Multidimensional Representation of a Dynamical System: Global Environment

Wasaburo Unno
RIST, Kinki University, Higashi-Osaka, Osaka-577, Japan

(Received December 3, 1990)

Abstract

Earth as a whole can be considered as a multi-dimensional dynamical system if not completely an isolated system. Global environment problem, for example, should be described in multi-dimensional phase space consisting of all observables including space and time. Multi-dimensional representation of one variable, e.g., temperature at one position, changing with time is first described, and is generalized to multi-dimensional representation of several variables, temperatures at several positions, and then is generalized to that of many variables, all the characteristic quantities of the system at various space-time positions. The multi-dimensional representation should be adopted to the dynamical system analysis described by a system of differential equations as well as the analysis of the observational data system of any dynamical system.
Key Words: Multi-dimensional Representation, Dynamical System, Global Environment

1 Introduction

Nothing is eternal. Although law of nature may be rather simple, things vary with time in very complicated manner because of coupling with other variables. To elucidate complicated connections, human beings are gifted of brain being able to construct multi-dimensional phase space in which the connections may be embedded without knots.

In a previous paper, a systematic way of constructing new science by means of multi-variable analysis has been introduced (Unno 1989). The idea comes from Takens (1981) who developed the method of determining the fractal dimension of a chaotic dynamical system in multidimensional phase space. Principal component analysis tells what are the essential physical entities of the system, and the application turns out to be very useful. For instance, the correction formula for the IRAS color dependence on the distance determination of mass-losing IR-stars is given from the comparison between the principal component analyses of IRAS colors with and without an additional entity measuring the distance (Unno 1990). Also, the cluster analysis of observed quantities in their phase space should be useful for the classification which is the first important step of science belonging to the natural history. From the cluster analysis, fractal dimension which measures number of active modes driving variations of the system can be obtained if sufficiently large number of entities are observed.

Global earth environment forms a very complicated dynamical system, if not entirely an isolated system. Astronomical, meteorological, geological, oceanographic, biological and human artificial entities are coupled to form the environment. Multi-dimensional representation is really called for for this problem. We will consider this representation problem in the present paper making use of the global environment as an example to be discussed.

2 One Variable Representation

The simplest case is the observation of one variable, temperature $T(t)$ at one position, varying with time t. We can construct n-dimensional phase space by defining a vector

$$(T(t), T(t+\tau), T(t+2\tau), \ldots T(t+(n-1)\tau)).$$

The time interval τ is taken to be of the order of the characteristic time of the system under consideration. The vector is represented by a point in the n-dimensional phase space, and a trajectory is formed by changing t continuously. If the dynamical system shows D- dimensional variability, the trajectory should be embedded in the $n(> 2D+1)$ -dimensional phase space for unambiguous embedding (Takens, 1981). Input data to analyze the dynamical behavior are positions in the $\mathbf{T}(t)$ space corresponding to the time of observation $t = t_0, t_1, t_2, \ldots.., t_N$.

In case of the earth environment, a problem is that the characteristic time can be day, week, month, year, 11 years (sunspot cycle), 160 years (Neptune orbital period), etc. up to 4.5×10^9 years (age of the earth). Also, observations are not obtainable in regular interval τ. The observational time interval $t_i - t_{i-1}$ can be of the order of one day to 10^8 years (geological observation). Such irregular time interval of observations has advantage of covering different characteristic time variations, if sufficient number of data points can be constructed for each characteristic time. Definition of the above vector space must be changed accordingly.

We now define a vector $\mathbf{T}(t)$ as follows,

$$\mathbf{T}(t_i) = (T(t_i), T'(t_i), T''(t_i), \ldots.., T^{(2n)}(t_i)), \quad (1)$$

where $T^{(2k-1)}(t_i)$ and $T^{(2k)}(t_i)$ denote the $(2k-1)$ and $2k$-th derivatives evaluated at $t = t_i$ after differentiating the following Lagrange interpolation formula,

$$T(t) = \sum_{j=-k}^{k} T(t_j) \frac{\Pi_{l=-k,\neq j}^{k}(t-t_l)}{\Pi_{l=-k,\neq j}^{k}(t_j-t_l)}. \quad (2)$$

The advantage of the above definition of vector $\mathbf{T}(t)$ is that all the components are evaluated at any one time t which may be taken close to one of the observation time, t_i, even if they are constructed from observations at several different times. Also, the physical interpretation of the result of the analysis will be easier with the use of the first and second derivatives that can be illustrated as velocity and acceleration or force. In the global environment problem, the temperature $T(t)$ should be measured at various positions on the earth. We may select 5 zones, i.e. tropical zone, north and south middle latitude zones, and north and south high latitude zones, and three types of surface, i.e. sea, green area and desert. Correlation exists among the temperatures of these 15 positions, but each of them is not entirely dependent on the others. They are, therefore, taken as different entities. In this way, we have 15 vector quantities in $(2n+1)$-dimensional space, or we trace the trajectory of temperature distribution $T_\alpha(t)$ $(\alpha = 1, 2, .., 15)$ in the 15-dimensional space with time t. For the dynamical system analysis, we plot ν points of $\mathbf{T}_\alpha(t_j)$ $(\alpha = 1, 2, ..., 15; j = 1, 2, ..., \nu)$ in $15(2n+1)$-dimensional space, where the vector $\mathbf{T}_\alpha$ is $(\mathbf{T}_1, \mathbf{T}_2, \ldots, \mathbf{T}_{15})$ of which $\mathbf{T}_l$ is $(T_l, T_l', T_l'', ..., T_l^{(2n)})$ as in equation (1). The dates of data points t_j can be assigned at will by use of the interpolation formula (2). This freedom will be used later to provide simultaneous data set for the principal component analysis of the dynamical system.

3 Physical Entities of the Global Environment Problem

Average surface temperature on the earth is said to be increasing with increasing CO_2 content in air arising from increased consumption of fossil fuels etc.. We see then that temperature, CO_2 content, consumption of fossil fuels, etc. are physical entities of the global environment problem. What else should be included in the list? For that we may consult a new book of the problem, (e.g. Masuda 1990). We may classify those entities in the book into 1) astronomical, 2) metheorological and oceanographic 3) botanical and 4) social and artificial entities. We obtain the following list.

1) astronomical:

solar radiation, UV intensity, solar activity, infall of minor planet,

(infall of comet, position of moon, Jupiter, etc. may be added.)

2) metheorological and oceanographic:

temperature, humidity, pressure, rain fall, water content in soil,

dust content in air, radioactive substance, density of sea water,

heights of troposhere, stratosphere, mesosphere and thermosphere,

infrared flux, cloud coverage, ice coverage, sea level,

concentration of CO, CO_2, N_2, NO_x, O, O_2, O_3, H_2O, HCl, SO_2, H, F_e, Al, C_a, M_g,

area covered with algae, coral, shells,

polar O_3 hole, polar stratospheric cloud, volcanic eruption, ice age

3) botanical:

forest area, tropical rainy zone forest, mangrove forest,

relative abundance of fauna and flora, tundra area, pH of rain,

neutralizability of alkaline soil,

4) social and artificial:

human population, life standard, skin cancer,

consumption of oil and coal, consumption of wood,

production of harocarbon (CCl_2F, CCl_2F_2, $CHClF_2$, $CClF_2CCl_2F$),

CH_4 , N_20, CC_4, methylchrolofolm

concentration of $ClONO_2$, HCl, ClO_x, HNO_3, HOCl, NO_2,

photochemical smog, acid rain, radioactive dust,

atomic and hydrogen bomb, nuclear engine and nuclear power station accident.

All of these entities should be given quantitatively. But, there are so many entities that it would not be practical to take them into account once for all. We may choose a subset of them depending on particular problem and time scale concerned. As the data base, however, we have to take them all into account.

Each of these entities measured at each of the 15 positions forms a vector similarly as the $\mathbf{T}(t_i)$ vector given in equation (1). If all the measurements are made at the same local time, these 15m (m being number of entities concerned) vectors constitutes the data base of which properties of dynamical system are to be analyzed. This, however, is not the case in general. Another problem is that measurements of these entities are not made in the same accuracy. In the following section, we will discuss about the construction of data base for analyzing a dynamical system , i.e., suitable choice of variables, normalization, interpolation and extrapolation of data, and suitable choice of instants at which data are given.

4 Multi-dimensional Representation

We have considered $15(2n+1)$-dimensional representation of one variable $T(t)$ in the form as $T^k_\alpha(t)$, $(\alpha = 1,2,..,15; k = 0,1,...,2n)$, in Section 1. The same formalism should be applied to all the variables listed in Section 2. Therefore, we may have $15m(2n+1)$-dimensional representation for the global environment problem, m being the number of variables listed in Section 2. We denote these m variables by $\mathbf{X}$ or by $(S, T, U, .., Z)$.

For later convenience, we measure T from its average observed values by shifting the zero point so that $\langle T\rangle = (1/N)\sum_i^N T_j = 0$ in units of its standard deviation $\sigma = [(1/N)\sum T^2]^{1/2}$ so that $\sigma = 0$. The choice of virtual dates of observations can be made arbitrary in principle with the help of the interpolation formula (2). But, daily, monthly, and yearly variations cannot be ignored because of its fairy large amplitude, while the time scale of our main interests is 10 to 100 years or more. Nonlinear couplings between the shorter and longer time variations play essential roles of the problem. We should, therefor, choose the virtual dates of observations to be of time intervals of cluster hierarchy such that three observing days belong to one observing week, three observing weeks belong to one observing season, three observing seasons belong to one observing year, and so forth. Number of observing days or number of points plotted on the phase space is then 3^c, if c steps of the clustering hierarchy are adopted. Adoption of three subclusters in one cluster is for convenience of computing the first and the second time derivatives in equation (1). It may, therefore, be advisable to take these lowest two time derivatives computed from data of different time span as the components of the vector $\mathbf{X}_\alpha(t_i)$ instead of higher order derivatives. Also, in order to account for the effect of daily variations, it may be more practical to take the amplitude of daily variations of $X^p_\alpha(t_i)$ as an additional variable. In this case, total dimensions of variable space becomes to be $2ma(2c+1)$, where 2 comes from the daily average and amplitude of variation, m the number of different observed quantities, a (=15) the number of different kinds of observing positions, and $(2c+1)$ the number of the vector components.

5 Principal Component Analysis

Purpose of the study of the global environment problem is to clarify causality relations and to predict future life conditions on the earth. Quantitative prediction is rather difficult even in weather forecast in which causality relations are well known. In the global environment, causes are numerous and intermixed with results, and no detailed causality relations are established yet. Numerical simulations and comparison with observation are useful to elucidate such relations. At present, however, natural historical approach which obtains information from the detailed numerical simulation made by the most comprehensive analog computer, the earth herself. The only problem is how to read the information.

Mathematical basis for modern natural history has been discussed by (Unno, 1989), in which principal component analysis is generalized for the case of incomplete data set. There are so many measurements in the global environment problem and accuracy of measurements varies in wide range. The generalized principal component analysis is, therefore, appropriate with the data set represented in multi-dimensional space given in the preceding section. Only the assignment of weight to each of the data is needed to apply the method. We put the weight w to be unity, if uncertainty of the measurement is less than say 0.1 σ, σ being the standard deviation of all the measurements of the variable in question. If the uncertainty, s, is larger, we take w to be $(0.1\sigma/s)^2$. The detail of the method is not repeated here except the following remark.

We consider that x_p value was observed to be $x_p(t_i)$ at time t_i in weight w and was distributed at the same time in normal distribution of standard deviation σ in weight $1-w$. In practice, the latter distributed values may be quantized to be the average, $\langle x_p \rangle$ (which is 0) $\pm\sigma$ with weight $(1/2)(1-w)$. This quantization does not alter $\langle x_p \rangle, \sigma$, and correlation coefficients with other variables. If there is no measurement, we put $w=0$. So, three values are associated to one measurements. The principal components that are orthogonal to each other are the vectors onto which the sum of the squared distances times the weights of data points in the multi-dimensional space is minimum.

The first principal component on which the projection of data points show the largest scatter should be the most influential driving agency of the system, and the physical meaning of the principal component could be judged from the direction cosines of the vector against the coordinate axes. On the contrary, principal components showing little scatter of data points represent redundancy of data. Physical mechanisms driving the dynamical system may well be represented by the distribution of data points in the main principal component space. Adjustment of uncertain data is also discussed in Unno(1989). Prediction of future global environment should follow the principal component analysis.

References

[1] Masuda, Y. 1990, Crisis of Earth Environment (in Japanese), Shin-Nihon Publ.Co.

[2] Takens, F. 1981, Lecture Notes in Math., **898** (Springer-Verlag), 366.

[3] Unno, W. 1989, Sci.& Tech., Kinki Uni., No.1, 41.