

# Freely available mean daily discharge series from Czechia: what can be inferred from them?

Ondrej Ledvinka<sup>1,\*</sup>

<sup>1</sup>Czech Hydrometeorological Institute, Hydrology Database and Water Budget Department, Na Sabatce 2050/17, 143 06 Prague 412, Czechia

**Abstract.** Most hydrometeorological data from Czechia are still provided for a fee. This especially applies to time series with a finer step than monthly. The fact that the data must be paid with respect to an expert's appraisal and that the proper licence agreement ratification has to be performed causes a considerable delay in the data transfer to potential customers, unfortunately including scientists as well. Naturally, this time-consuming process is unpleasant to the experts on both sides. Due to a substantial rise of university students' requests for data, the Czech Hydrometeorological Institute's hydrologists launched a website from which long mean daily discharge series representing ten selected water-gauging stations can be downloaded. Besides other assessments, the series may play an important role when studying climate change impacts on water resources in Czechia. Therefore, the objective here was to extract from these series some preliminary information on long-term changes such as abrupt and gradual trends caused either by the construction of reservoirs or by climate variability itself. The main tool used was nonparametric trend analysis. Mainly different months were of interest so as to determine if there have been recorded some changes in seasonality. The results may be easily expanded by students.

## 1 Introduction

Although some basic information can be found on the internet, regime hydrometeorological data from Czechia usually are not free-of-charge. The main provider of these data here is the Czech Hydrometeorological Institute (CHMI) where the process of data transfer to customers is still somewhat weird, perhaps inheriting the practice of older employees who are not accustomed to working with modern technologies. The process namely consists in inappropriate exploiting the capacities of experts working with databases rather than in the supporting staff being activated. It applies also to necessary licence agreements where a requested expert initiates their ratification and, essentially, must be a part of the whole cycle rather than only a decision maker regarding data accessibility and quality. It is more than evident that this situation is not sustainable. On the one hand the number of experts is decreasing and, consequently, each individual is becoming more responsible, on the other hand the number of requests for data substantially rises. Naturally,

---

\* Corresponding author: [ledvinka@chmi.cz](mailto:ledvinka@chmi.cz)

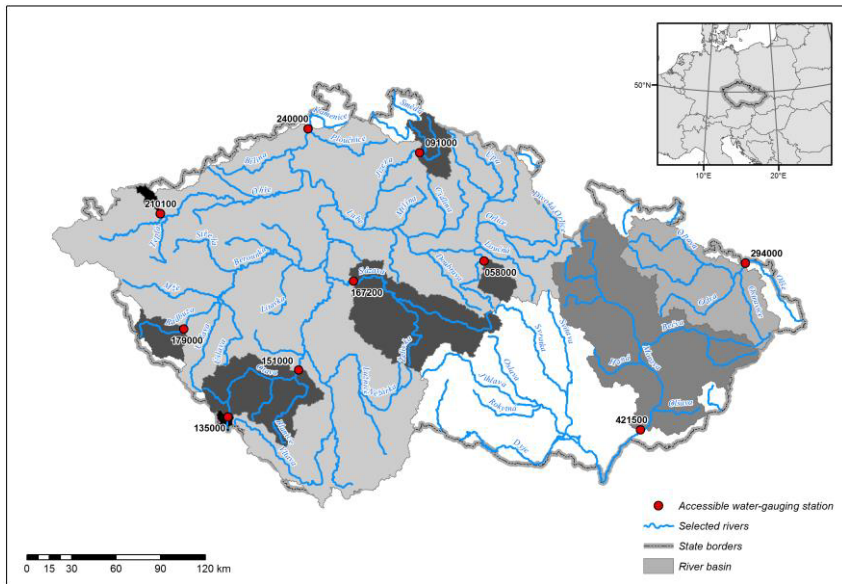
the resulting considerable delays in data transfer adversely influence the scientific needs and capacity building development.

At the turn of 2014 and 2015, the deputy director for hydrology at the CHMI decided that selected hydrological series should be freely available exactly for scientific purposes and students so that the staff of the CHMI could invest their time into addressing other issues. In 2015, some negotiation with branch/regional offices as well as with the deputy director for meteorology and climatology took place so as to determine which water-gauging and climatological stations should be represented in such a set of free data. Finally, thanks to the cooperation with the Czech National Committee for Hydrology (CNCH), Czech and English websites (see <http://cnvh.cz/>) devoted to the presentation of the data set were launched at the beginning of 2016. Although initially dedicated to small experimental basins in the Jizera Mts. in the north of Bohemia, the data set includes also ten long mean daily discharge series representing wider territory of Czechia (see Table 1 and Fig. 1). Such data are particularly valuable when dealing with possible climate change (CC) impacts on water resources, which are intensively studied now.

The aim of this study was to draw the attention of scientific community to the mentioned data set and to initiate the preferable use of it so that the number of requests for the regime hydrological data directed to the CHMI could drop. At the same time, through a nonparametric trend analysis, it is illustrated here what can be inferred from this data set regarding CC. The results here should be considered preliminary, while mainly the students are encouraged to continue such research, either by adopting other, more sophisticated, methodologies or by suggesting other issues that can be addressed using this data set. Note that the data representing the experimental basins are overlooked intentionally here since they will be a subject of another study.

**Table 1.** Closer information on water-gauging stations whose discharge series were studied here (adapted from <http://hydro.chmi.cz/ismnozstvi/>). The order of numbers (from south to north) corresponds to the *y*-axis in Fig. 2.

| Order<br>(from south<br>to north) | ID     | Name         | River      | Catchment<br>area [km <sup>2</sup> ] |
|-----------------------------------|--------|--------------|------------|--------------------------------------|
| 10                                | 240000 | Děčín        | Labe       | 51120.39                             |
| 9                                 | 091000 | Železný Brod | Jizera     | 791.27                               |
| 8                                 | 210100 | Stará Role   | Rolava     | 126.35                               |
| 7                                 | 058000 | Úhřetice     | Novohradka | 458.91                               |
| 6                                 | 294000 | Bohumín      | Odra       | 4663.77                              |
| 5                                 | 167200 | Nespeky      | Sázava     | 4038.65                              |
| 4                                 | 179000 | Staňkov      | Radbuza    | 701.54                               |
| 3                                 | 151000 | Písek        | Otava      | 2913.70                              |
| 2                                 | 135000 | Modrava      | Vydra      | 89.80                                |
| 1                                 | 421500 | Strážnice    | Morava     | 9144.80                              |

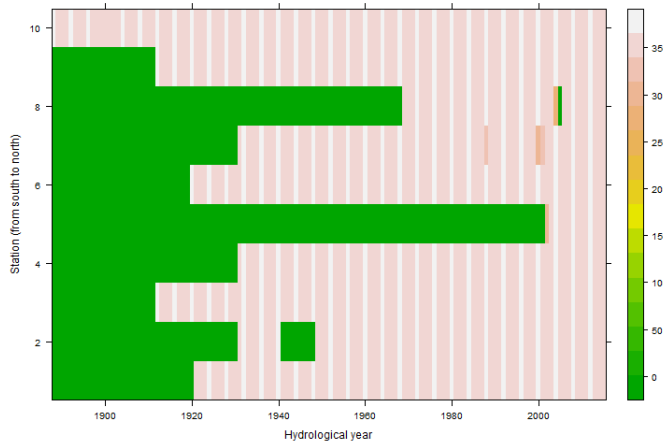


**Fig. 1.** Location of offered water-gauging stations with freely available long mean daily discharge series with respect to the borders of Czechia. All the geographical information, apart from the river network (source: Ministry of the Environment of the Czech Republic; other can be downloaded from <http://www.dibavod.cz/>), can be acquired at <http://hydro.chmi.cz/ismnozstvi/> upon registration.

## 2 Brief exploratory analysis of selected discharge series

In time series analysis (TSA), it is necessary first to know what time periods are covered by the data. It is not sufficient to know when a time series begins and ends, but also one has to be interested in the placement of missing values. Missing values (or gaps) are a common feature of time series and the daily discharge series offered online by the CHMI are not an exception. This fact is actually an advantage since the students can, at the very beginning of their career, face to this natural issue, where several reasons for the occurrence of gaps may be reflected (including political conflicts such as wars or issues connected to a momentary lack of employees at hydrological services needed for digitization of archived reports).

Figure 2 shows the availability of mean daily discharges at selected stations separately for all hydrological years (in Czechia defined as periods from 1 November of last year to 31 October) starting from 1888 (i.e. beginning of the longest series of station 240000) and proceeding to 2015. It may be deduced that the beginnings of the series differ. Irrespective of this fact, it may further be stated that stations 240000, 091000, 294000, 179000, 151000 and 421500 have uninterrupted time series composed of mean daily values. On the contrary, stations 210100, 058000 and 135000 have interrupted series, either somewhere close to the beginning or close to the end of the series. Only few daily values are missing in the series from station 058000. In the case of station 167200, it can be seen that it starts very late, which is caused by the fact that the measurement at this station succeeds the observations at another, closely located, station. It is questionable whether to merge this series with its predecessor. However, it is probable that it will happen in the future and the students will have another valuable longer series for their CC studies.



**Fig. 2.** An overview of the accessibility of mean daily discharges in every individual year from 1888 to 2015 representing ten water-gauging stations offered online by the CHMI. Leap years manifest themselves as the lighter stripes, meaning there are 366 instead of 365 values per year available.

Looking at Fig. 2, one may also notice several common periods that can be studied separately at the exploratory analysis stage. This can be useful, for instance, when investigating the influence of the presence of reservoirs. Thus, the whole period of hydrological years 1888–2015 was divided into four subperiods: entire period, 1931–2015, 1961–2015, and specifically 1981–2010 (i.e. current Czech hydrological reference period). It was not important how many stations measured at the beginning of the respective period. Rather it was aimed at possible changes in basic descriptive statistics. During the computation of the statistics, the missing values were ignored as well. Table 2 gives an overview of such pursuits.

Upstream reservoir operation may manifest itself in decreasing variation as well as in increasing lag-1 autocorrelation. Both may be apparent when assessing the series coming from station 294000 on the Odra River. At this station, also an increase in skewness can be observed, meaning that there have been less (or smaller) flood waves after the construction of reservoirs. On the other hand, rivers where the anthropogenic influence is believed to be very small experience a decrease in skewness (see e.g. station 135000 on the Vydra River). Of course, the descriptive statistics would change their values if monthly or annual series would be of interest. Particularly, apart from the long-term mean, they would decrease.

Furthermore, it is advisable, before performing a trend analysis, to decide whether the time series can be studied as a whole or it should rather be divided into several parts. Dividing the series is recommended especially when the entire series reveals inhomogeneity, which means that different parts of the record were collected under different (unequal) conditions. Therefore, a homogeneity check was carried out using an R tool prepared by climatologists in Canada. This tool is, to the author's knowledge, the only one that works with daily data as well as with missing values. The description of the method would not be trivial, and the readers are referred to [1] instead. Since the tool was designed for precipitation, discharge was transformed to runoff depth first.

Surprisingly, it was found that almost all the series are homogeneous regardless of the hypothesized change points connected to anthropogenic influences. The explanation may lie in the fact that the human impact on the investigated rivers started much earlier than the stations became operational. Anthropogenic influences may also be weaker than those of natural origin at larger rivers such as Labe. Nevertheless, there are two stations in the free data set where the null hypothesis of homogeneity had to be rejected at the level of

significance  $\alpha = 0.05$ . While the series of station 179000 revealed only one change point at the beginning of the hydrological year 1965, the series of station 167200 showed three change points during the 2000s. Since the reasons for these change points are not known without any detailed metadata, the two stations were no longer included in subsequent analyses.

**Table 2.** Basic statistics describing ten mean daily discharge series offered online by the CHMI. Four periods (entire period, 1931–2015, 1961–2015 and 1981–2010) are summarized in every single cell irrespective of missing values or the starting point of measurement. If the values are the same as those for the succeeding period (because there is no new information), they are not in bold.

| ID     | Long-term mean<br>[m <sup>3</sup> ·s <sup>-1</sup> ] | Coefficient of variation | Skewness     | Kurtosis      | Lag-1 autocorrelation |
|--------|--|--------------------------|--------------|---------------|-----------------------|
| 240000 | <b>312.00</b>  | <b>0.87</b>              | <b>3.23</b>  | <b>18.07</b>  | <b>0.968</b>          |
|        | <b>310.00</b>  | <b>0.84</b>              | <b>3.36</b>  | <b>20.17</b>  | <b>0.971</b>          |
|        | <b>313.00</b>  | <b>0.80</b>              | <b>3.51</b>  | <b>24.74</b>  | <b>0.974</b>          |
|        | <i>315.00</i>  | <i>0.81</i>              | <i>3.68</i>  | <i>27.70</i>  | <i>0.972</i>          |
| 091000 | <b>17.00</b>   | <b>1.09</b>              | <b>4.21</b>  | <b>31.87</b>  | <b>0.831</b>          |
|        | <b>16.70</b>   | <b>1.10</b>              | <b>4.18</b>  | <b>31.70</b>  | <b>0.839</b>          |
|        | <b>16.90</b>   | <b>1.10</b>              | <b>4.22</b>  | <b>33.78</b>  | <b>0.841</b>          |
|        | <i>17.20</i>   | <i>1.11</i>              | <i>4.30</i>  | <i>35.03</i>  | <i>0.839</i>          |
| 210100 | <i>2.32</i>  | <i>1.11</i>              | <i>5.08</i>  | <i>50.95</i>  | <i>0.881</i>          |
|        | <i>2.32</i>  | <i>1.11</i>              | <i>5.08</i>  | <i>50.95</i>  | <i>0.881</i>          |
|        | <b>2.32</b>  | <b>1.11</b>              | <b>5.08</b>  | <b>50.95</b>  | <b>0.881</b>          |
|        | <i>2.45</i>  | <i>1.06</i>              | <i>4.49</i>  | <i>37.85</i>  | <i>0.886</i>          |
| 058000 | <i>2.50</i>  | <i>1.47</i>              | <i>4.54</i>  | <i>33.37</i>  | <i>0.885</i>          |
|        | <b>2.50</b>  | <b>1.47</b>              | <b>4.54</b>  | <b>33.37</b>  | <b>0.885</b>          |
|        | <b>2.64</b>  | <b>1.41</b>              | <b>4.68</b>  | <b>36.75</b>  | <b>0.890</b>          |
|        | <i>2.54</i>  | <i>1.52</i>              | <i>5.34</i>  | <i>45.17</i>  | <i>0.888</i>          |
| 294000 | <b>43.10</b>   | <b>1.29</b>              | <b>6.58</b>  | <b>100.27</b> | <b>0.895</b>          |
|        | <b>43.00</b>   | <b>1.28</b>              | <b>6.93</b>  | <b>111.29</b> | <b>0.894</b>          |
|        | <b>43.30</b>   | <b>1.25</b>              | <b>8.08</b>  | <b>155.57</b> | <b>0.903</b>          |
|        | <i>41.60</i>   | <i>1.36</i>              | <i>10.41</i> | <i>221.00</i> | <i>0.904</i>          |
| 167200 | <i>21.30</i>   | <i>1.36</i>              | <i>6.74</i>  | <i>76.52</i>  | <i>0.941</i>          |
|        | <i>21.30</i>   | <i>1.36</i>              | <i>6.74</i>  | <i>76.52</i>  | <i>0.941</i>          |
|        | <b>21.30</b>   | <b>1.36</b>              | <b>6.74</b>  | <b>76.52</b>  | <b>0.941</b>          |
|        | <i>23.00</i>   | <i>1.38</i>              | <i>6.54</i>  | <i>70.65</i>  | <i>0.948</i>          |
| 179000 | <i>3.74</i>  | <i>1.29</i>              | <i>6.97</i>  | <i>94.48</i>  | <i>0.838</i>          |
|        | <b>3.74</b>  | <b>1.29</b>              | <b>6.97</b>  | <b>94.48</b>  | <b>0.838</b>          |
|        | <b>3.75</b>  | <b>1.27</b>              | <b>7.16</b>  | <b>106.35</b> | <b>0.826</b>          |
|        | <i>3.84</i>  | <i>1.32</i>              | <i>8.10</i>  | <i>132.68</i> | <i>0.803</i>          |
| 151000 | <b>23.60</b>   | <b>0.98</b>              | <b>7.16</b>  | <b>158.68</b> | <b>0.864</b>          |
|        | <b>23.40</b>   | <b>0.98</b>              | <b>7.55</b>  | <b>181.10</b> | <b>0.866</b>          |
|        | <b>23.90</b>   | <b>0.99</b>              | <b>8.47</b>  | <b>219.63</b> | <b>0.873</b>          |
|        | <i>24.40</i>   | <i>1.04</i>              | <i>10.26</i> | <i>283.80</i> | <i>0.854</i>          |
| 135000 | <i>3.38</i>  | <i>1.07</i>              | <i>5.01</i>  | <i>47.76</i>  | <i>0.750</i>          |
|        | <b>3.38</b>  | <b>1.07</b>              | <b>5.01</b>  | <b>47.76</b>  | <b>0.750</b>          |
|        | <b>3.35</b>  | <b>1.06</b>              | <b>4.56</b>  | <b>36.91</b>  | <b>0.760</b>          |
|        | <i>3.46</i>  | <i>1.08</i>              | <i>4.77</i>  | <i>40.96</i>  | <i>0.736</i>          |
| 421500 | <b>58.60</b>   | <b>1.09</b>              | <b>3.53</b>  | <b>18.57</b>  | <b>0.936</b>          |
|        | <b>59.00</b>   | <b>1.10</b>              | <b>3.58</b>  | <b>18.90</b>  | <b>0.932</b>          |
|        | <b>59.60</b>   | <b>1.05</b>              | <b>3.72</b>  | <b>21.64</b>  | <b>0.931</b>          |
|        | <i>59.30</i>   | <i>1.11</i>              | <i>4.01</i>  | <i>24.84</i>  | <i>0.938</i>          |

### 3 Trend analysis

Nowadays, investigators interested in trend analysis emphasize that the trend component having the lowest frequency [2] should be properly distinguished not only from short-term persistence (STP), but also from long-term persistence (LTP) effects. Both types of persistence cause most known trend tests to reject the null hypothesis of no trend too often. Therefore, where possible, so-called unit root tests were applied to find out if LTP could stand behind falsely detected trends. The possible effect of STP was considered directly in the trend test afterwards.

#### 3.1 Testing for stationarity

There are several types of stationarity relating to time series. First, what is usually forgotten, one may look at the stationarity in various moments such as mean and variance. Hydrologists sometimes apply tests devoted to heteroscedasticity (i.e. nonstationarity in the variance) and apply models that account for it (e.g. [3]), but, traditionally, they study possible nonstationarity in the mean. Second, one may distinguish between deterministic and stochastic stationarity. While stochastic nonstationarity in the mean can be modelled by differencing (see e.g. [4]) and is related to persistence, deterministic nonstationarity can be modelled by mathematical curves (among which one may select a line as well) and is often exactly what investigators want to detect when trying to prove CC impacts. Here, deterministic nonstationarity in the mean (i.e. deterministic trend) was distinguished from stochastic nonstationarity in the mean (i.e. stochastic trend) that can be modelled by fractional differencing.

For this purpose, hydrologists and climatologists (e.g. [4]) jointly apply one of the unit root tests and a stationarity test because their null hypotheses are set oppositely. In particular, the Phillips–Perron (PP) test that is designed to confirm nonstationarity and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) stationarity test were applied here to decide what type of stochastic processes generates the time series.

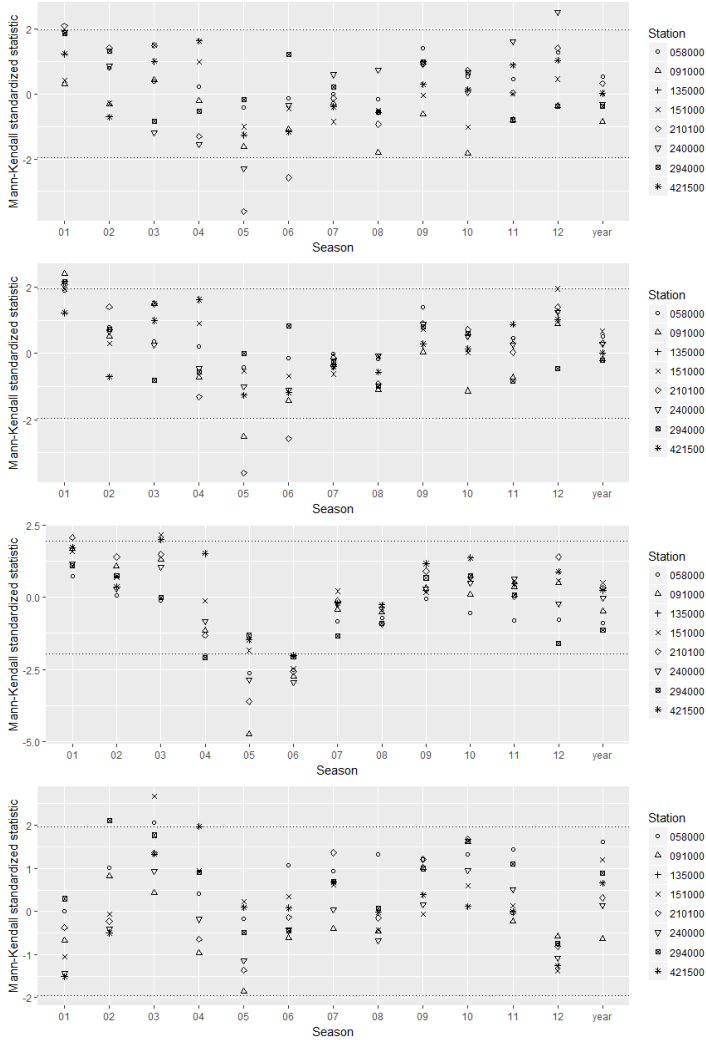
Note that prior to this testing, the series have to be deseasonalized. Therefore, a method based on wavelets suggested in [5] was adopted here. Also, during stationarity testing and the associated Hurst exponent estimation, missing values are usually not permitted. Thus, only the complete series (stations 240000, 091000, 294000, 151000 and 421500) were subjected to this procedure.

#### 3.2 Trend test employed

Trend analysis is a rapidly growing area in hydrology and there are a lot of tests that have been utilized by hydrologists. For their overview see, for example, [6]. In this paper, a methodology outlined in [7] was adopted, which means that the Yue–Wang (YW) modification of the nonparametric Mann–Kendall (MK) test was employed in particular. For more details the readers are referred to the original paper [8] or to [7]. Primarily, the test was applied to annual series, but also monthly series were of interest because they may reveal important changes in the seasonal course. Only the significance level  $\alpha = 0.05$  was selected here. Note also, that, this time, the missing values were allowed, so, again, the above four different periods for the selected stations could be examined.

### 4 Results and discussion

The procedure aimed at checking the stationarity suggested that the deterministic trends at five selected water-gauging stations could not be rejected. It means that the trends detected via the subsequent YW–MK test may very likely be of the deterministic origin.



**Fig. 3.** Resulting standardized statistics of the MK test corresponding to eight selected water-gauging stations and four selected periods (from top to bottom): entire period, 1931–2015, 1961–2015 and 1981–2010. Dotted lines indicate the 95% confidence interval.

Figure 3 summarizes the results of trend analysis. It may be deduced that the trends are really rare here. Indeed, looking at the annual scale, there are hardly any trends. This clearly corresponds to the findings regarding precipitation in Czechia published in [7]. However, more interesting insights were provided by the monthly time scale. Mainly, a sinusoidal course in the standardized MK test statistics is apparent. It also seems that this course is becoming stronger when the starting point of investigated series is shifted ahead. This phenomenon may be associated with the behaviour of snow that seemingly melts out earlier

now in Czechia and its mountains (e.g. [9]). Indeed, decreasing trends, if any, prevail at the turn of spring and summer, which is not so evident for the important period 1981–2010.

## 5 Conclusion and recommendations

In this paper, a freely available data set consisting of mean daily discharge series representing selected water-gauging stations located in Czechia was presented. Many of these series are relatively long, which enables students and other scientists, who are meant to be the main recipients of such data, to freely assess the possible response of Czech water resources to climate change. An example of trend analysis was suggested here, through which it was found that there are no changes in discharge when looking at the annual scale. However, when assessing monthly series, some trends were obvious. Notably, the strengthening decreases in discharge at the turn of spring and summer can pose some issues in water supply services in the future.

Nevertheless, not all aspects were studied here. The series that are now freely online may be assessed from other perspectives as well. For example, changes in seasonality should be evaluated more thoroughly. Also, a periodic behaviour of the time series could be studied, as well as the relationships among upstream and downstream places. Wavelet analysis and especially the methods dealing with the so-called wavelet coherence may be useful. From mean daily discharge series, one may also derive a lot of indicators in order to study hydrological drought patterns. Now, it depends on the students and other scientists if they will accept the offered chance to conduct some more rigorous work based on this data set.

Finally, it should be noted that Czechia is not the only country from Central Europe that currently provides at least some daily hydrometeorological data for free. In the case of Poland, similar data sets can be found at <https://dane.imgw.pl/>. German daily meteorological data can be downloaded from the official FTP server <ftp://ftp-cdc.dwd.de/pub/CDC/>. Probably, there will be more free data in the near future.

The author would like to thank the deputy director for hydrology at the Czech Hydrometeorological Institute. Mainly giving the chance to students from around the world to freely test their statistical skills on such a data set from Czechia is really appreciated.

## References

1. X. L. Wang, H. Chen, Y. Wu, Y. Feng, and Q. Pu, *J. Appl. Meteorol. Climatol.* **49**, 2416 (2010)
2. Y.-F. Sang, Z. Wang, and C. Liu, *Hydrol. Process.* **27**, 2021 (2013)
3. R. Modarres and T. B. M. J. Ouarda, *J. Hydrol.* **499**, 1 (2013)
4. S. Fatichi, S. M. Barbosa, E. Caporali, and M. E. Silva, *J. Geophys. Res.* **114**, D18121 (2009)
5. E. Szolgayová, J. Arlt, G. Blöschl, and J. Szolgay, *J. Hydrol. Hydromech.* **62**, 24 (2014)
6. D. Machiwal and M. K. Jha, *Hydrologic Time Series Analysis: Theory and Practice* (Springer Netherlands, Dordrecht, 2012)
7. O. Ledvinka, *Acta Hydrol. Slovaca* **16**, 199 (2015)
8. S. Yue and C. Wang, *Water Resour. Manag.* **18**, 201 (2004)
9. Z. Kliment, M. Matoušková, O. Ledvinka, and V. Královec, *J. Hydrol. Hydromech.* **59**, 36 (2011)