DESIGN OF DATABASE OF REGIONAL MODELS OF THE MOLECULAR ATMOSPHERE

S.A. Mikhailov

Institute of Atmospheric Optics,
Siberian Branch of the Russian Academy of Sciences, Tomsk
Received November 15,1993

Models of spatiotemporal distribution of thermodynamic quantities and gas constituents of the atmosphere are widely used as a priori *information in solving problems of atmospheric optics, environmental remote sensing, and simulation of the climate formation process. The results of designing the database of regional models of molecular atmosphere supported by personal computer are presented.*

The physical state and the chemical composition of the atmosphere as a molecular gas medium has a pronounced effect on the processes taking place in the earth's climate system and on the ways of propagation of various pollutants occurred as a by–products of the human activity. In addition, the atmosphere interacting with the electromagnetic radiation exerts essential influence on its propagation because of a number of processes, the main of which being the absorption, molecular–aerosol scattering, and refraction. In this connection it becomes evident that when solving the problems of the climatic–ecological monitoring, atmospheric optics, and remote sensing (including that based on space means) it is necessary to take into account the peculiarities of the spatiotemporal distribution of fields of basic thermodynamic quantities and the gas constituents of the atmosphere.

To solve the problem in question it is expedient to use the empirical models of the meteorological fields, representing the special features of their horizontal and vertical statistical structure, because the results of the original meteorological observations are, as a rule, not systematized, redundant and require the large magnetic data medium and resource expenses for their processing that makes it impossible to use these data in routine operations.

The information on the vertical structure of the atmosphere is of most importance for processing the satellite data on the earth's climate system state. Therefore for designing space equipment and for radiation correction of the satellite data at the initial stage of the development of the space sensing methods the models of the standard atmosphere and later the consulting models of the atmosphere have been developed.[1,2] The most known models are the LOWTRAN–type models[3] which incorporate the zonally averaged profiles of the vertical distribution of the air pressure and temperature and the concentrations of the water vapor, ozone, aerosol, and some other minor constituents as well within three latitudinal zones, namely, polar, middle, and tropical ones. As the data on the physical parameters and chemical composition of the atmosphere became available the number of consulting models increased. Now a series of such models is available, which represent the peculiarities of the vertical distribution of thermodynamic quantities and of minor gas constituents of the atmosphere in various regions of the globe.[4–10]

One of the most exhaustive and data provided models is the physical statistical model of the molecular atmosphere developed in the Institute of Atmospheric Optics (SB RAS, Tomsk) under the direction of V.S. Komarov.[11–16] The model is based on the results of the objective regionalization of the northern hemisphere according to the set of quantities "pressure–temperature–humidity–ozone".[11,12,17] For each of separated quasihomogeneous regions it includes the characteristics of the vertical statistical structure of the basic thermodynamic parameters (pressure, temperature, and humidity of the air, zonal and meridional components of the wind velocity) and optically active gas constituents of the atmosphere (O_3, CO_2, CO, CH_4, N_2O, NO, NO_2, SO_2, HNO_3), playing the leading role in forming such climate phenomena as the greenhouse effect, the depletion of the ozone content in the stratosphere in polar regions, and the formation of acid rains.

To describe the vertical statistical structure of physical parameters of the atmosphere the following statistical characteristics are used in the model of Institute of Atmospheric Optics :

a) For the set of the basic thermodynamic quantities and gas constituents of the atmosphere (pressure, air temperature, zonal, and meridional components of the wind velocity as well as water vapor, ozone):

– the mean value vector

$$\bar{\zeta}(p_i) = \frac{1}{N}\sum_{k=1}^{N}\zeta_k(p_i)\ ; \qquad (1)$$

– the standard deviation vector

0235-6880/94/02 140-06 $02.00

$$\sigma_\xi(p_i) = \sqrt{\frac{1}{N}\sum_{k=1}^{N} (\zeta_k(p_i) - \overline{\zeta}(p_i))^2}\ ; \qquad (2)$$

– the elements of the autocorrelation matrix (right triangle part)

$$\mathrm{cor}\ (p_i, p_j) = \frac{1}{N\ \sigma(p_i)\ \sigma(p_j)} \times$$

$$\times \sum_{k=1}^{N} (\zeta_k(p_i) - \overline{\zeta}(p_i))\ (\zeta_k(p_j) - \overline{\zeta}(p_j))\ ; \qquad (3)$$

– the eigenvectors of the autocovariation matrix $F_\alpha(p_i)$;

– the eigenvalues of the autocovariation matrix λ_α ;

– the sum of diagonal elements (trace) of the autocovariation matrix D .

b) For the set of the gas constituents of the atmosphere (CO_2, CO, CH_4, N_2O, NO, NO_2, SO_2, HNO_3):

– the mean value vector;

– the standard deviation vector.

Here $\zeta_k(p_i)$ is the k th value of the meteorological quantity ζ at the ith isobaric surface with the pressure p, GPa; N is the length of the data retrieval; α is the number of the eigenvector and corresponding eigenvalue.

In order to provide the routine use of the regional models of the molecular atmosphere in the process of solving the problem of designing the space equipment and of radiation correction of the satellite data the author has developed the design of the computer version of the database of these models, which provides the methodological base for the IBM PC/AT compatible implementation of the database. The programing version of the database of the regional models of the molecular atmosphere is intended for:

– systematization as well as optimal storage and access to the atmospheric models;

– model retrieval in accordance with the user requirement;

– presentation of the model characteristics in the form of tables and/or graphs;

– dissemination of the parameters of the model required into the systems of the automatized designing of the space equipment or radiation correction of the satellite data;

– providing the environmental protection organizations with the information on the values of thermodynamical quantities in the atmosphere and on the background concentrations of major pollutants in the physical–geographical region under consideration.

In this paper the results of designing the database of regional models of the molecular atmosphere on the basis of the methodological concepts DATAID – 1 developed in Milan Polytechnical Institute[18] are presented. In accordance with requirements of this methodology the design of the central database (DB) is subdivided into four stages: analysis of requirements, conceptual design, logical design, and physical design.

When designing the database of regional models of the molecular atmosphere at the stage of analysis of requirements we have developed the vocabulary of the design solutions which includes the data vocabulary and the operation vocabulary containing the specifications of data and that of applied problems, respectively.

At the next stage the conceptual model of the database of regional models of the molecular atmosphere was formulated. The conceptual model is typically considered as a formalized representation of statistical (the data) and dynamical (the operations) requirements to the DB, which pertain to the definite application domain.[19] Within the framework of the methodology DATAID – 1 the conceptual design is based on using the semantic model which makes it possible to represent the application domain as a model "entity – relationship" (ER–model), firstly proposed by Chen.[20] It should be noted that the means of the DATAID – 1 enable one to construct more complicated semantic models named the extended ER–models (EER–models) which include the generalized hierarchies and n–fold relationships.

At the stage of the conceptual design the global operational scheme of the database of regional models of the molecular atmosphere has been also constructed which represents the synthesis of local operational schemes. In doing so every of the local operational schemes represents one of the types of operations with the data which are pertinent to the described application domain and are included into the designed DB.

Figure 1 shows the global EER–model of a given application domain obtained at the stage of the conceptual design. The entities of the given application domain are enclosed in the double frames and the relationships – in the ordinary frames. It is worth–while to note that all relationships of the third and higher orders are n–fold ones, and the adjunct class of every of entities is required. The relationships A, B, C and D (they are also the off–spring entities) are semantically most significant. They contain the bulk of information on the parameters of regional models of the molecular atmosphere. Note that the relationships B and C have the same relational structure (see Table I), however, for the sake of convenience they are presented as two different relations. In doing so the relationship C (the basic model) corresponds to the most frequently used information on the vertical distribution of mean values and standard deviations of basic thermodynamic parameters and optically active gas components of the atmosphere.

The global operational schema of the database of regional models of the molecular atmosphere is depicted in Fig. 2. It has been constructed at the stage of the conceptual design as well. It can be seen from Fig. 2 that the control system of the DB (CSDB) supports three groups of operations connected with the DB updating, mapping of

its current state, and searching of the regional models. The entities (the DB relations) which take part in every of above mentioned operation groups are also shown.

At the stage of the logical design the tables of logical access containing the quantitative parameters of the access to the data were constructed. The simplification of the global scheme of data was also made whereby the hierarchies and multiple relationships were restructurized into nonhierarchy entities and binary relationships. Then on the basis of the simplified scheme the preliminary relational logical data model was formulated and the test of the schema of the database **R** for the correspondence to the normal Boys–Codd form (NBCF) was performed.

As is known, the normalization of the database scheme is the point of major importance of designing the relational DB. It enables one to eliminate a number of anomalies occurred in the production run of the DB (among them the update anomaly) and the informational redundancy in relationships. The normal Boys–Codd form is the strongest of the forms applied to designing the DBs.

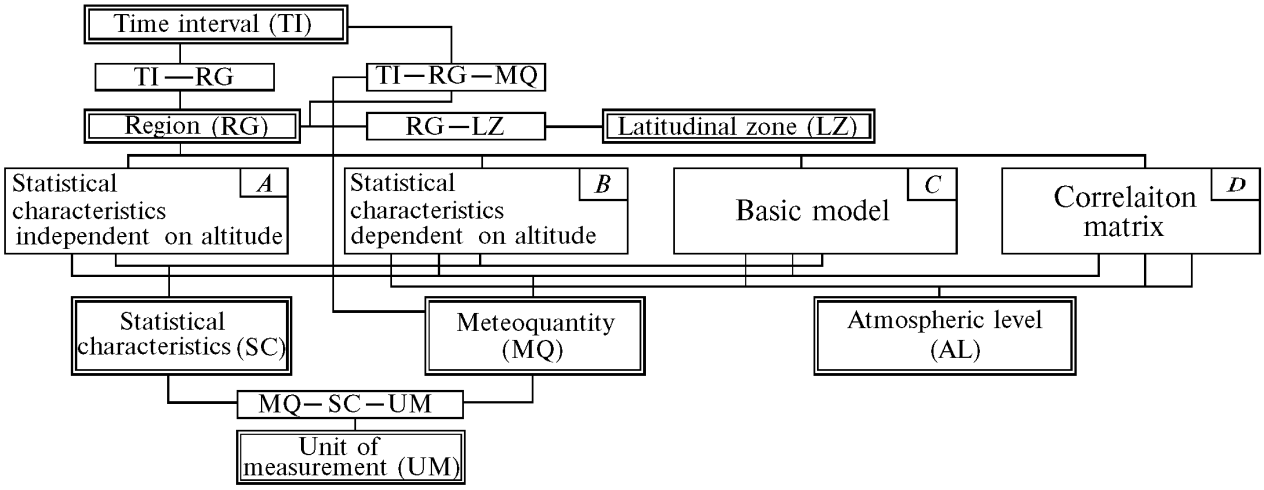

FIG. 1. Global EER model of the application domain "Regional models of the molecular atmosphere".

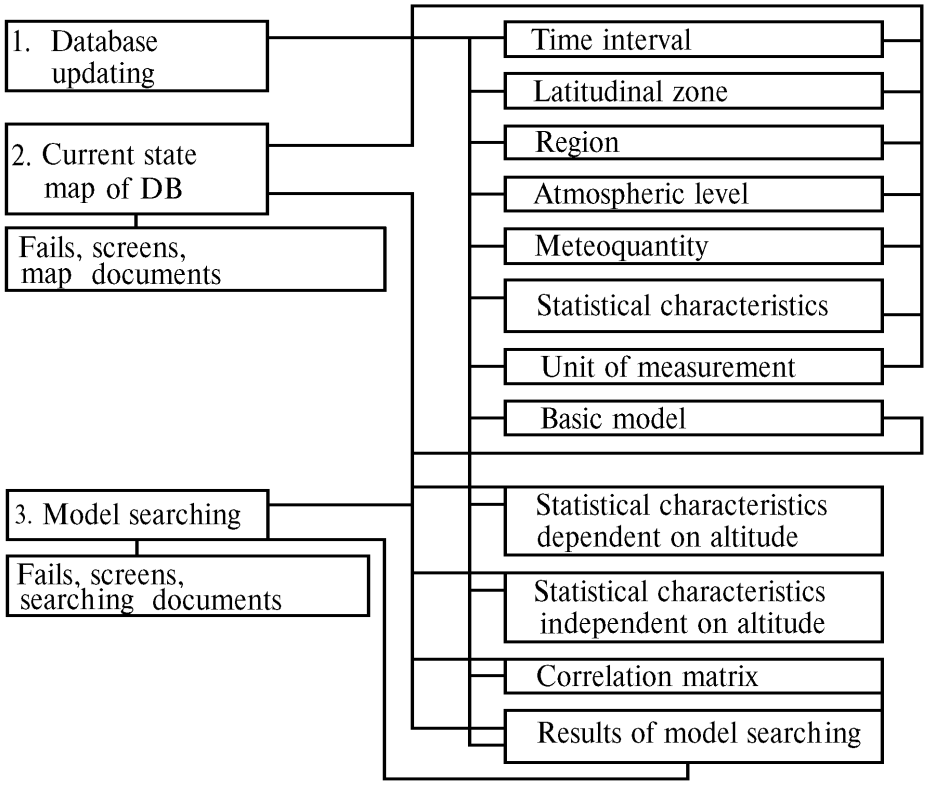

FIG. 2. Global operational scheme of the database of regional models of the atmosphere.

TABLE I. Relational scheme of the database of regional models of the molecular atmosphere.

Entity	Relationship	Key fields
1. Time interval		1. Code of time interval
2. Latitudinal zone		1. Code of latitudinal zone
3. Region		1. Code of region
4. Atmospheric level		1. Code of level
5. Meteorological quantity		1. Code of meteoquantity
6. Statistical characteristics		1. Code of statistical characteristics
7. Unit of measurement		1. Code of unit of measurement
8.	Meteorological quantity Statistical characteristics Unit of measurement	1. Code of meteoquantity 2. Code of statistical characteristics 3. Code of unit of measurement
9.	Time interval Region Meteorological quantity Statistical characteristics	1. Code of time interval 2. Code of region 3. Code of meteoquantity 4. Code of statistical characteristics
10–11.	Time interval Region Meteorological quantity Statistical characteristics Atmospheric level	1. Code of time interval 2. Code of region 3. Code of meteoquantity 4. Code of statistical characteristics 5. Code of atmospheric level
12.	Time interval Region Meteorological quantity Atmospheric level	1. Code of time interval 2. Code of region 3. Code of meteoquantity 4. Code of level (matrix row) 5. Code of level (matrix column)
13.	Time interval Region Meteorological quantity	1. Code of time interval 2. Code of region 3. Code of meteoquantity

According to Ref. 21 the database scheme **R** is in the normal Boys–Codd form with respect to the set of functional dependencies F, if every scheme of the relation $R \subseteq \mathbf{R}$ is in the NBCF with respect to F . On these grounds every scheme of relations $R \subseteq \mathbf{R}$ of the database of regional models of the molecular atmosphere was considered for the correspondence to the NBCF. In the case when this correspondence was violated the process of decomposition of the scheme of relation R was performed into two schemes of relations $R_1 \subseteq \mathbf{R}$ and $R_2 \subseteq \mathbf{R}$, which, generally speaking, may intersect each other. This process enabled us to remove the transitive dependence of the attributes from R and continued as long as all the relations obtained came into the correspondence with NBCF. The produced relational schema of the database of regional models of the molecular atmosphere is given in Table I. Note that only the key fields of relations of the DB are indicated in Table I, which allow one to identify unequivocally one or another tuple t of the relation R considered.

At the concluding stage of designing the database of regional models of the molecular atmosphere the relational scheme supported by the driver DBF–NTX was chosen as a physical scheme of data. This driver is supported by the language family XBase (dBase, Fox Pro, Clipper, etc.) that maintain the effective program implementation of the control systems of the relational DBs (CSDBs).

We now focus our attention upon the peculiarities of the space composition of data of the regional models of the molecular atmosphere. As it was mentioned above every model entering the DB is a set of statistical characteristics of the vertical distribution of basic thermodynamical parameters and gas constituents of the atmosphere corresponding to one of the quasihomogeneous regions in the northern hemisphere. In doing so every quasihomogeneous region is the contourous space object (polygon) of irregular form and the appropriate regional model is constructed as a set of the semantically significant descriptors.

To describe objects of such a type we use the approach earlier realized in Canadian geoinformational system CGIS.[22] In this approach the complete description of a polygon and of a regional model is provided by forming two sets of the schemes of relations $R_p \subseteq \mathbf{R}$ and $R_d \subseteq \mathbf{R}$, which are nonintersecting subsets of the scheme of the database **R**. The subset of the schemes of relations R_p contains the information on the space localization of polygons which

consists of the ordering values of geographical coordinates (with an accuracy of 2.5°) of their nodes, whereas the subset of the schemes of relations R_d includes the descriptive information and carries the major semantic load.

Note in conclusion that the above results of designing provide a basis for the program implementation of the first version of the database of regional models of the molecular atmosphere. The more elaborated versions of this program product are supposed to be developed starting from the present one. Improvements in the database of regional models are expected to be performed in the following directions:

– Elaboration of the graphical interface on the basis of the system of processing the cartographical information making use the electronic atlas including the maps of physical–geographical regions of the northern hemisphere in various projections;

– Improvement of regional models via expanding the number of physical parameters of the atmosphere and increasing the statistical provision of the model characteristics due to incorporation of the results of observations obtained in recent years;

– Elaboration of the network version of the database of models of the molecular atmosphere to operate in local and regional networks.

REFERENCES

1. U.S. Standard Atmosphere (U.S. Government Print Office, Washington, 1976), 289 pp.
2. Standards of USSR 440–64. Tables of Standard Atmosphere (Izdat. of Standards, Moscow, 1964).
3. J.I. Kneizys, et al., Atmospheric Transmittance / Radiance Code LOWTRAN 5 (Env. Res. Papers, AFGL–TR–80–0067, 1980), No. 697.
4. G.M. Keating and M.C. Pitts, Adv. Space Res. **7**, No. 9, 37–47 (1987).
5. J.J. Barnett and M. Corney, Handbook for MAP **16**, 47–85 (1985).
6. E.L. Fleming, S. Chandra, M.R. Schoeberl, and J.J. Barnett, *Monthly Mean Global Climatology of Temperature, Wind, Geopotential Height, and Pressure* (NASA Tech. Mem. TM–100697, 1988).
7. International Standard ISO 5878. Reference Atmosphere for Airspace Use (ISO/TS–20/SC–6(USA–10), 102E, 1982), 51 pp.
8. M.R. Ananthasayanam and R.A. Narasimha, Handbook for MAP **16**, 197–204 (1985).
9. A.E. Cole and A.J. Cantor, Air Force Surveys in Geophysics. AFGL–TR–78–0051, No. 382 (1978).
10. R.A. McClatchey, H.J. Bolle, and K.Ya. Kondratyev, *A Preliminary Cloudless Standard Atmosphere for Radiation Computation* (Boulder, Colorado, 1986), 53 pp.
11. V.E. Zuev and V.S. Komarov, *Statistical Models of Temperature and Gaseous Components of the Atmosphere* (D. Reidel Publishing Company, Dordrtecht–Boston–Lancaster–Tokyo, 1987), 306 pp.
12. V.S. Komarov and V.A. Remenson, Opt. Atm. **1**, No. 7, 3–16 (1988).
13. V.S. Komarov and S.A. Mikhailov, Meteorol. Gidrol, No. 3, 44–50 (1986).
14. V.S. Komarov, S.A. Mikhailov and N.Ya. Lomakina, in: *Optical–Meteorological Investigations of the Earth's Atmosphere* (Nauka, Novosibirsk, 1987), pp. 3–16.
15. V.S. Komarov, S.A. Mikhailov, and D.N. Romashov, *Statistical Structure of Vertical Atmospheric Ozone Distribution* (Nauka, Novosibirsk, 1988), 77 pp.
16. V.S. Komarov and S.A. Mikhailov, in: *Problems of Comprehensive Automation*, Proc. IV Int. Sci.–Techn. Conf., Kiev, October 17–20, 1990 (KPI, Kiev,1990), pp. 94–100.
17. S.A. Mikhailov, *"Spatiotemporal Structure of the Vertical Ozone Distribution and Its Few–Parameter Description"*, Author's Abstract of the Kand. Phys. – Math. Sci. Dissirt., Leningrad, Hydrometeorogical Institute (1988).
18. *Methodology and Tools for Database Design*, S. Ceri, ed. (Amsterdam, North–Holland, 1983).
19. C.J. Date, *An Introduction to Data Base System*, Vol. 1–2 (Menlo–Park, CA: Addison–Westley Publ. Co.,1986).
20. P.P. Chen, ACM TQDS **1**, No. 1, 9–36 (1976).
21. D. Meyer, *Theory of Relational Databases* (Mir, Moscow, 1987), 608 pp.
22. D.F. Mable, ed., *Computer Software for Spatial Data Handling*, Vol. 1–3 (Ottawa, Ontario: IGU, 1981), 1042 pp.