RESEARCH PAPER

EVALUATION OF TERRESTRIAL WATER STORAGE PRODUCTS FROM REMOTE SENSING, LAND SURFACE MODEL AND REGIONAL HYDROLOGICAL MODEL OVER NORTHERN EUROPEAN RUSSIA

Vadim Yu. Grigorev[1,2*], Inna N. Krylenko[1,2], Alexander I. Medvedev[3,4], Victor M. Stepanenko[1,3,4]
[1]Lomonosov Moscow State University, Faculty of Geography, Leninskie Gory, Moscow, 119991, Russia
[2]IWP RAS, Gubkina str., Moscow, 119234, Russia
[3]Lomonosov Moscow State University, Research Computing Center, Leninskie Gory, Moscow, 119991, Russia
[4]Hydrometeorological Research Center of the Russian Feleration, 123242, Moscow, Russia
***Corresponding author:** vadim308g@mail.ru
Received: May 20th, 2023 / Accepted: November 14th, 2023 / Published: December 31st, 2023
https://DOI-10.24057/2071-9388-2023-2899

ABSTRACT. Water storage is one of the key components of terrestrial water balance, therefore its accurate assessment is necessary for a sufficient description of hydrological processes within river basins. Here we compare terrestrial water storage using calibrated hydrological model ECOMAG forced by gauge observations, uncalibrated INM RAS–MSU land surface model forced by reanalysis and GRACE satellite-based data over Northern Dvina and Pechora River basins. To clearly identify differences between the datasets long-term, seasonal and residual components were derived. Results show a predominance of the seasonal component variability over the region (~64% of the total) by all datasets but INM RAS–MSU shows a substantial percentage of long-term component variability as well (~31%), while GRACE and ECOMAG demonstrate the magnitude around 18%. Moreover, INM RAS–MSU shows lowest magnitude of annual range. ECOMAG and INM RAS–MSU is distinguished by earliest begin of TWS decline in spring, while GRACE demonstrates latest dates. Overall, ECOMAG has shown the lowest magnitude of random error from 9 mm for Northern Dvina basin to 10 mm for Pechora basin, while INM RAS–MSU has shown largest one.

KEYWORDS: TWS, GRACE, LSM, hydrological model, cold climate, three-cornered hat method

CITATION: Grigorev V. Yu., Krylenko I. N., Medvedev A. I., Stepanenko V. M. (2023). Evaluation Of Terrestrial Water Storage Products From Remote Sensing, Land Surface Model And Regional Hydrological Model Over Northern European Russia. Geography, Environment, Sustainability, 4(16), 6-13
https://DOI-10.24057/2071-9388-2023-2899

ACKNOWLEDGEMENTS: The study was funded by the Russian Science Foundation, grant No. 21-47-00008 (terrestrial water storage change analysis), the statistical analysis was carried out under the Development Program of the Interdisciplinary Scientific–Educational School of Moscow State University "Cosmos". Part of data were collected and processed under Russian Science Foundation Project 21-17-00181

Conflict of interests: The authors reported no potential conflict of interest.

INTRODUCTION

Terrestrial water storage (*TWS*), containing surface water, soil moisture, groundwater, canopy, snowpack and glaciers, is one of the key components of the hydrological cycle. Accurate estimation of *TWS* at various temporal and spatial scales is important for global change research and water resources monitoring (Frolova et al. 2021). Many extreme hydrological events such as floods and droughts are accompanied and followed by extreme TWS magnitude (Tapley et al. 2019). Besides, *TWS* change (*TWSC*) can also be employed in the validation and calibration process of climate, land surface and hydrological models (Gupta and Dhanya 2021; Massoud et al. 2022; Scanlon et al. 2019; Wu et al. 2021). Occurrences of arid and humid periods in river catchments lead to fluctuations in *TWS*, whose changes, even over a multi-year period, can be significant compare to other water balance components. Thus, imbalance caused by ignoring storage change accounts for 7% (3%) of precipitation in arid (humid) catchments for a typical 10-year period (Han et al. 2020).

In situ measurements of *TWS* components, such as soil moisture, surface water, and snowpack, provide the most accurate and precise estimate of TWS at a local scale. However, the observations are often both sparse and heterogeneous in space and time. This applies especially to the Arctic region for which data are limited. The observations can be complemented by remote sensing or model output, but tracking of the variations of deep soil moisture and groundwater is still a challenge.

Variations of the Earth's gravity field on a century timescales are mainly driven by air and water mass

redistributions within the Earth's climate system. The Gravity Recovery and Climate Experiment (GRACE) and GRACE-FO satellite missions, since 2002 and 2018 respectively (Landerer et al. 2020), constitutes the variations and provide monthly, decadal and daily values of *TWS*. Hence GRACE retrieved data are independent of land surface properties, meteorological and hydrological observations, it provides an excellent opportunity for model evaluation. The GRACE derived *TWS* (TWS_{GR}) has been recognized to be reliable as the reference data to validate accuracy of models.

Different types of the models exist, such as global land surface models, whose main purpose is to sufficiently establish interconnection between atmosphere and land as the parts of climate system, or regional hydrological models, which have to derive accurate streamflow estimates under different conditions. As mentioned before, *TWS* plays an essential role in both climate system and streamflow generation process, so there are no clear signs of what kind of model should perform better. Besides, models differ in initial data, parametrization of river basin, meteorological forcing, etc. In the present study a regional hydrological model ECOMAG (EM) (Kalugin and Motovilov 2018) and the global land surface model INM RAS–MSU (IM) (Volodin and Lykosov 1998) were used as the main research tools.

So far, climate and land surface models showed larger range of *TWS* annual cycle in cold regions and smaller range in tropical and (semi)arid basins compared to GRACE (Scanlon et al. 2019; Wu et al. 2021). Moreover, the models have tendency to underestimate long-term *TWSC*, both positive and negative (Scanlon et al. 2018). However, in the absence of the ground truth TWS measurements there are great uncertainties for such assessments. On a basin scale (~10^5 km^2) the random errors of monthly TWS_{GR} have been estimated using different methods from 10 to 20 mm (Chen et al. 2021; Ferreira et al. 2016). An assessment of bias in TWS_{GR} over river basins can be performed for some areas and seasons by water balance equation. Thus, during snowmelt streamflow is the main source of *TWS* variation, and it can be measured in situ, while precipitation and evaporation assessment errors are neglectable. However, in general, precipitation and evaporation bias estimates are required to *TWS* estimate.

Current research on Arctic rivers is mainly aimed at investigating possible future changes in the water balance. Less attention is paid to the assessment of the current water balance. Thus, Nasonova et al. (2022) showed that the Northern Dvina basin (NDB) receives about 665 mm of precipitation annually, 295 mm of which leaves the river basin as river runoff. At the same time, the river runoff is likely to increase by 10% during the 21st century. The main component of the *TWS* variability over the Arctic region is seasonal snowpack formation and melting (Scanlon et al. 2019). Long-term *SWE* in late winter varies across the basin from 80-100 mm in the southwest to 200-220 mm in the east and northeast. Basin average *SWE* at the end of March (approximately maximum of *SWE*) in the NDB showed that they ranged from 126 mm to 221 mm according to ground measurements (Popova et al. 2021).

A GRACE-based assessment of the impact of precipitation *(P)*, evaporation *(E)*, and runoff *(R)* on *TWSC* was performed at a global scale earlier (Zhang et al. 2019). It was shown that *P, ET,* and *R* at the global scale explain 42.6%, 43.2%, and 4.2% of *TWSC*, respectively. However, over the Arctic region, *ET* and *R* are the key components explaining *TWSC*. Thus, in the NDB *ET* explains over 80% of *TWSC*. In the Pechora basin (PB) *ET* and *R* each explains 40-60% of *TWSC*. Comparison of seasonal *TWS* variation to global hydrological models and LSM was evaluated by Scanlon et al. 2019. A GRACE-derived estimate of seasonal amplitude in the NDB showed 191 mm with uncertainty around 15 mm. Similar results were obtained in the PB – 190±20 mm. On average models overestimated the seasonal amplitude of *TWS* by 22 and 27% in the NDB and PB, accordingly. However, there are significant variations in the ratio depending on the model and forcing. Thus, the ratio ranges between 1.17–1.26 in the NDB and between 1.12–1.49 in the PB.

The main objectives of this study were to: (1) evaluate GRACE, ECOMAG and INM RAS–MSU retrieved TWS variation over the Northern Dvina River basin (NDB) and the Pechora River basin (PB); (2) estimate systematic and random errors in the series. The overarching goal of this study is to provide an assessment that will allow more sophisticated validation of land surface and hydrological models against GRACE retrieved observations.

MATERIALS AND METHODS

The study area

The entire NDB and most part of the PB are covered with taiga. Northern part of the PB is characterised by tundra. Human impact on TWS variation is neglectable – there are no water reservoirs or significant water demand. Although flat terrain prevails in both catchments, a narrow stretch on the eastern edge of the PB belongs to the western slopes of the Nether-Polar Ural and Northern Ural Mountains. Mean annual air temperature in the NDB and PB is 1.2°C and -3.3°C accordingly. Magnitude of the annual precipitation is around 550 mm for the both basins. The northeastern part of the PB is occupied by permafrost. Flood period, which accounts for 2/3 of the runoff, in the NDB typically occurs in April to June, while in the PB occurs from May to July. During the summer-autumn period, 21% of the annual runoff is formed in the NDB, and 20% in the PB (Georgiadi and Groisman 2022).

Materials

We used the state of the art GRACE ITSG-Grace2018 daily dataset (Kvas et al. 2019). As in the standard processing of monthly GRACE gravity field models, not *TWS* mass variations are removed by subtracting the output of geophysical background models. The limited satellite ground track coverage during one day (15 tracks per day) does not allow for a stable global gravity field inversion so that additional information has to be employed. The time series is therefore processed by a Kalman smoother approach and auto-regressive model of order 3 which apply information about spatio-temporal variations of gravity field at sub-monthly time scale (Eicker et al. 2020; Kvas et al. 2019). Spatial resolution of the data is approximately 500 km, though it is provided as 0.5° grid.

ECOMAG is an integrated hydrological and channel routing model (Motovilov et al. 1998; Kalugin and Motovilov 2018). It calculates the characteristics of snow cover; soil moistening, freezing, and thawing; surface, subsurface, and groundwater flow; water motion through channel network with one-day time step and with a spatial resolution of the size of elementary watersheds. USSR soil and landscape maps as well as regional reference books of agrohydrological soil properties were used to estimate soil properties. Daily air temperature, precipitation and vapor pressure deficit series at 200 weather stations were used to force the model. Model parameters were calibrated and verified by daily water discharge series at 12 and 7

gauges within the NDB and PB accordingly. Depending on available observation data, NDB model was calibrated for 1994–2003 and verified for 2004–2013, while 1984–1993 and 1994–2003 were chosen for calibration and verification of PB model. ECOMAG showed good results in reproducing monthly mean runoff in both the Northern Dvina (Nash–Sutcliffe coefficient is 0.88) and Pechora (Nash–Sutcliffe coefficient is 0.76) basins.

INM RAS–MSU is a grid-based model with spatial resolution of 0.5° and hourly time stepping (Volodin and Lykosov 1998; Machul'skaya and Lykosov 2009). The model takes into account snow cover accumulation and melting, evaporation of intercepted water; surface, subsurface, and groundwater flow; freezing, and thawing of the soil; water uptake and transpiration by vegetation. The soil is discretized into 23 layers, resulting in a total depth of 10 m. The depth of the first soil layer is 1 cm and the last one is 5 m. GLCC was used to specify land cover properties (Loveland et al. 2000). Soil texture is derived from (Wilson and Henderson-Sellers 1985). Hourly ERA5 datasets were used to force the model (Hersbach et al. 2020). Computation was performed starting with 01.01.2000 and full saturated soil. No calibration was applied to the model.

Methods

Decomposition of TWS series

In order to separate a systematic and random components from TWS_{GR}, TWS_{EM} and TWS_{IM} signals we separated them into a long-term (TWS^{long}), seasonal (TWS^{seas}) and subseasonal/residual components (TWS^{res}). The decomposition was accomplished by STL method (seasonal-trend decomposition using locally estimated scatterplot smoothing). The key idea of the method is to consistently apply seasonal and long-term smoothing filters that are based on local regression (Cleveland et al. 1990). Since the method requires the length of seasonal cycle as constant, we have dismissed February 29 from further computations. The parameters of the decomposition were adopted from (Humphrey et al. 2016) as a polynomial of degree 2 was chosen to filter seasonal cycle and linear function was employed to obtain a long-term component. The width of smoothing window of local regression for seasonal cycle separation was set to 45 days. The parameter of long-term component separation was set to 548 days. The number of iterations in an inner loop have been 5 and we assumed that 3 iterations in an outer loop would be sufficient. We preferred STL over simple seasonal averaging because STL allows to reduce weight of outliers in series so they have limited effect on seasonal and long-term components. Thus, the weights of TWS^{res} values after the first inner loop decrease from 1 to 0 as their absolute values increase from the median to 6 medians.

The relative variation of each component can be calculated as the ratio of mean absolute deviation (MAD) of a single component to the sum of the MAD. This metric was proposed in (Kim et al. 2009) and is called the component contribution ratio (CCR). As an example, for TWS^{long}

$$CCR^{long}, \% = \frac{MAD\,(TWS^{long}) \times 100}{MAD\,(TWS^{long}) + MAD\,(TWS^{seas}) + MAD\,(TWS^{long})}$$

Seasonal and long-term components of *TWS* and *TWSC* were estimated based on their deviation from the ensemble mean.

Three-cornered hat method

The residual component of *TWS* contains a short periodic signal and random noise. Assuming that all three datasets include a short-period component we applied the generalized three-cornered hat method (TCH) to estimate the magnitude of short periodic signal and noise. The generalized three-cornered hat method (TCH) is similar to the classical TCH or triple collocation (TC) approach. TC as well as TCH allows to estimate random error in time series if there are at least three time series with common signal but different noises. Consider TWS series to be $\{TWS_i: i \in (1, 2, 3)\}$ and split each time series as $TWS_i = TWS_{true} + \varepsilon_i$ where TWS_{true} denotes a true value of *TWS* and ε_i is a random error. Let's also denote

$$TWS_i - TWS_j = \varepsilon_i - \varepsilon_j = \varepsilon_{ij}\, i,j \in (1,2,3) \;\; and \;\; i \neq j$$

with standard deviation of ε_{ij} and ε_i as σ_{ij} and σ_i accordingly.

Then if all ε_i are independent, TC is follows as

$$\sigma_i^2 = 0.5\left(\sigma_{ik}^2 + \sigma_{ij}^2 - \sigma_{jk}^2\right) i,j,k \in (1,2,3) \; and \; i \neq j \neq k$$

The system includes 3 unknown variables and 3 algebraic equations. However, if ε_i are dependent, so as $cov(\varepsilon_i, \varepsilon_j) \neq 0$, there are 6 unknown variables and 3 algebraic equations. TCH, unlike TC, under some assumptions can be used to solve the underdetermined problem. TCH as proposed in (Premoli and Tavella, 1993) used three key assumptions/requirements: ε_i is normally distributed $\varepsilon_i \sim N(0, \sigma_i^2)$ and the covariance matrix of ε_i is positive definite. The third assumption is the minimum of the sum

$$\sum cov\left(\varepsilon_{ij}^2\right) i,j,k \in (1,2,3) \; and \; i \neq j \neq k$$

TCH was previously used to estimate accuracy of several monthly TWS_{GR} and *ET* products (Ferreira et al. 2016; Xu et al. 2019).

RESULTS

Over the NDB TWS_{GR}^{long} and TWS_{EM}^{long} have shown good agreement with Pearson correlation coefficient *(r)* equals to 0.82, while *r* between TWS_{IM}^{long} and TWS_{GR}^{long} and TWS_{EM}^{long} have been 0.77 and 0.51 accordingly (Fig. 1). The PB has similar magnitudes of *r*, from 0.42 (TWS_{GR}^{long}-TWS_{IM}^{long}) to 0.69 (TWS_{EM}^{long}-TWS_{IM}^{long}).

There are no clear signs of decrease or increase *TWS* over the basins in 2003–2014. Negative anomalies of *TWS* took place over both basins in 2005–2006 and in 2011 over the PB as well. Each dataset showed pronounce seasonal cycle of *TWS* with one phase of growth and one phase of decline (Fig. 2).

Generally, the maximum of *TWS* in NDB occurs at the end of March, almost a month early than over the PB. The dates of *TWS* minimum in both basins lay in the first half of August (Tab. 1).

There are a few patterns in the dates of seasonal peaks obtained from different datasets. Firstly, GRACE extremums occur later than the others. The tendency is most pronounced for the maximums over NDB and minimums over PB. Despite up to one month lag between the maximums around the basins, minimums occur almost at the same time. Over the NDB, ECOMAG and GRACE have showed largest difference between the maximum dates reaching 20 days, whereas the minimums showed 7 days range. Over the PB, GRACE showed minimum of TWS 15 days later than ECOMAG and IM did. INM RAS - MSU demonstrate earliest begin of *TWS* spring decline over the PB, more than three weeks earlier than the other datasets.

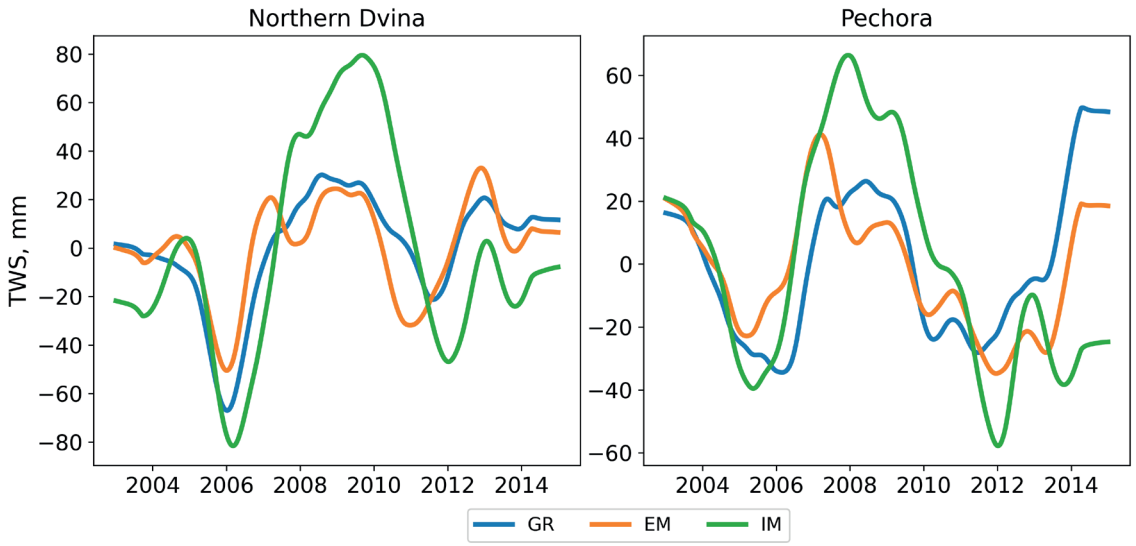

Fig. 1. Change of *TWS*^{long} over Northern Dvina and Pechora basins, mm.

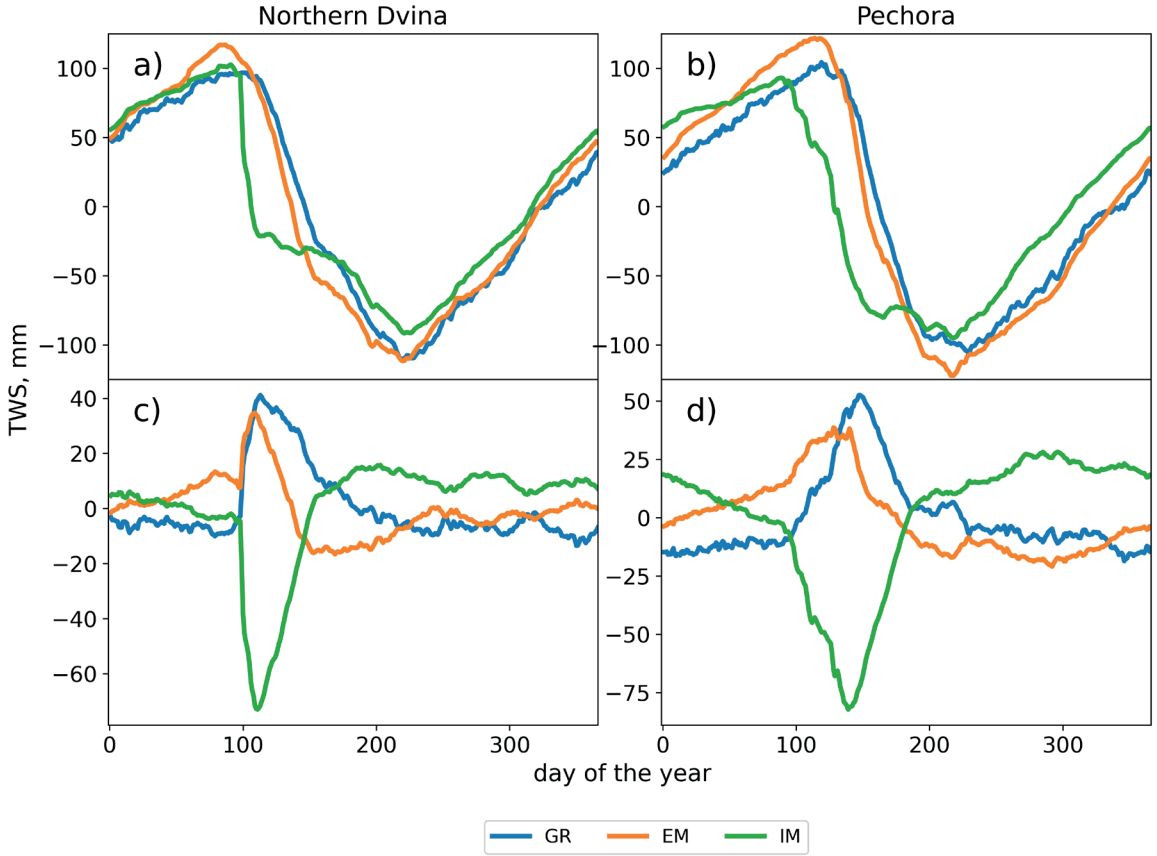

Fig. 2. Mean annual cycle of total water storage (*TWS*^{seas}) over Northern Dvina (a) and Pechora (b) basins and deviations from ensemble mean for Northern Dvian c) and Pechora d), mm.

Table 1. The average dates of maximum/minimum *TWS* and their respective standard deviation (in days) in Northern Dvina and Pechora basins according to ECOMAG, GRACE and INM RAS – MSU

The average date of max/min	ECOMAG	GRACE	INM RAS – MSU
Northern Dvina	24.03 (1.7)/7.08 (3.4)	13.04 (4.3)/ 14.08 (9.3)	28.03 (5.4)/ 11.08 (5.6)
Pechora	24.04 (7.0)/4.08 (1.4)	26.04 (6.6)/ 19.08 (6.5)	3.04 (6.1)/ 4.08 (1.1)

ECOMAG showed the least variability in the date of the extremums occur and GRACE showed the most. Also, *TWS* extremums are rather flattened - for about 7 days around them *TWS* change for less than 4 mm.

In the NDB, in terms of magnitude, TWS_{IM}^{seas} and TWS_{EM}^{seas} differ the most, while GRACE has intermediate position between them. Relative to average over three datasets, ECOMAG overestimates *TWS* from the beginning of a year to mid-April and underestimates from mid-April to the early August. IM shows opposite to ECOMAG signs of bias. From mid-August to late December ECOMAG and IM have shown similar magnitude with difference not more than 15 mm. The minimum of TWS^{seas} is in range from -93 mm (IM) to -113 mm (ECOMAG and GRACE). Maximums of TWS_{GR}^{seas} and TWS_{IM}^{seas} are close (99 mm and 104 mm accordingly) while TWS_{EM}^{seas} maximum is significantly higher – 118 mm. Consequently, the range of TWS_{EM}^{seas} variations exceeds GRACE by 8% and IM by 17%.

The dynamic of TWS_{EM}^{seas} differs most from TWS_{IM}^{seas} and TWS_{GR}^{seas} during period of rapid *TWS* decline from April to May. Thus, maximum of a 10-days average rate of TWS_{EM}^{seas} decline is 4.2 mm/day against 3.2 mm/day and 10.4 mm/day for GRACE and IM accordingly. Every TWS^{seas} series shows that rate of TWS decline starts to decrease from April 24th (TWS_{IM}^{seas}) to June 3rd (TWS_{EM}^{seas} and TWS_{GR}^{seas}), so it is less pronounce for TWS_{GR}^{seas}. From the occurring of *TWS* minimum until the second half of winter, all three datasets show almost identical trajectories.

The differences among the datasets are more distinguish within the PB. Maximum difference was retrieved for pair TWS_{IM}^{seas} - TWS_{GR}^{seas} at the second half of May when it reaches 128 mm. The range of TWS^{seas} variate from 193 mm to 247 mm according to IM and ECOMAG. GRACE has shown magnitude and the range close to average between IM and ECOMAG. Unlike the NDB, TWS_{IM}^{seas} within the PB does not show the highest 10-day *TWS* decline rate (5.1 mm/day). The TWS_{EM}^{seas} and TWS_{IM}^{seas} both shows significantly higher decline rate than in NDB – 6.1 mm/day and 4.7 mm/day, respectively.

Overall, the datasets do not differ much in MAD TWS^{res} magnitude (Fig. 3). The seasonal component is the most pronounced in EM dataset and the least pronounced in IM's. IM has though shown some differences with less pronounced seasonal component and more substantial long-term component.

Within the NDB, GRACE and ECOMAG have shown similar results as CCR^{seas} is around 67–70%, whereas CCR^{long} is 14.5% and CCR^{res} is roughly 17%. GRACE and ECOMAG also showed similar ratio of $MAD\ TWS^{seas}/MAD\ TWS^{res}$ (the ration of seasonal component variation to residual component variation) – 4.6 and 4.8, while the ratio of IM is significantly less - 3.8. The most noticeable difference of INM RAS - MSU from other datasets is the significant TWS^{long} component (34 mm), which MAD is more than twice as large as GRACE and ECOMAG. The datasets' performance over the PB is similar to NDB one, but IM shows a less pronounced TWS^{long} component (29.3 mm) and GRACE more pronounced (20.5 mm).

Overall, the difference, between how dominant particular components are, among the datasets can be seen. GRACE is taking middle position having CCR^{long} more than ECOMAG, but less than IM. However, GRACE has the lowest $MAD\ TWS^{seas}/MAD\ TWS^{res}$ ratio, as GRACE ratio is less IM one by 6% and less than ECOMAG one by 28%.

Over both basins, ECOMAG showed smallest random error (Tab. 2). GRACE and INM RAS – MSU presented similar performance with error (σ_{err}) more than that of ECOMAG by 33-42% and 52-53% in NDB and PB respectively.

Table. 2. Random error of TWS estimation by GRACE, ECOMAG and INM RAS – MSU datasets, in mm, and its ratio to variation of TWSres, %.

ECOMAG also showed smallest σ_{err} / σ_{res} ratio in both basins. Over the PB GRACE and IM have showed similar ratio, while over the NDB GRACE appeared to be worse than IM.

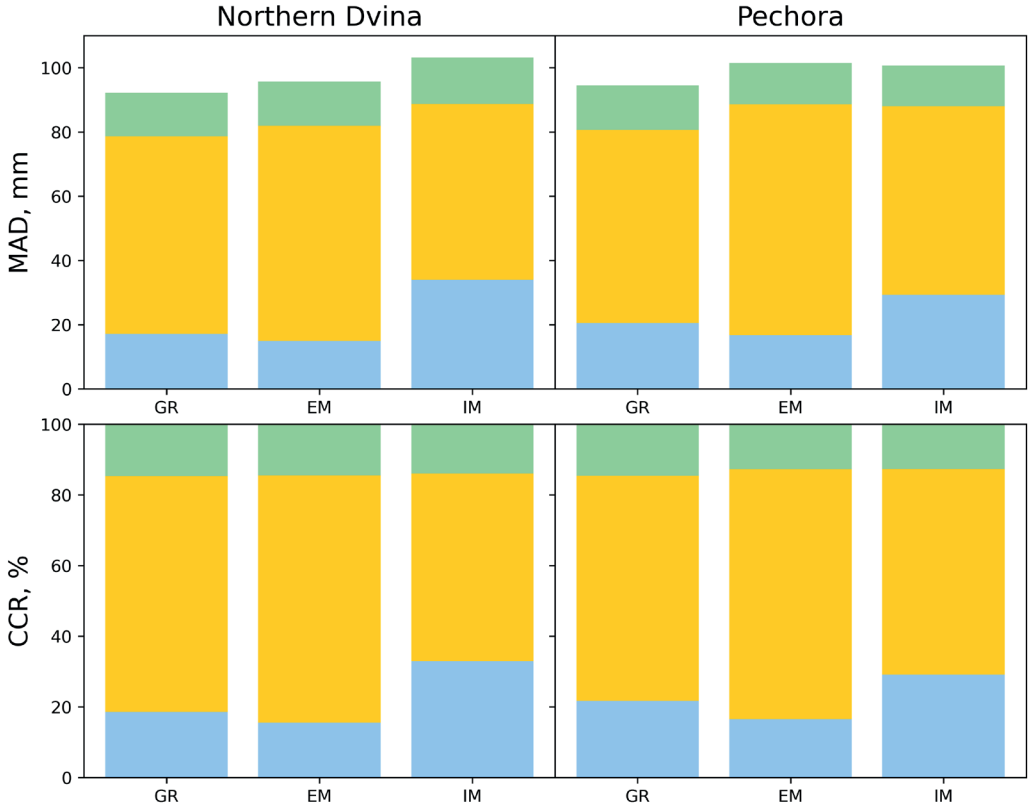

Fig. 3. Contribution of each temporal component to TWS variation. Blue – long-term, orange – seasonal, green – residual

Table 2. Random error σ_{err} of TWS estimation by GRACE, ECOMAG and INM RAS – MSU datasets, in mm, and its ratio to variation of TWS^{res}, %

	Basin	GRACE	ECOMAG	INM RAS - MSU
σ_{err}, mm	NDB	12.8	9.0	13.8
	PB	13.4	10.1	15.4
$\sigma_{err}/\sigma_{res}$, %	NDB	68	48	65
	PB	73	56	81

DISCUSSION

Predominance of seasonal component in *TWS* variation over cold regions based on the monthly series was evaluated worldwide in previous works (Humphrey et al. 2016; B. R. Scanlon et al. 2019). However, *TWS* variations over the Northern Dvina and Pechora basins differ from each other both in magnitude and timing. PB has later date of the maximum and faster rate of *TWS* decline during spring flood. However, despite higher snowfall magnitude, rate of *TWS* accumulation during cold period in the PB is close to NDB rate – less than 80 mm. Occasionally, it can be explained by higher runoff during the period.

In (Scanlon et al. 2019) it was shown that LSM have tendency to overestimate (compared to GRACE) variations of TWS^{seas} over cold regions. Our results are rather opposite. IM showed smallest range of TWS^{seas} in both basins, while hydrological model ECOMAG showed largest range. It's probably not related to the difference in forcing data. Thus, the annual precipitation in the NDB (PB) basin used in ECOMAG was 605 (704) mm, while in the INM RAS – MSU it was 706(719) mm. INM RAS-MSU obtained *TWS* begins to differ from that of ECOMAG in February-March, which is associated with overestimation of river discharge during the winter low flow period in INM RAS-MSU, which is more noticeable in the PB. A much sharper decline in *TWS* compared to other data sets is characteristic of INM RAS-MSU in the NDB. This is due to both overestimation of water discharge during the flood period (maximum discharge is overestimated by 50%) and annual water discharge (by 13%). The sharp decline in the rate of *TWS* decrease in the first decade of June is associated with the beginning of low flow period. INM RAS - MSU shows a short period of flood recession, about 45 days, while ECOMAG shows an average duration of 75 days. One reason for these differences may be the use of the water-holding capacity of snow in ECOMAG, which delays the start of flood season, as well as the use of more detailed vegetation and land surface properties maps over Russia, which caused the snowmelt process in the model to be less synchronous compare to INM RAS - MSU. Overestimation of river discharge during winter may be related to overestimation of the groundwater supply, which may also be related to the high magnitude of MAD TWS^{long}.

Over the PB according to INM RAS – MSU spring flood period starts approximately 16 days early and more smoothly compared to ECOMAG. In spite of this, over the PB the models perform more similar in terms of shape and volume of flood runoff than over the NDB. Therefore, the dynamics of *TWS* in the PB according to the INM RAS - MSU dataset differs from the others mainly by the presence of a shift by 15-20 days.

Magnitudes of MAD TWS^{res} over the basins are similar according to all datasets. That is rather surprising since daily resolution of GRACE series is the result of an approximation that should lead to partially loss of *TWS* signal, that was not eliminated in EM and IM datasets.

Over the NDB the datasets have shown the smaller random error σ_{err} as well as its ratio to σ_{res}. Probably it is related to a large number of weather stations and less complicated features of the basin like absence of permafrost and mountain regions. GRACE also shows higher random error in the PB, which is probably related to a less than the NDB area. In the PB, the random error of GRACE is larger than that in the NDB by only 0.6 mm, while for other datasets it is 1.1 mm and 1.6 mm.

The results obtained here have a number of limitations. Two catchments selected in this study are located in the cold climates. For these catchments, the relative contribution of river runoff and evapotranspiration is approximately equal, which is not the case for most of the Earth. Therefore, the effect of calibrating a hydrological model to discharge data may not have as much effect on the *TWS* in other regions. Also, the ECOMAG calibration was aimed at improving the accuracy of spring flood, which for the area under consideration means also accuracy in reproducing the seasonal *TWS*. Moreover, the Northern Dvina and Pechora basins do not have a significant economic impact such as irrigation and pumping that models cannot always take under consideration. Therefore, the choice of the study area is mostly a favorite of ECOMAG. However, the large size of the basins and the absence of large-scale intense rains are mostly favorable to GRACE, which is not able to catch short-term variations of *TWS*.

CONCLUSION

Results show a predominance of the seasonal component variability over the region (64% of the total) by all datasets but INM RAS–MSU shows a substantial percentage of long-term component variability as well (~31%), while GRACE and ECOMAG demonstrate the magnitude almost twice as low. Hydrological model ECOMAG showed the highest magnitude of the seasonal maximum and minimum, while LSM INM RAS – MSU showed the lowest. However, INM RAS – MSU showed most rapid decline of *TWS* over the NDB during approximately the first half of spring flood period, while GRACE's rate was lowest. ECOMAG is distinguished by earliest begin of TWS decline in spring, while GRACE demonstrates latest dates.

ECOMAG has shown the lowest magnitude of random error from 9 mm for Northern Dvina basin to 10 mm for Pechora basin, while INM RAS – MSU has shown the highest magnitude. However, none of the datasets showed a significantly higher than 1 signal-to-noise ratio in the residual *TWS* component. The question remains open to what extent the obtained errors are related to the models themselves and to what extent the accuracy of input precipitation data since short-period changes in *TWS* are related to precipitation.

Over three datasets ECOMAG showed the best performance considering the lowest rate of random error and the most accurate spring flood hydrograph. However, calibration to discharge data may not be so effective in

evaporation-predominant regions without snowmelt floods. The region also lacks significant economic impacts on water resources, which greatly simplifies model development for this area. The large size of the basins and predominance of seasonal component in *TWS* variation are favorable to GRACE.

It should be noted that the advantages of using GRACE data in hydrology are related not so much to the accuracy of this product as to its independence from ground observation data and physiographic features of river basins.

REFERENCES

Chen J., Tapley B., Tamisiea M.E., Save H., Wilson C., Bettadpur S. and Seo K.W. (2021). Error Assessment of GRACE and GRACE Follow-On Mass Change. Journal of Geophysical Research: Solid Earth, 126(9), e2021JB022124, DOI: 10.1029/2021JB022124

Cleveland R.B., Cleveland W.S., McRae J.E. and Terpenning I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess (with Discussion). Journal of Official Statistics, 6, 3–73.

Eicker A., Jensen L., Wöhnke V., Dobslaw H., Kvas A., Mayer-Gürr T. and Dill R. (2020). Daily GRACE satellite data evaluate short-term hydro-meteorological fluxes from global atmospheric reanalyses. Scientific Reports 2020 10, 4504, DOI: 10.1038/s41598-020-61166-0

Ferreira V.G., Montecino H.D.C., Yakubu C.I. and Heck B. (2016). Uncertainties of the Gravity Recovery and Climate Experiment time-variable gravity-field solutions based on three-cornered hat method. Journal of Applied Remote Sensing, 10(1), 015015, DOI: 10.1117/1.JRS.10.015015

Frolova N.L., Grigorev V.Y., Krylenko I.N. and Zakharova E.A. (2021). State-of-the-art potential of the GRACE satellite mission for solving modern hydrological problems. Vestnik of Saint Petersburg University. Earth Sciences, 66(1), 107–122 (in Russian with English summary), DOI: 10.21638/SPBU07.2021.107

Georgiadi A.G. and Groisman P.Y. (2022). Long-term changes of water flow, water temperature and heat flux of two largest arctic rivers of European Russia, Northern Dvina and Pechora. Environmental Research Letters, 17(8), 085002, DOI: 10.1088/1748-9326/AC82C1

Gupta D. and Dhanya C.T. (2021). Quantifying the Effect of GRACE Terrestrial Water Storage Anomaly in the Simulation of Extreme Flows. Journal of Hydrologic Engineering, 26(4), 04021007, DOI: 10.1061/(ASCE)HE.1943-5584.0002072

Han J., Yang Y., Roderick M.L., McVicar T.R., Yang D., Zhang S. and Beck H.E. (2020). Assessing the Steady-State Assumption in Water Balance Calculation Across Global Catchments. Water Resources Research, 56(7), e2020WR027392. DOI: 10.1029/2020WR027392

Hersbach H., Bell B., Berrisford P., Hirahara S., Horányi A., Muñoz-Sabater J., Nicolas J., Peubey C., Radu R., Schepers D., Simmons A., Soci C., Abdalla S., Abellan X., Balsamo G., Bechtold P., Biavati G., Bidlot J., Bonavita M., … Thépaut J. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999–2049, DOI: 10.1002/qj.3803

Humphrey V., Gudmundsson L. and Seneviratne S.I. (2016). Assessing Global Water Storage Variability from GRACE: Trends, Seasonal Cycle, Subseasonal Anomalies and Extremes. Surveys in Geophysics, 37(2), 357–395, DOI: 10.1007/S10712-016-9367-1

Kalugin A.S. and Motovilov Y.G. (2018). Runoff Formation Model for the Amur River Basin. Water Resources, 45(2), 149–159, DOI: 10.1134/S0097807818020082

Kim H., Yeh P.J.F., Oki T. and Kanae S. (2009). Role of rivers in the seasonal variations of terrestrial water storage over global basins. Geophysical Research Letters, 36(17), DOI: 10.1029/2009GL039006

Kvas A., Behzadpour S., Ellmer M., Klinger B., Strasser S., Zehentner N. and Mayer-Gürr T. (2019). ITSG-Grace2018: Overview and Evaluation of a New GRACE-Only Gravity Field Time Series. Journal of Geophysical Research: Solid Earth, 124(8), 9332–9344, DOI: 10.1029/2019JB017415

Landerer F.W., Flechtner F.M., Save H., Webb F.H., Bandikova T., Bertiger W.I., Bettadpur S. V., Byun S.H., Dahle C., Dobslaw H., Fahnestock E., Harvey N., Kang Z., Kruizinga G.L.H., Loomis B.D., McCullough C., Murböck M., Nagel P., Paik M., … Yuan D.N. (2020). Extending the Global Mass Change Data Record: GRACE Follow-On Instrument and Science Data Performance. Geophysical Research Letters, 47(12), e2020GL088306, DOI: 10.1029/2020GL088306

Loveland T.R., Reed B.C., Ohlen D.O., Brown J.F., Zhu Z., Yang L. and Merchant J.W. (2000). Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. International Journal of Remote Sensing, 21(6–7), 1303–1330, DOI: 10.1080/014311600210191

Machul'skaya E.E. and Lykosov V.N. (2009). Mathematical modeling of the atmosphere-cryolitic zone interaction. Izv. Atmos. Ocean. Phys., 45, 687–703, DOI: 10.1134/S0001433809060024

Massoud E.C., Bloom A.A., Longo M., Reager J.T., Levine P.A. and Worden J.R. (2022). Information content of soil hydrology in a west Amazon watershed as informed by GRACE. Hydrol. Earth Syst. Sci, 26, 1407–1423, DOI: 10.5194/hess-26-1407-2022

Motovilov, Yu., Gottschalk, L., Engeland, K. and Belokurov, A. (1998). ECOMAG — regional model of hydrological cycle. Application to the NOPEX region. Department of Geophysics, University of Oslo.

Nasonova O.N., Gusev Y.M. and Kovalev E. (2022). Climate change impact on water balance components in Arctic river basins. Geography, Environment, Sustainability, 15(4),148–157, DOI: 10.24057/2071-9388-2021-144

Popova V.V., Turkov D.V. and Nasonova O.N. (2021). Estimates of recent changes in snow storage in the river Northern Dvina basin from observations and modeling. Ice and Snow, 61(2), 206–221 (in Russian with English summary), DOI: 10.31857/S2076673421020082

Premoli A. and Tavella P. (1993). A Revisited Three-Cornered Hat Method for Estimating Frequency Standard Instability. IEEE Transactions on Instrumentation and Measurement, 42(1), 7–13, DOI: 10.1109/19.206671

Scanlon B.R., Zhang Z., Rateb A., Sun A., Wiese D., Save H., Beaudoing H., Lo M.H., Müller-Schmied H., Döll P., van Beek R., Swenson S., Lawrence D., Croteau M. and Reedy R.C. (2019). Tracking Seasonal Fluctuations in Land Water Storage Using Global Models and GRACE Satellites. Geophysical Research Letters, 46(10), 5254–5264, DOI: 10.1029/2018GL081836

Scanlon Bridget R., Zhang Z., Save H., Sun A.Y., Schmied H.M., Van Beek L.P.H., Wiese D.N., Wada Y., Long D., Reedy R.C., Longuevergne L., Döll P. and Bierkens M.F.P. (2018). Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. Proceedings of the National Academy of Sciences of the United States of America, 115(6), E1080–E1089, DOI: 10.1073/PNAS.1704665115

Tapley B.D., Watkins M.M., Flechtner F., Reigber C., Bettadpur S., Rodell M., Sasgen I., Famiglietti J.S., Landerer F.W., Chambers D.P., Reager J.T., Gardner A.S., Save H., Ivins E.R., Swenson S.C., Boening C., Dahle C., Wiese D.N., Dobslaw H., … Velicogna I. (2019). Contributions of GRACE to understanding climate change. Nature Climate Change, 9(5), 358–369, DOI: 10.1038/s41558-019-0456-2

Volodin E.M. and Lykosov V.N. (1998). Parametrization of heat and moisture transfer in the soil-vegetation system for use in atmospheric general circulation models: 1. Formulation and simulations based on local observational data. Izvestiya, Atmospheric and Oceanic Physics, 37(4), 405–416.

Wilson M.F. and Henderson-Sellers A. (1985). A global archive of land cover and soils data for use in general circulation climate models. Journal of Climatology, 5(2), 119–143, DOI: 10.1002/JOC.3370050202

Wu R.J., Lo M.H. and Scanlon B.R. (2021). The Annual Cycle of Terrestrial Water Storage Anomalies in CMIP6 Models Evaluated against GRACE Data. Journal of Climate, 34(20), 8205–8217, DOI: 10.1175/JCLI-D-21-0021.1

Xu T., Guo Z.X., Xia Y.L., Ferreira V.G., Liu S.M., Wang K.C., Yao Y., Zhang X. and Zhao C. (2019). Evaluation of twelve evapotranspiration products from machine learning, remote sensing and land surface models over conterminous United States. Journal of Hydrology, 578, 124105. DOI: 10.1016//J.JHYDROL.2019.124105

Zhang Y., He B., Guo L., Liu J. and Xie X. (2019). The relative contributions of precipitation, evapotranspiration, and runoff to terrestrial water storage changes across 168 river basins. Journal of Hydrology, 579, 124194. DOI: 10.1016/J.JHYDROL.2019.124194