reviewed paper

Climate Data Analysis on IGIS

Filipp Galyano, Vasily Popovich, Manfred Schrenk, Natalia Zhukova

(Dr. Filipp Galyano, SPIIRAS, 39, 14th Line, V.O., St. Petersburg, 199178, Russia, galiano@oogis.ru)
(Dr Sc, Vasily Popovich, SPIIRAS, 39, 14th Line, V.O., St. Petersburg, 199178, Russia, popovich@oogis.ru)
(Dr Manfred Schrenk, CEIT, Concorde Business Park 2/F, A-2320 Schwechat, Austria, office@ceit.at)
(Dr, Natalia Zhukova, SPIIRAS, 39, 14th Line, V.O., St. Petersburg, 199178, Russia, gna@oogis.ru)

1 ABSTRACT

The paper deals with a system of algorithms used in the intelligent GIS for solving problem of environment monitoring based on integrated data from various data sources, statistical and operative data on meteorology and oceanography in the first place. The described set of algorithms is developed for solving tasks that arise at three main stages of data analyses: data verification, regularization and recognition. Derivative parameters calculation algorithms that allow to compute relevant parameters of the environment in case of limited computational power without computationally complex models solving recognition and forecasting problems play discrete role. Example of the software implementation of the system is considered for the problems of climate monitoring in the Arctic.

2 INTRODUCTION

Exponential increase of the information content in all spheres of human activity at the end of the twentieth and in the beginning of the twenty first centuries called for principally new approaches in information treating and apprehending. Critical breakthrough in informational technology (IT) related to geo-referenced information was achieved due to the development of high-tech intelligent geographical information systems (IGIS) [19]. Although the existing systems provide wide opportunities to be adapted to specific objectives and/or geographical areas they are not sufficiently effective to meet the scope of challenges constantly arising as a result of changing climate conditions. The research overall goal is to develop an algorithm set to empower IGIS-based decision making support system that will synthesize ocean/ice/atmosphere observation-based and model-based products for the purpose of fast access to the available information about the given region.

3 GENERAL STRUCTURE OF DATA PROCESSING

The developed system is expected to handle large volumes of heterogeneous information, at that, taking into account the constraints imposed on available computing resources. To overcome the difficulties encountered in these problems solving three areas of scientific and technological research are formed: data harmonization, integration and fusion [10].

(1) Harmonization in the broadest sense can be interpreted as the standardization of data. Its distinguishing feature is an orientation to a great number of consumers.

(2) Integration, whose hallmark is the orientation to solving some particular class of problems, is supposed to have a specific data model.

(3) Fusion is aimed at obtaining information of higher quality, the exact definition of "quality" depends on the application. Data fusion is accompanied by a decrease in data content.

In view of the described paradigm the data processing within the developed system can be divided into the following stages:

(1) Pre-processing stage (verification).

(2) Regularization stage.

(3) Analysis stage including recognition and forecasting.

3.1 Verification stage

Main purpose of data verification stage is to structure storage, analysis and processing of received data meant for preparing them to meet the challenge of building a regular data grid.

The main requirements to verified data are:

(1) data should not contain duplicating values;

(2) data must be converted to a common format for data submitting and integrated into a single data storage;

(3) metadata should be described in terms of subject domain;

(4) data, if possible, should not contain noise or outliers;

(5) data, if possible, should not contain gaps;

(6) should be formed statistical evaluation of data quality, including:

(6a) techniques of processing and analysis for each type of data received from each source;

(6b) results of statistical processing and analysis of data;

(6c) comparison of results of statistical processing with statistical background.

The considered requirements are provided by:

(1) developing a single model of subject domain and single model of measurements and results of their processing representation;

(2) using expanded set of methods and models of measurements processing and analysis;

(3) using unified analysis methods of measurements;

(4) using a set of self-trained algorithms that provide a choice of adaptive processing parameters;

(5) fuzzy measures application intended for compliance of processing results with the admission values limits and with statistical background.

The stage of data verification received from various sources assumes the consecutive solution of the following key tasks:

(1) development of algorithms for the harmonization of data generated at different times by different data sources;

(2) development of algorithms for data integration, including search for and elimination of duplicating values, usage of specialized techniques for processing data from each data source, statistical data processing;

(3) development of algorithms for data fusion, including algorithms for data interpolation, creation of data formalized description, creation of statistical data models for different regions.

3.2 Regularization stage

The main objective of the regularization is development of a regular grid using gathered measurements and estimation of data accuracy in knots of the regular grid.

The main requirements to the formed regular grid are as follows:

(1) high efficiency of new data assimilation;

(2) application to development of regular grids' methods and models should be universal, and specifics of the area on which data are processed should be considered;

(3) the highest possible resolution of regular grids should be used;

(4) flexibility of a regular grid provided by changing resolution in time, height/depth and coordinates depending on data volume;

(5) ensuring absence of data accuracy loss in presence of large data volume at the expense of low grid resolution;

(6) providing accompanying data on reliability of provided data;

(7) possibility of rebuilding grids locally for an area the new data are received for.

Providing the considered requirements is possible with an allowance for:

(1) applications of multiresolution approach to grids' creating;

(2) use of statistical data models as bases for grids' development;

(3) use of the data processing complex method based on currently existing methods.

The stage of data regularization assumes the consecutive solution of the following key tasks:

(1) development of a general statistical model based on all collected measurements;

(2) development of a regular grid based on collected data and using statistical model;

(3) recalculation of a regular grid through the reanalyze procedures, grid specification when acquiring new data.

3.3 Analysis. Forecast and recognition tasks

Two important problems solved at the stage of analysis is the task of forecasting the given variable (or set of variables) values and the problem of recognizing objects of a given class. Here and below, by pattern is meant an n-dimensional column vector $X = [x_1, \dots, x_n]^T$, where $x_1, \dots, x_n$ are non-negative numbers, and "T" is the symbol of matrix transposition. Under the data recognition is meant a mapping of arbitrary pattern X, where integers 1, ..., c represent the class. The feature i is an element of the vector X with the index i. The set of classified patterns is denoted as U.

The problem of pattern recognition is formulated as follows. Given:

(1) The number of classes c;

(2) A set of m training patterns: $X_1, \dots, X_m$;

(3) Class of any training pattern: $f(X_1) = c_1, \dots, f(X_m) = c_m$;

(4) Arbitrary n – dimensional vector u from U.

(5) A set of weights for classes $\beta_i \in [0;1]$, i=1,…,c.

(6) A set of weights for features $\gamma_j \in [0;1]$, j=1,…,n.

(7) On the set of vectors U is set a metric $\rho(u_1, u_2)$, where $u_1, u_2 \in U$.

(8) The distance d is set – the value of the metric, that determines the size of the neighborhood to be considered for classification.

Search for:

The class of vector u: f(u)=?

The prediction problem can be formulated as follows. In a given area in some way are chosen points p(φ, λ, h) with the coordinates φ, λ and h, where φ and λ are geographic coordinates of a point (latitude and longitude), and h is the depth / height. For these points are known the measured values of various parameters (temperature, pressure, humidity, etc.) for a sufficiently long period of time T, obtained with a discrete time ΔT. The values of these parameters x_i at each point p(φ,λ,h) at time t, t = 0, ...,ΔT forms a row-vector X(φ,λ,h,t)=[x1,...,xn], where n is the total number of investigated parameters, xi=xi(φ,λ,h,t). It is necessary to determine the value of vector X at all points p(φ, λ, h) in a given time T + mΔ, where m – is a constant. The parameters of the forecast are:

The forecast depth. History of the variables' values at a given point taken into account when calculating the future value.

The forecast horizon. Time when the forecast is carried out; it is set by the value of the constant m.

3.3.1 SVD Classification Method

As a unified approach to recognition and forecasting algorithms we use SVD-classification algorithm [21] as fast, reliable and simple method of data orthogonalization and dimension reduction. The method is also can be used as pattern recognition technique for forecasting purposes.

3.3.2 SVD Classification Algorithm

The idea of SVD classification has been previously discussed in several papers like [3, 20] published by the authors. Let us outline only basic concepts of the proposed idea. Two main stages of SVD classification are being studied: training phase; and – recognition one. The training phase is reduced to forming a new space of smaller dimension properties and displaying the training sample elements in it. Recognition consists of classified vector display in the received space and search for the next nearest element out of the training sample.

3.3.3 Training Phase

Brief description of the algorithm training phase is given below:

For each property $j=1,\dots,n$ find and store maximal value from all elements of training set - E_j .

Normalize each property (element X_i) to $[0;1]$ via dividing it by E_j .

Multiply every element of every training vector X by its weight γ_j .

Create training matrix $A=[X_1 \dots X_m]^T$ with dimension $m \times n$.

Calculate maximal singular value s , as well as left and right singular vectors L and R jf training matrix (can be done using different algorithms).

Set s = s(k) , L = L(k) , R = R(k) . Store singular value s and right singular vector R.

For each $i=1,\dots,m$ store component l_i of the left singular vector L and class c_i that corresponds to training element X_i .

3.3.4 Recognition Phase

(1) Normalize each element u to $[0;1]$ via dividing it by E_j .

(2) Multiply each element u by its weight γ_j .

(3) Calculate direct image of u in new feature space:

$$w=\frac{u^T R}{s} \qquad (1)$$

(4) Find max value from $l_i, i=1 \dots m$:

$$M=max\{l_1, \dots l_m\} \qquad (2)$$

(5) Calculate P_i :

$$P_i=\frac{|w-l_i|}{M} \qquad (3)$$

(6) Calculate values t_i for each $i=1 \dots m$ and find minimal value; where t_i are calculated as:

$$t_i = P_i (1-\beta_i) \qquad (4)$$

(7) Set the class c_i as a calculated class for u .

The properties weights and classes are entered for the contribution accounting, at that, for separate properties into recognition results and change of errors' ratio of the first and second kind. For classification an accuracy increase, the proximity of classified objects in space and/or in time is also considered [20].

3.4 The derivative parameters calculation task

Derivative parameters (DPs) are oceanic/atmospheric/ice parameters; they are not directly measured in the field or can be received remotely. Calculation of the most of these parameters generally requires an application of high-resolution 3D modeling systems. Technically these modeling systems are too complicated and computationally too costly. Another approach is to estimate DPs with reasonable accuracy by theoretical/empirical algorithms (models). Models take measured parameters as the input data and produce DP at the output. For the developed prototype (see below) list of the used DPs is given below.

3.4.1 Calculation of water density

Water density (WD) is an important parameter of the sea water primarily determined by thermohaline properties (water temperature and salinity) and hydrostatic pressure. Two types of WD are used for oceanographic applications: (i) potential WD that only depends on thermohaline properties and (ii) in situ WD that additionally depends on hydrostatic pressure and geographical latitude. Importance of WD for oceanographic conditions is dictated by the fact that WD determines distribution of mass around the ocean and therefore controls density (geostrophic) currents [4].

3.4.2 Calculation of sound speed in the water

The sound speed (SS) is the distance traveled in unit of time by a sound wave propagating through an elastic medium. SS in the seawater depends on hydrostatic pressure, temperature, and salinity, and empirical equations have been derived to accurately calculate the sound speed variables. For calculation of the sound speed Wilson's empirical formula proposed in 1960 is commonly used. Wilson's formula is accepted by the National Oceanographic Data Centre (NODC), USA for computer processing of hydrological information.

3.4.3 Geostrophic current

Dynamics of horizontal flows (synoptic and large scale ocean currents) on the rotating Earth is substantially determined by the balance between the horizontal pressure gradient and Coriolis force [4]. This balance in oceanography is known as the geostrophic balance, which determines geostrophic currents (GC). Far from the boundaries (ocean surface, bottom and coasts) the GC provide rather precise approximation of real currents, if the latter are averaged over time scale of the order of few days or more.

3.4.4 Wind current

Dynamics of the upper layer in the sea (down to about 20-30 m) is mainly controlled by the wind. In accordance with Kamenkovich's theory [6] the wind forcing is transfered to the water column through the wind stress at the ocean-air interface. Wind stress initiates motion of water, affected by the Coriolis force. These two forcing equilibrate in about pendulum day. As a result the WC distribution arrives to a steady state.

3.4.5 Wave height

Wave height (WH) is one of the most important parameters which determines the navigational conditions in the sea. Therefore reliable information about WH is highly practical for planning the optimal navigation routes, selection of the locations for running scientific and industrial operations in the sea and providing for safety and economic efficiency of such operations. Currently sophisticated models are used in prognostic centers around the world for WH operational forecasting. However, these models often may not be used because of time-consuming computations. This is why the empirical models based on statistically obtained relationship and providing reasonable accuracy are applied.

3.4.6 Ice drift

Ice drift (ID) is important navigational parameter in the ice-covered seas and in seas with seasonal ice cover, like the Barents Sea. The major forcing, affecting ID is the surface wind. The direction of ID in the low concentrated ice massive coincides with the direction of geostrophic wind (along sea level pressure contours). Simple empirical relationship between the surface wind speed and ID speed was suggested by Shuleikin [13]. In consolidated ice zones this relationship is more complicated and depends on other factors, including total ice concentration, ice thickness, rafting etc. [2]. At present for practical purposes ID could be estimated from sequential satellite images with high resolution.

3.4.7 Ice thickness

Ice thickness (IT) is an important climatological parameter and together with ice extent it allows for estimating the total ice volume and its variability. This parameter is also crucial for navigation and planning of various operations in the ice covered seas. So far IT could be accurately measured only by ice drilling, or by remote sensing measurements from ships (submarines) or aircrafts. It is expected that new satellite-based sensors, like CryoSat [17], will be able to provide the detailed information on IT over entire Arctic in the nearest future. Indirect method of estimating the ice thickness is based on the concept that ice of different type (nilas, grey ice etc.) and different age (fast ice, one year old or over one year old) has different thickness [2]. Empirical relationship between ice type, age and thickness are generally known.

3.4.8 Heat flux at the water-air interface

Turbulent heat flux (HF) at the water-air interface is the major factor that governs energy exchange between the ocean and the atmosphere in high latitudes. HF determines the heat content of the upper ocean layer and through it controls ice freeze/thaw processes. Due to large horizontal inhomogeneity, precise calculation of HF can only be done via direct measurements in the field [11]. The latter can only be executed under the aegis of special targeted experiments from board of research vessels, or at ice camps. Well established empirical method of approximate HF calculation is widely used in oceanography and meteorology. This method is based on Monin-Obukhov theory of atmospheric boundary layer, which allows for linking HF with temperature difference between water and air and surface wind speed through the so-called bulk-formula (see [1]).

4 ROLE OF THE ARTIFICIAL INTELLIGENCE BASED SCENARIO APPROACH

The peculiarity of IGIS project is the need to integrate the existing heterogeneous software into a single system of scientific computations, as well as rapid development and embedding into the system of a set of new software for the emerging research in the additional computational tasks.

An adequate development environment software for the design of complex research system would be environment, allowing for visual, in the form of block diagrams, representation of complex data processing algorithms, that include existing as well as newly developed computation blocs, automatic execution of these algorithms, their easy modification and correction in the process of execution. Such an environment exists. This is a visual environment for scenario simulation DroolsTab [18].

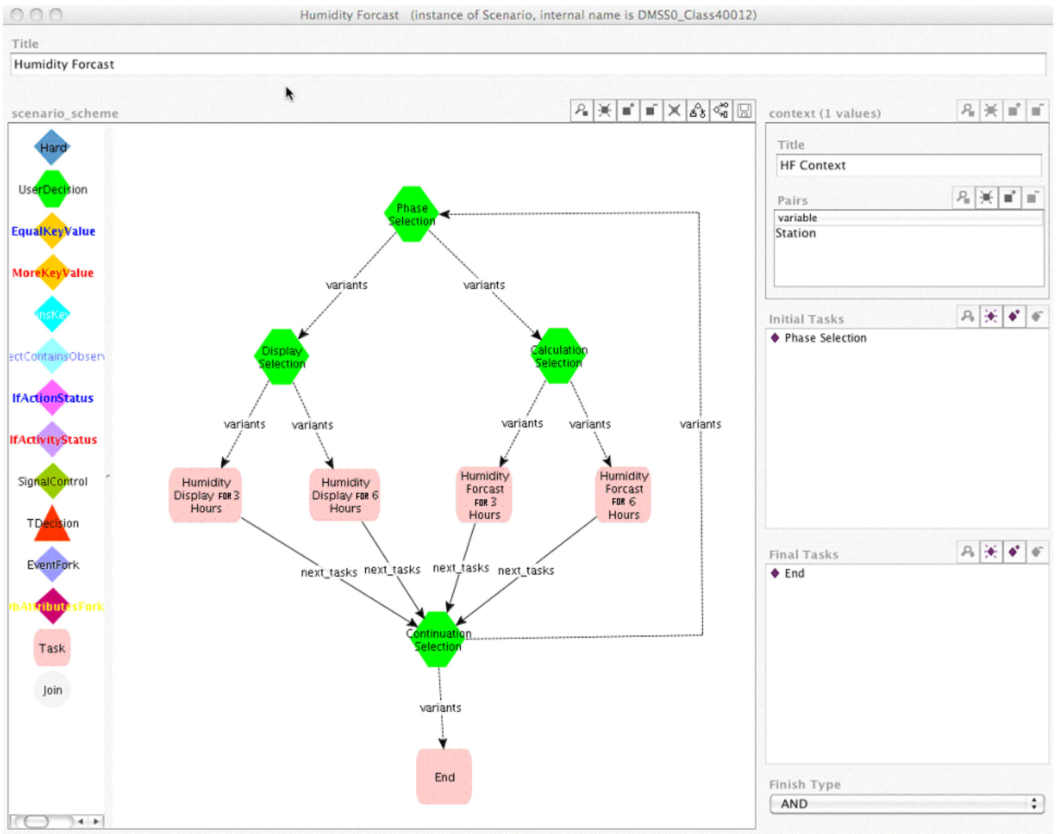

Fig. 1. Scenario development Window

The expert system is used to design, visualize and execute scenarios (algorithms) of data processing. Herewith during the execution of scenarios are automatically invoked existing data processing algorithms and also executed newly developed computational blocks, embedded in the scenarios. The scenarios also contain blocks of decision-making in which data processing branches depending on various conditions.

5 CASE STUDY

Computer prototype of the system was developed by OOGIS research lab.[1] Technically the system will consist of two major components: IGIS-engine, i.e. the program package capable of storage, manipulation and visualization of various types of geo-referenced data, and the data[2] itself. The novel feature of the proposed system, compared with the existing analogues (e.g., Ocean Data Viewer (ODV), http://odv.awi.de/) is its multidisciplinary character.

Here are some screenshots of working system.

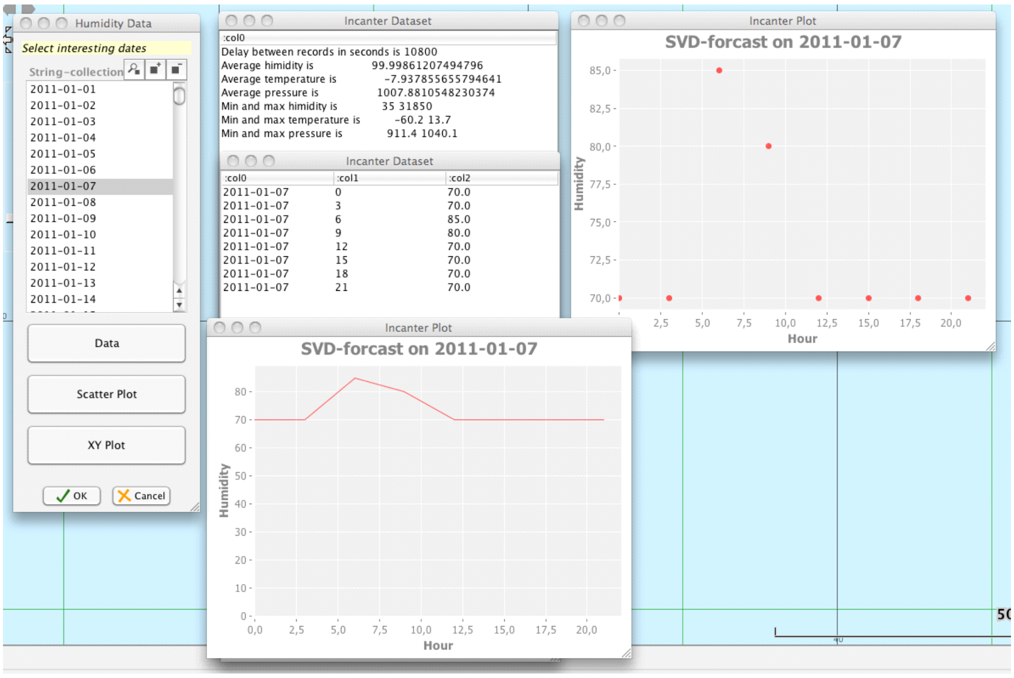

Fig. 2. Windows show the results of calculations at scenario execution

In Fig. 3 the examples of regularization of basic data for temperature are presented.

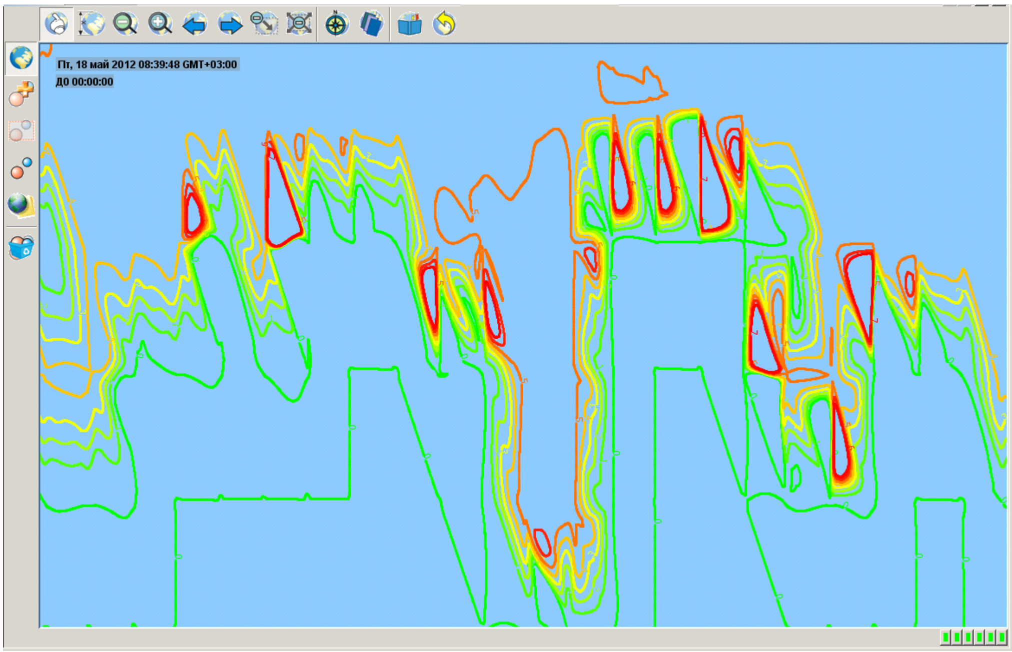

Fig.3. Examples of regularization of temperature values. Color represents temperature value.

[1] According to ONRGrant # 62909-12-1-7013.

[2] Under the term “data” are meant digital arrays used by the internal software for generating the end-products (maps, graphs, diagrams, vertical sections, etc.)

5.1 Ice recognition example

The example of ice situation assessment according to MODIS satellite for the Barents Sea region is given in Fig. 4.

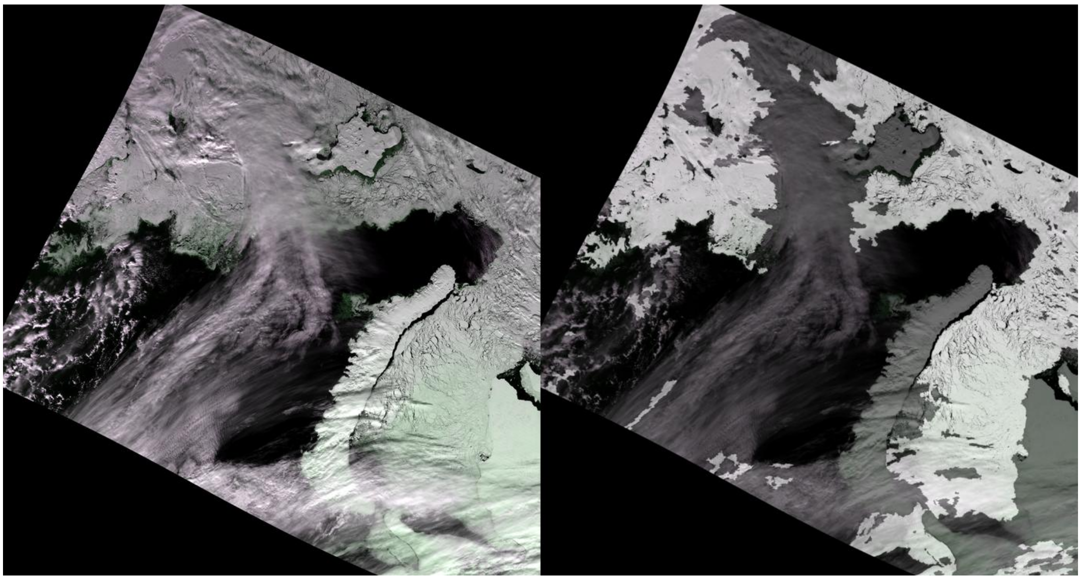

Fig. 4. Example of ice situation assessment according to remote sensing. Input data – at the left, result of ice distribution – to the right. Areas of recognized ice are marked by white color.

5.2 Forecast example

In the system SVD method is applied to the atmosphere parameters forecasting. In the presented example forecasting of temperature, pressure and humidity is executed based on the data received within the previous stages of the project. For comparison the method of the next nearest neighbor is chosen.

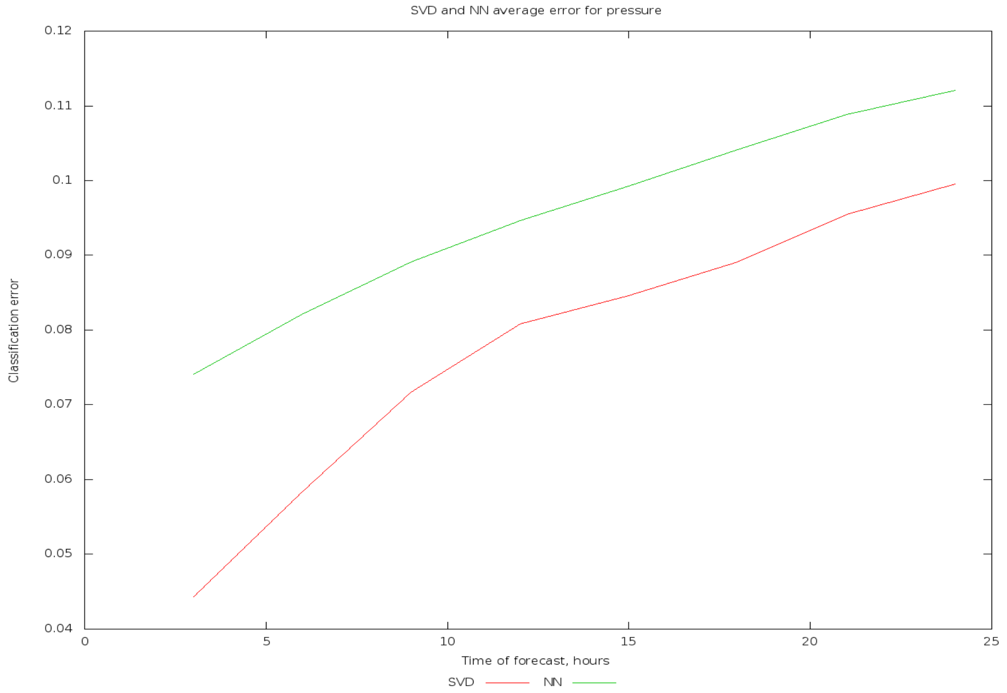

Fig. 5. Comparison of SVD classification and a method of the next nearest neighbor for the pressure values forecast.

At a forecast by the next nearest neighbor method for this station the row site (stories of values), being characterized, the maximum proximity to the current values is chosen. Results of comparison of SVD classification with a method of the next nearest neighbor in forecast accuracy on the basis of over 50000 values of atmospheric parameters (humidity, pressure, temperature) for forecasting for the period from 3 to 24 hours in atmospheric surface layer are presented in Fig.5.

The dependency diagram presents an average dependence between accuracy (Y – axis) and the forecast horizon (X – axis) for all WMO stations. As follows from the diagram SVD-classifier (the red line) shows a steady increase in the forecast accuracy over the next nearest neighbor method (the green line) in the whole range of values. The highest increase of SVD accuracy (approximately 3 %) is observed for short term forecast (under 10 hours).

6 CONCLUSION

Presented basic technology allows to cover the entire range of tasks confronting the system for data processing and storing, as well as visualization of the results and system adaptation for particular current task. Using a scenario approach and unified ontology allows to configure the system according to the peculiarities of the problem. Unified approach to the problems of classification and prediction, based on SVD-classification allows to optimize the detection accuracy or running time. Together with the use of algorithms for calculating the derivative properties that allows to solve the full range of data analysis problems with limited computational resources. Further development of the proposed system can be used (1) to increase the set of avianle algorithms and (2) to develop a system of evaluation criteria for data input to automate the selection of algorithms optimal for the problem solving.

7 REFERENCES

1. Andreas E.L., R.E. Jordan, and A.P. Makshtas, Parameterizing turbulent exchange over sea ice: The ice station Weddell results, Boundary Layer Meteorol., 114, 439-460, 2005.
2. Doronin and Heisin, Sea ice, Hydrometeoizdat, 319 p., Leningrad, 1975 (in Russian).
3. Galliano F.R. Earth surface sites classification Algorithm based on singular valley decomposition method, «Information Technologies». №12, pp. 35-37, 2010 (in Russian).
4. Gill, A.E., Atmospheric-Ocean Dynamics, Academic Press, 662 p, NewYork, 1982.
5. Hall E.L. Computer image processing and recognition.-NY.:Academic Press, 1979.
6. Kamenkovich, V.M., Basics of large-scale ocean circulation theory, In: Oceanology, Nauka, V.2, P.359-400, Moscow, 1978 (in Russian).
7. Persson P.O.G., Fairall C.W., Andreas E.L., Guest P.S., Perovich D.K. Measurement near the Atmospheric Surface Flux Group tower at SHEBA: near surface conditions and surface energy budget. J. Geophys. Res., 107 (C10), DOI: 10.1029/2000JC000705 Moscow, 2002.
8. Popovich, V., Ermolaev, V., Hovanov, N., Chirov, D., Konuhovsky, P., Vanurin. S., Acoustics Field's Calculation Optimization on the Combined Approach Basis.//ECUA 2010, July 25-29, Istanbul, Turkey, 2010.
9. Popovich, V., Vanurin, S., Kokh, S., & Kuzyonny,V., Intellectual Geographic Information System for Navigation Safety // IEEE Aerospace and Electronic Systems Magazine, Volume Twenty Five (Number Eight), NY, 2011.
10. Popovich, V., Voronin, M. Data Harmonization, Integration and Fusion: three sources and three major components of Geoinformation Technologies.//Proceedings of IF&GIS, September 25-27, St. Petersburg, 2005.
11. Repina I.A., Smirnov A.S., Heat and momentum exchange between the atmosphere and ice from the observational data obtained in the region of Franz Josef land, Atmospheric and Oceanic Physics., 36, 5, 618-626, 2000.
12. Sarkisyan,A.S., Numerical analysis and forecast of ocean currents, Hydrometeoizdat, 182 p., 1977 (in Russian).
13. Shuleikin, V.V., Physics of the sea, Nauka, 1083 p, Moscow, 1968. (in Russian)
14. UNESCO: The practical salinity scale 1978 and the international equation of sea water 1980. Tenth report of the joint panel on oceanographic tables and standards. UNESCO technical papers in Marine Sciences, Paris, N36, 1981.
15. Zaslavsky, M,M. and A.S. Monin, Wind waves, In: Oceanology, Nauka, V.2, P.146-179, Moscow, 1978. (in Russian)
16. http://www.aari.ru/news/text/2012/ %D0 %92 %D0 %90 %D0 %AD1605.pdf
17. http://esamultimedia.esa.int/docs/Cryosat/Mission_and_Data_Descrip.pdf
18. http://oogis.ru/component/option,com_remository/Itemid,34/func,fileinfo/id,15/lang,en/
19. Popovich, V.V. Concept of Geoinformatic Systems for Information Fusion// Proceedings of 1st International Workshop on Information Fusion and Geographic Information System. St. Petersburg, September 17-20, 2003.
20. Galliano F.R. Earth surface parts classification algorithm based on singular value decomposition method. «Information Technologies». №12, pp. 35-37, 2010 (in Russian).
21. Gene H. Golub, Charles F. Van Loan. Matrix Computations (3rd ed.) Baltimore: Johns Hopkins University Press, 1996.