ENVIRONMENTAL RESEARCH LETTERS

LETTER

Advancing AI-based pan-European groundwater monitoring

Yueling Ma\textsuperscript{1,2,*}, Carsten Montzka\textsuperscript{1}, Bibi S Naz\textsuperscript{1,2} and Stefan Kollet\textsuperscript{1,2}

\textsuperscript{1} Institute of Bio and Geosciences (Agrosphere, IBG-3), Forschungszentrum Jülich, 52428 Jülich, Germany
\textsuperscript{2} Centre for High-Performance Scientific Computing in Terrestrial Systems, Geoverbund ABC/J, 52428 Jülich, Germany
* Author to whom any correspondence should be addressed.
E-mail: y.ma@fz-juelich.de

Keywords: groundwater, water table depth anomaly, artificial intelligence (AI), long short-term memory (LSTM) network, transfer learning, LSTM-TL, Europe

Supplementary material for this article is available online

Abstract

The main challenge of pan-European groundwater (GW) monitoring is the sparsity of collated water table depth (\textit{wtd}) observations. The \textit{wtd} anomaly (\textit{wtd}_a) is a measure of the increased \textit{wtd} due to droughts. Combining long short-term memory (LSTM) networks and transfer learning (TL), we propose an AI-based methodology LSTM-TL to produce reliable \textit{wtd}_a estimates at the European scale in the absence of consistent \textit{wtd} observational data sets. The core idea of LSTM-TL is to transfer the modeled relationship between \textit{wtd}_a and input hydrometeorological forcings to the observation-based estimation, in order to provide reliable \textit{wtd}_a estimates for regions with no or sparse \textit{wtd} observations. With substantially reduced computational cost compared to physically-based numerical models, LSTM-TL obtained \textit{wtd}_a estimates in good agreement with \textit{in-situ} \textit{wtd}_a measurements from 2569 European GW monitoring wells, showing \( r \geq 0.5 \), root-mean-square error \( \leq 1.0 \) and Kling-Gupta efficiency \( \geq 0.3 \) at about or more than half of the pixels. Based on the reconstructed long-term European monthly \textit{wtd}_a data from the early 1980s to the near present, we provide the first estimate of seasonal \textit{wtd}_a trends in different European regions, that is, significant drying trends in central and eastern Europe, which facilitates the understanding of historical GW dynamics in Europe. The success of LSTM-TL in estimating \textit{wtd}_a also highlights the advantage of combining AI techniques with knowledge contained in physically-based numerical models in hydrological studies.

1. Introduction

Groundwater (GW) is the dominant source of drinking water in Europe, with about 75% of European Union residents relying on GW for their water supply, and recognized as an important contributor to industry (e.g. cooling water) and agriculture (irrigation) (European Commission 2008). In addition, GW plays a critical role in sustaining surface water bodies, adapting to climate variability and supporting biodiversity (van der Gun 2020). In recent years, however, Europe has experienced several severe summer droughts and heat waves (e.g. Fink et al 2004, Stahl et al 2016, Van Lanen et al 2016, Bastos et al 2020, Boergens et al 2020), affecting all compartments of the water cycle. Droughts often affect large areas across national boundaries (Brauns et al 2020). High ambient atmospheric temperature, continuous precipitation deficits, and large evapotranspiration losses translate into delayed and attenuated soil moisture (\( \theta \)) deficits, which subsequently reduce GW recharge and increase water table depths (\textit{wtd}) and ultimately, cause low streamflow or dried-up rivers (Tallaksen et al 2009, Van Loon 2015, Hellwig et al 2020). More frequent and severe droughts anticipated in future changing climate will exacerbate the vulnerability of European GW systems, emphasizing the necessity of GW monitoring in European GW management (Guerrero et al 2018, Ault 2020).

The major challenge for pan-European GW monitoring is the sparsity of \textit{wtd} observations, which is attributed to the large spatial gaps and temporal inconsistency in the \textit{wtd} measurement network (Brauns et al 2020). Sparse \textit{wtd} measurements pose
difficulties in understanding GW dynamics at the continental scale and limit insight into extreme events (e.g. droughts) and their impacts on GW over Europe. To overcome the challenge, standardized meteorological drought indices over extended time scales and terrestrial water storage anomalies derived from the Gravity Recovery and Climate Experiment (GRACE) satellite observations have been widely used to predict and quantify GW anomalies (e.g. Bloomfield and Marchant 2013, Kumar et al 2016, Van Loon et al 2017, Boergens et al 2020). However, their reliability is often questioned, especially in small watersheds, mainly due to the unrealistic assumption about the linear translation of changes in meteorological signals (e.g. precipitation) into GW anomalies or the coarse spatial resolution (0.5°, about 55 km) of the GRACE observations (Kumar et al 2016, Van Loon et al 2017). Physically-based numerical models enable a realistic representation of processes in the terrestrial water and energy cycles. While they may provide more accurate GW anomaly estimates than the above methods, the models require extensive physical background knowledge and become too time and computationally demanding to be applied for high-resolution long-term simulations at larger scales (Wunsch et al 2018, Hauswirth et al 2021). The wtd anomaly (wtda) is a measure of the increased wtd due to droughts. In this study, a methodology is proposed that offers reliable high-resolution long-term estimates of wtda at the European scale with low computational cost, which allows effective and efficient GW monitoring over Europe.

With the emergence of deep learning (DL), increase in computational power, and availability of large datasets, the application of artificial intelligence (AI) has again been attracting considerable attention across scientific disciplines, including hydrological sciences. Rajaee et al (2019) and Tao et al (2022) summarized over 200 papers on the use of AI for GW resource modeling from 2001 to 2020. AI can produce GW predictions as accurately as physically-based numerical models with much less background knowledge and computational cost (Govindaraju 2000, Shen 2018). As a subset of AI, DL techniques such as long short-term memory (LSTM) networks have achieved great success in predicting GW changes in previous studies (e.g. Zhang et al 2018, Hauswirth et al 2021, Vu et al 2021). Specifically, LSTM networks are able to detect long-term dependencies between GW anomalies and input hydrometeorological forcings without explicitly defining spatially varying time lags, thereby simplifying the data preprocessing process (Hochreiter and Schmidhuber 1997, Gers et al 2000). In the field of hydrology, such DL techniques are generally trained using a supervised training algorithm with a supplementary teacher signal (here wtda observations, wtdao) to guide the training process, as shown in figure 1(a). As a result, it is challenging to implement DL techniques to estimate wtda without sufficient wtdao used for training.

Transfer learning (TL), which is a machine learning technique (belonging to the broad field of AI) that applies the knowledge gained in a data-rich domain into a related data-scarce domain, can address the issue of sparse wtda observations to train DL techniques (Tan et al 2018). The increased knowledge often significantly improves network performance in the data-scarce domain and/or accelerates computing progress (Goodfellow et al 2017), and thus, TL has become very popular in image classification (e.g. Quattoni et al 2008, Lu et al 2015, Lee et al 2017) and natural language processing tasks (e.g. Lu et al 2015). A recent study (Ma et al 2021) transferring hydrological parameters across continents has also demonstrated the usefulness of TL in improving hydrological predictions. In contrast with sparse in-situ wtdao (the data-scarce domain), anomalies derived from modeling results, such as modeled wtda (wtdam), are spatiotemporally continuous throughout the European continent and can be regarded as a data-rich domain and used for knowledge transfer.

Here, we propose an AI-based methodology that integrates LSTM networks with TL, LSTM-TL, to estimate monthly wtda at the continental scale in order to facilitate pan-European GW monitoring. The methodology (illustrated in figure 1(b)) uses monthly precipitation and θ anomalies (pθ, and θa) as input and relies on the close connection between GW and other compartments in the water cycle to produce wtda estimates. We evaluated the obtained wtda estimates by LSTM-TL (wtda_{LSTM-TL}) with collated in-situ wtdao from over 2500 European GW monitoring wells (figure 2) in order to explore the reliability of LSTM-TL. Using pr, θ, and θa from various observational datasets (pθao, and θao) as input to LSTM-TL, we reconstructed monthly wtda data at the European scale from the early 1980s to the near present and investigated GW changes across Europe in recent drought years. Finally, seasonal wtda trends were derived for different European regions from the reconstructed data. To the best of our knowledge, these trends are shown for the first time.

2. Methods

2.1. LSTM-TL

The study proposes LSTM-TL to estimate wtda at individual 0.11° pixels over Europe in the absence of collated wtd measurements, thereby mitigating the negative impact of sparse measurements on European GW management. The methodology combines LSTM networks and TL. LSTM networks are utilized to capture the time-varying and time-lagged relationship between wtda and input hydrometeorological variables. The use of TL is based on an
established LSTM architecture which was already successfully tested for predictions of \( wtd_{a,m} \) over Europe (see the supplementary section). In contrast to other studies that transfer knowledge between different regions, such as Ma et al. (2021), this study transfers knowledge between modeled and observed data of the same spatial region. As previously shown (Ma et al. 2021a), among a number of hydro-meteorological variables with large-scale remotely sensed observations and reanalysis products, the
combination of $p_a$ and $\theta_a$ played the most important role in estimating $wd_{a,m}$ over Europe based on the test performance of the LSTM networks. Thus, here we select $p_a$ and $\theta_a$ as input. At pixels without modeled $\theta_{a,m}$ ($\theta_{a,m}$) information (where surface water bodies exist), $p_a$ is the only input variable. In contrast to typical feedforward networks, LSTM networks utilize not only single data points, but also sequences of data, in our case time series. This kind of ‘memory’ includes previously useful input (Shen 2018) such as antecedent $p_a$ and $\theta_a$ to estimate $wd_a$, although not explicitly shown in the inputs.

Figure 1(b) illustrates the workflow of LSTM-TL, encompassing two steps. Firstly, the LSTM networks are constructed at individual pixels on published modeling results (Furusho-Percot et al 2019) from an uncalibrated fully coupled atmospheric-land-surface-GW modeling system (i.e. the Terrestrial Systems Modeling Platform, TSMP). The internal LSTM networks are trained on modeling data from January 1996 to December 2012 (204 time steps, about 80% of the total data), validated on modeling data from January 2013 to December 2014 (24 time steps, about 10% of the total data) and tested on modeling data from January 2015 to December 2016 (24 time steps, about 10% of the total data). The reader is referred to Ma et al (2021a, 2021b) or the supplementary section for detailed information on the network training, validation and testing processes. Secondly, without additional training, the LSTM networks trained in the previous step are utilized to estimate $wd_a$ based on $p_{a,o}$ and $\theta_{a,o}$ thereby transferring the modeled input–output relationship ($f_m$) to the observation-based estimation. The implementation of LSTM-TL is based on the assumption that $f_m$ agrees well with the observed input–output relationship ($f_o$). Due to the lack of monthly GW pumping data, the current application of LSTM-TL does not account for the anthropogenic impact on GW.

LSTM-TL has three salient characteristics. First, the algorithm is independent of $wd_{a,o}$ to estimate $wd_a$, which allows its usage in large regions even without $wd$ observations. Second, the methodology can produce longer-term $wd_a$ estimates than modeling data used for training, and the time length of the obtained $wd_a$ estimates depends on input $p_{a,o}$ and $\theta_{a,o}$. This is useful for reconstructing historical $wd_a$ and predicting future $wd_a$ at the continental scale. Third, once the internal LSTM networks are successfully trained, the methodology can be used without additional training, and thus, requires small computational cost to generate new $wd_a$ estimates.

2.2. Datasets

All anomalies here are relative to the 1996–2012 period (i.e. the training period of the LSTM networks in LSTM-TL), to avoid future information from leaking into the LSTM networks during training. The $wd_a$ was computed by equation (1) for each calendar month and pixel individually to ensure the spatial comparability and to account for the seasonality. The calculation of $p_a$ and $\theta_a$ is similar to $wd_a$:  

\[ wd_a = (wd_m - wd_{p,a}) / wd_{sd}, \]

where $wd_m$ is the monthly data of $wd$, $wd_{p,a}$ is the climatological average of $wd_m$ (i.e. averages of $wd_a$ in January, February, …, December), and $wd_{sd}$ is the climatological standard deviation of $wd_m$.

The monthly modeled $p_a$ ($p_{a,m}$) and $\theta_{a,m}$ as well as $wd_{a,m}$ were derived from published modeling results of TSMP (the supplementary section) from 1996 to 2016, with a spatial resolution of 0.11° (Furusho-Percot et al 2019). The $\theta_{a,m}$ was calculated from the modeled surface layer $\theta$ valid for 0–5 cm soil depth. The monthly $p_{a,o}$ were derived from the 0.5° gridded ERA5 bias-corrected ‘Rainfall flux’ and ‘Snowfall flux’ data from 1979 to 2019 (Muñoz Sabater 2021a), the 0.055° gridded COSMOS-REA6 'TOT_PRECIP' data from 1995 to August 2019 (Bollmeyer et al 2015) and the 0.1° gridded ERA5-Land ‘Total precipitation’ data from 1981 to May 2021 (Muñoz Sabater 2021b) at the individual pixel level, and the monthly $\theta_{a,o}$ were obtained from the 0.1° gridded ERA5-Land ‘Volumetric soil water layer 1’ data (i.e. volume of soil water at 0–7 cm below the land surface) from 1981 to May 2021 (Muñoz Sabater 2021b) and the 0.25° gridded GLEAM v3.5a ‘SMsurf’ ($\theta$ at 0–10 cm below the land surface) from 1980 to 2020 (Miralles et al 2011, Martens et al 2017) at the individual pixel level. Detailed information of the observational datasets is given in the supplementary section. Before calculating anomalies, all input observational datasets were re-gridded to 0.11° × 0.11° via the first-order conservative interpolation method (Jones 1999).

The monthly $wd_{a,o}$ were derived from consecutive monthly $wd$ measurements at 2569 GW monitoring wells (figure 2) located in European unconfined aquifers from 1996 to 2016. The $wd$ measurements were obtained either from web services or by request from governmental authorities; detailed information is provided in supplementary table 2. We calculated $wd_{a,o}$ from the $wd$ measurements to remove errors related to the differences in reference surface elevations that were used by different countries. The 0.11° gridded $wd_{a,o}$ data used for evaluation were estimated by averaging $wd_{a,o}$ from all the wells that lie within the same 0.11° pixels. The number of wells within each 0.11° pixel varies from 1 to 49, with a median value of 2 (supplementary figure 1). The pixels are unevenly distributed in eight hydrometeorologically different regions over Europe (named PRUDENCE regions) (Christensen and Christensen 2007), and their numbers and associated well numbers are provided in supplementary table 3. The yearly averaged $wd_{a,o}$ generally ranged from −0.5 to 0.5 at
individual pixels for the 1996–2016 period, as illustrated in figure 2.

3. Results

3.1. LSTM-TL performance

LSTM-TL produced $wtd_{a,LSTM-TL}$ at the European scale with a spatial resolution of 0.11° (about 12.5 km) from January 1996 to December 2016 based on anomalies from modeling results of TSMP and averaged $pr_{a,o}$ and $\theta_{a,o}$ from observational datasets. To assess the performance of LSTM-TL, we evaluated the obtained $wtd_{a,LSTM-TL}$ with a 0.11° gridded $wtd_{a,o}$ data from January 1996 to December 2016. The evaluation metrics used here are the Pearson correlation coefficient ($r$), root-mean-square error (RMSE), Nash–Sutcliffe efficiency (NSE) and Kling-Gupta efficiency (KGE) (the supplementary section). For pixels with $wtd_{a,o}$, LSTM-TL achieved $r \geq 0.5$, RMSE $\leq 1.0$, $NSE \geq 0.3$ ($NSE \geq 0.0$) and KGE $\geq 0.3$ at 42%, 56%, 28% (57%) and 57% of the pixels for the 1996–2016 period, respectively (figures 3(a)–(d)), showing reliable performance in estimating $wtd$ over Europe, especially when considering the generation of $wtd_{a,LSTM-TL}$ without involving $wtd_{a,o}$. Regional differences were observed in the scores (figure 3(e)). LSTM-TL always obtained the best scores in British Isles (BI), with a median $r$ of 0.63, a median RMSE of 0.77, a median $NSE$ of 0.35, and a median KGE of 0.60, and the worst scores in Mediterranean (MD), with a median $r$ of 0.37, a median RMSE of 1.15, a median $NSE$ of $-0.05$, and a median KGE of 0.29. We focused on the medians of the scores to reduce the influence of outliers on the overall performance evaluation. In addition to the limitations caused by the sparse number of observation wells (supplementary table 3), the poor performance in MD might be due to: (a) the presence of intensive GW abstraction for irrigation in MD, the impact of which on GW is neglected in the current implementation of LSTM-TL; (b) the fact that the wells investigated are predominantly located in deeper aquifers where the internal LSTM networks tended to underperform. As shown in figure 3(e), in general, the scores of $wtd_{a,LSTM-TL}$ were improved compared to $wtd_{a,LSTM(0)}$ (outputs of the LSTM networks using modeling results from TSMP as input) and $wtd_{a,m}$, and were deteriorated compared to $wtd_{a,LSTM(0)}$ (outputs of the LSTM networks trained on observations), which is expected. The $wtd_{a,LSTM(0)}$ is the only estimate that involved $wtd_{a,o}$ in its generation. The difference in the agreement of $wtd_{a,LSTM-TL}$ and $wtd_{a,LSTM(0)}$ with $wtd_{a,o}$ decreased for the 2015–2016 period (i.e. the test period when the LSTM networks were applied to previously unobserved data in the training, supplementary figure 2), which lends confidence in applying LSTM-TL in areas with limited or without $wtd$ observations. More findings and discussion on the comparison of different $wtd_a$ estimates are presented in the supplementary section.

3.2. Recent GW drought analysis using reconstructed pan-European long-term water table depth anomaly data

The time length of $wtd_{a,LSTM-TL}$ is determined by input $pr_{a,o}$ and $\theta_{a,o}$, which allows the generation of $wtd_{a,LSTM-TL}$ beyond the time period of $wtd_{a,m}$ used for training (i.e. the 1996–2016 period). This is an advantage of LSTM-TL over using a physics-based model alone. Here we adopted LSTM-TL to reconstruct pan-European monthly $wtd_a$ data RD1-6 (table 1) from the early 1980s to the near present based on various $pr_{a,o}$ and $\theta_{a,o}$ pairs from observational datasets. Different input $pr_{a,o}$ and $\theta_{a,o}$ data result in different $wtd_{a,LSTM-TL}$ data. The internal LSTM networks were only trained once on modeling results and then used for data reconstruction without additional training.

Figure 4 illustrates the spatial distribution of yearly averaged $wtd_{a,LSTM-TL}$ over the European continent from RD1-6 for 2003, 2015, 2018 and 2019, constituting drought years (Fink et al 2004, Van Lanen et al 2016, Bastos et al 2020, Boergens et al 2020). The spatial distribution presents similar patterns of positive and negative $wtd_{a,LSTM-TL}$ over Europe across all reconstructed data for the investigated years. The reconstructed data by LSTM-TL with GLEAM $\theta_{a,o}$ as input (RD2, RD4 and RD6), figures 4(b), (d) and (f) generally showed larger positive values (i.e. drier events) than the ones with ERA5-Land $\theta_{a,o}$ as input (RD1, RD3 and RD5, figures 4(a), (c) and (e). We found positive anomalies (i.e. dry events) in all European regions except Iberian Peninsula (IP) for 2003 and in Mid-Europe (ME), Eastern Europe (EA), parts of France (FR) and IP for 2015, which agree well with the reported regions affected by meteorological droughts (Fink et al 2004, Van Lanen et al 2016). During the 2018–2019 period, the central and northeastern parts of Europe experienced persistent GW droughts. The GW deficits were triggered by extreme spring warming and brightening in May 2018 and further exacerbated during consecutive summer heat waves in 2018 and 2019 (Bastos et al 2020, Boergens et al 2020, Hari et al 2020). A recovery of GW storage took more than one year in ME, EA and FR (figure 5). In these regions, the strongest impact of GW drought (manifested by the maximum $wtd_{a,LSTM-TL}$) was observed in October and November 2018 after summer precipitation deficits, reflecting the lagged response of GW to precipitation. Here we only considered the hydroclimatic effects on GW storage. GW droughts can be intensified by increases in GW withdrawals during drought periods when surface water supplies are low, which might prolong the recovery period of GW storage after droughts.
3.3. Seasonal trends of water table depth anomalies in various European regions

We derived seasonal $\Delta wtd_a$ trends in different PRUDENCE regions by fitting seasonal averaged $\Delta wtd_a^{\text{LSTM-TL}}$ from the reconstructed data RD1-6 into first-order polynomials $y = mx + b$. The $m$ is the change rate of $\Delta wtd_a$ [yr$^{-1}$], where $m > 0$ corresponds to a drying trend and $m < 0$ corresponds to a wetter condition. The Wald test (Hauck and Donner 1977) was utilized to determine the statistical significance of the trends, with 95% as the threshold. For verification, we compared seasonal averaged $\Delta wtd_a$ and
Table 1. Sources of input $p_{ra,o}$ and $\theta_{a,o}$ data utilized to reconstruct $wd_a$ data RD1-6 as well as time periods of RD1-6.

| Reconstructions | Data source of input $p_{ra,o}$ | Data source of input $\theta_{a,o}$ | Time period              |
|-----------------|----------------------------------|-------------------------------------|--------------------------|
| RD1             | ERA5 bias-corrected              | ER5-Land                            | January 1981–December 2019 |
| RD2             | ERA5 bias-corrected              | GLEAM                               | January 1980–December 2019 |
| RD3             | COSMO-REA6                       | ER5-Land                            | January 1995–August 2019  |
| RD4             | COSMO-REA6                       | GLEAM                               | January 1995–August 2019  |
| RD5             | ER5-Land                         | ER5-Land                            | January 1981–May 2021     |
| RD6             | ER5-Land                         | GLEAM                               | January 1981–December 2020 |

Figure 4. Spatial distribution of yearly averaged $wd_a$, LSTM-TL over the European continent for the recent drought years 2003, 2015, 2018 and 2019. The anomalies were calculated from (a) RD1; (b) RD2; (c) RD3; (d) RD4; (e) RD5; and (f) RD6. The $wd_a$, LSTM-TL $> 0$ corresponds to a dry event while $wd_a$, LSTM-TL $< 0$ corresponds to a wet event.

their trends in ME (i.e. the PRUDENCE regions with the most pixels with $wd_{a,o}$) obtained from RD1-6 and $wd_{a,o}$ for the 1996–2016 period (figure 6). The seasonal averaged $wd_{a,o}$,LSTM-TL from RD1-6 showed a good match with $wd_{a,o}$ in ME except overestimated values in the Summer-Winter 2002 (a record flood over Europe, Ulbrich et al. 2003). Both RD1-6 and $wd_{a,o}$ revealed that there was no significant $wd_{a}$ trend in ME for the 1996–2016 period, with $r$-squared values ranging from 0 to 0.13.

Figure 7 maps seasonal $wd_a$ trends at the European scale in RD1-6 for the time periods listed in table 1, with pixels without significant trends masked in gray. Inspecting the maps, ME, EA and FR tended to experience intensified GW droughts while parts of MD tended towards a wetter GW pattern during the study periods. This finding is consistent with the trend pattern observed in the GRACE terrestrial water storage data for the 2002–2017 period, that is, a distinct pattern of mid-latitude drying and of high- and low-latitude wetting (Reager et al. 2016, Tapley et al. 2019). However, the statistical significance of the GRACE-observed trend pattern was not assessed. The seasonal $wd_a$ trends are strongly influenced by the LSTM-TL input $p_{ra,o}$ and $\theta_{a,o}$. The reconstructed data by LSTM-TL using GLEAM $\theta_{a,o}$ as input (figures 7(b), (d) and (f)) exhibited more pronounced and coherent $wd_a$ trends in four seasons than the ones using ERA5-Land $\theta_{a,o}$ as input (figures 7(a), (c) and (e)). The latter showed weaker $wd_a$ trends in spring and winter. In general, the seasonal $wd_a$ trends in the reconstructed data were in good agreement with the trends in the LSTM-TL input $p_{ra,o}$ and $\theta_{a,o}$ data used to produce the data (supplementary figures 4–9).

4. Discussion

Efficient GW monitoring is essential for European GW management, especially under increasing pressure with changing climate. Yet, it is challenging to monitor GW at large scales, mainly due to the sparsity of $wd$ observations. To address this challenge, we
propose the LSTM-TL methodology, which combines LSTM networks and TL, to produce spatiotemporally continuous \( w_{tda} \) estimates at the European scale in the absence of \( w_{td} \) observations. Thus, LSTM-TL is useful for regions with no or sparse \( w_{td} \) observations but spatiotemporal continuous modeling results, which can be generated over any region worldwide, in principle. Note, this work does not conclude that the role of physically-based models in the hydrometeorological sciences can be replaced with AI technologies. On the contrary, in the absence of \textit{in-situ} measurements, physically-based models are required to construct \( f_m \) in the application of LSTM-TL.

The core rational of the methodology is to transfer the model-based relationship (\( f_m \)) between \( w_{tda,m} \) and input hydrometeorological variables (here \( pr_{a,m} \) and \( \theta_{a,m} \)) to the observation-based estimation to predict \( w_{tda,LSTM-TL} \) from \( pr_{a,o} \) and \( \theta_{a,o} \) (e.g. from satellites or reanalysis data sets) also for regions without \( w_{td} \) observations. The approach is based on the major assumption that the model-based relationship \( f_m \) agrees with the observed one (\( f_o \)). This assumption is indeed valid, since the agreement of \( w_{tda,LSTM-TL} \) with \( w_{tda,o} \) was improved by input correction compared to the agreement of \( w_{tda,m} \) with \( w_{tda,o} \). This comparison constitutes a strong and independent test of the underlying assumption.

Compared with physically-based numerical models, LSTM-TL requires orders of magnitude less computational resources in generating \( w_{tda} \) estimates (i.e. \( w_{tda,LSTM-TL} \)) at the same temporal and spatial scales, thereby facilitating early warning and fast decision making for extreme GW events. The reconstructed European monthly \( w_{tda} \) data successfully reproduced GW droughts in recent drought years. The slightly different drought distributions derived from different reconstructed data reveal the impact of input \( pr_{a,o} \) and \( \theta_{a,o} \) on the obtained \( w_{tda,LSTM-TL} \). Based on the European monthly \( w_{tda} \) data reconstructed by LSTM-TL, we provided the first estimates of seasonal \( w_{tda} \) trends in different European regions, which show that central and Eastern Europe had significant drying trends during the study periods. This is consistent with the trend of the GRACE terrestrial water storage data presented in Reager \textit{et al} (2016) and Tapley \textit{et al} (2019).

In addition, in the application, limitations of the derived LSTM-TL should be recognized. First, in this study, LSTM-TL neglects anthropogenic
Figure 6. Seasonally averaged $wtd_a$ and their trends in ME from RD1-6 and $wtd_{a,o}$ for the 1996–2016 period (MAM: March, April and May, i.e. spring; JJA: June, July and August, i.e. summer; SON: September, October and November, i.e. autumn; DJF: December, January and February, i.e. winter). The bold lines show the seasonal trends of $wtd_{a,LSTM}$ in RD1-6 while the bold dashed lines present the seasonal trends of $wtd_{a,o}$. The gray trend lines indicate no significant trend. The black dashed lines represent $wtd_a = 0$. A longer period with $wtd_a > 0$ corresponds to a dry event while a longer period with $wtd_a < 0$ corresponds to a wet event.

Figure 7. Seasonal $wtd_a$ trend maps over Europe (MAM: March, April and May, i.e. spring; JJA: June, July and August, i.e. summer; SON: September, October and November, i.e. autumn; DJF: December, January and February, i.e. winter). The seasonal $wtd_a$ trends were derived from (a) RD1; (b) RD2; (c) RD3; (d) RD4; (e) RD5; and (f) RD6. The pixels without significant trends are shown in gray.
interventions, such as GW pumping, on GW dynamics, because the modeling results used for training the internal LSTM networks give a near-natural representation of the water cycle (Furusho-Percot et al 2019). Thus, the methodology may underestimate \( wtd_a \) in regions with strong human interventions such as IP and MD (de Graaf et al 2019). Second, while the utilized modeling results guarantee good spatial coverage of the European domain, time series are limited at the individual pixel level (<300 time steps for training), which may reduce the skill of LSTM-TL in estimating \( wtd_a \). Improved performance of LSTM-TL can be expected as training data increase, which would require extending \( f_m \). Also, the 0.11° (~12.5 km) resolution is not high enough to resolve the topographic impact of individual hillslopes on \( wtd \) (Fan et al 2019). This can be solved by the improved resolution of the modeling results used for training. Third, the behavior of LSTM networks is more difficult to interpret than other AI techniques because of the time-varying weights and biases in the connection of neurons. Here, the input variables were selected based on the network performance i.e. the contribution of individual input variables to the improvement of test performances (Ma et al 2021a). Using more interpretable AI methods (e.g. XGBoost) to facilitate the selection of input variable may provide additional insight into the response of GW to other hydrometeorological variables, as shown in Hadi et al (2019) and Ravindran et al (2021). Last but not least, we found the LSTM networks in LSTM-TL usually had poor test performance in deep aquifers, which might be attributed to the inability to capture very long-term dependences between \( wtd_a \) and input hydrometeorological variables. Some alternative DL techniques that are superior in time series forecasting can be considered, such as transformers (Vaswani et al 2017), which can effectively capture dependences within the time steps used for training.

In summary, our results show that LSTM-TL provided the best spatiotemporally continuous \( wtd_a \) estimates over Europe and may serve as a useful alternative to \( \text{in-situ} \ wtd_a \) until collated \( wtd \) data sets are available. In addition to data reconstruction described in this study, LSTM-TL can be deployed in online GW monitoring, which would be useful in European GW management.

Data availability statement

The TSMP-G2A data set is available online at 10.17616/R31NJMH3 (Furusho-Percot et al 2019). The \( \text{pr}_{a,m} \) and \( wtd_{a,m} \) data are available online at 10.26165/JUELICH-DATA/WPRA1F (Ma et al 2021c). The \( \theta_{a,m} \) data are available online at 10.26165/JUELICH-DATA/AMQ6N1 (Ma et al 2021d).

The data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/10.26165/JUELICH-DATA/ZBLDIR (Ma et al 2021c).

Acknowledgments

This work was supported by the European Commission HORIZON 2020 Program ERA-PLANET/GEOEssential project [Grant Number: 689443]. The authors gratefully acknowledge the computing time granted through JARA-HPC and the VSR commission on the supercomputer JUWELS at Forschungszentrum Jülich.

Author contributions

Y M conceived and designed the experiments with the support of S K; Y M conducted all the experiments and analyzed the results with feedback from C M, B S N and S K; Y M prepared the manuscript with contributions from all co-authors.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Code availability

A Jupyter Notebook showing an example about the implementation of LSTM-TL is available online at 10.26165/JUELICH-DATA/ZBLDIR.

ORCID iDs

Yueling Ma  https://orcid.org/0000-0002-1869-7702
Carsten Montzka  https://orcid.org/0000-0003-0812-8570

References

Ault T R 2020 On the essentials of drought in a changing climate Science 368 256–60
Bastos A et al 2020 Direct and seasonal legacy effects of the 2018 heat wave and drought on European ecosystem productivity Sci. Adv. 6 eaba2724
Bloomfield J P and Marchant B P 2013 Analysis of groundwater drought building on the standardised precipitation index approach Hydrol. Earth Syst. Sci. 17 4769–87
Boergens E, Günther A, Dobslaw H and Dahle C 2020 Quantifying the central European droughts in 2018 and 2019 with GRACE follow-on Geophys. Res. Lett. 47 e2020GL087285
Bollmeye C et al 2015 Towards a high-resolution regional reanalysis for the European CORDEX domain Q. J. R. Meteorol. Soc. 141 1–15
Brauns B et al 2020 The groundwater drought initiative (GDI): analysing and understanding groundwater drought across Europe Proc. Int. Assoc. Hydrol. Sci. 383 297–305
Christensen J H and Christensen O B 2007 A summary of the PRUDENCE model projections of changes in European climate by the end of this century Clim. Change 81 7–30

de Graaf J F E M, Gleseson T, Rens van Beek L P H, Sutanudjaja E H and Bierkens M F P 2019 Environmental flow limits to global groundwater pumping Nature 574 90–94

European Commission 2008 Groundwater protection in Europe: the new groundwater directive—consolidating the EU regulatory framework (Brussels)

Fan Y et al 2019 Hillslope hydrology in global change research and Earth system modeling Water Resour. Res. 55 1377–72

Fink A H, Brücher T, Krüger A, Leckebusch G C, Pinto J G and Ulbrich U 2004 The 2003 European summer heatwaves and drought-synoptic diagnosis and impacts Weather 59 209–16

Furusho-Percot C, Goergen K, Hartick C, Kulkarni K, Keune J and Kollet S 2019 Pan-European groundwater to atmosphere terrestrial systems climatology from a physically consistent simulation Sci. Data 6 1–9

Gers F A, Schmidhuber J and Cummins F 2000 Learning to forget: continual prediction with LSTM Neural Comput. 12 2451–71

Goodfellow I, Bengio Y and Courville A 2017 Transfer Learning and Domain Adaptation Deep Learning (Cambridge, MA: MIT Press) pp 526–31 (available at: www.deeplearningbook.org)

Govindaraju R 2000 Artificial neural networks in hydrology. I: preliminary concepts J. HYDROL. Eng. 5 115–23

Guerreiro S B, Dawson R J, Kilsby C, Lewis E and Ford A 2018 Future heat-waves, droughts and floods in 571 European cities Environ. Res. Lett. 13 034009

Hadi S J, Abba S I, Sammen S S, Salih S Q, Al-Ansari N and Yaseen Z M 2019 Non-linear input variable selection approach integrated with non-tuned data intelligence models for streamflow pattern simulation IEEE Access 7 141533–48

Hari V, Rakovec O, Markonis Y, Hanel M and Kumar R 2020 Increased future occurrences of the exceptional 2018–2019 central European drought under global warming Sci. Rep. 10 12207

Hauck W W and Donner A 1977 Wald’s test as applied to estimate groundwater table depth anomalies across Europe Environ. Res. Lett. 17 58

Hauswirth S M, Bierkens M F P, Beijk V and Wanders N 2021 The PRUDENCE model projections of changes in European climate by the end of this century Clim. Change 81 7–30

e e Graaf J F E M, Gleseson T, Rens van Beek L P H, Sutanudjaja E H and Bierkens M F P 2019 Environmental flow limits to global groundwater pumping Nature 574 90–94

European Commission 2008 Groundwater protection in Europe: the new groundwater directive—consolidating the EU regulatory framework (Brussels)

Fan Y et al 2019 Hillslope hydrology in global change research and Earth system modeling Water Resour. Res. 55 1377–72

Fink A H, Brücher T, Krüger A, Leckebusch G C, Pinto J G and Ulbrich U 2004 The 2003 European summer heatwaves and drought-synoptic diagnosis and impacts Weather 59 209–16

Furusho-Percot C, Goergen K, Hartick C, Kulkarni K, Keune J and Kollet S 2019 Pan-European groundwater to atmosphere terrestrial systems climatology from a physically consistent simulation Sci. Data 6 1–9

Gers F A, Schmidhuber J and Cummins F 2000 Learning to forget: continual prediction with LSTMM Neural Comput. 12 2451–71

Goodfellow I, Bengio Y and Courville A 2017 Transfer Learning and Domain Adaptation Deep Learning (Cambridge, MA: MIT Press) pp 526–31 (available at: www.deeplearningbook.org)

Govindaraju R 2000 Artificial neural networks in hydrology. I: preliminary concepts J. HYDROL. Eng. 5 115–23

Guerreiro S B, Dawson R J, Kilsby C, Lewis E and Ford A 2018 Future heat-waves, droughts and floods in 571 European cities Environ. Res. Lett. 13 034009

Hadi S J, Abba S I, Sammen S S, Salih S Q, Al-Ansari N and Yaseen Z M 2019 Non-linear input variable selection approach integrated with non-tuned data intelligence models for streamflow pattern simulation IEEE Access 7 141533–48

Hari V, Rakovec O, Markonis Y, Hanel M and Kumar R 2020 Increased future occurrences of the exceptional 2018–2019 central European drought under global warming Sci. Rep. 10 12207

Hauck W W and Donner A 1977 Wald’s test as applied to estimate groundwater table depth anomalies across Europe Environ. Res. Lett. 17 58

Hauswirth S M, Bierkens M F P, Beijk V and Wanders N 2021 The PRUDENCE model projections of changes in European climate by the end of this century Clim. Change 81 7–30
van der Gun J 2020 Groundwater resources sustainability Global Groundwater ed A Mukherjee, B R Scanlon, A Aureli, S Langan, H Guo and A A McKenzie (Amsterdam: Elsevier) pp 331–45
Van Lanen H A J et al 2016 Hydrology needed to manage droughts: the 2015 European case Hydrol. Process. 30 3097–104
Van Loon A F 2015 Hydrological drought explained WIREs Water 2 359–92
Van Loon A F, Kumar R and Mishra V 2017 Testing the use of standardised indices and GRACE satellite data to estimate the European 2015 groundwater drought in near-real time Hydrol. Earth Syst. Sci. 21 1947–71
Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I 2017 Attention is all you need (arxiv:1706.03762)
Vu M T, Jardani A, Massei N and Fournier M 2021 Reconstruction of missing groundwater level data by using long short-term memory (LSTM) deep neural network J. Hydrol. 597 125776
Wunsch A, Liesch T and Broda S 2018 Forecasting groundwater levels using nonlinear autoregressive networks with exogenous input (NARX) J. Hydrol. 567 743–58
Zhang J, Zhu Y, Zhang X, Ye M and Yang J 2018 Developing a long short-term memory (LSTM) based model for predicting water table depth in agricultural areas J. Hydrol. 561 918–29