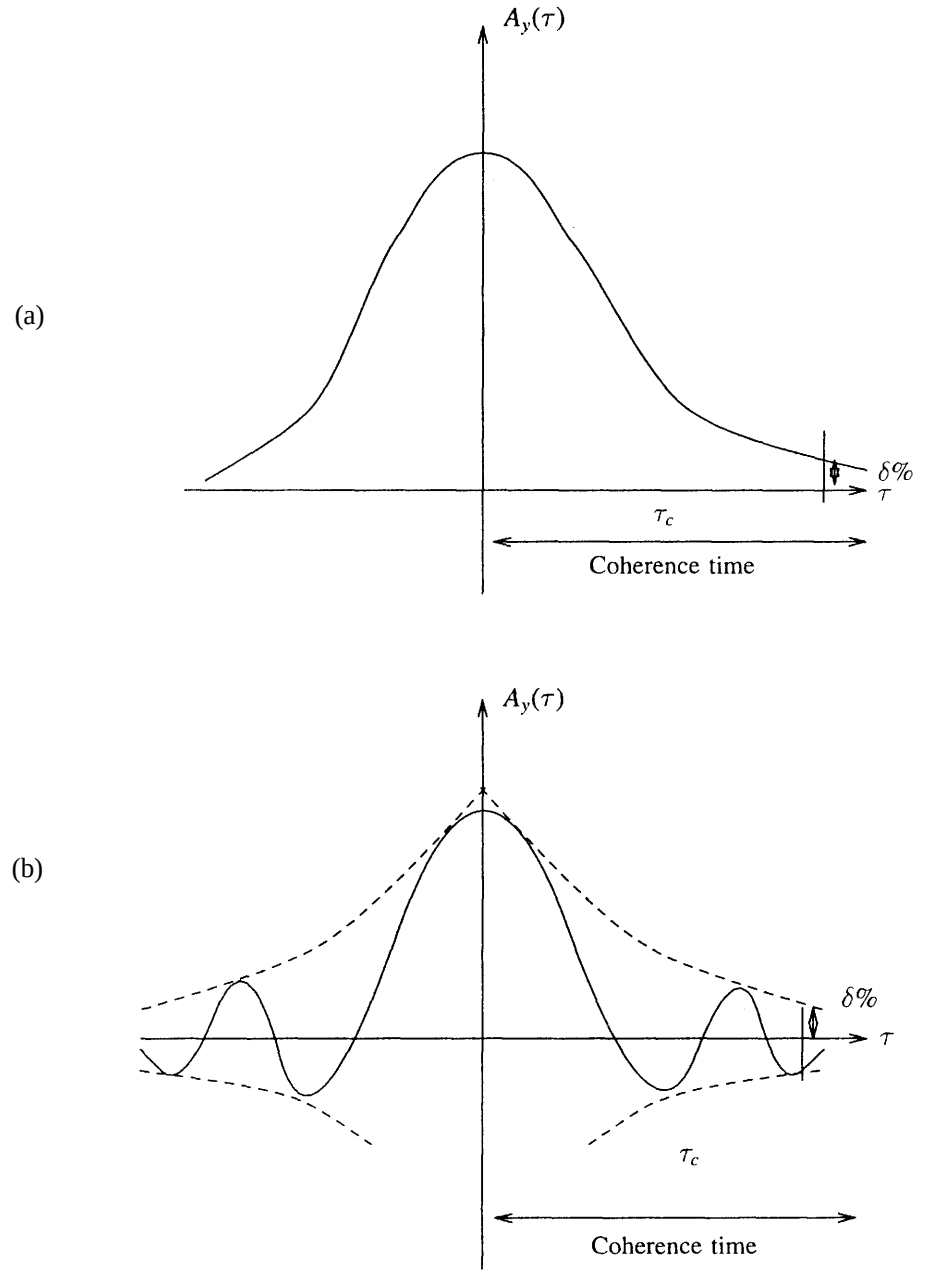

Figure 2.9. Autocorrelation functions; (a) monotonically decreasing and (b) oscillating

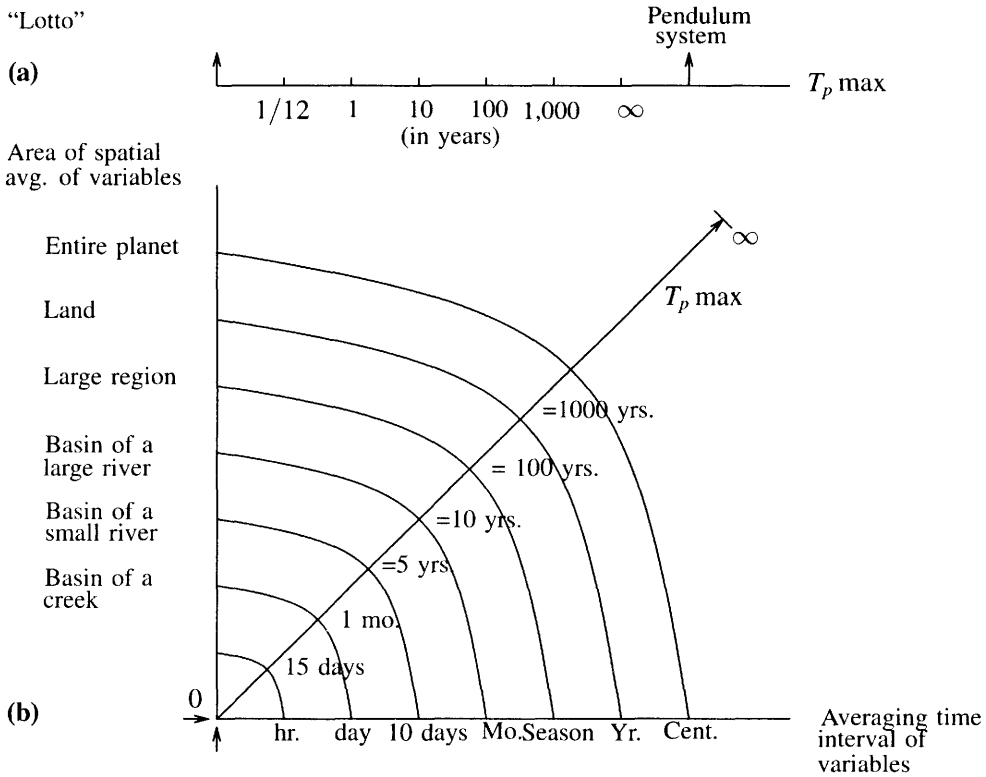

Figure 2.10. Qualitative variation of maximum prediction validity time T_{pmax} as a function of object properties and averaging interval of variables; (a) axis of maximum prediction time with constant averaging, (b) location of axis (a) in the plane of time and space averages

It is appropriate to remember that the achievable prediction time of a forecast depends not only on the averaging interval of variables, but also on physical properties of the process being predicted, as well as on the quality and characteristics of the mathematical prediction apparatus. If an exact deterministic description of the process is known, then prediction is reduced to detailed calculations.

For example, the motions of planets can be predicted exactly for long time intervals in advance. Outputs of a generator of random numbers or the results of a "lotto" game cannot be predicted as a matter of principle. These two examples are extreme cases corresponding to "purely" deterministic objects and "purely" random objects with equiprobable outcomes. In actual physical problems we are always located somewhere between these two extremes (Figure 2.10a).

The autocorrelation function of a process with its coherence time contains some information on its predictability (the degree of determinancy or randomness). The analysis of autocorrelation functions indicates that by increasing the averaging interval of variables in time or space we can, so to speak, shift the process from the region of unpredictability into the region of exact and long-term calculability. Figures 2.11a and b demonstrate the autocorrelation functions for one with calendar averaging and another with moving averages on the empirical data of river outflow.

One can see that with the increase in the interval of averaging of the data, the correlation function for a single time scale becomes ever more sloping, and the correlation interval increases. In the moving average case, a smaller step of sampling the initial data enables

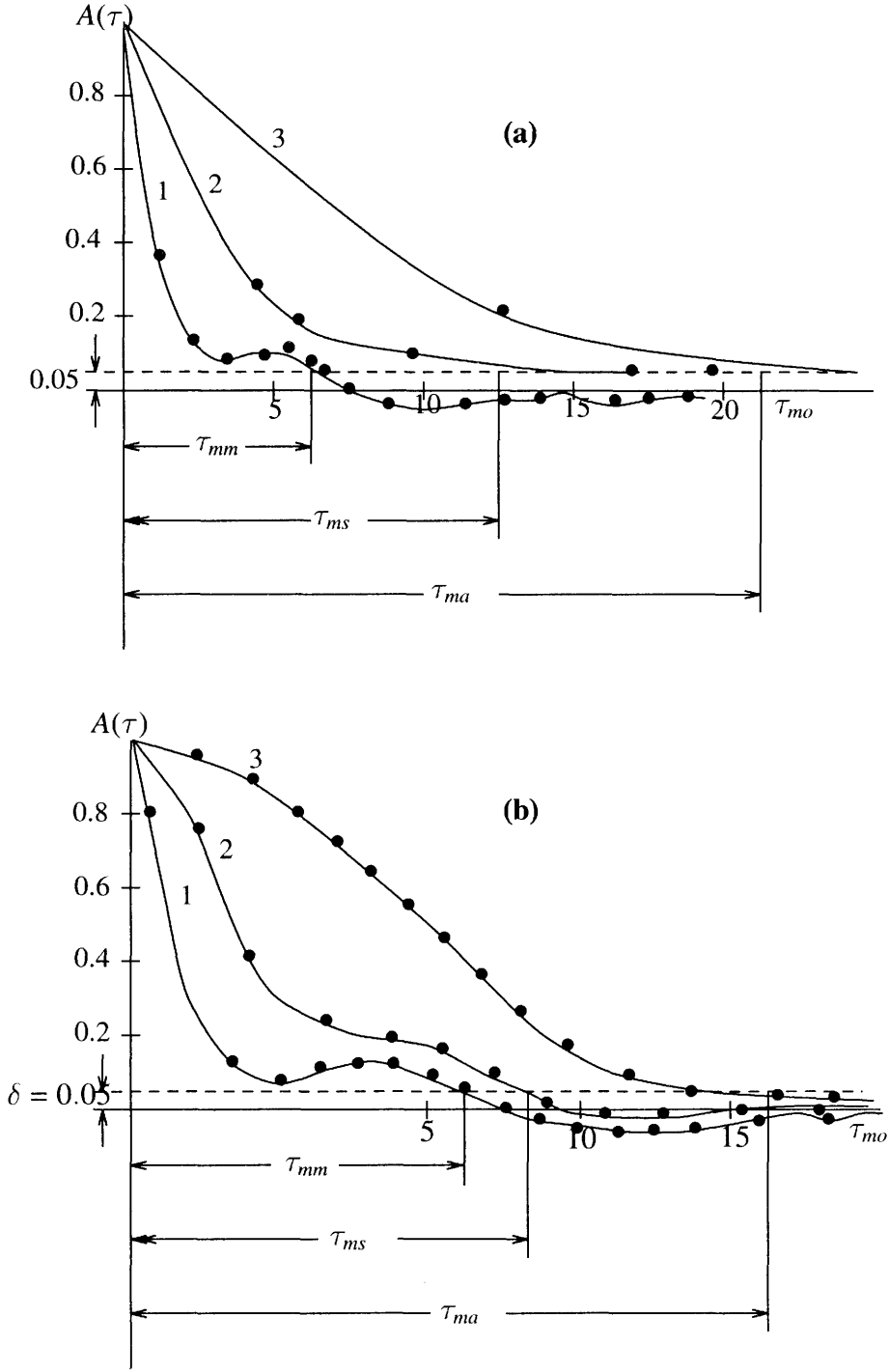

Figure 2.11. Autocorrelation functions of a river outflow; (a) with calendar averages and (b) moving averages on (1) monthly data, (2) seasonal data, and (3) annual data

us to keep unchanged the number of sample data (all monthly values), which leads to a broadening of the spectrum of the original signal and to a corresponding narrowing of its correlation function. The correlation function obtained in the case of moving averages occupies an intermediate position between the correlation functions of unsmoothed data and the data of calendar smoothing. Thus, the correlation time can serve not only as a measure of the limit of predictability of the process, but also as a measure of detailedness of a number of modeling languages.

Example 2. In the harmonic algorithm the trend is represented as a sum of a finite number of harmonic components (usually the optimal number of components does not exceed $m = 20$).

$$q_{mo}(t) = \sum_{i=1}^{m} (A_i \cos w_i t + B_i \sin w_i t), \tag{2.84}$$

where $q_{mo}(t)$ is the mean monthly data.

Running moving average is an approximation of the operation of integration over a given interval of time.

$$q_{run}(t) = \frac{1}{3}\int_0^3 q_{mo}(t)dt \cong \frac{1}{3}\sum_{j=0}^{3}\sum_{i=1}^{m}(A_i \cos w_i t + B_i \sin w_i t), \tag{2.85}$$

where $q_{run}(t)$ is the running average of three on the mean monthly data.

Integration does not change the number of harmonics to be added or their frequencies, but it does decrease the amplitudes by a factor of $1/w_i$. As a result, the components with comparatively high frequencies decrease more than the others, and the curve $q_{mo}(t)$ becomes much smoother than the original curve. This also explains the smoothing effect shown in the figures above. The same reasoning holds true for the curves of correlation functions for seasonal and yearly data.

Example 3. This is demonstrated using the same data by constructing algebraic, differential (difference) and integral type models on an interval of 20 years.

algebraic model:

$$q = a_0 + a_1 t + a_2 t^2;$$

differential model and its difference analogue:

$$\frac{dq}{dt} = a_1 + a_2 t,$$
$$\Delta q = q_{+1} - q_0 = b_0 + b_1 t;$$

integral model and its discrete summation analogue:

$$\int_0^t q\,dt = a_0 t + a_1 \frac{t^2}{2} + a_2 \frac{t^3}{3} + C,$$
$$\sum_0^n q = b_0 t + b_1 \frac{t^2}{2} + b_2 \frac{t^3}{3} + C_1. \tag{2.86}$$

The autocorrelation functions for these three types of models over the 80-year period are shown in Figure 2.12.

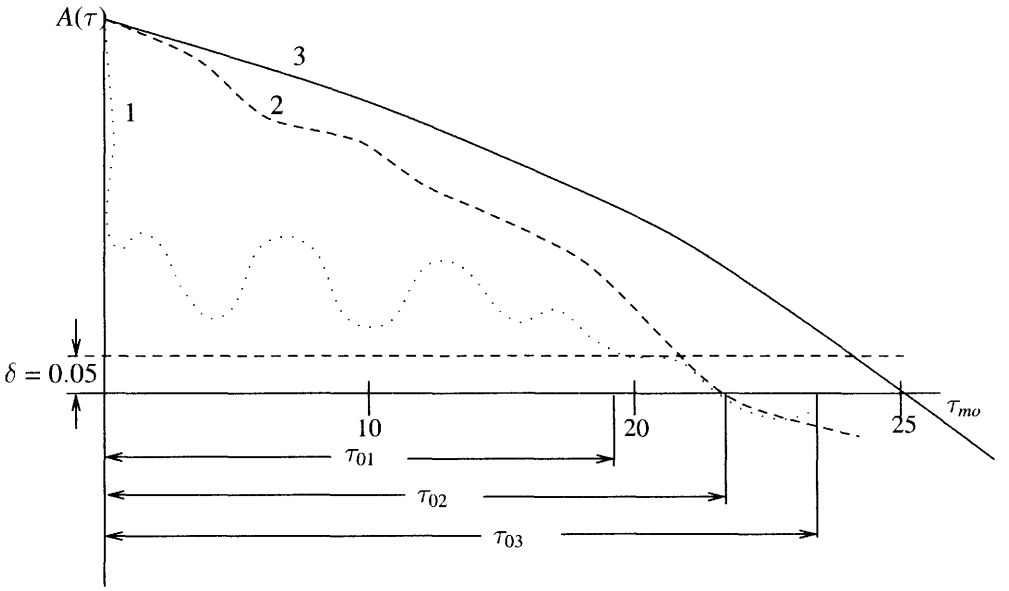

Figure 2.12. Autocorrelation functions for languages of (1) integral, (2) algebraic and (3) differential equations

One can see that the language of differential equations is the most diffuse of the three modeling languages; it is more suitable for long-range predictions. This explains the widespread use of differential equations in the equivalent analogue of finite-difference equations in modeling as compared with algebraic and integral models.

Let us take the problem of *weather forecasting.* Weather forecasters use data gathered by satellite in order to predict the weather quite successfully over an extended period of time, but this prediction is only possible in terms of a very general language. They convey the future weather picture qualitatively ("it will be warmer," "precipitation," "cold," etc.). More quantitative predictions require the use of mathematical models. As per various studies it is indicated that the daily prediction interval cannot exceed 15 days and practical predictions have even shown for a much shorter interval of time (not more than 3 to 4 days). The mean monthly values of variables are less correlated than the average daily variables; the maximum length of the prediction interval of mean monthly values does not exceed 3 to 4 months. Average yearly values of variables have an intermediate degree of correlation, and the maximum achievable prediction interval of average yearly values is 8 to 10 years. It is important to point out that the limit imposed on the interval of prediction, measured in the same units of time, increases together with the interval over which the variables are averaged. In other words, the interval span for average daily values is 15 days, the span for average monthly values is 4 x 30 = 120 days, and the interval for average yearly values is 10 x 365 = 3650 days, etc.

Reliable long term predictions of weather are frequently related to the idea of analogues. This idea is simple and interesting: one must find an interval in the prehistoric measured data whose meteorological characteristics are identical to the currently observed data. The future of this interval (observed in the past) will be the best forecast at present. Nevertheless, attempts to apply the idea of analogues always produced results that were not very convincing. The fact is that for such a large number of observed variables (and also many unobserved ones) it is impossible to find exact analogues in the past. Resorting to group

analogues, introduction of weighing coefficients for each measurement, and other measures first bring us to regression analysis and then, after further improvements, to the inductive approach algorithms. Therefore, inductive learning can be interpreted as an improved method of group analogues in which the analogues of the present state of the atmosphere are selected by using special criteria and summed up with specific weighing coefficients to produce the most probable forecast. Weather forecasting is an object whose structure switches when a new type of circulation is established randomly at the time of equilibrium. Nevertheless, it is possible to investigate an optimum method for overcoming the predictability limit applicable to some weather variables (temperature and pressure at surface layer, etc.). This will be discussed in later chapters. Further research is needed on this subject.

It seems that insurmountable barriers have been established for quantitative predictions. However, the self-organization method enables one to overcome these limitations and to solve the problem of long-term predictions, because the limit of predictability depends on the time interval of averaging. Self-organization uses two or three averaging intervals for correcting the variable under study; for example, the daily prediction is corrected according to a 10-day prediction, the 10-day prediction is corrected according to the mean monthly prediction, and the mean monthly prediction is corrected in accordance with the average yearly prediction. In this way we can achieve a breakthrough in methods of long-term and very long-term prediction which has heretofore not been achievable by any other method.

3.3 Principal characteristics for predictions

The principle characteristic of achieving an objective goal is for detailed (sharp) predictions in a low-level language which contain the greatest amount of detail while maintaining the prediction lead time that is typically obtained by using the most general high-level language. The more general the language, the longer the achievable prediction lead time (Figure 2.10).

Let us give here some examples indicating the levels of languages:

(i) *Prediction of processes in economic and ecological systems.*

A language which preserves probabilistic moments of the process is used at the upper level to select quantitative predictions by using the mean annual values of variables and the mean seasonal or monthly values. The middle-level language consists of modeling mean annual values and the lower level (detailed) consists of modeling average seasonal or monthly values.

(ii) *Prediction of river flows,*

The upper level uses the language which preserves the nature of probability of distributions, the middle level consists of predictions of average annual run-off, and the lower level involves predictions of average seasonal or monthly values. The conversion from statistical to quantitative predictions should be performed by taking into account the principle—that is, by using rationalized (multilevel) scanning of quantitative predictions.

(iii) *Long-term weather forecasting.*

The upper level can be a language which preserves the weather forecast for a large region (or a long averaging time). The middle level will then consist of predictions for small parts of the region (or medium averaging time), and finally the lower level will give predictions for a specific point and specific time.

The examples given above contain three levels of detailedness of the modeling language, which is obviously not required for all problem-solving tasks.

As we know, the principle of self-organization is realized in single-layer (combinatorial) and multilayer inductive learning algorithms. Using the basic structures of these algorithms, multilevel prediction algorithms are operated in several different languages simultaneously,

within which the predictions expressed in a more general language are used for selection of an optimum quantitative prediction in the more detailed language. Several levels are needed to overcome the "limit of predictability" of detailed predictions, and also to eliminate the multivalued choice of a prediction on the basis of general criteria. Let us go through different cases of self-organization modeling for clarity in multicriterion analysis.

Case of exact data

In case of exact data, exact computation takes place for prediction (for example, motion of heavenly bodies, prediction of eclipses, etc.) from the solution of a system equations as mathematical models of the cosmic system of bodies.

Under the conditions of exact empirical data, self-organization modeling can only have as its purpose the discovery of laws hidden in the data. It is sufficient to use any one internal or external criterion like regularity or minimum bias criterion in sorting out the models. It is important to note that we do not require multicriterion choice of a model. More complex problems arise within the field of noisy data.

Case of noisy data

It is sufficient to impose on one of the variables (usually the output) a very small additive or multiplicative noise so that the position of the variable is changed cardinally. If we try to obtain an optimal model using only internal criteria, we always end up with a more complex model, that will be more accurate in the least squares sense; only external criteria provide a model with optimal complexity. Let us consider various systems of equations describing an object; they are not equally valuable since they are connected with measurement of different variables. The optimal system with the fewest excessively noisy variables can be sorted out among variants of the system of equations using the system criterion of minimum bias:

$$\eta_{s_{(bs)}} = \frac{1}{s}(\eta_{bs_1} + \eta_{bs_2} + \cdots + \eta_{bs_s}), \tag{2.87}$$

where $\eta_{s_{(bs)}}$ is the system criterion for the system of equations and the η_{bs_i}, $i = 1, 2, \cdots, s$ are the criteria for each equation in the system of s equations.

As we know from the information theory point of view, increasing the noise stability decreases the transmission capacity; this means that with an increase in the noise level, a model simpler than a physical model becomes optimal. (Here physical model means a model corresponding to the governing law hidden in the noisy data.) It is expedient to distinguish two kinds of models: (i) a physical or identification model which is suitable for analysis of interrelations and for short-range predictions, (ii) a nonphysical or descriptive model for long-range predictions. One can discover a physical model with various concepts of modeling, but detailed long-range predictions are impossible without the help of inductive learning.

If the data are noisy, even to obtain a physical model requires one to organize rational sorting of physical models by self-organization using several criteria which have definite physical meanings. Usually one needs a model which is not only physical but also easy to interpret instantaneous unaveraged values of the variables; that means the model is chosen based on the simultaneous selection of minimum bias criterion and short-range prediction criterion.

$$\eta_{bs}^2 = \sum_{i=1}^{N} \frac{(\hat{y}_A - \hat{y}_B)_i^2}{y_i^2}, \quad \Delta^2(C/W) = \sum_{i=1}^{N_c} \frac{(y - \hat{y})_i^2}{(y_i - \bar{y})^2}, \tag{2.88}$$

where y *is* the output variable, y_A and y_B are the estimates of the models obtained based on the sets A and *B,* respectively, y is the estimated prediction, and y is the average value of y.

In the plane of two criteria, each model corresponds to its own characteristic point; the point corresponding to the model of optimal complexity lies closer to the coordinate origin than do the points of other models participating in the sorting. Here we can say that one can find a physical model using both deductive reasoning of man and self-organization of machine with respect to choice of many criteria.

In obtaining nonphysical models for long-range detailed predictions, the role of man, as he remains the author of the model, consists of supplying the most efficient set of criteria for sorting the models. The dialogue between man and machine is in the language of criteria and not in the language of exact instructions. In addition, to use the minimum bias criterion on two sets of data A and *B,* the step-by-step prediction criterion is to be included for calculating the prediction error on entire interval *(W = A + B)* of data. The above short-range prediction criterion $\Delta(C/W)$ is used as long-range prediction criterion $i(W)$ as per notation by replacing N_C with N_W for the entire range of data points. This criterion is desirable to use not only for choosing the structure of the model but also for removing the bias of the estimates of the coefficients in the model. In addition to these criteria, in multicriteria choice of an optimal nonphysical model for long-range predictions, stability criteria of moments (upper and lower) and probabilistic characteristics of correlation functions are used; these will be explained later in the chapter. This means that multicriterion choice is one of the basic methods of increasing noise stability of inductive learning algorithms.

The physical and nonphysical models differ not only in their purpose but also in their informational basis because of reasoning of the objective criteria. The arguments of physical model can be all input variables and their lagged values (for dynamic models). The arguments of nonphysical predicting models can only include different intervals of averaging and the time variables which are known on the entire interval of long-range prediction. Physical models that are obtained are usually linear and nonphysical models are nonlinear with respect to time.

Case of time series data

If an algorithm is used for obtaining a single "optimum" prediction (according to any criteria) using pre-history data, then such algorithm is meant for only short-range or average-term prediction (for one to two or three to five time intervals in advance respectively). If the algorithm envisions the use of empirical data in order to obtain a single prediction over a large averaging interval (for example, one year), and several predictions (in accordance to multicriteria) over a small averaging interval of variables (for example, seasonal) in order to use the balance criterion over the interval of predictions (ten to 20 years in advance), then the choice of seasonal models on the basis of yearly model is done on the basis of balance-of-predictions criterion [58], [65].

$$B^2_{season} = \sum_{i=1}^{N_C} b_i^2$$

$$b_i = \hat{Q}_{yr} - \frac{1}{4}(\hat{q}_w + \hat{q}_{sp} + \hat{q}_{su} + \hat{q}_f)_i, \qquad (2.89)$$

where N_c is the number of prediction points, $\hat{Q}_{yr}$ is the prediction based on the yearly model (a single prediction), $\hat{q}_w, \hat{q}_{sp}, \hat{q}_{su}$, and $\hat{q}_f$ are predictions based on different variants of the set of seasonal models for winter, spring, summer, and fall correspondingly, and N_c is at the range of prediction interval of ten years.

In the same fashion one can build an algorithm which envisions over a very long averaging interval (for example, 11 years) and at the same time several predictions over shorter averaging intervals (for example, one year or one season); if the algorithm uses a two-level balance-of-predictions criterion, then that would be successful for very long-range predictions (40 or more years in advance) [58]. The choice of the yearly models and the model which uses the averaging interval of 11 years is based on the following balance-of-predictions criterion:

$$B^2_{11yrs} = \sum_{i=1}^{N_c} b_i^2$$

$$b_i = \hat{Q}_{11yrs} - \frac{1}{11}(\hat{q}_1 + \hat{q}_2 + \hat{q}_3 + \cdots + \hat{q}_{11})_i, \tag{2.90}$$

where $\hat{Q}_{11yrs}$ is the prediction based on the model which uses the averaging interval of 11 years (a single prediction); $\hat{q}_1, \hat{q}_2, \hat{q}_3, \cdots \hat{q}_{11}$ are predictions based on various versions of the set of yearly models.

The rules for building up such algorithms realize the principle of "freedom of choice of decisions" formulated by Gabor [22]. The basic long-term prediction is harmonic or polynomial prediction of variables when the averaging interval is of maximum length. The criterion of prediction balance "pulls up" the accuracy and the averaging time of predictions for small averaging intervals to the accuracy and prediction time obtained when the averaging interval is long.

Another issue where the self-organization stands firm is when a decision is to be made in case of two or more contradictory requirements, which is called "Pareto problem." The "Pareto region" is the region where the solutions contradict each other and which requires the use of experts. This is achieved by the self-organization method yielding a new problem formulation of multicriterion control selection done heuristically on the basis of physical properties of the system to be predicted. The lead time of prediction interval usually reaches the time of interval used for validity of the criterion. In order to eliminate multivalued selection, scanning of forecasts for different intervals is replaced by multilevel algorithm development as scanning of algorithms and models, generating a variety of predictions on the basis of their external criteria.

4 DIALOGUE LANGUAGE GENERALIZATION

Complex systems analysis is based on modeling of a system with interactive elements in order to identify the system structure and parameters, to perform various tasks like short- and long-term predictions of processes, and to optimize the control task. Usually during algorithm development, the computer has a passive role; that is, it is unable to participate in creative modeling. Interpolation problems are multi-solution problems; additional data set or *a priori* testing set is necessary to obtain a unique solution. Commonly used simulation methods are based on a large volume of *a priori* information that is difficult to obtain.

Self-organization modeling is directed to reduce *a priori* information as much as possible. The purpose of self-organization is not to eliminate human participation (it is impossible unless a complete intelligence model is developed), but to make this participation less laborious, reduce some specific problems, and avoid expert participation. This can be achieved in ergatic information systems by using more generalized "man-machine" meta-language, which uses general criteria given by man—the learning is done by the computer. In addition to the generalized criteria, man provides the empirical data. In some cases man may be involved in final model corrections. Here it is shown that many things still can

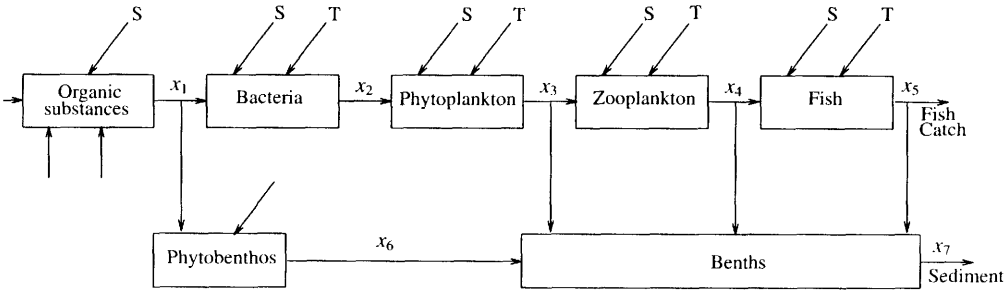

Figure 2.13. Usual (subjective) system analysis (example)

be done to reduce human involvement in the creative modeling process and make it much easier.

4.1 Regular (subjective) system analysis

The regular (subjective) system analysis shown in Figure 2.13 is a system imitation model for the northwest region of the Black Sea [33] (details of the model are not shown to simplify description).

An organic substance is formed (variable x_1 is substance production) from biogene substances P and N that were exposed to the sun. Bacteria (biomass x_2) eat the organic substance, and the phytoplankton (x_3) eat the bacteria. Zooplankton (x_4) eat the phytoplankton and the fish (ichthyomass x_5) eat the zooplankton. If we are interested in analyzing fish catch, the following equations (for the surface layer of water) are used.

$$\begin{aligned}
\tau_1 \frac{dx_1}{dt} + x_1 &= a_0 + a_1 S + a_2 P + a_3 N, \\
\tau_2 \frac{dx_2}{dt} + x_2 &= b_0 + b_1 S + b_2 T + b_3 x_1, \\
\tau_3 \frac{dx_3}{dt} + x_3 &= c_0 + c_1 S + c_2 T + c_3 x_2, \\
\tau_4 \frac{dx_4}{dt} + x_4 &= d_0 + d_1 S + d_2 T + d_3 x_3, \\
\tau_5 \frac{dx_5}{dt} + x_5 &= l_0 + l_1 S + l_2 T + l_3 x_4.
\end{aligned} \tag{2.91}$$

By excluding the intermediate variables, we derive a linear differential equation of the fifth order for the output variable x_5 (fish catch) and analyze its solutions for given initial conditions. If the equations are nonlinear, then we can substitute the derivatives by finite-differences and find the results using simultaneous step-by-step integration of the system of nonlinear equations.

This example shows the basic characteristics of imitation modeling and commonly used system analysis: (i) this model requires in-depth knowledge of the subject; this knowledge is based on a large volume of information that is entered in the computer by the modeler; (ii) empirical data are not needed but may be used for scaling the coefficients by using the least squares method. One may design and analyze this model using a calculator; (iii) the results of such knowledge are subjective because the model is based on the author's subjective understanding (there may be as many different models as many modelers). The model does not resolve scientific disputes between experts on the subject; and (iv) only physical models can be obtained, but these are not suitable for long-term predictions.

4.2 Multilevel (objective) analysis

The idea of sorting many variants using some set of external criteria in the form of an objective function in order to find a mathematical model of a given complex subject seems unreal. Self-organization method tries to rationalize such sorting so that an optimal model is achieved. Multilevel algorithgms of inductive learning serve just this purpose. They allow changes of large number of variables to be considered. The model structure, which is characterized by the number of polynomial elements and its order, is found by sorting a large number of variants and by estimating the variants according to specific first level selection criteria (regularity, minimum bias, balance of variables and others). If the objectivity of the model is not achieved, then the high level criteria are used.

Here we give the concept of multilevel objective analysis under various conditions of multicriteria. The single-level analysis using one of the basic network structures like combinatorial, multilayer or harmonic is sometimes not sufficient for detailed analysis and we go for multistage analysis which is described as a multilevel algorithm. These prediction algorithms operate in separate different languages simultaneously as the predictions at a general language are used for obtaining a more detailed model at the next detailed language. Several levels are very essential, as one is to overcome the limit of predictability of detailed predictions and another is to avoid the multivalued choice of a model using the general criteria. Thus, in the stages of these algorithms, three basic directions of dialogue language are preserved; (i) the self-organization principle, which asserts that with gradual increase in the complexity of model, the external criteria pass through their minima, enabling us to choose a model of optimum complexity, (ii) an algorithm for multilevel detailed long-range predictions, and (iii) an algorithm for narrowing the "Pareto region" in case of multi-criterion choice of decisions.

4.3 Multilevel algorithm

The multilevel system is subjected to all the general laws governing the behavior of multilevel decision-making systems which realize the principle of incomplete induction. As in multilayer algorithm, here there is possibility of losing the best predictive model; an increase of the "freedom of choice" decreases the possibility of such loss. Various principles related to selection and optimization of "freedom of choice" in multilayer algorithm also apply to the multilevel system of languages having different levels of details.

If we had a computer with large capacity, then the problem of selecting detailed models could be solved by simply scanning all versions of partial models using combinatorial algorithm with a large ensemble of criteria. Since the capacity is limited, it is necessary to expose the basic properties of the models step by step.

In order to reduce the volume of scanning and to achieve uniqueness of choice, the principle discussed above is realized in several levels whose schematic structure for one version is shown in the Figure 2.14.

Let us explain the operations performed during these levels.

Objective systems analysis

The purpose of this level is to divide the system variables into output, input variables and variables which have no substantial effect on the outputs. Here structure of and number of equations is to be chosen in such a way that the overall model is consistent. The structure as well as number of equations must not be changed significantly when a new data set is added. The estimation of coefficients should not be changed. This type of sifting for

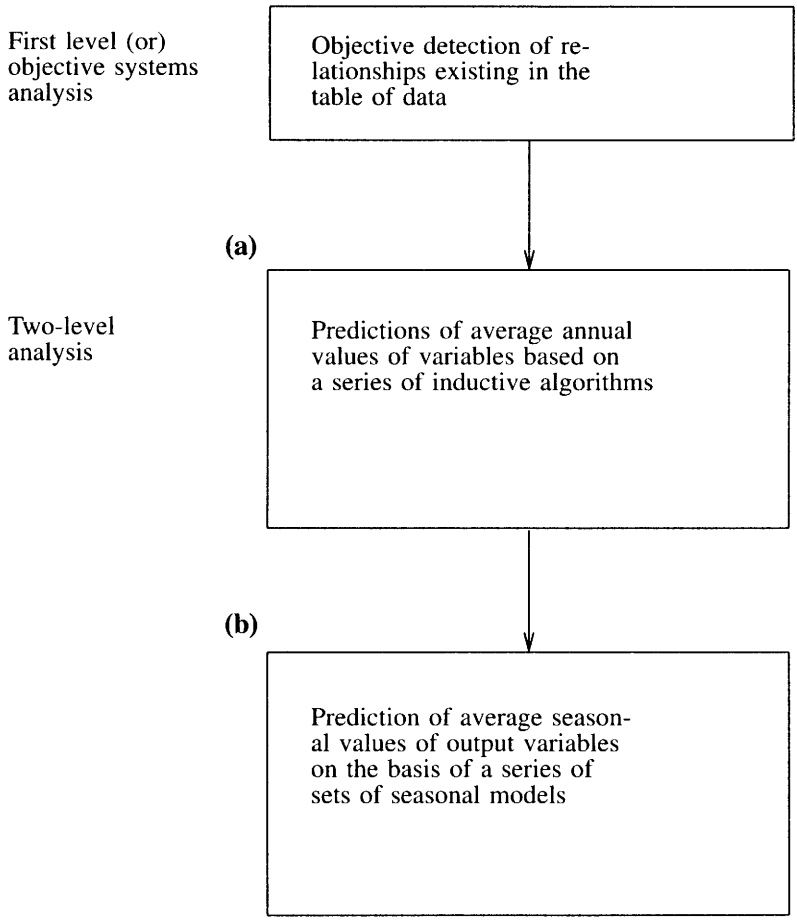

Figure 2.14. Schematic structure of a multilevel algorithm for longterm prediction

systems of equations is done systematically; that's why this level is called an objective systems analysis (OSA). An objective model is identified as a set of output variables and the connections between the system components as a result of learning. The analysis involves testing of several hypotheses about the model structure below:

First layer: The following hypothesis is tested for a single equation using empirical data. There are $M_1 = n$ equations formed, one for each variable.

$$x_i^t = a_0 + a_1 x_i^{t-1} + a_2 u, \tag{2.92}$$

where $x_i, \quad i = 1, \cdots, n$ are the state variables, u is an external influence, and superscripts t and $t-1$ indicate arguments with no delay and one-step delay, correspondingly. There are two methods to determine the variables of external influences u. In the first one, experts specify the disturbances *a priori* before execution of the program and in the second, the suitable control disturbances are chosen from the variables, which are already sorted out in the program. However, the role of these variables is considered as less important.

F_1 best models are obtained using the minimum-bias criterion.

$$\eta^2_{bs_i} = \sum_{p \in N} (\hat{x}^A_i - \hat{x}^B_i)^2_p, \quad i \in n, \tag{2.93}$$

where N is number of observations of empirical data, $\hat{x}^A_i$ is the estimated output of the model based on the training set A, and $\hat{x}^B_i$ is the estimated output of the model based on the testing set B.

Second layer: The hypothesis is tested for the structure of models described by two equations including the delayed and nondelayed arguments (one can use more delayed arguments).

$$\begin{aligned} x^t_i &= a_0 + a_1 x^{t-1}_i + a_2 x^t_j + a_3 x^{t-1}_j + a_4 u, \\ x^t_j &= b_0 + b_1 x^t_i + b_2 x^{t-1}_i + b_3 x^{t-1}_j + b_4 u, \end{aligned} \tag{2.94}$$

where $i, j = 1, 2, \cdots, n$; $i \neq j$. There are $C^1_{n-1} = n - 1$ equations for each variable, and overall there are a total of $n.C^1_{n-1} = n(n-1)$ equations. The system obtains F_2 best models of optimal complexity from among all two-set models (M_2) of state variables i, j using the system criterion of minimum-bias

$$\eta_{s_{(bs)}} = \frac{1}{2}(\eta_{bs_i} + \eta_{bs_j}), \tag{2.95}$$

where η_{bs_i} and η_{bs_j} are minimum-bias estimates of ith and jth equations.

Third layer: The system models consisting of three equations is found at this layer.

$$\begin{aligned} x^t_i &= a_0 + a_1 x^{t-1}_i + a_2 x^t_j + a_3 x^{t-1}_j + a_4 x^t_k + a_5 x^{t-1}_k + a_6 u, \\ x^t_j &= b_0 + b_1 x^{t-1}_j + b_2 x^t_i + b_3 x^{t-1}_i + b_4 x^t_k + b_5 x^{t-1}_k + b_6 u, \\ x^t_k &= c_0 + c_1 x^{t-1}_k + c_2 x^t_i + c_3 x^{t-1}_i + c_4 x^t_j + c_5 x^{t-1}_j + c_6 u, \end{aligned} \tag{2.96}$$

where $i, j, k = 1, 2, \cdots, n$; $i \neq j \neq k$. There are C^2_{n-1} equations for each variable and there are a total of $n.C^2_{n-1}$ equations. All three-set models (M_3) of variables (i, j, k) are evaluated using the system criterion of minimum-bias.

$$\eta_{s_{(bs)}} = \frac{1}{3}(\eta_{bs_i} + \eta_{bs_j} + \eta_{bs_k}). \tag{2.97}$$

Better sets of models (F_3) are obtained from this layer based on the criterion measure.

It proceeds further and tests for four-, five-set models, and so on until the system criterion of minimum bias starts increasing. Ultimately, the overall process determines the set of variables for the complex object and its linearized structure. Usually the system consists of three to five equations. The variables in the selected set of equations are called "characteristic" variables. Figure 2.15 shows how the minimum bias error of system criterion is reduced as the number of equations increases; each column of the points in the figure corresponds to group of models having similar structure. The approximate limit for successful analysis of modeling is established as $\eta_{s_{(bs)}} \leq 10^{-5}$ on the practical use of the objective system analysis.

If one of the equations has high minimum bias value, then such an equation is considered inconsistent and is excluded from the analysis. If none of the equations is good, then the analysis fails. This can happen if the state variables are too noisy or if the given state variables do not contain any characteristic variables. Noise immunity can be improved by designing specific criteria; the noise immunity depends on the mathematical form of the criterion and on the method of convolution of the criteria into general form. The second level of such criteria are given below; the multicriteria analysis, symmetrical, and combined criteria significantly improve the noise immunity of the algorithm.

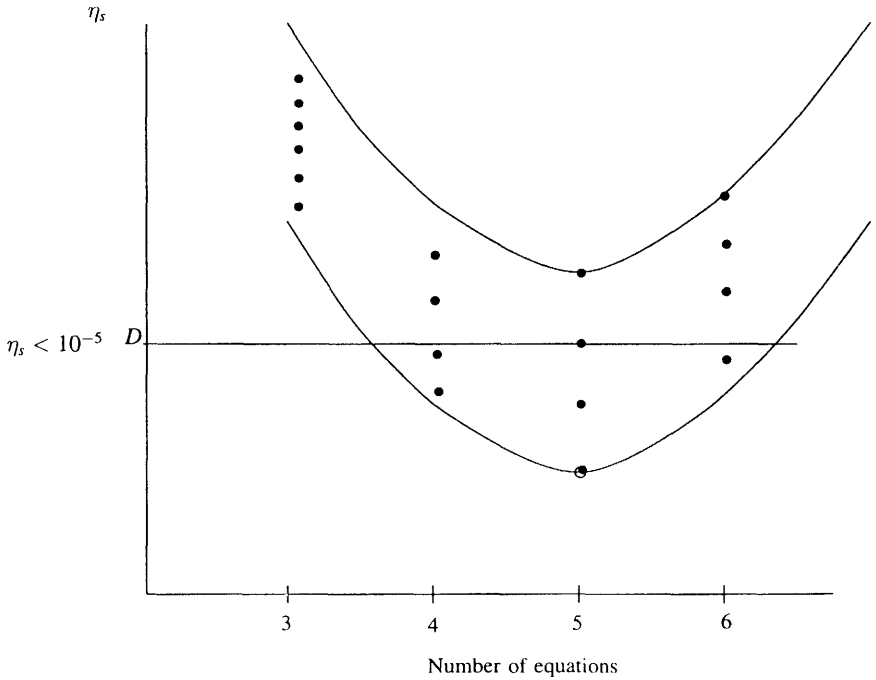

Figure 2.15. Objective system analysis where D is the depth of the minimum and "." is the characteristic point of the model

Another form of minimum-bias criterion is

$$\eta_{bs}^2 = \sum_{p=1}^{N_W} \frac{(\hat{x}_A - \hat{x}_B)_p^2}{x_p^2}. \tag{2.98}$$

Symmetric regularity criterion:

$$\Delta^2(AB) = \sum_{p=1}^{N_B} (\hat{x}_A - x)_p^2 + \sum_{p=1}^{N_A} (\hat{x}_B - x)_p^2 = \Delta^2(B/A) + \Delta^2(A/B). \tag{2.99}$$

It is equal to the sum of two regularity criteria, which represent the usual case, when N_A and N_B are used as data points in training and testing sets, alternatively.

Another form of regularity criterion is

$$\Delta^2(AB) = \sum_{p=1}^{N_W} (\hat{x}_A - x)_p^2 + \sum_{p=1}^{N_W} (\hat{x}_B - x)_p^2 = \Delta^2(W/A) + \Delta^2(W/B), \tag{2.100}$$

where $N_W = N_A + N_B$.

Combined criterion:

$$\Delta^2(AB) = \eta_{bs}^2 + 2 \sum_{p=1}^{N} (\hat{x}_A - x)_p (\hat{x}_B - x)_p. \tag{2.101}$$

Sometimes this type of convolution may lead to additional problems in selecting the final optimal set of equations, but one must use them with care. All equations with the characteristic points below a certain confidence level D are considered equivalent. The final set of equations determines the input-output variables and connection diagram for such set of system components as shown in Figure 2.13.

In this algorithm, the result of calculations on each consecutive layer (with increased number of equations with increased complexity) does not cancel variables of the previous layer, but only adds new "characteristic" variables. Overall, the total number of equations generated are $M = n.\sum_{s=1}^{n} C_{n-1}^{s-1}$ in maximum of n layers; sets of models are formed in each layer and evaluated among them. It is also possible to reduce significantly the scope of calculations at each layer as follows: the first layer uses the computer capacity in full, the next layer uses the set of output variables determined on the basis of first layer, and so on. The system of equations which corresponds to the minimum of the system criterion is chosen as the optimal system. Variables which do not appear in this system are excluded from further consideration. The results from the OSA are passed on to the next levels and used to solve two types of problems: (a) identification of the physical model which is suitable for short-term predictions, (b) identification of nonphysical models for long-term predictions (two-level analysis).

Physical model for short-term predictions

From the above objective system analysis, the set of characteristic (output) variables are identified. Based on the set, nonlinear physical models are developed for the system and its components. The multilayer algorithm with redenoted variables are used for obtaining the optimal model even with very short data samples. We call this model a physical model because of its characteristic variables and its evaluation from a single-level analysis. The physical models obtained this way are not suitable for long-term predictions even though the noise level is relatively low.

Another fact is that not all characteristic variables resulting from OSA can be predicted with the same success. One can use one of the following accuracy criteria for evaluating short-term predictions of the variables either short-range prediction criterion

$$\Delta_i^2(C) = \sum_{p \in N_C} \frac{(x_i - \hat{x}_i)_p^2}{(x_i - \bar{x}_i)_p^2}, \tag{2.102}$$

or the criterion of step-by-step prediction

$$\Delta_i^2(W) = \sum_{p \in W} (x_i - \hat{x}_i)_p^2. \tag{2.103}$$

The variable that has the least convolution value for these criteria is called the "leading" variable. Considering the prediction for the "leading" variable, we find predictions for all other variables which are not even characteristic variables.

Nonphysical model for long-term predictions (two-level analysis)

The *first stage* of two-level analysis is to divide the set of predictions of the average annual values of variables (those not discarded during the objective system analysis) into "good", "satisfactory" and "unsatisfactory" predictions, and to select the best predictions (one for each variable). Input variables that are not satisfactory are excluded for further cosideration.

Output variables are retained regardless of the quality of their annual predictions because the ultimate goal of the entire algorithm is to predict the output quantities. Predictions of models which are obtained because of various algorithms like multilayer, combinatorial, and harmonic algorithms are subjected to comparison as they use different reference functions. The choice of prediction models in all algorithms is made with reference to two criteria; the minimum bias η_{bs} and the prediction criterion $\Delta(C)$, or in the case of small number of data points, the regularity $\Delta(B)$ and the prediction criterion $\Delta(C)$. The models which are more predictive as per these criteria (one prediction for each algorithm) are evaluated further with reference to two other criteria—prediction criterion $\Delta(C)$ and criterion of preservation of first two moments $\rho(m)$. The criterion $\Delta(C)$ is used on examin set C; predictions are assumed to be "good" for $0 < \Delta(C) < 0.5$, "satisfactory" for $0.5 < \Delta(C) < 0.8$, and "unsatisfactory" for $0.8 < \Delta(C)$. Input variables whose predicted annual values are below some threshold are excluded from further consideration.

$$\rho(m) = [(\frac{x_{iav} - \hat{x}_{iav}}{x_{iav} + \hat{x}_{iav}})^2 + (\frac{\sigma_i - \hat{\sigma}_i}{\sigma_i + \hat{\sigma}_i})^2], \tag{2.104}$$

where x_{iav} and σ_i are the mean value and the variance of the variable x_i according to the test set B—i.e., on the interpolation interval N_B—and $\hat{x}_{iav}$ and $\hat{\sigma}_i$ are the mean value and the variance of estimated predictions of $\hat{x}_i$ on the interpolation and prediction intervals $N_B + N_C$. These are computed as below:

$$x_{iav} = \frac{1}{N_B} \sum_{p \in N_B} x_{ip}; \quad \sigma_i = \frac{1}{N_B} \sum_{p \in N_B} (x_i - x_{iav})_p^2 \tag{2.105}$$

$$\hat{x}_{iav} = \frac{1}{(N_B + N_C)} \sum_{p \in (N_B + N_C)} \hat{x}_{ip}; \quad \hat{\sigma}_i = \frac{1}{(N_B + N_C)} \sum_{p \in (N_B + N_C)} (\hat{x}_i - \hat{x}_{iav})_p^2 \tag{2.106}$$

The criteria $\Delta(C)$ and $\rho(m)$ are used in sequence. Algorithms under consideration are first examined on the basis of $\Delta(C)$ and in the next scanning based on $\rho(m)$ they are identified as "good" and "satisfactory." One or more better algorithms are selected for each variable for small values of $\rho(m)$ and for $\Delta(C) < 0.8$. The reliability of annual prediction estimated according to the criterion $\rho(m)$ normally improves if the average prediction is better as per $\Delta(C)$. If none of the algorithms provides satisfactory predictions, then it is necessary to introduce one more level of detailedness—for example, the averaging interval is longer than one year. The output variables that have performed good predictions of annual values ($\Delta(C) \leq 0.8$ and $\rho(m) \leq 0.01$) are hereafter called the "leading" output variables.

The *second stage* is to identify the system of seasonal models using the long-term predictions of average annual and average seasonal values of variables. This means that, the levels of detail contained in various predictions are analyzed such that the average seasonal values of variables are corrected on the basis of average annual values, which are evaluated as per the first stage. The main purpose of this stage is to obtain long-term predictions of the average seasonal values of the output variables.

It was indicated that optimum seasonal (detailed) predictions are not obtained by scanning a large number of competing random predictions, but rather by scanning a relatively small number of models, each of which generates one prediction according to its own criteria. In case of cyclic models, the scanning must include all sets of seasonal models which preserve their natural sequence. Here, usually, cyclic means we consider the seasonal models in the sequence of seasonal changes; i.e., winter, spring, summer and autumn, but in a number of cases the cyclic behavior can be created artificially. Using harmonic algorithm we find

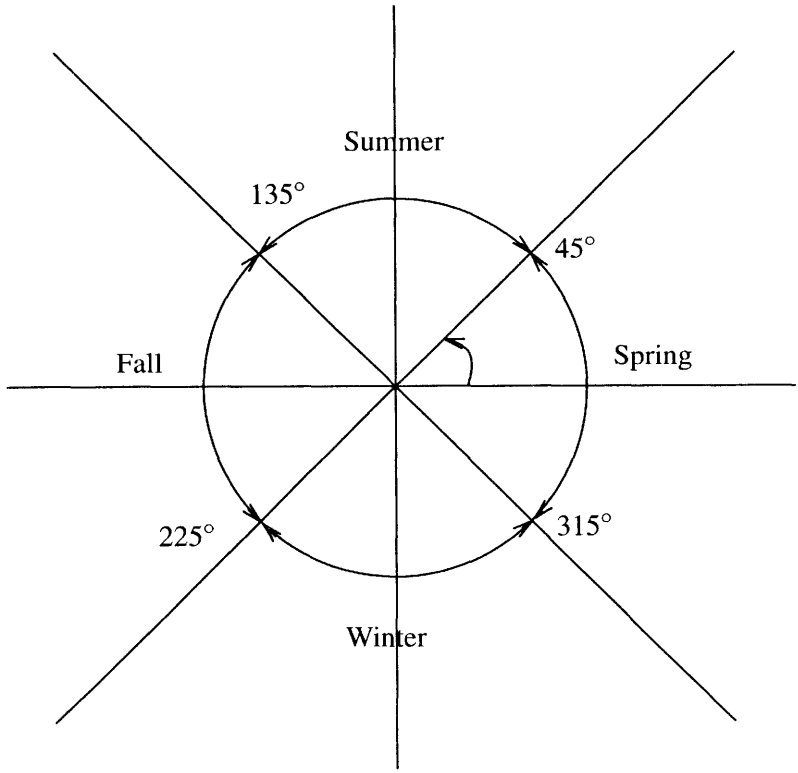

Figure 2.16. Cyclic pattern for four seasons

harmonic series containing only harmonic components which approximate a given process. We select the fundamental term of the process with the largest amplitude and we divide its period into four seasons as shown in Figure 2.16: "summer" 45°–135°, "winter" 225°–315°, "spring" 315°–45°, and "fall" 135°–225°. Before this, the polynomial trend is estimated and subtracted from the data to leave only the cyclic oscillatory part.

For noncyclic processes, the balance criterion is expressed by the sum of squares of the differences with the system of algebraic equations obtained from the previous stage.

Seasonal models are obtained using combinatorial or multilayer algorithms by scanning through a large number of competing models using the minimum bias criterion η_{bs} and prediction criterion $\Delta(C)$. When there is only one "leading" variable, then select up to ten models with different structures for this particular variable and select only one model for every other variable. The scanning of the sets of seasonal models is organized to find the optimum set. Here it is necessary to use different data bases for yearly and seasonal data in the algorithm. The yearly predictions are performed based on the one-dimensional pattern, and the seasonal predictions use the Γ-shape pattern with two-dimensional time count (refer chapter 4 for details). Best models are selected from both the levels. The balance-of-predictions criterion is used to determine the optimal model.

$$B^2 = \sum_{p \in N_C} b_p^2 \rightarrow \min; \quad b_p = \hat{Q}_{yr} - \frac{1}{4}(\hat{q}_w + \hat{q}_{sp} + \hat{q}_{su} + \hat{q}_f), \tag{2.107}$$

where $\hat{Q}_{yr}$ is the yearly prediction value of the leading variable x_i on an examin set N_C, and

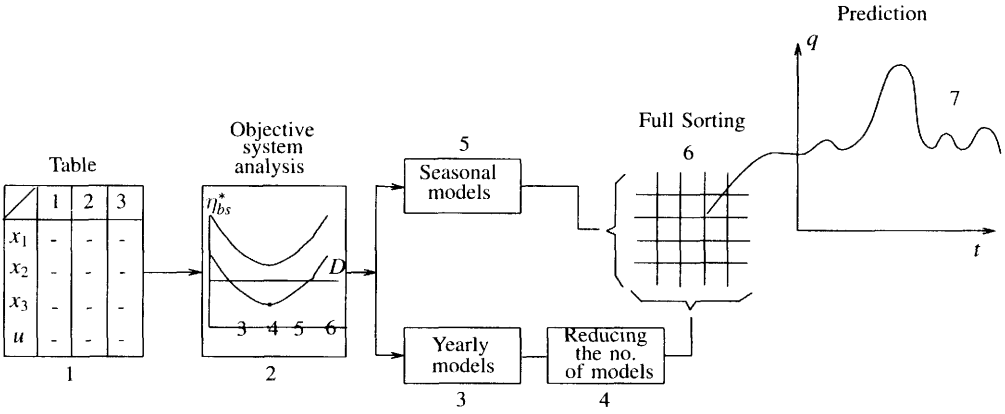

Figure 2.17. Use of OSA for long-range predictions

$q_w, q_{sp}, q_{su}, q_f$are seasonal predicted values of same variable for winter, spring, summer, and fall respectively.

Step-by-step integration of optimum system equations gives the desired long-term predictions simultaneously for all output variables. When there are several "leading" output variables, the better set of models is selected on the basis of system criterion of balance of predictions:

$$B^* = \sqrt{\frac{1}{s}\sum_{i=1}^{s} B_i^2}, \tag{2.108}$$

where s is the number of leading variables that have good and satisfactory annual predictions.

Some practical examples are presented in later chapters. The general scheme of the multilevel algorithm is given in Figure 2.17: the first block indicates the supply of initial data table, the second block denotes first-level analysis which is called an objective system analysis (output variables are determined here), then onwards to two-level analysis; the third and fourth blocks show the first stage of the two-level analysis, and fifth and sixth blocks show the second stage of the analysis. In the first stage of two-level analysis, the third block denotes the selection of $F\backslash$ systems of equations for mean annual values of the output variables. The fourth block denotes the choice of $F_2(< F_1)$ systems of equations according to an external criterion. In the second stage of two-level analysis the fifth block denotes the selection of F_3 systems of equations for mean quarterly or seasonal values of the output variables. The sixth block denotes the sorting of the variants of the predictions in the space of system structures according to the criterion of balance of predictions, and the seventh block indicates the long-range predictions of a specific output variable.

The models used for two-level prediction with two-dimensional time count are considered as nonphysical; for example, they include both yearly and seasonal values of the variables simultaneously. The parameters of two-dimensional time coordinates (t and T) can also be considered into the systems of equations for mean annual and mean seasonal data.

The reliability of choice of a better set of models will increase when the number of scanned predictions is increased. Let p be the number of intervals of the detailed prediction within a year (months, seasons, etc.), let s be the number of leading output variables, and k be the number of models selected for each leading variable in accordance with the combinatorial algorithm. Then the number of compared model sets will be C - $(k^p)^s$.

The freedom of choice can be increased by four to five times in the same length of computer time by changing the averaging intervals to "season-year"; i.e., one can scan through eight model versions for each season. The number of compared predictions (for a single "leading" variable) will be $C_{season-year} = k^8 = 8^4 = 4096$. Therefore seasonal prediction models are preferred over monthly prediction models whenever they are adequate.

The improvement of ergatic or man-machine systems is based on the gradual reduction of human participation in the modeling process. The human element involves errors, instability, and undesired decisions. One approach to this problem is to specify the objectives, or—using technical language—determine the set of criteria. Based on such objective criteria, inductive learning algorithms are able to learn the complexities of the complex system. In self-organization processing the experts must agree on the set of criteria of lower level (regularity, minimum bias, balance of variables, and prediction criteria). If for some reason they cannot come to an agreement, then the solution is to use second-level criteria based on improvement of noise immunity. However, the important problems of sequential decision making, (such as the set of criteria determining their sequence, level of "free choice" and so on), are solved during this decade. Man still participates in the process but his task is made easier. The second area is multicriteria decision making in the domain of more "efficient solutions," where the criteria contradict each other. The solution is to use a number of random process realizations for each probability characteristic like transition graph, correlation function, probability distributions, etc. Additional *a priori* information is needed in order to choose one realization. One may have to balance the realizations of two processes that have two different averaging intervals for the variables (balance of seasonal and yearly, etc).

We conclude this section by saying that the ergatic information systems do not have any "bottle-neck" areas in which the participation of man, needed in principle, cannot be reduced or practically eliminated by moving the decision-making process on the level with a higher degree of generalization, where the solutions are obvious.