Computer clustering technology in dendroecological diagnostics problems

V.N. Popov,[1] I.A. Botygin,[2] and V.A. Tartakovskii[1]

[1] *Institute of Monitoring of Climatic and Ecological Systems, Siberian Branch of the Russian Academy of Sciences, Tomsk*
[2] *Tomsk Polytechnic University*

Received November 12, 2004

Multicomputer software system is presented. It implements a distributed solution of problems of dendroecological diagnostics. The architecture of the system includes calculation cluster and communication server, providing access to information on the dendroecology and to the instrumental means of analysis of annual tree rings, aimed at revealing the changes in the parameters of the environment in which the trees grew. The annual tree rings are simulated based on the key properties of radial tree growth, namely monotonicity of growth in space and time and finiteness of tree growth rate, while variations of wood density are considered as a spatiotemporal oscillation process.

Introduction

Presently, new information technologies are actively used to solve problems in diagnostics of the environment and ecology. In particular, the role of long-term ecological prediction and dendroecological studies has recently increased, being implemented and accessible in computer networks. Basic academic studies in the field of dendroecology call for development of new methods of obtaining experimental data that are more correct in relation to the changes in the corresponding climatic data.

The specific feature of problems of dendroecological diagnostics is that they require mathematical processing of a huge amount of data (sometimes 1000-yr and longer time series) and collaborative (simultaneous) effort of many individuals at practically all stages of a dendroecological study, as well as that they involve storage and systematization of large volumes of inhomogeneously structured information (proper chronological observation time series, processing results, and pertinent meteorological, geologic, geophysical, aerospace etc. observation time series). Thus, the large amount of calculations, usually required to process dendrochronological data, necessitates the use of efficient calculation systems and development of new distributed methods of data processing. The methods of distributed processing involve physical distribution of data processing in the space on several computers, connected by channels of the data transfer.

The technological principles of the distributed processing are, in fact, the principles of open systems,[1,2] such as the performance, scalability, software portability, software reliability, data security, and computer security. It is just realization of these principles that provides for meeting practical demands in terms of not only collective work of users and mobility (portability) of software, but also joint operation with already developed software, and also protection of confidential information, among many others. The solution of problems of processing large amounts of data is the subject of research in clustering of computer components (computer clusters)[3] and grid-technologies.[4,5] Studies in the virtualization of the centers of data processing[6] are aimed at ensuring collective (multiple) access to computer resources. At the same time, the concept of data storage capacities is directed toward solution of the problem of storage and use of inhomogeneous information.[7,8]

This paper describes a multicomputer system of dendroecological data processing, designed for distributed processing and analysis of images of annual tree rings with the purpose of revealing changes in the environmental parameters reflected in tree growth. The traditional denrochronological time series are oriented toward dating events and contain information only on annual ring width. The known methods of measurement of tree growth are too laborious and have no sufficient algorithmic and software aids. Within the seasons of growth, a single tree may allow only a small number of discrete samples, so the quality of the obtained realization is determined only after tree growth ends. It is also difficult to compile uniform sample with sufficient volume. To study statistical interrelations of the growth process and weather, it is necessary to fix these random processes as functions of time during the season of growth.[9,10] It is suggested to consider the variations of wood density along the radius of a tree disc as a certain oscillation whose phase is a strictly increasing function of the radius. The radial growth is defined as a monotonic function of time, inverse with respect to phase. At the same time, the influence of the environment and genetic program of the tree are manifested in the form of amplitude-phase modulation of the cellular structure. The

0235-6880/05/01–02 105-06 $02.00

mathematical tools used are based on introduced condition of dispersion causality and formalism of analytical signal.[11]

1. Architecture of the computer cluster

The aforesaid features of the dendroecological problems lead to some problems in creation of massive professional centers of dendroecological data processing and storage at a separate institution. This requires too large computer capabilities and powerful databases and, naturally, very large costs. A natural solution in this situation is the concept of information, software, and hardware aids, solution of dendroecological problems in several specialized supercomputer centers, and virtualization of the access for researchers to these information-processing resources.

The basic concept of implementation of multi-computer system of dendroecological data processing is the operation of the system in a network variant. Given that such processing centers have international significance, as well as that the telecommunications in Russia have poor infrastructure, a web-interface is chosen as a basis for implementation of the network-based system operation.

We have performed experimental study of required computer capabilities for such a dendroecological problem as processing of images of annual tree rings (scanning and filtering). Depending on the size of the image processed, a few tens of minutes of processor time was required on Celeron CPU of 1.7 MHz and DDR RAM 256 Mb. The processor burden increased practically proportionally to the number of connected users. One of the solutions of the problem of lack of computer resources is the connection of several computers into a multi-computer processing system (Fig. 1), since calculational capabilities of a single computer cannot be increased up to infinity.

To ensure inter-server links (communication server ↔ calculation server, communication server ↔ server of database, calculation server ↔ server of database) inside the calculation cluster, when a network filter is used at any network it is necessary that IP addresses of the corresponding servers be included in the list of enabled entry and exit TCP connections of this network filter.

Planning and managing of the processes of dendroecological data processing are assigned to a special communication server. Interaction of all users with processing system occurs only via communication server. This server itself performs no mathematical calculations. Mainly it is required to ensure optimal loading of calculation servers managed by it and on-line work of the users.

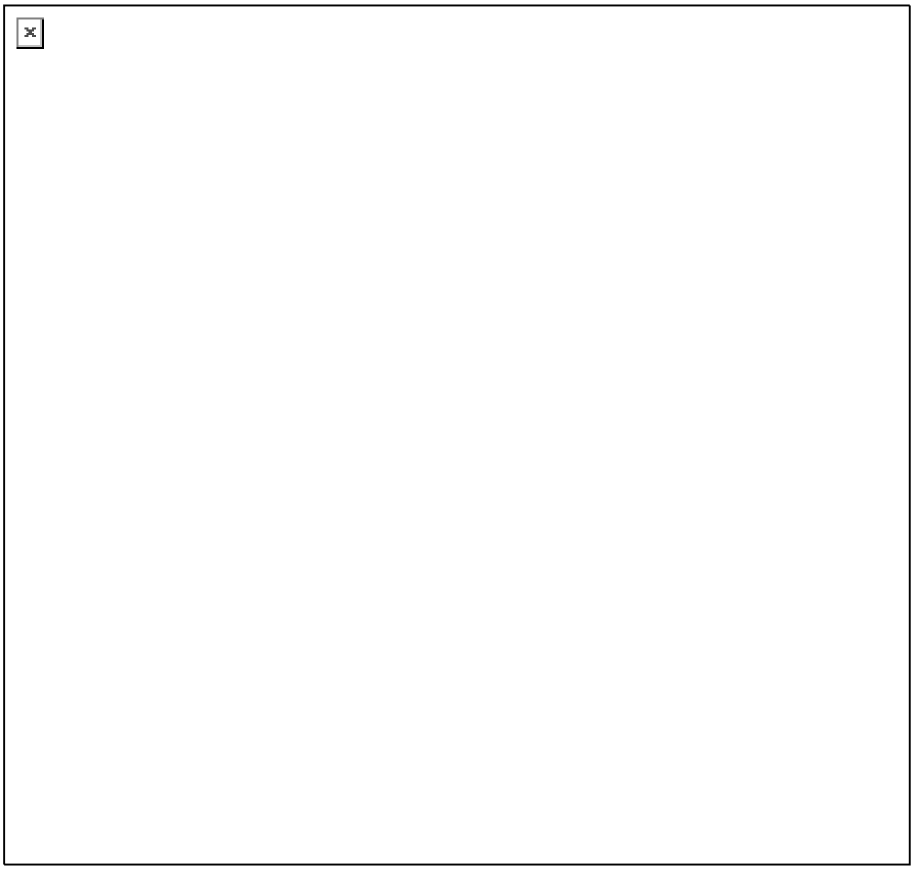


Fig. 1. Architecture of multi-computer system of dendroecological data processing.

The problems to be solved by the algorithms of mathematical processing, run on the calculation servers of the cluster, may be very different. For instance, calculation of radial tree growth and wood density in a growth season, construction of long-term wood-ring chronologies, etc. It is important that these systems of dendroecological data processing keep a common interface for a proper connection to communication server.

For full-productivity operation of dendroecological data processing system, it is necessary to arrange automated acquisition, systematization, and storage of information of dendroecological monitoring, as well as creation and accumulation of dendroecologic databases. None of these problems can be solved without application of database management system (DMS). For versatility of the system software, it is necessary that the DMS understands language of SQL queries. Although there are certain nuances in use of SQL commands of different databases, they often reduce to the difference in syntax of some commands and their options. At the same time, the key commands such as SELECT, INSERT, UPDATE, and DELETE, serving to select, add, update, and delete information, are usually standard. In principle, the system is required to insure interaction with any servers of databases understanding the language of SQL queries. Data types to be stored in DMS relations must include integer and real values of different digit capacity, fixed and dynamic strings, dates, Blob-fields for storage of arbitrary data such as binary files of images.

Like many other applied multi-user systems, the processing of dendroecological data more and more urgently calls for a change to "user–server" calculation model and distributed processing. With regards to dendroecological data analysis system, the "user–server" architecture is interesting and urgent primarily because it provides simple system programmatic solutions and relatively economic solution of the problem of collective remote user access to both computational and information resources of the system. We note that most optimal implementation of the user-related part is based on the use of standard means of navigation in the Internet (such as Internet Explorer, Netscape Navigator, and other similar browsers), and not on the development of individual user-related modules.

2. Mathematical model of annual tree rings

The mathematical model of annual tree rings can be represented as interferograms, created by "ecological" $U(\rho,\theta)$ and "biological" $B(\rho,\theta)$ fields:

$$R(\rho,\theta) = |B(\rho,\theta) + U(\rho,\theta)|^2 = \\ = u^2 + b^2 + 2ub \cos\Phi, \qquad (1)$$

where ρ, and θ are polar coordinates; $b = |B|$; $u = |U|$; and $\Phi = \arg B - \arg U$. This difference $\Phi(\rho,\theta)$ contains information on environmental parameters.

A basic property of the function $\Phi(\rho,\theta)$ is its monotony within the interval [0, ρ], giving rise to concentric ring-shaped structures about the center at $\rho = 0$. In addition, it is assumed that the absolute values of $b(\rho,\theta)$ and $u(\rho,\theta)$ slowly change within the mean period of circular structure. Moreover, from these properties it follows that the interference component [third term in Eq. (1)] is a real part of analytical signal of the variable ρ.

A traditional method of demodulation of circular structures (reconstruction of function $\Phi(\rho,\theta)$) consists in monitoring of rings, determination of their order, or numbering. However, there are also some other algorithms not requiring explicit analysis of the structure of rings. In the presence of analytic signal, the function $\Phi(\rho,\theta)$ is reconstructed using Hilbert transform **H** and bandpass filtering **F** with respect to ρ for all θ (Ref. 12):

$$\Phi(\rho,\theta) = \arg[\mathbf{F}\, R(\rho,\theta)] = \\ = \arctan\{[\mathbf{HF}\, R\,(\rho,\theta)]/\mathbf{F}\, R(\rho,\theta)\}. \qquad (2)$$

The meaning of the function $\Phi(\rho,\theta)$ in dendroecology is given by its components:

$$\Phi(\rho,\theta) = 2\pi\rho^n/\tau + g_b(\rho) + g_u(\rho) + s(\rho,\theta), \qquad (3)$$

where the first term determines the power-law tree growth with average ring width τ in the absence of any perturbations, the second term g_b describes age-specific trend and other isotropic endogenous factors, the third term g_u stands for isotropic exogenous factors, and, finally, the function $s(\rho,\theta)$ quantifies the effect of Sun, inclination of the earth's axis, latitude, and wind pattern. To distinguish among these components and determine their characteristics in the presence of noise, filtering is applied: median, polynomial, and in trigonometric basis, as well as calculation of integral moments of the function $R(\rho,\theta)$.

3. Programmatic implementation of multi-computer system of dendroecological data processing

The above-mentioned scheme of clustering of computer resources was implemented to solve the problem of analysis of images of annual tree rings. For faster calculations, all required mathematical transformations were implemented in the medium of server application. The Object Linking and Embedding (OLE) technology for distribution of processing among network components (for linking and running programs in other computation media) was not used.

The calculation cluster is chosen to achieve the following tasks.

Acquisition, systematization, and storage of scientific information in the region of dendroecological monitoring. The information contains both references to the corresponding materials, including web sites where such information can be found, and full-text documents. Also there is an option for a remote sending of files directly to the server and transfer them to the database.

Creation and management of experimental dendroecological database. The database is created by means of model experiments, performed on the server by processing the images of tree cuts (graphic files). Together with the output information, obtained from the model experiment, the modeler also enters into the database the information about processed tree disc: tree species, tree orientation, place of acquisition, latitude and longitude of location, and climatic data. The database can be immediately accessed through the server interface, specially developed for these purposes. Also, this database can be autonomously managed via interface available for database manager.

Analysis of dendroecological data. From the entire spectrum of problems, here we will consider the problem of analysis of annual tree rings, and in particular, the calculation of azimuth and root-mean-square width of the region of maximum growth of annual tree rings, and calculation of indices of growth of annual tree ring width. From the information viewpoint, the annual rings on cross-sectional tree cut are a modulated signal, which depends on two variables (azimuth and distance from biologic center of tree disc). It is noteworthy, that this is precisely the two-dimensional analysis of annual tree rings that allows us to obtain new results concerning isotropy of climate-ecological environmental state.

Estimate of ecological environmental state. The estimate is based on the information contained in annual wood layers, decoding of which allows one to reveal seasonal, annual, and multiyear variations of climate on timescales considerably exceeding the period of meteorological observations. Simultaneously, it is possible to reconstruct many important climatic characteristics, most important of which are air temperature in different seasons and yearly, precipitation amount in different seasons and yearly, anomalies of atmospheric pressure, frequency and intensity of droughts, frequency of frost events during growing season and severe frosts during winter.[13]

In the structure of calculation cluster there is a dedicated communication server. The main management mission of the communication server is the creation of the basis of calculation servers and the problems to be solved by them. The information on calculation servers is stored in a special table of database (table of link vectors). The fields in the table are: a universal locator of calculation server (DNS name or IP address), identifier of tasks completed by the server, number of users contacting this calculation server, maximum number of users working simultaneously, state of calculation server (ON/OFF or the presence of connections).

For information support of the work of multi-computer system of dendroecological data processing, the Oracle DMS is used. Both calculation servers and communication server are interacting with servers of Oracle databases. Synchronization and correct handling of all queries for database are performed using standard (routine) means of the Oracle DMS. Server-related parts of calculation cluster, accomplishing main mathematical processing, and that of communication cluster are based on Apache HTTP Server, while user-related part, maintaining interface of remote interaction with user, is based on MS Internet Explorer.

4. Analysis of annual tree rings

In the experiment, testing performance of multi-computer system of dendroecological data processing, we used three calculation servers. They deployed the above-mentioned and actually operational server applications for analysis of annual tree rings. Figure 2 illustrates interaction and data flows between the user and particular calculation server.

The user chooses the class of problems to be solved by selecting the corresponding menu item while trying to connect to communication server.

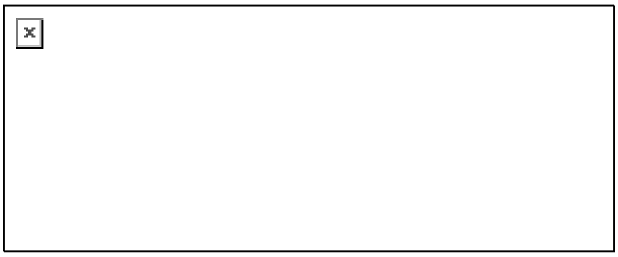


Fig. 2. Illustration of interaction and data flow between user and particular calculation server

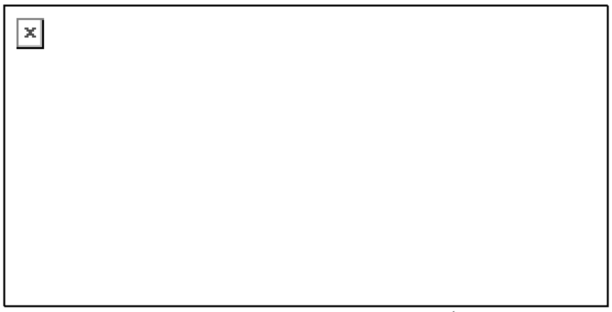
The calculation server is chosen by communication server according to the following algorithm: (1) based on identifier of the class of problems indicated by the user, the link vectors from the table are used to compose a list of calculation servers capable of handling this task; (2) this compiled list of calculation servers is reduced by excluding switched-off servers and servers, currently inaccessible for technical reasons; (3) this remaining list of calculation servers is then checked to delete servers for which the number of users utilizing this calculation server reaches maximum possible number of simultaneously working users; (4) the working calculation server is decided to be the one having maximum difference between number of users utilizing the given calculation server and maximum possible number of simultaneously working users. When non-unique, the calculation server is chosen to be the first from the list.

After the calculation server is identified, control is transferred to it, and the field of the number of users, utilizing this calculation server, is incremented in the table of link vectors.

The user operations included:

Calculation of azimuth and root-mean-square width of the region of maximum growth of annual tree rings: (1) determination of the center of rings; (2) scanning of the image along all the azimuths; (3) conversion of the image from polar to Cartesian coordinates[14]; (4) processing of the image using median (Fig. 3*a*) and polynomial[15] (Fig. 3*b*) filters; (5) conversion of the image from grey-scale to black and white; (6) calculation of annual ring width and integrated growth along all the azimuths; (7) normalization of the width of each ring; and (8) determination of the azimuth of maximum growth and root-mean-square width of the tree growth.

Figure 3*a* shows the effect of median filter with five-element window on radial section of tree disc. Figure 3*b* shows the effect of polynomial filter (fifth-order polynomial) on radial section of tree disc.



a *b*

Fig. 3. Median (*a*) and polynomial (*b*) filtering: before (—) and after (▬▬) filtering.

Calculation of growth indices of annual tree ring width: (1) scanning of the image across the chosen section; (2) image processing using polynomial filter; (3) image conversion from grey-scale to black and white; (4) calculation of growth indices for annual tree ring width; (5) calculation of average growth index of the annual tree ring growth.[16]

Figure 4 illustrates calculation of the growth indices according to the study of annual tree rings over 50 years.

The growth indices were calculated in order to standardize the data representation. The standardization removes age-induced trend from initial measurements of annual ring width, converting measurement sequence to dimensionless indices of annual ring width. The variations of the indices are most modulated by external factors, and the procedures of subsequent analysis allow one to identify the leading environmental factors and estimate their relative contributions. Here, the growth index (index of annual ring width) is a transformed value of annual ring width, obtained by dividing it by the value corresponding to it on the smoothed curve.

After the end of work or 10-min outage, the user is disconnected from the system and the field of the number of users, utilizing this calculation server, is decremented in the table of link vectors.

Actuality and confidence of the field of the state of calculation server (ON/OFF or presence of connection) in the table of link vectors was verified by a special program, run on communication server once every 30 min.

Conclusion

In this paper we have present an architecture of multi-computer system; it includes a calculation cluster and communication server, ensuring access to both dendroecological information and instrumentation tools for mathematical analysis of annual tree rings to retrieve variations of environmental parameters, reflected in tree growth. Mathematical processing of dendroecological data (annual tree rings) is based on representation of wood density as spatiotemporal oscillation process and includes the calculation of azimuth and root-mean-square width of the region of maximum growth of annual tree rings, as well as the growth indices of annual tree ring width.

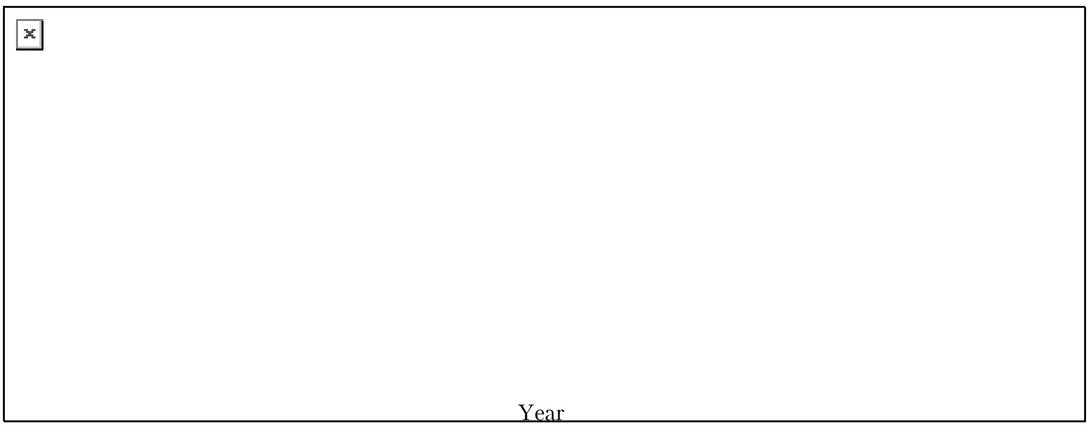


Fig. 4. Calculation of the growth indices of annual tree rings: growth (squares); growth indices (circles); polynomial (vertical bars).

Acknowledgments

This work is supported by the Russian Foundation for Basic Research (grant 05–07–98009).

References

1. Yu.V. Gulyaev, and A.Ya. Oleinikov, Informatsionnye Tekhnologii i Vychislitel'nye Sistemy, No. 3, 4–12 (2003).
2. V.K. Batovrin, Yu.V. Gulyaev, Yu.L. Izhvanov, A.Ya. Oleinikov, A.S. Sigov, and A.N. Tikhonov, http://www.cplire.ru/win/casr/os/ 3_12/8/sp2.htm.
3. V. Mitrofanov, A. Slutskin, K. Larionov, and L. Eisymont, http://www.osp.ru/os/2003/05/029.htm.
4. A. Lizun, http://www.comizdat.com/3/17/ 19/2524.
5. M.A. Kamenshchikov, and V.N. Kornienko, Informatsionnye Tekhnologii i Vychislitel'nye Sistemy, No. 3, 45–50 (2003).
6. L. Chernyak, http://www.osp.ru/os/2004/05/ 031.htm.
7. A.A. Sakharov, Sistemy Upravleniya Bazami Dannykh, No. 4, 55–70 (1996).
8. V. L'vov, Sistemy Upravleniya Bazami Dannykh, No. 3, http://www.isuct.ru/~ivt/books/. DBMS/DBMS7/dbms /1997/03/30.htm.
9. E.A. Vaganov, and A.V. Shashkin, *Growth and Structure of Annual Rings of Coniferous* (Nauka, Novosibirsk, 2000), 232 pp.
10. S.G. Shiyatov, E.A. Vaganov, A.V. Kirdyanov, V.B. Kruglov, V.S. Mazepa, M.M. Naurzbaev, and R.M. Khantemirov, *Methods of Dendrochronology.* Part 1. *Foundations of Dendrochronology. Acquisition and Obtaining of Wood-Ring Information* (Publishing House of Krasnoyarsk State University, Krasnoyarsk, 2000), 80 pp.
11. V.A. Tartakovskii, Yu.N. Isaev, V.D. Nesvetailo, Yu.V. Volkov, and V.N. Popov, Avtometriya **38**, No. 5, 118–127 (2003).
12. V.A. Vitrichenko, V.P. Lukin, L.A. Pushnoi, and V.A. Tartakovskii, *Problems of Optical Monitoring* (Nauka, Novosibirsk, 1990), 351 pp.
13. S.G. Shiyatov, Zap. Sverdlov. Otdel. Vsesoyuz. Botan. Obshchestva, No. 6, 53–81 (1973).
14. P.V. Veltmander, http://algolist.by.ru/graphics/ index.html.
15. G. Forsight, M. Malcolm, and K. Mouler, *Computer Methods of Mathematical Calculations* [Russian translation] (Mir, Moscow, 1980), 243 pp.
16. V.N. Popov, I.A. Botygin, Yu.V. Volkov, and V.A. Tartakovskii, in: *Proceedings of IV International Symposium on Monitoring and Rehabilitation of Environment*, Tomsk (2002), pp. 103–104.