Analysis of accumulated precipitation based on satellite data in Central Siberia

O E Yakubailik¹,³ and T V Yakubailik²
¹ Federal Research Center Krasnoyarsk Science Center of the SB RAS, Akademgorodok, 50, Krasnoyarsk, 660036, Russia
² Institute of Computational Modelling SB RAS, Akademgorodok, 50/44, Krasnoyarsk, 660036, Russia
³E-mail: oleg@icm.krasn.ru

Abstract. Central and southern regions of the Krasnoyarsk Territory are chosen for the study of precipitation in this investigation. Precipitation using satellite data taken from the NASA IMERG-V06 dataset has been explored. It has been compared with the daily data from weather stations received from the NOAA GSOY archive and with the CPC Unified V1.0/RT dataset based on the interpolation of data of the ground-based weather stations. High degree correspondence between the studied data sets has been revealed using statistical indicators.

1. Introduction
Territory considered in this article located in central and southern areas of the Krasnoyarsk Territory. It geographically occupies a central position in Russia so it is of significant interest for agroclimatic meteorology, fire hazard assessment and other relevant regional tasks [1, 2]. It includes four watersheds of the right tributaries of the Yenisei River - Kahn, Mana, Tuba, Oja. They cover both agricultural territories and mountain ranges, steppes and forests. The total area of this territory is about 8.9 million hectares, approximately 450 km from north to south and 350 km from west to east [3].

The specificity and difficulty for the meteorological analysis of this territory is that the network of weather stations is very sparse on it. In particular, they are completely absent in its central zone at a distance of more than 200 km from north to south and from west to east. Mountainous terrain is about half of the territory. In addition, the time series of these weather stations have significant omissions, especially in the last 30 years.

In this context, significant interest is evident in the recently emerged new detailed precipitation information based on satellite measurements. IMERG-V06 dataset – NASA Global Precipitation Measurement (GPM) Integrated Multi-satellite Retrievals for GPM (IMERG) is considered one of the best possible alternatives available recently [4, 5]. Precipitation data are presented on a 0.1-degree lat/lon grid, based on microwave and infrared (IR) satellite estimates, precipitation gauge analyses, and other precipitation estimators at fine time and space scales over the entire globe [6]. This is a new dataset, one of the most accurate at present. IMERG version V06 was published in August 2019.

According to data obtained from the NASA GPM IMERG archive, a series of average annual accumulated precipitation is formed for the period from 2001 to 2019 for the territory under consideration.
2. Materials and methods

Data processing procedure consists of several stages:

- **Downloading data.** IMERG 2-byte GeoTIFF files with monthly precipitation data are used for the research presented in this article. These files have the following filenames (yyyy – year, mm – month). For example:
  
  3B-MO-GIS.yyyymm01-S000000-E235959.mm.V06B.zip

  The total size of all files is 5.6 GB.

- **GeoTIFF files with data on accumulated monthly precipitation are unpacked from all received zip archives.** Their filename patterns have the following format (yyyy – year, mm – month). For example:
  
  3B-MO-GIS.MS.MRG.3IMERG. yyyyymm01-S000000-E235959. mm.V06B.total.accum.tif

- **All files are grouped into different directories by year.** All data for each specific year are stored in its own directory. Then they are summed. The total accumulated precipitation is calculated using bat-file scripting tools and GDAL command line raster calculator gdal_calc.py.

- **GDAL gdalwarp utility is used to extract data by area of interest.**

- **Excel is used to compare precipitation data from weather stations received from the archive NOAA GSOY and corresponding CPC_Unified V1.0/RT and IMERGE-V06 grid data cells.**

- **An auxiliary small program is written, which simultaneously computed the statistical indicators of time series of the studied datasets with the accumulation of precipitation for the year.**

3. Results and discussion

An attempt is made to comparatively analyze data from various sources on the average annual rainfall in the study area, containing the catchment areas of four rivers – the right tributaries of the Yenisei River, whose total area is about 8.9 million hectares. Three different data sources are considered:

- **Daily data from weather stations received from the archive NOAA GSOY [7].**

- **CPC Unified V1.0/RT – daily precipitation data presented on a 0.5-degree lat/lon grid, constructed on the basis of weather station data by interpolation with consideration of orographic effects [8, 9].**

- **IMERG-V06 data – NASA Global Precipitation Measurement (GPM) Integrated Multi-satellite Retrievals for GPM (IMERG) – precipitation data presented on a 0.1-degree lat/lon grid, based on microwave and infrared (IR) satellite estimates, precipitation gauge analyses, and other precipitation estimators at fine time and space scales over the entire globe [6, 10].**

As a first step, a map of the difference in the average annual precipitation over two data sets–IMERG-V06 and CPC Unified V1.0/RT is designed.

These data have been previously rescaled on a regular grid with the same 0.01-degree resolution using spline interpolation. Figure 1 shows average annual precipitation by IMERG-V06 data computed for 2001-2019 period as described above.

Because of CPC Unified V1.0/RT data is available for the time period from 1979 to the present, and IMERG-V06 data is available from 2001 to 2019, the maximum possible range is chosen from 2001 to 2019.

A map of the degree of variability of the amount of precipitation for the considered period of time from 2001 to 2019 is presented in figure 2. As we can see, the territory under consideration is characterized by a rather weakly varying amount of precipitation from year to year. Overwhelming part of it has a coefficient of variation (ratio of Standard deviation to Average precipitation) less than 13%.

Unlike the figure 1, the data in the Figure 2 are given in the original spatial resolution, i.e. without smoothing spline interpolation. The observed "squares" correspond to the cells of the initial 0.1-degree grid, which is approximately 7×11 km. Using these "squares" you can visually assess the degree of detail of the data used.
Figure 1. Average annual precipitation by IMERG-V06 data computed for 2001 – 2019 period.

Figure 2. Variability of the annual amount of precipitation for years 2001 to 2019.
Figure 3 shows the percentage difference between the average precipitation according to CPC_Unified V1.0/RT and IMERGE-V06. It is evident that this difference in most of the considered territory does not exceed 10-20%.

![Map showing precipitation difference](image)

**Figure 3.** Relative error of average annual precipitation in CPC_Unified V1.0/RT data comparing to NASA IMERG-V06 data.

Next, we compare all three considered datasets. Only 6 weather stations from the NOAA GSOY archive fall into this area of interest, but not one of them is provided with complete data for recent years. In the period from 2001 to 2019, only one of the six weather stations has a series of data lasting 13 years. It is SOLYANKA station with observations from 2001 to 2013. All the rest totaled no more than 4 years with data.

The period from 2001 to 2019 is considered as the base, since it is the only period of time when IMERG data is available at the moment.

A comparison of the three data sources considered shows that the results are generally close to each other. Histogram columns in figure 4 contains the values of the grid data CPC_Unified V1.0/RT and IMERGE-V06 for the period 2001-2019, as well as points with the values at the weather station for the period 2001-2013.

From the graph in figure 4 it follows that in those years when a large amount of precipitation is observed, the data for all sources used were high. The data are similar to each other in periods with minor precipitation. Insufficient data on weather stations make it difficult to conduct a detailed comparison of the results.

It should be noted that according to expert estimates of meteorologists, the measurement error on rain gauges can be up to 20 percent or more, of which up to 10% can be due to unevenness and gusts of wind, and up to 10% due to wetting effects on the walls of the rain gauge [11]. In case of solid precipitation (snow) the error may be even greater. A comparative analysis of ground-based measurements and satellite data has been the subject of much work in recent times [4, 5].
Figure 4. Precipitation data from the SOLYANKA weather station (56°10’N, 95°15’E) and its corresponding CPC Unified V1.0/RT and IMERGE-V06 grid data cells.

Table 1 with statistical indicators presents the average values, standard deviation and coefficient of variation. The variation does not exceed 15%, which indicates a good agreement between these data.

Table 1. Table of statistical indicators of the studied data sets for the location point of SOLYANKA weather station.

|               | Average of 13 years | Standard deviation | Coefficient of variation |
|---------------|---------------------|--------------------|--------------------------|
| SOLYANKA weather station | 500                 | 75                 | 15%                      |
| CPC Unified V1.0/RT      | 502                 | 47                 | 9%                       |
| IMERG-V06             | 564                 | 64                 | 12%                      |

Paired correlation coefficients demonstrate a high degree of correlation between the studied data sets (table 2).

Table 2. Correlation between the average precipitation values over 13 years for the location point of SOLYANKA weather station.

| Correlation pair                        | Correlation coefficient |
|-----------------------------------------|-------------------------|
| IMERG-V06 – SOLYANKA weather station    | 0.87                    |
| IMERG-V06 – CPC Unified V1.0/RT         | 0.75                    |
| SOLYANKA weather station – CPC Unified V1.0/RT | 0.69                    |

4. Conclusions

The lack of precipitation data from weather stations and the lack of monitoring networks based on meteorological radars in Central Siberia make us pay attention to satellite information on this topic.

The conducted studies are devoted to assessing the characteristics of the satellite-based NASA IMERG-V06 dataset, which could potentially be the backbone of the planned agro-climatic studies. A comparative analysis with data available on weather stations and reanalysis data shows a high degree of
correspondence between the datasets considered. This allows us to conclude that there are good prospects for using the NASA IMERG-V06 dataset and recommend it for use in further studies.

References

[1] Erunova M, Shpedt A, Trubnikov Y and Yakubailik O 2019 IOP Conference Series: Earth and Environmental Science 315(3) 032022
[2] Shaparev N, Tokarev A and Yakubailik O. 2020 Int. J. of Sustainable Development & World Ecology 27 349-57
[3] Erunova M and Yakubailik O 2020 IOP Conference Series: Earth and Environmental Science 421(6) 062039
[4] Beck H, Pan M, Roy T, Weedon G, Pappenberger F, Dijk A, Huffman G, Adler R and Wood E 2019 Hydrol. Earth Syst. Sci. 23 207-24
[5] Kubota T. et al. 2020 Global Satellite Mapping of Precipitation (GSMaP) Products in the GPM Era ed Levizzani V, Kidd C, Kirschbaum D, Kummerow C, Nakamura K and Turk Satellite Precipitation Measurement. Advances in Global Change Research (Springer, Cham) 67
[6] Huffman G, Bolvin D, Braithwaite D, Hsu K, Joyce R, Kidd C, Nelkin E, Sorooshian S, Tan J and Xie P 2019 NASA Global Precipitation Measurement (GPM) Integrated Multi-satellite Retrievals for GPM (IMERG), Algorithm Theoretical Basis Document (ATBD) (Retrieved from https://pps.gsfc.nasa.gov/Documents/IMERG_ATBD_V06.pdf.)
[7] Lawrimore J, Ray Ron, Applequist S, Korzeniewski B and Menne M 2016 Global Summary of the Year (GSOY), Version 1. (NOAA National Centers for Environmental Information) The Climate Data Guide: CPC Unified Gauge-Based Analysis of Global Daily Precipitation 2018 ed National Center for Atmospheric Research Staff (Retrieved from https://climatedataguide.ucar.edu/climate-data/cpc-unified-gauge-based-analysis-global-daily-precipitation)
[8] Xie P, Yatagai A, Chen M, Hayasaka T, Fukushima Y, Liu C and Yang S 2007 Journal of Hydrometeorology 8 607
[9] Skofronick-Jackson G et al. 2017 Bull. Amer. Meteor. Soc. 98 1679-95
[10] The Guide to Hydrological Practices (World Meteorological Organization) Retrieved from http://www.whycos.org/hwrp/guide/