Comparative analysis of data on air temperature based on current weather data sets for 2007-2019

A V Dergunov and O E Yakubailik

1 Federal Research Center Krasnoyarsk Science Center of the SB RAS, 660036, Krasnoyarsk, Russia, Akademgorodok, 50
2 Institute of Computational Modelling SB RAS, 660036, Krasnoyarsk, Russia, Akademgorodok, 50/44
3 E-mail: oleg@icm.krasn.ru

Abstract. Modern tasks in the field of agriculture require meteorological information of high spatial and temporal resolution. In this study, air temperature was compared using the CRU_TS, ERA5-Land, and GFS datasets in the study area for the period from 2007 to 2019. The information obtained showed a high level of correlation (0.9) of the considered data sets for the study period. However, a more detailed analysis revealed that there may be months when the air temperature values of different data sets can vary significantly for several consecutive years, and the magnitude of these deviations relative to the compared data sets can vary depending on the time of year. In addition, features of the underlying surface, for example, the presence of extensive mountain ranges, can affect the final data and the results of their comparisons. It is shown that modern ERA5-Land and GFS data sets are not inferior to CRU_TS data, and their advantages in the form of high spatial and temporal resolution can provide better results for solving problems.

1. Introduction

Modern tasks in the field of agricultural industry require the use of various meteorological information of high spatial and temporal resolution. Climate change can lead to drought, pest outbreaks, and fires. Changes in the thermal regime can also lead to a redistribution of water resources, the conditions for the development of living organisms, and a decrease in the bio-productivity of agroecosystems [1, 2]. According to calculations, the climate change in Central Siberia in the next hundred years will be great, and therefore it will take a long time to adapt the natural systems. [3].

The study of the delicate relationships between meteorological parameters (temperature, air humidity deficit, rainfall) and changes in seasonal crop yields requires long-term historical weather data [4]. The use of such data can provide an understanding of existing climate change trends, contribute to the development of long-term and short-term yield forecasts, in order to determine the most suitable for growing crops [5].

Meteorological information can be obtained from ground-based weather observation stations data. However, in the tasks of exploring territories of several thousand square kilometers, the spatial resolution of the available network of ground-based weather stations may not be enough. An alternative is special weather data sets, for example, the Climatic Research Unit gridded Time Series (CRU_TS), and weather forecast models, for example, the Global Forecast System (GFS). These sources provide
layers of various meteorological information deposited on regular rectangular grids covering the entire globe. The spatial resolution of such data can vary from 2.5 degrees to 0.25 and higher. Developers are constantly improving the accuracy of such information.

In this work, the research area is the region that covers the southern territory of the Krasnoyarsk Territory, the Republic of Khakassia, and the northern part of the Republic of Tyva with coordinates 88-98°E, 51-57°N (figure 1).

The study area is characterized by the presence of an extensive system of mountain ranges in the south of the Krasnoyarsk Territory and the Republic of Khakassia and in the north of the Republic of Tyva.

Figure 1a shows the spatial distribution of the July average air temperature in 2019 at an altitude of 2 meters from the earth’s surface of the study area according to the ground component of the 5th generation reanalysis model of the European Center for Medium-Range Weather Forecasts (ECMWF) ERA5-Land. The spatial resolution of the data presented is 0.1 degrees (about 9 km) per pixel. The difference between air temperatures in mountainous regions and flat terrain is up to 10-12 degrees. This image shows the capabilities of modern meteorological datasets.

The aim of this work is to compare the average monthly air temperature in the study area according to three meteorological data sets: CRU TS, ERA5-Land and Global Forecast System for a 13-year period from 2007 to 2019.

2. Materials and methods
The meteorological data sets used in the work are free and distributed freely.

2.1. Global Forecast System Model
The Global Forecast System (GFS) model is a global numerical weather prediction system containing a global computer model and variational analysis performed by the United States National Meteorological Service (NWS). This model is a combination of four separate models: atmosphere, ocean, land/soil and sea ice. This dataset contains dozens of atmospheric and ground-based variables from temperature, wind and rainfall to soil moisture and atmospheric ozone concentrations. This is one of the most famous world meteorological models. Global data analysis and forecasting is carried out 4 times a day. Weather forecast is available up to 16 days in advance [6].
The accuracy of the GFS model is constantly improving. In particular, data with a horizontal resolution of 1 degree has been available since March 2004, with a resolution of 0.5 degrees since January 2007. The horizontal resolution at the moment is 0.25 degrees (about 25 km at the latitude of the city of Krasnoyarsk) since January 2015.

GFS model data is presented in *.grib2 format. This data format is a standardized World Meteorological Organization (WMO) and is designed to store historical and forecast weather data [7]. Each individual file contains more than 500 layers of various meteorological information at more than a hundred vertical levels.

In this work, we used the actual data of temperature analysis at the surface level and at a height of 2 meters from the surface of the earth. Their horizontal resolution was: 0.5 degrees for the period from 2007 to 2014 and 0.25 degrees for the period from 2015 to 2019.

2.2. CRU_TS dataset
The Climatic Research Unit gridded Time Series (CRU_TS) dataset is a widely used climate dataset developed by the British Council for Environmental Research (NERC). CRU_TS provides high spatial resolution information based on data from an extensive network of observations of ground-based weather stations from 1901 to the present [8].

The CRU_TS dataset presents dozens of different meteorological variables averaged over months. The current spatial resolution is 0.5 degrees (about 50 km at the latitude of the city of Krasnoyarsk). The spatial coverage of CRU_TS data is the surface of all land except Antarctica [8].

The data from the CRU_TS set is presented in the form of files of two formats *.dat and *.nc. One file contains data for 10 years of observation or 120 months. The work used data on air temperature from 2007 to 2019 (156 months).

2.3. ERA5-Land dataset
ERA5-Land is a fifth-generation reanalysis dataset of the European Center for Medium-Range Weather Forecasts (ECMWF) with higher spatial resolution compared to ERA5. The ERA5-Land kit was obtained using the ECMWF ERA5 ground-based climate reanalysis component. The ERA5-Land provides a consistent view of water and energy cycles at the surface level over several decades, from 1981 to the present. The native spatial resolution of the ERA5-Land dataset is 0.1 degrees (about 9 km at the latitude of the city of Krasnoyarsk) [9].

The data from the ERA5-Land set provides hourly or monthly average information and is presented in the form of files with the resolution *.grib or *.nc. One file contains data for the selected time interval. The work used average monthly data on air temperature at a height of 2 meters from the surface for the period from 2007 to 2019 (156 months).

2.4. Data processing
GFS model data was downloaded from the archives of the US National Center of Atmospheric Research (NCAR) and the US National Climatic Data Center (NCDC) [10, 11].

The first step in processing the downloaded GFS data archive was to trim the source files into the layers “TMP: surface” and “TMP: 2m above ground” and the coordinates of the selected study area. To accomplish this task, the wgrib2 program was used. This program is specially designed by the manufacturer of data from the GFS model for reading and converting data in *.grib2 format [12]. To automate the process, a special script was written in the Python programming language.

The second processing step was to convert the received data into geo-referenced raster GeoTIFF files. Next, the process of averaging the received raster images per day was carried out. Then, the monthly average temperature data of the study area were made at two levels: on the surface and 2 meters above the earth’s surface. To automate the data processing, special scripts were written in the command-line language of the Windows OS using the GDAL library for processing geospatial data.

The CRU_TS data set was downloaded in *.nc format from the Center for Environmental Data Analysis (CEDA) archive [13].
The ERA5-Land data set was downloaded in *.nc format from the Climate Data Store (CDS) archive [14].

The processing of *.nc format files of the CRU_TS and ERA5-Land data sets consisted in extracting the monthly average data and cutting them according to the coordinates of the study area. To automate the processing, a special script was written in Python.

3. Results and discussion

As a result of processing the initial data, four archives were obtained containing information on the air temperature of the study area for a 13-year period from 2007 to 2019:

- GFS model data archive at the surface level.
- GFS model data archive at a level of 2 meters from the ground.
- archive of the CRU TS data set.
- ERA5-Land dataset archive.

Figure 2 shows the mid-January and mid-July variations in air temperature in the territory of the city of Krasnoyarsk and its immediate vicinity (coordinates: 92.25 – 93.75°E, 55.5 – 56.5°N) according to information from the analyzed data sets for the period from 2007 to 2019.

![Figure 2. January (a, b) and July (c, d) average air temperature variations according to information from 4 data sets for a 13-year period.](image)

Figures 2a and 2b show the variations in January average air temperature in the territory of the city of Krasnoyarsk and its immediate surroundings. The data obtained show that the profile of changes in the air temperature graphs corresponding to the GFS model data, in both cases, is more consistent with the data of the ERA5-Land set.

Figures 2c and 2d show the variations in July average air temperature in the territory of the city of Krasnoyarsk and its immediate environs. It is shown here that in the period from 2007 to 2011 the given data sets demonstrate different air temperatures, the difference is on average 2.5 degrees.
Table 1 shows the correlation values of the January and July average air temperatures from the analyzed data sets for the territory of the city of Krasnoyarsk and its immediate vicinity for the period from 2007 to 2019.

**Table 1.** Correlation of the January and July average air temperatures from the analyzed data sets over a 13-year period.

| Data sets               | Correlation R | January | July |
|-------------------------|---------------|---------|------|
| CRU_TS – GFS:2m        |               | 0.9     | 0.5  |
| CRU_TS – GFS:surface   |               | 0.9     | 0.5  |
| ERA5-Land – GFS:2m     |               | 0.97    | 0.94 |
| ERA5-Land – GFS:surface|               | 0.97    | 0.91 |
| CRU TS – ERA5-Land     |               | 0.86    | 0.6  |

Table 2 shows the correlation values of average monthly air temperatures from the analyzed data sets over the territory of the city of Krasnoyarsk and its immediate vicinity for the period from 2007 to 2019.

**Table 2.** Correlation of the average monthly air temperatures from the analyzed data sets over a 13-year period.

| Data sets               | Correlation R |
|-------------------------|---------------|
| CRU TS – GFS:2m        | 0.93          |
| CRU TS – GFS:surface   | 0.93          |
| ERA5-Land – GFS:2m     | 0.99          |
| ERA5-Land – GFS:surface| 0.99          |
| CRU TS – ERA5-Land     | 0.94          |

Figure 3 shows the difference between the average January temperature according to ERA5-Land and GFS and the average July temperature according to CRU TS and GFS, averaged over the period from 2007 to 2019, for the study area.

**Figure 3.** The difference between: a) January average temperature data of ERA5-Land and GFS; b) July average temperature data of CRU TS and GFS data averaged over 13 years.
The figures show that the difference in the average monthly temperature for different data sets in a significant part of the territory does not exceed 1-2 degrees. Slightly higher differences are observed in the highlands.

4. Conclusions
Data analysis showed a high level of correlation (0.9) of the considered data sets for the entire studied time period. However, a detailed analysis revealed that there may be months when the air temperature from different data sets can vary significantly over a period of time, for example, the July CRU_TS data from 2007 to 2011.

It was found that the value of the air temperature with which the analyzed data sets differ from each other depends on the time of year. In winter, this figure is higher than in summer.

The overall correlation of the data sets CRU_TS, ERA5-Land and GFS was at a high level. Starting in 2015, the GFS model data has twice the spatial resolution (25 km), compared to the CRU TS data, as well as a much higher time resolution (4 times a day) and hundreds of different meteorological data layers. At the same time, the ERA5-Land dataset provides hourly or monthly average data with a spatial resolution of about 9 km, but it has a much more limited set of meteorological information compared to the GFS model.

Modern ERA5-Land and GFS datasets are not inferior to CRU TS data, and their advantages, in the form of high spatial and temporal resolution, can provide better results for solving tasks.

References
[1] Panova M L 2011 Assessment of the possible impact of climatic factors on agriculture in the south of the Tyumen region Bulletin of the Tyumen State University 4 66-72
[2] Dergunov A V and Yakubaylik O E 2019 Analysis of climatic characteristics of the distribution area of the Siberian silkworm Spatial data processing for monitoring natural and anthropogenic processes (SDM-2019) 349-53
[3] Chebakova N M, Reifeldt D and Parfenova E I 2003 Redistribution of plant zones and populations of Siberian larch and Scots pine in Central Siberia during climate warming Siberian Ecological Journal 10(6) 677-86
[4] Kurdyukov U F, Levitskaya N G, Loshchinina L P, Shubitidze G V and Vasilieva M U 2014 Dependence of the spring wheat crop on the type of crop rotation and meteorological conditions Agriculture 1 41-3
[5] Bryl S V 2015 Adapting agriculture to global climate change Environment. management 5 83-7
[6] The Global Forecast System (GFS) Documentation Retrieved from: https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs.php
[7] WMO GRI2 Edition 2 Documentation Retrieved from: https://www.wmo.int/pages/prog/www/WMOCodes/Guides/GRIB/GRIB2_062006.pdf
[8] Harris I et al. 2020 Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset Scientific data 7(1) 1-18
[9] ERA5-Land: data documentation Retrieved from: https://confluence.ecmwf.int/display/CKB/ERA5-Land%3A+data+documentation
[10] NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive Retrieved from: https://rda.ucar.edu/datasets/ds084.1/
[11] Global Forecast System (GFS) Retrieved from: https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs
[12] Wgrib2 Utility to read and write grib2 files Retrieved from: https://www.cpc.ncep.noaa.gov/products/wesley/wgrib2/
[13] Centre for Environmental Data Analysis (CEDA) CRU TS v. 4.04 Retrieved from: https://crudata.uea.ac.uk/cru/data/hrg/#current
[14] Climate Data Store. ERA5-Land monthly averaged data from 1981 to present Retrieved from: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land-monthly-means