Atayan A. M. (Don State Technical University, Rostov-on-Don) **Calculation of the operating time of a computing system based on correlation analysis**

In the modern world, the issues of ecology and the preservation of the quality of coastal (especially fresh) and commercial waters are becoming more and more relevant. In order to preserve water complexes and maintain their integrity, it is important not only to take organizational, engineering and technical solutions, but also to have highly effective methods for modeling various potential and actual mechanisms of primary and secondary pollution of coastal systems, which make it possible to quickly and efficiently, based on interrelated high-precision models of hydrophysics and hydrobiology predict the processes of pollution spreading and the occurrence of hazardous phenomena in coastal systems [1, 2].

To increase the accuracy of mathematical models based on solving diffusion-convection problems, it is necessary to include factors that have a significant impact on hydrobiological processes: parameterizable microturbulent diffusion and advective transport in various directions [3]. The calculation of data on a multiprocessor computer system can significantly reduce the computation time. However, the time efficiency of a computing system may not always be expected. In this case, it is correct to carry out a theoretical analysis of the calculation of the computation time based on the correlation analysis.

Let there be a certain variable i, which represents the ith observation of the dependent variable y_i, and wee denote the explanatory factors by a vector x_i. Then we can express the multiple regression model in the following form:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i, \tag{1}$$

where $i = 1, 2, \ldots, p$; β – free member, ε_i – member containing the error.

Vector final regularity n is a matrix of values of explanatory factors, dimension n on the $(p+1)$. The model in matrix form will look like

$$Y = X\beta + \varepsilon. \tag{2}$$

The estimate of this model for some sample will be an equation in which $\beta = (\beta_0, \beta_1 ... \beta_p)'$, $\varepsilon = (\varepsilon_0, \varepsilon_1 ... \varepsilon_n)'$.

The condition for minimizing the residual sum of squares can be represented as:

$$S = \sum_{i=1}^{n} (y_{x_i} - y_i)^2 = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon\varepsilon' = (Y - X\beta)'(Y - X\beta) \to min. \tag{3}$$

Carrying out transformations in (3) we obtain

$$S = Y'Y - \beta' X' Y' - Y' X \beta + \beta' X' X \beta. \tag{4}$$

This implies

$$S = Y'Y - 2\beta' X' Y' + \beta' X' X \beta \to min, \tag{5}$$

where $X'X$ – matrix of sums of first powers, squares and pairwise products n observations of explanatory factors; $X'Y$ – vector of products, dimension n observations of explanatory factors and dependent variables.

The solution of the matrix equation will be the vector

$$\beta = (X'X)^{-1}(X'Y), \tag{6}$$

where $(X'X)^{-1}$ – matrix inverse to the matrix of system coefficients; $X'Y$ – vector, free members of the matrix.

To calculate the operating time of the computing system, y_i acts as the final time, and the explanatory factors indicated by the x_i vector are: the size of the computational grid, the number of computing nodes used. Thus, it seems possible to calculate the average running

*The reported study was funded by RFBR according to the research project 20-31-90105

time of the entire system. On the basis of the presented regression analysis, a graph (Fig. 1) was obtained of the dependence of the running time of a parallel program on the amount of transmitted data and the number of involved computing nodes of a multiprocessor system.

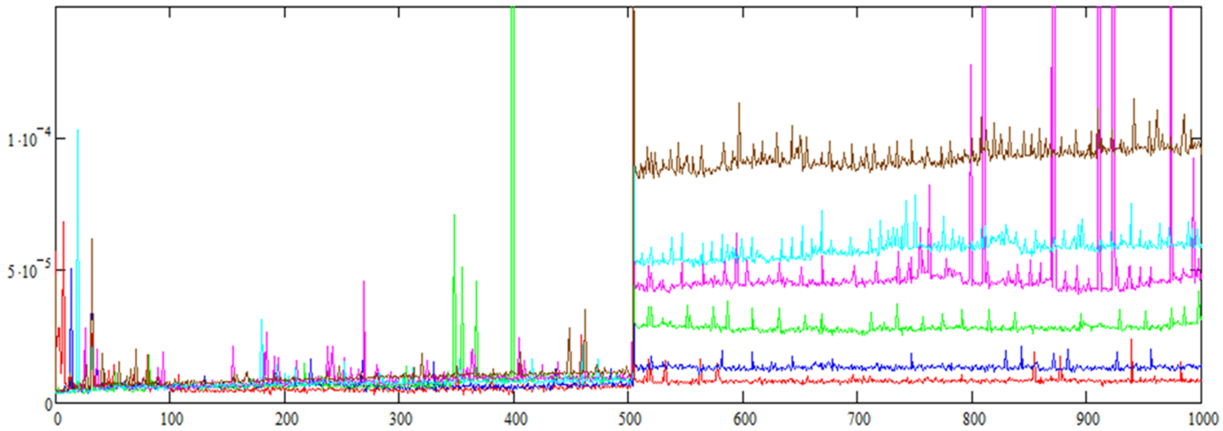

Figure 1 – time dependence on the volume of transmitted data and the number of computing nodes involved.

Latency and data transfer times are calculated using linear regression. The formula for latency is:

$$t_k = \begin{cases} 5.21 \cdot 10^{-6} + 1.53 \cdot 10^{-7} k, \ if \ n \leq 512, \\ 6.733 \cdot 10^{-6} k, \ if \ n > 512, \end{cases} \quad (7)$$

where k – number of nodes. Transmission time per data $t_x = 3.3 \cdot 10^{-9}$.

Regression analysis makes it possible to give a final theoretical estimate of the time of operation of a computing system when parameterizing model data. In the case of developing mathematical models and designing algorithms for the parallel solution of the problem of mathematical modeling of the transfer of pollutants in the aquatic environment, the correlation analysis will make it possible to predict the time for calculating the data.

REFERENCES

1. *Alekseenko, E., Thouvenin B., Tronczynsky J., Carlotti F., Garreau P., Tixier C., Baklouti M.* Modeling of PCB trophic transfer in the Gulf of Lions; 3D coupled model application // Mar Pollut Bull. – 2018. – p.140–155. DOI: 10.1016/j.marpolbul.2018.01.008.

2. *Badenko, V. L.* High-performance computing // St. Petersburg: Publishing House of the Polytechnic University. – 2010. – p. 180

3. *Sukhinov, A. I., Chistyakov, A. E., Shishenya, A. V., Timofeeva, E. F.* Modeling of Coastal Hydrophysical Processes in Multiple-Processor Systems Based on Explicit Schemes // Math. Models Comput. Simul. – 2018. – Vol. 10, Is. 5. – p. 648–658. DOI: 10.1134/S2070048218050125 (*in Russian*)