

# Applied Tasks in Virtual Research Environment based on a web GIS platform ‘Climate+’

© Fazliev A.Z.                      © Privezentsev A.I.

Institute of Atmospheric Optics SB RAS, Tomsk, Russia

faz@iao.ru                      remake@iao.ru

© Gordov E.P.    © Okladnikov I.G.    © Ryazanova A.A.    © Titov A.G.

Institute of Monitoring of Climatic and Ecological Systems SB RAS, Tomsk, Russia

gordov@scert.ru    okladnikov@scert.ru    raa@scert.ru    titov@scert.ru

© Bart A.A.                      © Starchenko A.V.

Tomsk State University, Tomsk, Russia

bart@math.tsu.ru                      starch@math.tsu.ru

**Abstract.** Two types of applied tasks used in the thematic virtual research environment (VRE) based on the “Climate+” platform are considered. Tasks of both types use significant amount of climatic or meteorological data. The first type of applied tasks whose solutions describe quantitatively climate of chosen territory are on-line computed and mapped using GIS technologies. The second type of applied tasks includes tasks used for decision making. Those along with the computational component, includes tools for expert selection of the initial conditions for these tasks, tools for determining the semantic homogeneity of physical quantities used in the calculations, and software for forming the A-box of the knowledge base of a decision support system (DSS). Presented are several first-type tasks and the second-type task about changing the depth of the active soil layer in the northern regions of Western and Eastern Siberia (the interfluvium of the Ob and Yenisei rivers) for a period of 60 years. The solution of this task and the structure of a typical ontology individual used for the decision making are presented. The role of the ontology description of solutions of applied tasks in the VRE based on the “Climate+” platform is discussed.

**Keywords:** Virtual Research Environment, climate and meteorological applied tasks, platform “Climate+”.

## 1 Introduction

The creation of a virtual research environment (VRE) to deal with large data arrays becomes popular in data intensive domains [1,2]. Climate and meteorology is one of such subject domains [3], within which a thematic VRE was created on the basis of the “Climate+” platform [4]. The platform uses a client-server technology, where information resources of the server component are represented by three layers (data, metadata, and ontologies); applications used these resources are divided into three groups (computational software, geoportal, and ontology application). The client component is connected with a GIS-client and applications used databases, which are currently developed.

Necessity of ontologies usage in geophysical sciences and their role have been demonstrated in the papers [5-9].

In this paper we describe new steps in development of the ‘Climate+’ platform, namely applications of different

types and ontology that refer to three layers (Data, Metadata (Information) and Ontology (Knowledge) layers) of data processing services of the VRE under development. Then we describe the Ontology layer and the applied problem. Special attention will be paid to semantic heterogeneity and formalization of thematic domains related with applied tasks, in particular, a solution of a reduction problem.

There are two types of tasks which should be solved using the thematic VRE. The first type tasks are simple tasks of calculation of some physical value characterizing climatic processes. Examples of those solved within the previously developed “Climate” platform can be found in paper [10]. Some additional examples are also described in the third section of the report. Tasks of the second type which initiated the development of the “Climate+” platform are oriented to the solution of applied tasks, the results of which can be used to make practical decisions in domains crucially depending on climatic conditions. Such tasks require thematic numerical modeling of rather complicated processes occurring under influence of climatic conditions. This modeling involves data obtained by the first type tasks solving. An example of the second type task appears in process a road infrastructures development planning in Northern regions subjected to

climate changes. One of the problems of global warming is an increase in the active soil layer depth due to the increase in the temperature in the northern latitudes and melting ice in the soil. To take into account these processes one needs in a thematic decision support system (DSS). The construction of DSS is connected with the construction of a knowledge base on the possible evolution of the active layer depth and the restrictions imposed on the road infrastructure in permafrost areas. The A-box of this knowledge should contain facts about the change in the active soil layer depth for tens of years ahead.

The report discusses the ways of forming the DSS knowledge base using OWL-ontologies that describe solutions of applied tasks and their properties.

## 2 General architecture

The web GIS platform ‘Climate+’ developed at the IMCES SB RAS is aimed at processing and analysis of geospatial gridded datasets in Earth system science, and online visualization of results [4]. Its architecture is shown in Figure 1. It represents a typical client-server structure, where in general the server is a set of geographically distributed standalone nodes providing common interface (API), and client applications (basically, Web-GIS client). The server part of the architecture includes a high-performance computing system with a data storage attached. It is presented by two tiers:

- resources tier, including data and metadata;
- server applications (middleware) tier.

The client part of the architecture is based on modern graphical web browser. It is presented by a single ‘Client applications’ tier, respectively.

The data layer contains netCDF datasets and PostGIS databases while the metadata layer represents the Metadata database (MDDb) describing geospatial datasets and their processing routines framework of the computational backend.

The computational backend developed contains data processing and visualization software components based on GNU Data Language (GDL, <http://gnudatalanguage.sourceforge.net/>) and Python. Geospatial datasets are processed by a specialized set of validated software modules running within the

The visualization component of the backend generates files in the following formats: GeoTIFF, ESRI Shapefile, Encapsulated PostScript, CSV, XML, netCDF, float GeoTIFF. The final results are represented by raster and vector cartographical layers accompanied by corresponding binary netCDF data.

The geoportal provides cartographical web services such as WMS, WFS, WPS, as well as server-side part of the Web-GIS client applications which comply with general INSPIRE (INfrastructure for SPatial InfoRmation in Europe, <https://inspire.ec.europa.eu>) requirements to geospatial data visualization.

The results of the first type applied tasks (computational problems describing changes of states of spatio-temporal objects) solving are added to the data and metadata layers and might be used later. To solve the problem of semantic heterogeneity in the ‘Climate+’ platform, an ontology layer characterizing the properties

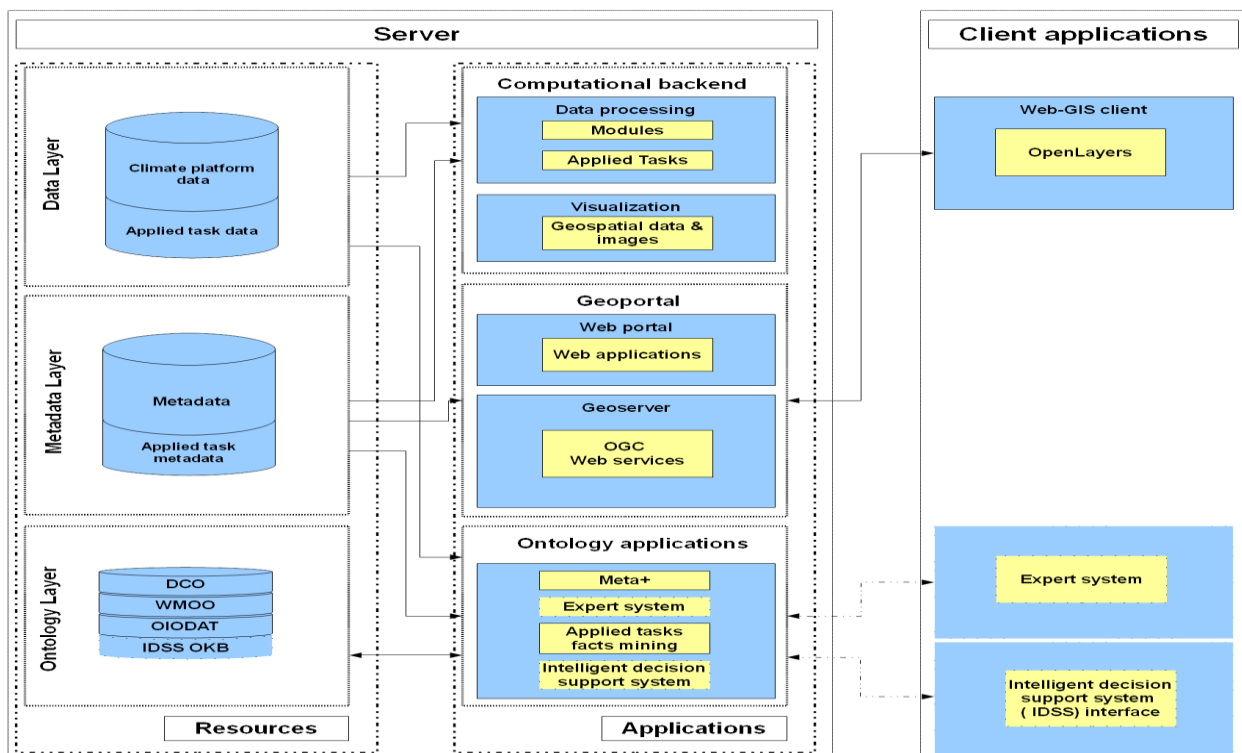


Figure 1 Platform Climate+ general architecture outline

of the data collections (Reanalysis, Observations, Modelling Data) is created. This ontology is used to select input data for Applied Tasks applications.

### 3 Developed software modules

To describe extreme climate events statistics dedicated analytical software tools were integrated into the web GIS platform 'Climate+'. Packages written in R language (<https://www.r-project.org/>, [11]) "extRemes" [12, 13], "quantreg" [14] and "copula" [15, 16] were used as a basis [17].

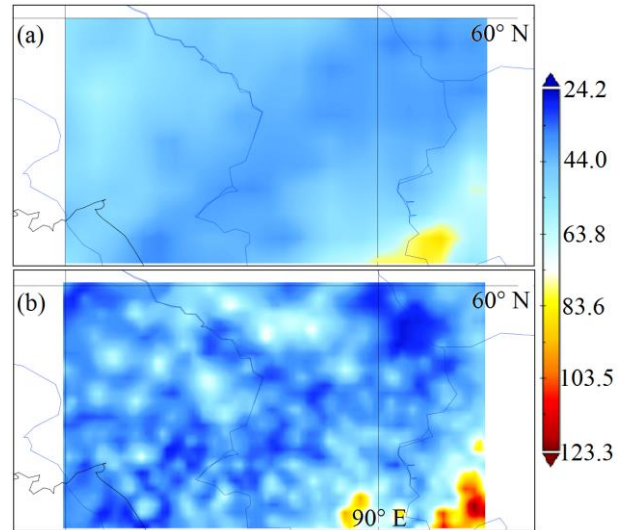
At present, the system allows to calculate basic statistical characteristics and indicators of the temporal structure of meteorological series, describing patterns of changes in time and space. The functionality of the system includes a calculation of trends, an assessment of their statistical significance and a degree of correlation of meteorological quantities. The IPCC recommended climatic change indices are also calculated: extreme values of daily temperature and daily rainfall and their probabilistic characteristics. In addition, the system calculates characteristics of nonstationary distributions of various extreme values. The implemented set of computational procedures makes it possible to get a complete picture of peculiarities of occurring changes in climatic characteristics for the region of study. Modular organization of the system allows to expand its functionality by adding new software components developed by both developers and users of the system.

#### 3.1 Time-dependent statistics of extremes

A statistical description of extreme precipitation and temperature can be achieved using the concepts of extreme value statistics (EVS).

Software implementation of EVS in R language (package "extRemes") allows statistical modelling of maximum values based on a non-stationary generalized extreme value distribution. Required for risk assessment quantities can be calculated using this distribution function. In particular, it is the probability of the observed variable to exceed a certain level. These levels are frequently expressed as return levels  $r_T$  for a certain return period  $T$ .  $r_T$  is defined as the level which is exceeded on average every  $T$ , i.e., with probability  $\frac{1}{T}$ .

This functionality was used to calculate 100-years return levels of July maximum precipitation based on ECMWF ERA Interim [18] (Fig. 2a) and APHRODITE JMA [18] data (Fig. 2b) for the Southern Siberia region (52.5-60° N, 75-95° E).



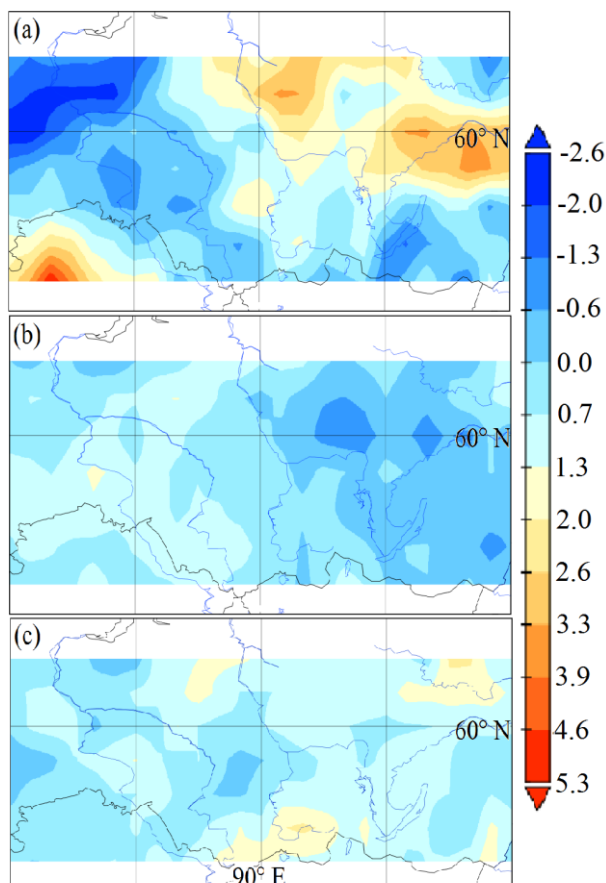
**Figure 2** 100-years return levels of July maximum precipitation for the Southern Siberia region: a - ECMWF ERA Interim data, 0.75x0.75 horizontal grid, 1979-2007 years, b - APHRODITE JMA data, 0.25x0.25 horizontal grid, 1979-2007 years

Results obtained demonstrate similar behavior of the calculated characteristic, but the APHRODITE JMA results have more details and higher values in some regions.

#### 3.2 Quantile regression

The analysis of trends in meteorological observations is one of the most common activities in climate change studies. Quantile regression provides a well-defined statistical framework for estimating the rate of change not only in the mean as in ordinary regression, but in all parts of the data distribution. Given a random variable  $Y$  with cumulative continuous distribution function  $F_Y(y)$ , the quantile function  $Q_Y(\tau)$  is defined from the  $F_Y(y)$  as  $Q_Y(\tau) = F_Y^{-1}(\tau)$ . The quantile is defined as the value  $Q_Y(\tau)$  such that  $P[Y \leq Q_Y(\tau)] = \tau$ ,  $0 \leq \tau \leq 1$ . Then, considering the conditional distribution of  $Y$  given  $X = x$ , the conditional quantile function  $Q_{Y|X}(\tau; x)$  verifies  $P[Y \leq Q_{Y|X}(\tau; x) | X = x] = \tau$ . Whereas ordinary regression is based on the conditional mean function  $E[Y|X = x]$  and minimization of the respective residuals, quantile regression is based on the conditional quantile function and minimization of the sum of asymmetrically weighted absolute residuals  $\sum_{i=1} \rho(\tau) |y_i - Q_{Y|X}(\tau; x = x_i)|$ , where  $\rho$  is the tilted absolute value function.

Quantile regression calculation is implemented in R language by the software package "quantreg" [14]. Quantile values of interest are set between 0 and 1.

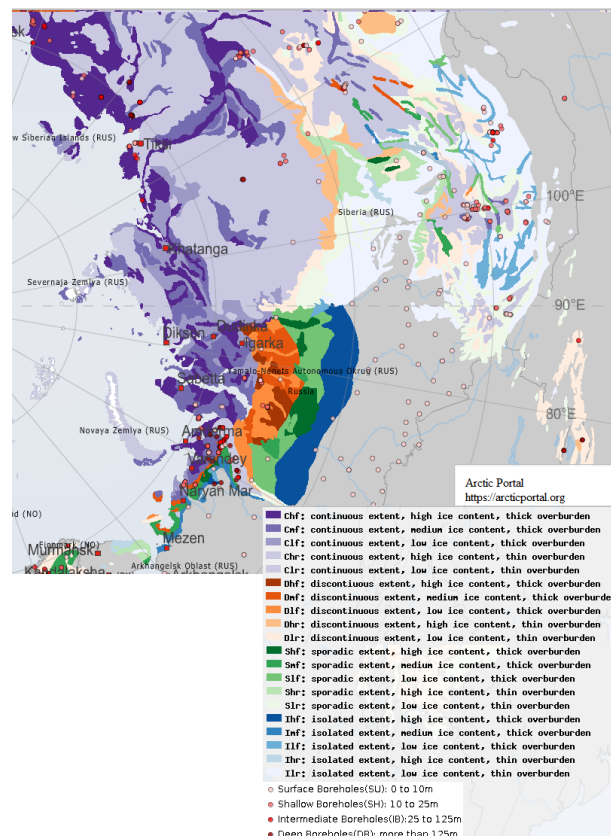


**Figure 3** Maximum January temperature trends based on ECMWF ERA 40 data (2.5x2.5 horizontal grid, 1961-2002 years): a - at quantile 0.05, b - at quantile 0.5, c - at quantile 0.95

The computational backend developed contains data processing and visualization software components based on GNU Data Language (GDL, <http://gnudatalanguage.sourceforge.net/>) and Python. Geospatial datasets are processed by a specialized set of validated software modules running within the framework of the computational backend. Based on ECMWF ERA 40 data [20] trends of maximum January temperature for the Southern Siberia region (50-65° N, 60-120° E) are shown on Fig. 3. Results obtained show that maximum January temperature at quantile 0.05 is changed (both decreased and increased) to the greater extent in comparison with (b) and (c) almost everywhere. Temperature at quantile 0.95 is changed to the less extent.

#### 4 Applied task “Permafrost evolution in the Northern Part of the Ob–Yenisei Interfluve”

The general formulation of the task is considered in a number of publications [21]. The region between 60 to 75° N and 73 to 93° E, related to the European–Western Siberian permafrost sector, is a subject of the study, see Fig. 4. It is covered by a 1.5°x2° grid.



**Figure 4** Permafrost coverage of the Russian Federation territory. [Map provided by [www.arcticportal.org](http://www.arcticportal.org)]

Time evolution of temperature of the frozen soil layer up to 20 meters depth caused by heat conductivity is considered for 1970 - 2030 years. Consideration is based on the simplified model of permafrost thaw which was tested using the measurement data from the station at Cape Marre-Sale [22]. Several admissions and simplifications have been accepted in the model formulation. In view of significant differences in the horizontal and vertical scales of the region under study, where the heat conductivity is non-stationary, the process is studied in a 1D formulation. It is also assumed that the permafrost layer is a homogeneous medium with effective thermophysical properties, which are considered constant and invariable with depth. Thermal effects due to changes in the permafrost phase state are ignored. The monthly average temperatures of the upper soil layer 1 cm thick calculated within a climate model are used in the description of the atmosphere forcing the thermal regime of the soil as boundary conditions at the "atmosphere–frozen layer" interface. At the lower boundary of the region under study, no heat flow is assumed. This approach allows qualitative assessments of the thermal regime of permafrost in the Arctic regions depending on climate changes without excessive detail. The condition of the soil temperature excess over the ice point is used to determine the thawing boundary. To simulate a time variation in the vertical temperature profiles in the permafrost layer, the following heat conductivity equation is used:



$$\rho C \frac{\partial T}{\partial t} = \frac{\partial}{\partial z} \left( \lambda \frac{\partial T}{\partial z} \right), 0 < z < H; t > 0.$$

Here  $T$  is the temperature;  $\rho, C, \lambda$  are the density, specific heat, and the thermal conductivity of the surface layer;  $H$  is the depth of the region under study. The initial and boundary conditions are stated as follows:

$$t = 0: T = \varphi(z), 0 \leq z \leq H;$$

$$z = 0: T = \chi(t), t > 0;$$

$$z = H: \frac{\partial T}{\partial z} = 0, t > 0.$$

Here the functions  $\varphi(z)$  and  $\chi(t)$  are the calculation results in the soil model of Institute of Numerical Mathematics RAS. Our task statement is a simpler than the statement suggested earlier in the work [23]. However, it requires less input parameters, related to the simulation subject, to be specified (e.g., density, specific heat, thermal conductivity of the frozen soil, and soil humidity), does not require specification of boundary conditions for the soil temperature at depth, and is less demanding of computational resources. In other words, the model suggested in this report is more appropriate for climate scales.

Output data of the model are monthly mean vertical profiles of the soil temperature from the surface to a depth of 20 m on an inhomogeneous vertical grid (0.01, 0.02, 0.04, 0.08, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95, 1.05, 1.15, 1.25, 1.35, 1.45, 1.55, 2.0, 3.0, 5.0, 10.0, and 20.0 m) in each cell of a horizontal grid.

Several Circumpolar Active Layer Monitoring (CALM, [https://www2.gwu.edu/~calm/data/north.html]) stations which measure soil temperature and frozen depth are situated in Western Siberia (Table 1). This resource provides for annual data on the depth of frost penetration in the soil (see) by the end of the thaw season.

**Table 1** Sites of thaw-depth measurements in Western Siberia

Code	Name	Location	
		LAT	LONG
R1	NWS	65° 20' N	72° 55' E
R3	MSYP	69° 43' N	66° 45' E
R4	PGP	70° 07' N	75° 35' E
R5	VDYP	70° 17' N	68° 54' E
R5 A	VDYP	70°16'31.8" N	68°53'29.9" E
R5 B	VDYP	70°17'43.8" N	68°53'00.5" E
R5 C	VDYP	70°18'05" N	68°50'28.7" E
R5 D	VDYP	70°16'27" N	68°53'26.8" E
R50a	UGF GP5	66.31537 N	76.90772 E
R50b	UGF GP15	67.477910 N	76.6952900 E
R53	HPU	66.723483 N	66.080488 E

Designation used in Table 1: NWS is used for Nadym, West Siberia; MSYP, for Marre-Sale, Yamal Peninsula;

PGP, for Parisento, Gydan Peninsula; VDYP, for Vaskiny Dachy, Yamal Peninsula; UGF, for Urengoy Gas Field; and HPU, for Harp, Polar Urals.

## 6 Ontology description of solution

The results of the numerical solution of the task are numerical arrays of temperatures at different depths (24 levels) for 720 months (60 years) for each of 100 horizontal cells (10 in latitude and 10 in longitude). Based on the values of the numerical arrays, parameters p1 and p2 are calculated, which are the values of the properties of the applied ontology individuals generated. The properties of these individuals are listed in Table 2.

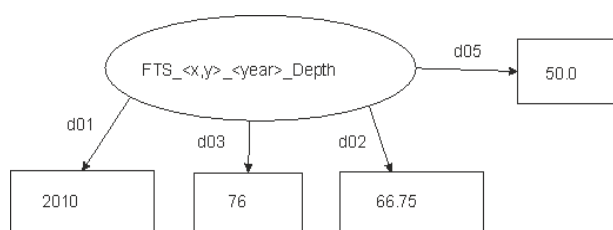
**Table 2** Properties of individuals that characterize the solution of the applied task

Domain	Property	Range	
t3:OutputData	t3:hasTime	string	d01
t3:OutputData	t3:hasLatitude	float	d02
t3:OutputData	t3:hasLongitude	float	d03
t3:OutputData	t3:hasDepth	float	d04
t3:OutputData	t3:has_Site_Average_of_the_Annual_End-of-Season_Thaw_Depth	float	d05
t3:OutputData	t3:hasAverage_Monthly_Thaw_Soil_Temperature	float	d06

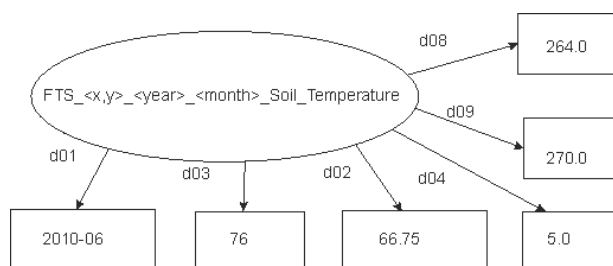
Based on the numeric arrays, the characteristics selected, which are the values of the properties of applied ontology individuals generated, are calculated. The property t3:hasAnnual\_End-of-Season\_Thaw\_Depth is valuable only in the case where the measurement station coordinates fall in a 1.5°×2° cell centered at a point that corresponds to values of the properties t3:hasLatitude and t3:hasLongitude. No individuals are generated for cells with negative soil temperatures.

## 7 A-Box of DSS knowledge base

Two types of individuals are constructed by 1728000 soil temperature values calculated. Examples of the structure of such individuals are given in Figs. 5 and 6. Using these temperature values, its annual means are calculated, and the number of a level at which the soil temperature changes sign is found. The value of the maximum thaw depth allows one to compare the numerical simulation results with the measurement data. The mean temperature and depth values can be used to analyze the permafrost structure (transition from continuous to discontinuous propagation) and the character of the phenomenon (periodicity or trend).



**Figure 5** Scheme of the individual describing the calculated value of the annual end-of-season thaw depth in the fixed point



**Figure 6** Scheme of the individual describing the calculated value of the soil temperature at fixed depth (5 meters) in the fixed point

FTS is the prefix that shows that an individual of the solution of Freeze-up and Thawing of Soil task; <x,y> are the geographical coordinates of the center of a computing cell; <year> is the year; <month> is the month; the ending Depth/Soil\_Temperature designates which physical parameter is described.

## 8 Conclusion

We presented description of the two types of application used 80 Tb IMCES SB RAS collections of climatic and meteorological data. In the first type of applications statistical characteristics of climatic characteristics are on-line calculated and relevant results are presented to user as maps of their fields. The second type applications are aimed at short- or long-term prognosis of physical values evolution important for decision making. As an example of such application the problem of long-term active soil layer evolution in Northern regions is considered. On the basis we created the database which is used to form knowledge base about evolution of physical values determining behavior natural objects and a planned transport infrastructure in the Northern part of West Siberia.

To expand web GIS platform 'Climate+' functionality we plan development of a new software module utilizing the powerful package "copula". It will allow one to calculate probability distributions of multivariate random variables and determine a structure of dependence between different climatic variables.

**Acknowledgments.** The authors thank the Russian Science Foundation for the support of this work under the grant No16-19-10257.

## References

- [1] European Virtual Environment for Research - Earth Science Themes: a solution (EVER-EST), Horizon 2020, grant agreement no 674907, <http://ever-est.eu/>
- [2] Fulvio Marelli, Helen Glaves, Mirko Albani (2017) EVER-EST: a virtual research environment for the Earth Sciences; Geophysical Research Abstracts Vol. 19, EGU2017-17847, EGU General Assembly 2017
- [3] L. Kalinichenko, A. Fazliev, E. Gordov, N. Kiselyova, D. Kovaleva, O. Malkov, I. Okladnikov, N. Podkolodny, N. Ponomareva, A. Pozanenko, S. Stupnikov, A. Volnova, New Data Access Challenges for Data Intensive Research in Russia, CEUR Workshop Proceedings, v. 1536, 2015, P.215-237, 17-th International Conference on Data Analytics and Management in Data Intensive Domains, DAMDID/RCDL 2015; Obninsk; Russian Federation; 13 - 16 October 2015; Code 118237.
- [4] A.A. Bart, A.Z. Fazliev, E.P. Gordov, I.G. Okladnikov, A.I. Privezentsev, A.G. Titov, Virtual Research Environment for Regional Climatic Processes Analysis: Ontological Approach to Spatial Data Systematization, Data Science Journal, 2018 (in press)
- [5] Athanasios, N, Kalabokidis, K, Vaitis, M and Soualakellis, N 2009 Towards a semantics-based approach in the development of geographic portals. Computers & Geosciences, 35(2): 301-308. DOI: <https://doi.org/10.1016/j.cageo.2008.01.014>
- [6] Bogdanovic, M, Stanimirovic and A, Stoimenov, L 2015 Methodology for geospatial data source discovery in ontology-driven geo-information integration architectures. Web Semantics: Science, Services and Agents on the World Wide Web, 32: 1-15. DOI: <https://doi.org/10.1016/j.websem.2015.01.002>
- [7] Brodaric, B, Fox, P and McGuinness, D L 2009 Geoscience knowledge representation in cyberinfrastructure. Computers & Geosciences, 35 (4): 697-699. DOI: <https://doi.org/10.1016/j.cageo.2009.01.001>
- [8] Husain, M F, Al-Khateeb, T, Alam, M and Khan, L 2011 Ontology based policy interoperability in geo-spatial domain. Computer Standards & Interfaces, 33(3): 214-219. DOI: <https://doi.org/10.1016/j.csi.2010.03.011>
- [9] Lutz, M, Sprado, J, Klien, E, Schubert, C and Christ, I 2009 Overcoming semantic heterogeneity in spatial data infrastructures. Computers & Geosciences, 35(4): 739-752.
- [10] Riazanova A A, Voropay N N, Okladnikov I G and Gordov E P. Development of

- computational module of regional aridity for web-GIS "Climate" // IOP Conf. Series: Earth and Environmental Science. 2016. V. 48. 012032. doi:10.1088/1755-1315/48/1/012032DOI: <https://doi.org/10.1016/j.cageo.2007.09.017>
- [11] Gilleland E 2011 Using R to Analyze Extremes National Center for Atmospheric Research (Boulder, Colorado, U.S.A)
- [12] Gilleland E 2016 Package "extRemes" The Comprehensive R Archive Network (CRAN) <https://cran.r-project.org/web/packages/extRemes/extRemes.pdf>
- [13] Gilleland E and Katz R W 2016 extRemes 2.0: An Extreme Value Analysis Package in R Journal of Statistical Software 72 8
- [14] Koenker R., Portnoy S., Tian P., Zeileis A., Grosjean P. and Ripley B. D., 2017 Package "quantreg" The Comprehensive R Archive Network (CRAN) <https://cran.r-project.org/web/packages/quantreg/quantreg.pdf>
- [15] Hofert M, Kojadinovic I., Maechler M., Yan Y., 2017 Package "Copula" The Comprehensive R Archive Network (CRAN) <ftp://cran.r-project.org/pub/R/web/packages/copula/copula.pdf>
- [16] Yan J 2007 Enjoy the Joy of Copulas: With a Package copula Journal of Statistical Software 21 4
- [17] A A Ryazanova, I G Okladnikov and E P Gordov 2017 Integration of modern statistical tools for the analysis of climate extremes into the web-GIS "CLIMATE" IOP Conference Series: Earth and Environmental Science 96 doi:10.1088/1755-1315/96/1/012014
- [18] Dee D P et al. 2011 The ERA-Interim reanalysis: configuration and performance of the data assimilation system Quarterly Journal of the Royal Meteorological Society 137 Issue 656 Part A pp 553–597
- [19] APHRODITE JMA, [http://www.chikyu.ac.jp/precip/data/APHRO\\_V1003R1\\_readme.txt](http://www.chikyu.ac.jp/precip/data/APHRO_V1003R1_readme.txt)
- [20] Kallberg P, Simmons A, Uppala S, Fuentes M 2007 ERA-40 Project Report Series. The ERA-40 Archive Report of European Centre for Medium Range Weather Forecasts, England
- [21] E.E. Machul'skaya, Vasily N. Lykossov, Simulation of the thermodynamic response of permafrost to seasonal and interannual variations in atmospheric parameters, Izvestiya Atmospheric and Oceanic Physics 38(1):15-26, January 2002
- [22] Sudakov I.A., Bobylev L.P., Beresnev S.A. Modeling permafrost thermal regime under ongoing climate change. Vestnik of SpBGU. Earth Sciences. (1), 81-88.
- [23] Alipova K.A., Bart A.A., Fazliev A.Z., Gordov E.P., Okladnikov I.G., Privezentsev A.I., Titov A. G., "Systematization of climate data in the virtual research environment on the basis of ontology approach", Proc. SPIE v.10466, 23-rd International Symposium on Atmospheric and Ocean Optics: Atmospheric Physics, 1046675 (2017); <https://doi.org/10.1117/12.2289761>