Vietnam Journal of Earth Sciences, 44(1), 144-164, https://doi.org/10.15625/2615-9783/16980

Vietnam Academy of Science and Technology
Vietnam Journal of Earth Sciences
http://www.vjs.ac.vn/index.php/jse

Identifying hydrologic reference stations to understand changes in water resources across Vietnam - a data-driven approach

Hong Xuan Do[1*], Manh-Hung Le[2,3], Hung Thanh Pham[4], Tu Hoang Le[5], Binh Quang Nguyen[4]

[1]Faculty of Environment and Natural Resources, Nong Lam University - Ho Chi Minh City, Ho Chi Minh City, Vietnam
[2]National Center for Water Resources Planning and Investigation, Hanoi, Vienam
[3]Department of Engineering Systems and Environment, University of Virginia, Charlottesville VA, USA
[4]The University of Danang - University of Science and Technology, Da Nang, Vietnam
[5]Research Center for Climate Change, Nong Lam University - Ho Chi Minh City, Ho Chi Minh City, Vietnam

Received 01 October 2021; Received in revised form 15 January 2022; Accepted 10 March 2022

ABSTRACT

Among the impacts of climate change and variabilities, change in the terrestrial water cycle has been investigated extensively due to the importance of water resources to the Earth System. The detection and attribution of changes in river flows, however, are usually complicated by the impacts of on-ground human activities such as urbanization or dam construction that have altered flow regimes. As a result, hydrological reference stations – gauges measuring river flows of catchments that are unregulated - have been identified in many countries to provide streamflow records that are suitable for climate studies. Such a network, to the best of our knowledge, has not been promoted widely in Vietnam, making it difficult to determine the actual impact of climate change and variability to river flow regimes at the national scale. To address this limitation, this study uses a data-driven approach to identify stations that have not been influenced substantially by human activities for Vietnam. Specifically, we have carefully assessed streamflow records at 68 locations that are not influenced by tidal regime to identify stations with relatively good data quality. The drainage area associated with each of these stations was then delineated following international standards, and then used to identify catchments that were not associated with large dams. The catchment boundary was also used to extract land surface information, available through an ensemble of different satellite-based data products, to further identify catchments that are not featured by large urban areas, or experienced a substantial change in land cover during the 2000-2019 period. Using these criteria, this study suggested two subsets of stations distributed across the countries, providing a good starting point for future investigations into the impacts of climate change on water resources in Vietnam. The investigation also suggests that more investments are required to maintain and expand the hydrologic reference network for Vietnam.

Keywords: watershed hydrology, statistical analysis, large-sample hydrology, catchment characteristics.

1. Introduction

Among the implications of climate change, changes in the terrestrial water cycle are one of the most compelling (IPCC, 2021), owing to the importance of water resources to almost all components of the Earth System. To assess the impacts of climate change on the global water cycle, numerous studies have investigated changes in hydrologic regimes - including droughts (Marengo and Espinoza,

---

[*] Corresponding author, Email: doxuanhong@hcmuaf.edu.vn

2016; Le et al., 2019; Gudmundsson et al., 2021), floods (Burn, et al., 2016; Bloschl et al., 2020; Do et al., 2020a) and identify the dominant factors modulating these changes (Mallakpour and Villarini, 2015; Wasko and Nathan, 2019; Do et al., 2020b; Fowler et al., 2021). The backbone of these investigations is high-quality streamflow data that are appropriate for detecting trends in streamflow indices and attributing these changes to external factors (Addor et al., 2020). However, not all streamflow records are suitable for climate-related research (Burn et al., 2012; Whitfield et al., 2012), as streamflow is sensitive to many catchment attributes that are subject to human intervention, such as changes in channel morphology, urbanization, or water engineering. As a result, data requirements for climate research are usually strict, and streamflow information needs to be recorded at near natural or unregulated basins (Slack and Landwehr, 1992; Archfield et al., 2016; Beck et al., 2017; Hodgkins et al., 2017; Do et al., 2017).

Due to the strict data criteria, reference hydrologic networks (RHN) - stations recording streamflow over catchments that are not heavily affected by human activities - are crucial to support investigations into long-term changes and variabilities of streamflow as a response to climate change. In many countries, substantial efforts have been invested in RHNs, particularly through systematic efforts of water agencies to operate and maintain such networks. In the United States (US), the “reference network” concept was proposed in the middle of the 20th century (Langbein and Hoyt, 1959) and has since been realized into the Hydrologic Benchmark Network (HBN) by the US Geological Survey (Mast et al., 1984). The same concept has also been adopted in Canada (Whitfield et al., 2012), the United Kingdom (Harrigan et al., 2018), Australia (Turner et al., 2012), and many other countries to support climate studies. The World Meteorological Organization (WMO, 2005) has also promoted this practice to establish a global network of streamflow stations that are suitable for climate studies. The common approach to establish these RHNs is to select stream gauges with a minimal human footprint in that catchment characteristics. For instance, RHNs usually include catchments that are not associated with either large dams (and reservoirs) or established urban areas.

In Vietnam, a general consensus among the hydrology community is that there is no such “reference station”. It is because the country has experienced significant changes in land use, land cover, as well as rapid urbanization since the 1980s (Truong et al., 2018). The intensive development of dams and reservoirs in the last two decades has further exacerbated the situation, making it seems unrealistic to have any catchments that were not regulated by human activities. However, there was no systematic effort (to the best of our knowledge) has been invested to assess whether the above-mentioned notation (i.e., Vietnam does not have any gauged catchment that is unregulated) is valid. This limitation is one of the primary reasons that has led to a lack of large-scale investigations into climate change impacts on hydrological regimes for Vietnam, as it is challenging to identify catchments with streamflow records unaffected by human interventions.

This study aims to fill this gap and adopt a data-driven approach to identify streamflow stations that could be classified as “reference” among active stream gauges in Vietnam. Such a network, should it be identifiable, is crucial for future investigations into changes in Vietnam’s water resources. To achieve this objective, we first synthesized the metadata of all streamflow stations in Vietnam obtained from the Vietnam Meteorological and

Hydrological Administration (VHMA) through multiple projects. Stations that are active (as of June 2021) and not subject to lunar tidal influence were then identified, and discharges recorded at those stations were then synthesized. We also put substantial efforts into developing catchment boundaries associated with each of the selected stations. We then identified two subsets of potential "reference stations" as those with (i) high-quality streamflow records (i.e., those that do not show spurious changes) and (ii) were not influenced substantially by human activities.

2. Data and Method

2.1. Data

2.1.1. A nationwide streamflow archive for surface water

In Vietnam, the Ministry of Natural Resources and Environment (MONRE) is responsible for planning the observational networks for hydro-meteorological variables. The VHMA specifically operates and maintains stream gauges across Vietnam. As described in Decision 90, Vietnam has a total number of 354 hydrological stations, including both traditional and automated hydrological stations. Traditional stations are those that have been installed for several decades and thus can provide long-term data suitable for any investigations into changes in river discharge and water resources. This characteristic is aligned with the purpose of our study, and thus we have restricted our search to only traditional stations.

We then adopted three criteria to narrow down the number of stations selected for this investigation: (i) stations that still operate as of June 2021; (ii) stations that are classified as level 1 and level 2 (i.e., the ones that measure both streamflow and water level); and (iii) non-tidal hydrological stations. The rationale of the first two criteria is to select stations that still operate (thus are useful for future investigation), while the last criterion was chosen to remove gauges that are influenced by reverse flow, which could create substantial noises in the signal of changes in streamflow regimes.

Of all stations listed in Decision 90, only 74 hydrological stations met the proposed criteria. We further removed six stations that do not have drainage area information in their metadata (this information is crucial to get a reliable catchment boundary). Discharge data available across those 68 stations (see Appendix 1 for station metadata) were then synthesized for this investigation. The first year of data entrance was 1956, when Vietnam had reliable measurement equipment (Tran, 2006; Tran et al., 2012), and the last data entrance was in 2019.

As shown in Fig. 1a, the number of stations with available data has gradually increased over time - indicating substantial investment of water authorities to hydrological observation network - and there are 57 stations that have more than 30 years of data (Fig. 1b; data is not necessarily continuous) available. We note that, due to data entrance criteria, the Mekong Delta was not included (Fig. 1c). This region is influenced by not only tidal effect but also upstream human activities, and thus is not suitable for this investigation.

We also note that, according to international standards, a reference station usually has data records for at least 30 years. However, we decided to investigate all 68 stations that have been identified from our initial screening process. The rationale of our approach is to provide a holistic perspective about Vietnam's observational network for river discharge, which is still not available to the public at large from our knowledge.

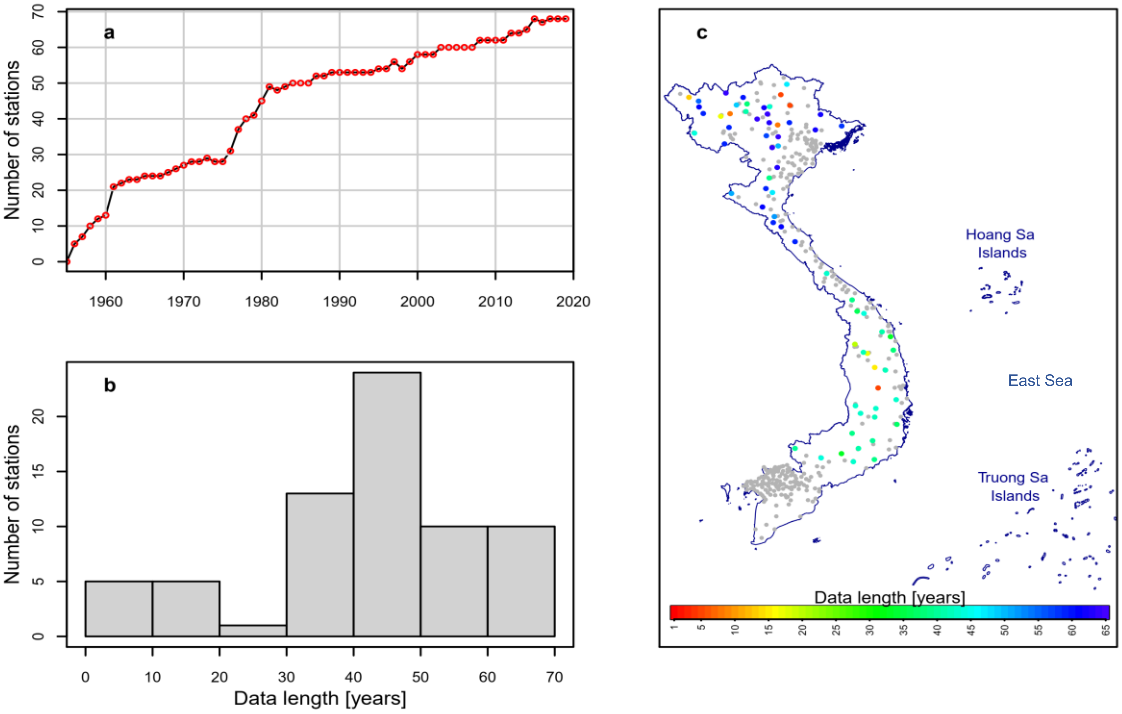


Figure 1. Overview of streamflow stations collected for this investigation: (a) the number of gauges with available data (at least 350 data points for any specific year) per year during the 1955-2019 period; (b) histogram of data length (in years) for gauging stations; (c) spatial distribution of 68 active stations that were used for this investigation (color dots) as well as stations that were filtered prior to this investigation (grey dots; more information is available in the "Data" section)

2.1.2 Digital elevation model (DEM) and land cover information derived from satellite imagery

The 90m DEM has been obtained from the Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales (HydroSHEDS, available from hydrosheds.org). HydroSHEDS is based on high-resolution elevation data obtained by the Space Shuttle for NASA's Shuttle Radar Topography Mission (SRTM). HydroSHEDS is a mapping product that provides consistent hydrological information for global and regional scale applications (river networks, drainage directions, flow accumulations, and watershed boundaries). The HydroSHEDS DEM was used to generate catchment boundaries for the selected stations.

This study used the European Space Agency Climate Change Initiative Land Cover (ESA CCI LC; hereafter referred to as ESA LC) to represent land cover across Vietnam. This product provides consistent annual global land cover maps at 300 m spatial resolution from 1992 to 2019. The ESA LC is based on various satellite data sources, including Envisat MERIS, AVHRR, SPOT-VGT, and PROBA-V datasets. It provides different land categories that are suitable at the scale of the entire world. The ESA LC dataset has been validated in several regions in Asia and Africa, demonstrating its good agreement with ground observation (ESA, 2017). The ESA LC dataset was used in our investigation to show the magnitude of land cover change over each catchment between 2000 and 2019. In this study, we grouped 37 land cover types into 7 classes of agriculture, forest, wetland, water, built-up, barren land, and snow (Table 1).

Table 1. Land cover classes that were used in this investigation and corresponding land cover types of ESA CCI LC data

Classes (code)	ESA CCI LC code
Agriculture (1)	10, 11, 12, 20, 30, 40, 120, 121, 122, 130, 140, 150, 151, 152, 153
Forest (2)	50, 60, 61, 62, 70, 71, 72, 80, 81, 82, 90, 100, 110, 160, 170
Wetland (3)	180
Built-up (4)	190
Barren land (5)	200, 201, 202
Water (6)	210
Snow (7)	220
See http://maps.elie.ucl.ac.be/CCI/viewer/download/CCI-LC_Maps_Legend.pdf for complete nomenclature of ESA CCI LC data	

2.1.3. A compiled database of major dams in Vietnam

In this investigation, we have synthesized a database of water infrastructure presence across Vietnam, using the Mekong Infrastructure Tracker database – developed by the Stimson Center (available at https://www.stimson.org/2020/mekong-infrastructure-tracker-tool/; last access: Sep 2021) – as the starting data product. The data underlying the Stimson database is collected from various sources (from water infrastructure developers to the community), and the information has undergone a careful verification process.

This dataset, however, does not include all major water infrastructures within Vietnam and thus we have consulted extra data sources in this study. To ensure that all required information is freely accessible, we have limited the database to only major dams (i.e., those associated with a reservoir of at least 10 million cubic meters). Specifically, the following sources were used to generate a list of all major infrastructures for Vietnam, including (i) hydrological reports for river basins provided by VHMA; (ii) public data available on disaster prevention website (http://dulieu.phongchongthientai.vn; last access: Sep 2021); (iii) public data available on irrigation database system (http://thuyloivietnam.vn; last access: Sep 2021); and (iv) public data available on the open data portal of the Ministry of Industry and Trade (https://thuydienvietnam.vn; last access: Sep 2021). We also consulted websites of investors, managers, and operators of reservoirs as well as the association of large dams and water resources development in Vietnam (http://www.vncold.vn; last access: Sep 2021) to fill in the required information for this investigation (i.e., geographical coordinates and volume storage) whenever the information is not available from the listed data sources.

Figure 2 shows the spatial distribution of 87 major dams included in our water infrastructure database to support our investigation. These dams are associated with a total storage volume of at least 10 million cubic meters. Our database, therefore, provides a better representation of human intervention to local hydrology compared to other international databases (e.g., most dams in the Global Reservoir and Dam Database (Lehner et al., 2011) have a storage capacity of at least 100 million cubic meters). It is apparent from Fig. 2 that most major dams in Vietnam have been constructed within the last two decades (note that there are 17 dams that do not have information about construction time), indicating the need for identifying stations with streamflow records unaffected by human intervention through dam construction and operation.

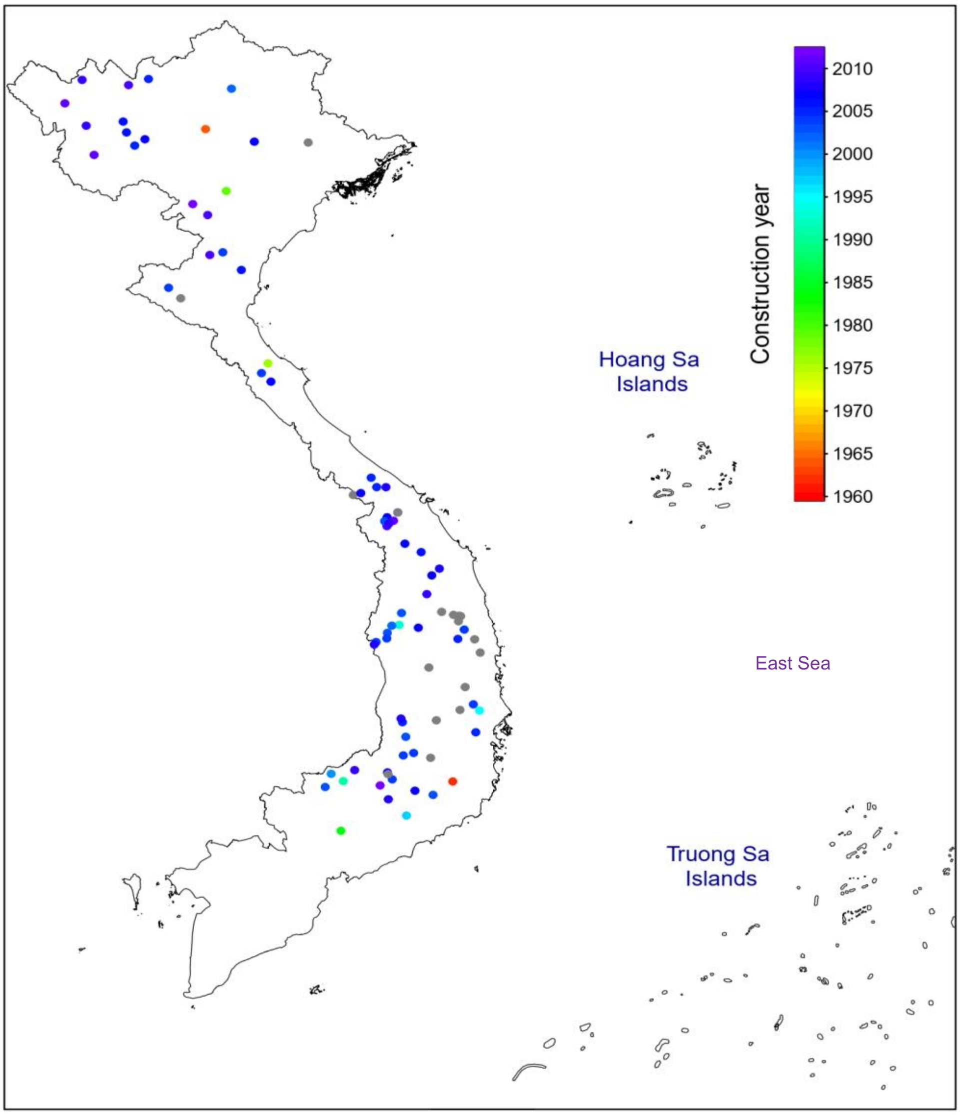


Figure 2. Spatial distribution of major dams (associated with a total storage volume of at least 10 million cubic meters) across Vietnam. The "construction year" was shown for reference (grey dots are dams with no information of construction year)

2.2. Method

2.2.1. Assessing streamflow homogeneity

There is a consensus among Vietnamese hydrologists that all streamflow records across Vietnam's River have been impacted by many factors, including river regulations, water abstraction, and human activities in river basins. Many other factors can also lead to changes in streamflow records that cannot be explained by natural variabilities, such as changes in gauging instruments, relocation of stations, or revision of the rating curve whenever new data is available (Gudmundsson et al., 2018). However, there is no publication (to the best of our knowledge) that systematically documents these factors for the selected stations thus it is not possible to explicitly remove time series associated with abrupt changes that could mislead detection and attribution analysis.

To ensure that catchments identified as “reference” through this investigation are accompanied by streamflow records that do not exhibit spurious patterns, we adopted an international standard (ECA & D Project Team and Royal Netherlands Meteorological Institute, 2013) and assessed the homogeneity in streamflow records across the selected stations. The approach of this assessment comprises of two steps: (i) calculating streamflow indices (see Appendix 2) that are usually used in trend detection and attribution studies; and (ii) conducting statistical tests to assess whether the time series of the calculated indices exhibit spurious changes that are unlikely climate-induced.

Specifically, we first screened all daily data points and provided a “suspect” flag if these data points were either negative, outliers, or consecutively constant for more than 10 days. The remaining data points were then assigned a “reliable” flag. We then assessed the homogeneity of streamflow data recorded at each station separately, following the procedure described in Gudmundsson et al. (2018). We first derived, from the quality-controlled records, a range of different indices that are regularly used for trend detection studies at the monthly, seasonal, and annual time steps (e.g., annual maximum, annual mean streamflow; see Gudmundsson et al. (2018) for a full list of these indices). The indices were only calculated for months/seasons/years that have at least 25/85/350 “reliable” daily observations, respectively.

We then applied four independent homogeneity tests - including (i) the standard normal homogeneity test (Alexandersson, 1986), (ii) the Buishand range test (Buishand, 1982), (iii) the Pettitt test (Pettitt, 1979), and (iv) the von Neumann ratio test (von Neumann, 1941) - to each of the indices (e.g., the four mentioned tests were applied to time series of annual mean discharge of a specific catchment). The time series of a specific streamflow index was then assigned into one of the following four categories:

- “Useful”: if only one or no homogeneity test rejects the null hypothesis (the time series is homogeneous) at the 1% level;
- “Doubtful”: if two tests reject the null hypothesis at the 1% level;
- “Suspect”: if three or four tests reject the null hypothesis at the 1% level;
- “Not sufficient data”: if less than 20 reliable data points (at either the yearly, seasonal, or monthly resolution) are available for that specific index.

We note that this procedure was discussed in more detail in Gudmundsson et al. (2018), and the readers are recommended to consult this publication for more in-depth information. To be classified as “reference” in our assessment, stations must be associated with “useful” time series across all calculated streamflow indices.

2.2.2. Identifying best outlet and delineating catchment boundary

An important input for our investigation is a reliable drainage boundary corresponding to each of the selected stations. Such information, however, is not readily available from data providers, and thus an ad hoc effort was adopted to delineate catchment boundaries. To fulfill this task, outlet identification is one of the most important steps, and we have adopted a consistent procedure across all catchments to manually identify the locations representing the best outlet.

Specifically, we first reviewed station reports (obtained from VHMA) which provide geographical coordinates, administrative address, river name, and describe the river section where the stations are located. We then used Google Earth to verify whether the geographical coordinates are consistent with the physical locations described in the reports. For cases where there was substantial inconsistency, we carefully reviewed the area surrounding the provided geographical coordinates and identified locations that fit the descriptions in station reports.

Station locations were then imported into ArcGIS software for watershed delineation. As HydroSHEDS was used to represent catchment topography, several outlets were slightly relocated (from the station location) to ensure it resides on the correct stream in the HydroSHEDS river network. After the catchment was delineated, we used the reported catchment area as the benchmark to assess the quality of the delineated catchments, similar to the approach described in Do et al. (2018). Specifically, if the delineated catchment has a discrepancy in catchment area within ±10% relative to the reported catchment area, the boundary was flagged as "reliable". Otherwise, we made a search for an outlet (within a 10-km radius and that is closest to the correct geographical coordinates of the station) that would generate a catchment boundary with a discrepancy rate within ±10%. For cases where such an outlet (i.e., the one that would yield a discrepancy within ±10%) is not identifiable, we flag the catchment boundary as "suspect".

Figure 3 provides an example of this procedure for the Dong Nai River at Thanh Binh station. The coordinates reported in the metadata (denoted in Fig. 3 as point A) are inconsistent with station description (available in station report). We have manually relocated the station location to the new one (denoted as point B) to fix this error in the metadata. An adjustment was then made to ensure the outlet (used for catchment delineation) is on the HydroSHEDS river network, and the area calculated from the delineated catchment boundary yields a small area discrepancy relative to the reported catchment area.

2.2.3. Identifying human influence across upstream drainage area

To investigate human influence, we used the delineated catchment boundary to extract - from the dams and land cover databases - the following information: (i) number of dams located within a specific catchment, (ii) total storage capacity of all these dams, (iii) percentage of catchment associated with an "urban" land cover type, and (iv) percentage of the catchment that has undergone land cover conversion.

To generate the last metric, we calculated the percentage change of land use/cover types between 2000 and 2019 (as illustrated in Fig. 4). We first assigned a code varying from 1 to 7 (described in Table 1) to each land use/cover type. We then overlayed the land cover data of the two timesteps and calculated the percentage of cells exhibiting changes following equation 1.

$$P = \frac{\sum_{j=1}^{i} A_{ij}}{A} \times 100 \quad \text{with } i \neq j \qquad (1)$$

where, P is the percentage of land area exhibiting land cover change over the 2000-2019 period; A_{ij} is the area of the land cover category changes from i to j during the period (i, j has a value from 1 to 7) and A is the total area of the catchment.

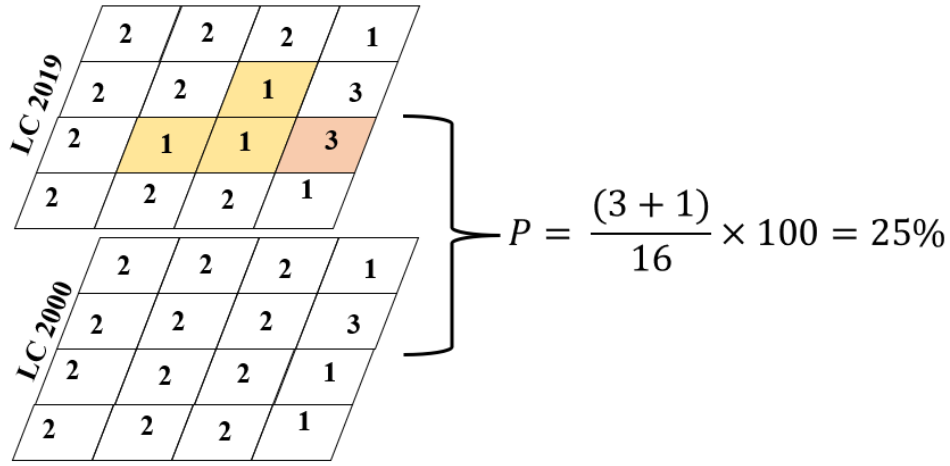
Having all these metrics processed for selected catchments, we then applied the following data-driven filters to get the final subset of "reference stations":

- Time series of all indices are classified as "useful" from the homogeneity assessment (i.e., the hydrograph does not show abrupt changes that are unlikely climate-induced; ECA & D Project Team and Royal Netherlands Meteorological Institute, 2013);

- The total storage capacity of upstream dams is less than 10% of the recorded annual flow volume, following established standards of RHNs (Marsh and Hannaford 2008; Whitfield et al., 2012).
- Catchment has less than 2% of its area classified as "urban". This type of land cover could introduce substantial impacts to flow regimes - even when it only represents a relatively small proportion of the total catchment - that cannot be explained by changes in climate conditions (Lins, 2009).
- Catchment has less than 10% of land cover change rate (Whitfield et al., 2012) during the 2000-2019 period when Vietnam has undergone one of the most dynamic eras in terms of land-use change (World Bank Group, 2012).
- Catchment has more than 50% of its area within the Vietnam boundary.

We note that these criteria are generally consistent with the approach used in previous investigations to identify natural or near-natural catchments. The rationale of these criteria is to identify (and remove) stations that have either abrupt changes in their time series or substantial presence of human intervention over their drainage area. The final criterion, on the other hand, was proposed to reduce the uncertainty in our evaluation, as we have limited information of dam constructions outside of Vietnam border to support this investigation.

Figure 3. Example of the procedure to identify the catchment outlet that was used to delineate the catchment boundary associated with Thanh Binh station located on the Dong Nai River. The geographical coordinates reported in the metadata (point A) were first manually relocated to point B, which was then used as the initial location to search for the best outlet (point C). The best outlet is the location that lies on HydroSHEDS river network and would generate a catchment boundary with a low area discrepancy relative to the reported catchment area (obtained from station documents)



Note: Agriculture (1); Forest (2); Residential (3)

Figure 4. Illustration of land-cover change assessment. Each land cover type was assigned a numeric code. Cells exhibiting change from one code to another one (colored) between the two timesteps were identified. The percentage of catchment showing changes in land cover were then calculated

3. Results and Discussions

3.1. Homogeneity assessment

Figure 5 summarizes the result of the homogeneity assessments for all indices generated at the yearly (panel a), seasonal (panel b), and monthly (panel c) resolutions. At the yearly resolution, there are nine time series that do not have sufficient data for the homogeneity assessment (i.e., there are less than 20 years with reliable data over those 9 stations). Of all 59 remaining stations, almost all indices time series were classified as "useful", indicating that the null hypothesis of homogeneity was not rejected in at least three out of the four statistical tests. At the monthly and seasonal resolutions, there was sufficient data for the homogeneity tests across all selected stations, although it is important to note that the 20-month threshold is still relatively short. As a result, the assessment should be consulted with care for cases with less than 10 years of records (see Fig. 1c for these locations).

It is apparent from Fig. 5 that there were many cases when the null hypothesis of homogeneity was rejected in two or more statistical tests (i.e., those time series classified as "doubtful" and "suspect") when the analysis was conducted at either the seasonal or monthly resolution. Among the analyzed indices, the minimum (MIN) and 7-day minimum (MIN7) were the ones having the most inhomogeneous time series (over up to 20 locations, or 29% of all stations), indicating the high sensitivity of this index to external factors that could modulate streamflow regimes. On the other hand, the standard deviation (SD) index tends to be insensitive to external factors, as less than 5 stations exhibit inhomogeneity in both seasonal and monthly resolution.

The results of the homogeneity assessment suggest the importance of considering a wide range of streamflow indices to reliably conclude about the quality of streamflow records that are of interest (as some indices might be more sensitive to external factors than the others). As a result, our study also identified the stations exhibiting homogeneity across all of the assessed indices to ensure that the stations recommended for future investigations are generally not subject to abrupt changes that cannot be attributed to climate changes and variabilities.

Figure 6 shows the distribution of stations exhibiting homogeneity across all indices. Of all 68 stations that were assessed, there were 57, 42, and 40 stations that have all indices classified as "useful" at the yearly (Fig. 6a), seasonal (Fig. 6b), and monthly (Fig. 6c) resolutions. Even when we applied a stricter criterion, of which all indices (regardless of temporal resolutions) must be classified as "useful", there were still 29 stations that satisfied the condition (Fig. 6d). In general, these stations are distributed across all climate regions of Vietnam (except for the southern part, which was not included in this investigation as mentioned in our Data section), indicating it is possible to investigate changes in Vietnam's water resources and attribute any detected changes to climate influence using a large sample of catchments with good data quality.

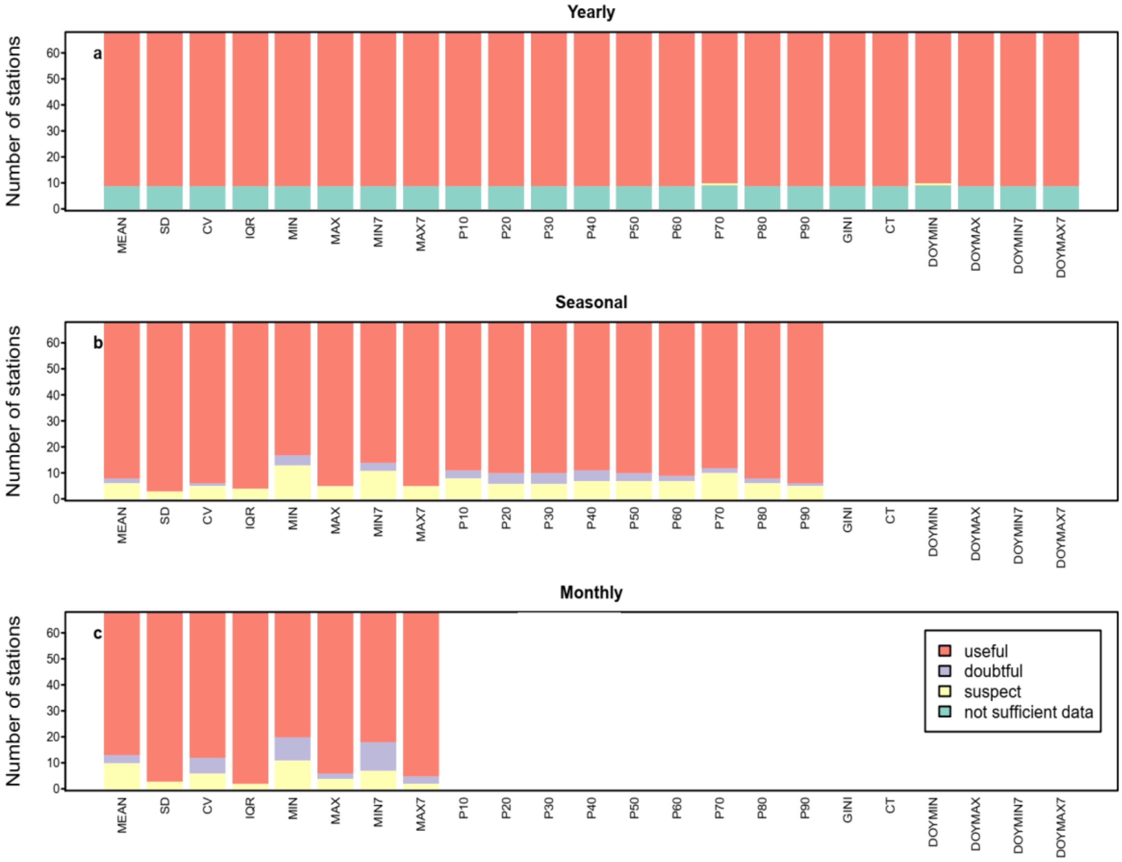


Figure 5. Summary of the homogeneity analysis for all generated indices at yearly, seasonal and monthly resolution. The number of stations that are classified as (1) useful, (2) doubtful, (3) suspect, and (4) not sufficient data are shown

We also note that this assessment only considers spurious patterns from the hydrograph and thus might miss other factors that still yield substantial influence on the hydrological regimes. To address this limitation, the next sections will describe our efforts in identifying catchments that are likely influenced substantially by human activities using data derived from satellite imagery.

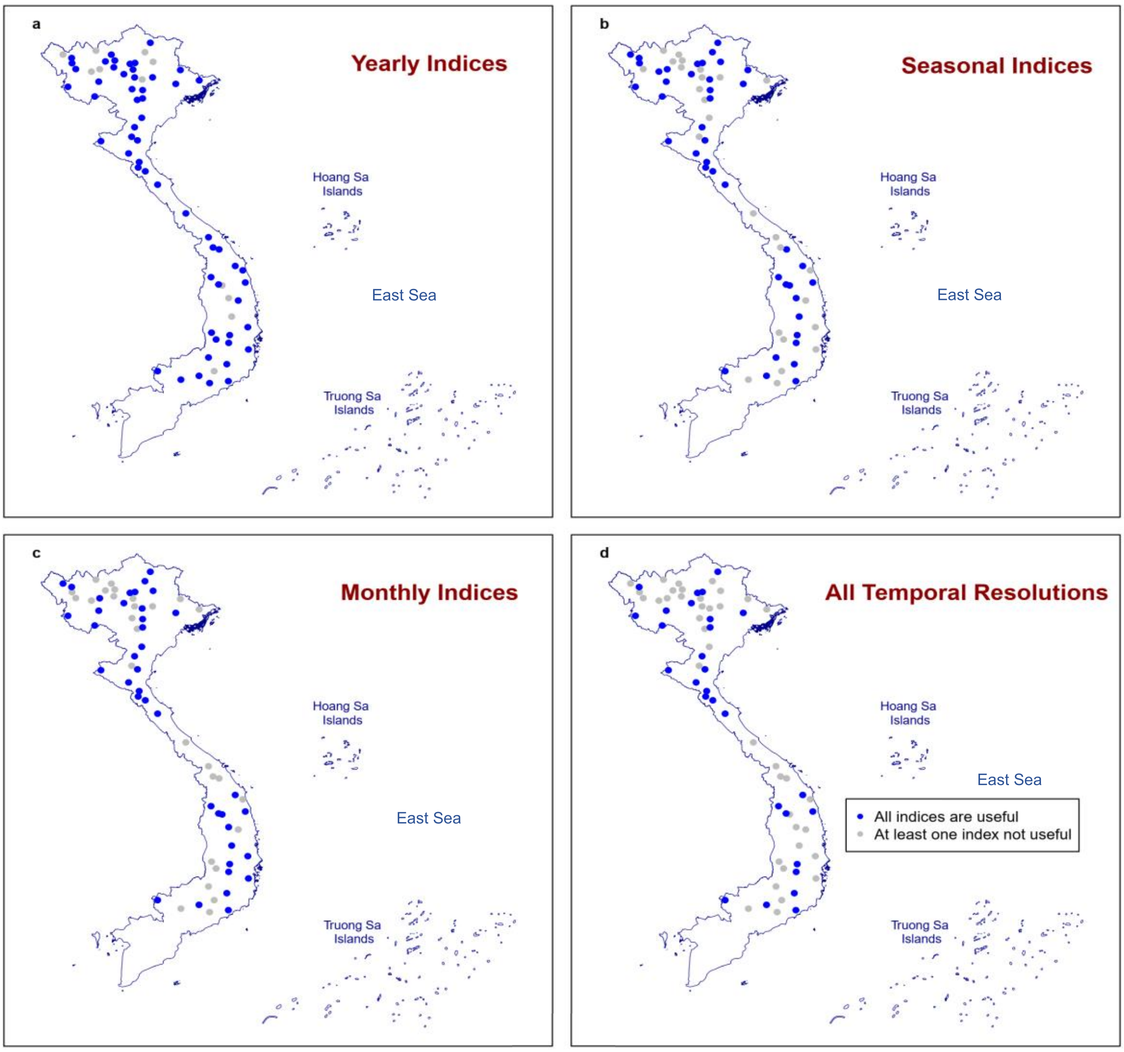


Figure 6. Maps showing the distribution of useful/suspect stations. Red dots represent time series that have all indices classified as useful at (a) the yearly resolution, (b) the seasonal resolution, (c) the monthly resolution, and (d) all temporal resolutions. Grey dots represent time series that have at least one index classified as either "suspect", "doubtful", or "not sufficient data"

3.2. Catchment boundaries associated with selected stations

To assess human intervention in the streamflow regime, it is essential to reliably delineate the drainage boundary upstream of each selected station. To achieve this objective, we have manually reviewed all station locations and selected the best outlet that generates a catchment boundary with a delineated catchment area highly consistent with the area reported in station documents.

Figure 7 illustrates the result of this effort and shows the boundaries of all 68 catchments that will be used to extract human intervention information associated with the assessed stations. There were 57 stations associated with an area discrepancy of less than 10%, indicating the relatively high quality of the delineated catchment boundary. Figure 7 also highlights that many of Vietnam's catchments, particularly over the Hong - Thai Binh River and the Ma River, are transboundary catchments, with a significant portion of their catchment boundaries belonging to the neighboring countries. We note that, due to data accessibility, we only collected information on dams located within Vietnam's administrative boundary. This approach will potentially lead to high uncertainty in our evaluation of transboundary catchments. We thus have applied a filter to separate all catchments into two groups:

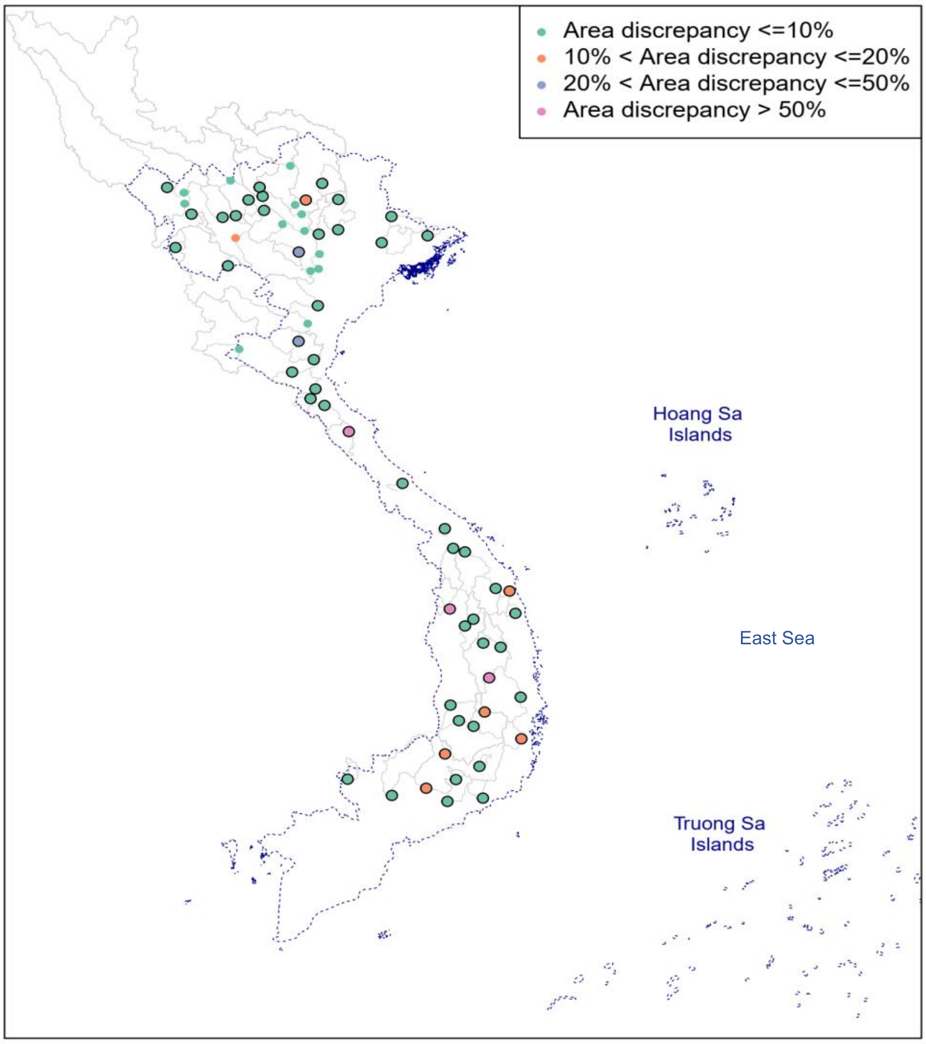


Figure 7. Catchment boundaries of all selected stations (grey lines) and the associated stations (colored dots). The colors represent the discrepancy between the area calculated from the delineated catchment and information reported in station documents. Circled dots represent stations that were classified as domestic

- "Domestic" catchments are those with more than 50% of their catchment boundary inside the Vietnam administrative borders.
- "International" catchments are those with at least 50% of their catchment boundary outside the Vietnam administrative borders.

Of all 68 catchments, there were 54 classified as "domestic" (circled dots in Fig. 7), and this flag will then be used as a condition to identify hydrologic reference stations.

3.3. Evaluation of human influence across catchments

The catchment boundaries were then used to extract information of human intervention across the upstream of each selected streamflow station. We first identified the major dams (those with a designed total storage capacity of at least 10 million m^3) located within each catchment boundary. The sum of storage capacity of all those dams was then calculated and compared to long-term average annual flow volume. We then made a pragmatic decision and identified stations with relatively low influence by dam operations as those with the ratio between storage capacity and annual flow of no more than 10%. Figure 8 shows the results of this assessment and highlights the locations of 20 stations that were classified as being influenced substantially by dam operation in our investigation (orange dots). There are 35 stations that do not have any large dams located in their upstream drainage area and another 13 stations that have the sum storage capacity of less than 10% of annual flow.

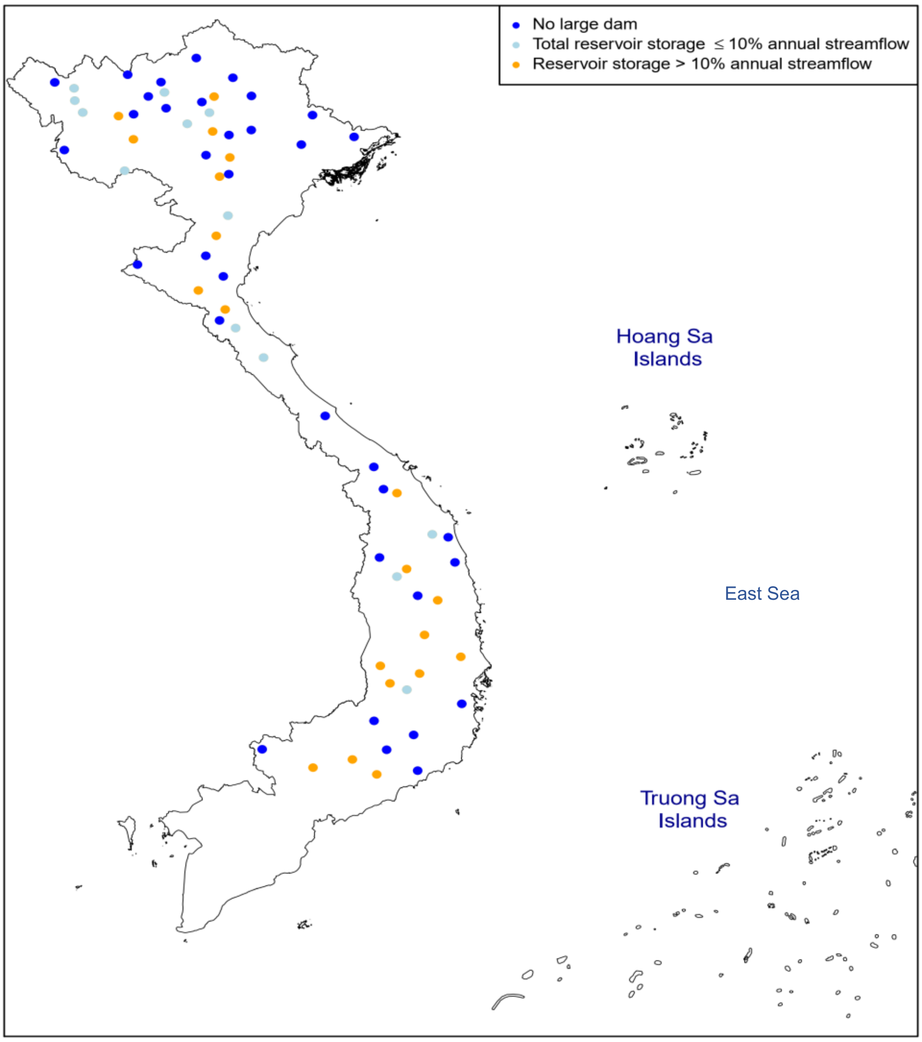


Figure 8. Large-dam presence across 68 hydrological stations. Each station was classified into one of the three classes representing the potential influence of large-dam on the catchment's streamflow regime. Only dams with the total storage capacity of at least 10 million m^3 are considered in this assessment

Figure 9 shows two important factors representing human intervention on hydrologic regimes of the assessed catchments: the presence of urban areas (calculated from the 2019 land cover map) and land cover changes rate across each of the catchments during the 2000-2019 period. Of all 68 catchments, the vast majority (67) has less than 2% of the catchment area classified as built-up (or urban), as shown in Figure 9a. This result is reasonable as our investigation only focuses on gauges that are not subject to lunar tidal influence; thus, the major coastal cities (i.e., Ho Chi Minh City, Hai Phong, Da Nang, and Can Tho) are not present in any of the assessed catchments.

Figure 9b shows the result of our investigation into changes in land cover across each of the catchments. Our approach is similar to that discussed in Whitfield et al. (2012), and consider that catchments with a large agricultural area can still be classified as "undisturbed" if the land cover has not exhibited substantial changes (and thus the physical attributes of land cover remain relatively stable) during the assessed period. This assessment reviews an interesting pattern in the land cover change in Vietnam: catchments located in the Central Highland exhibit the highest rate of land cover changes during the 2000-2019 period, reflecting the implication on the land cover of national policies in recent decades to boost the regional economy. There are 40 catchments experiencing changes in land cover types over no more than 10% of their drainage area (blue dots).

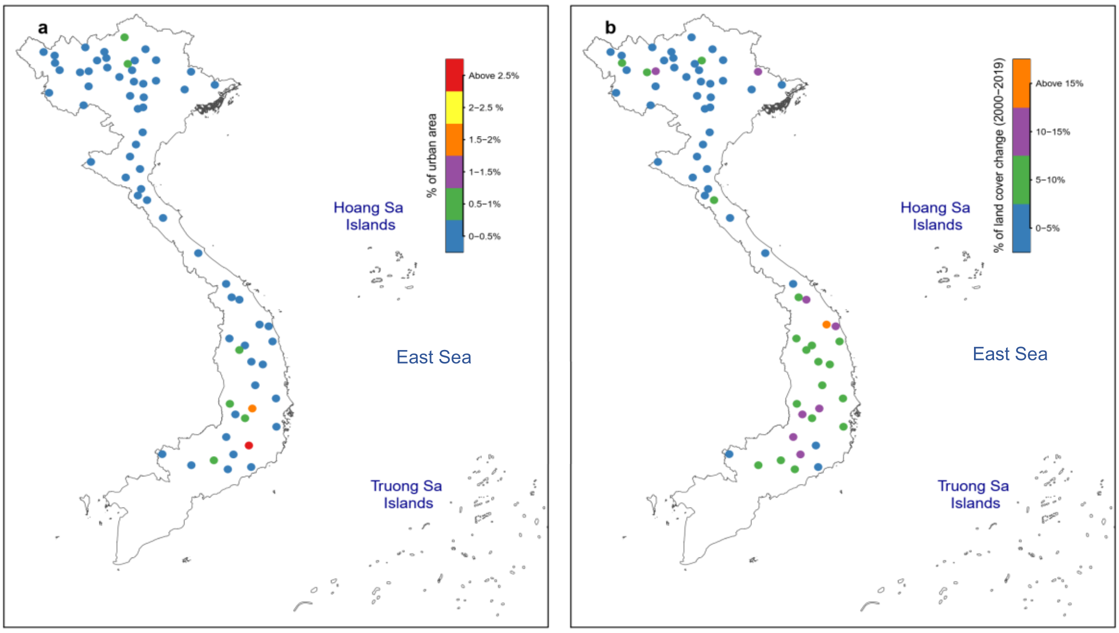


Figure 9. Land cover characteristics across 68 catchments: (a) the percentage of catchment area classified as "built-up" (or "urban") land cover type; (b) the percentage of catchment area exhibiting a conversion from one land cover type to another type during the 2000-2019 period

3.4. Spatial distribution of potential RHNs to support climate studies in Vietnam

Figure 10 shows the final subsets of stations that could be identified as "reference" using the data-driven approach described in this article. Fig. 10a provides a more "generous" set of criteria (denoted as "Set A"), as the stations only need to have yearly streamflow indices showing homogeneity to be classified as "useful". The subset showed in Fig. 10b (denoted as "Set B"), on the other hand, represents stations that met stricter data quality criteria, as all indices at the seasonal and monthly resolutions need to be classified as "useful" to be included.

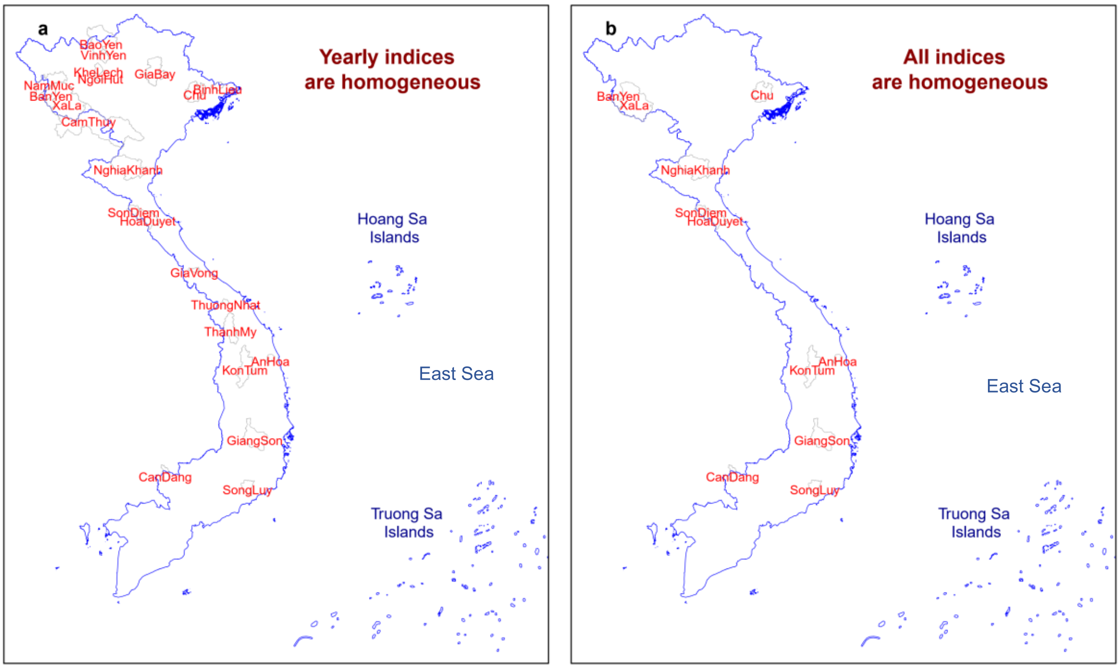


Figure 10. Spatial distribution of potential RHNs that met all data-driven criteria described in this study. The same criteria for catchment characteristics were used, but different homogeneity criteria were considered: (a) only yearly indices need to be homogeneous (corresponding to blue dots in Fig. 6a), and (b) all indices must be homogeneous (corresponding to blue dots in Fig. 6d)

The rationale for providing these two subsets lies in the potential applications of these catchments in climate studies. For studies that focus on long-term changes in the normal condition of water resources, homogeneous indices at the yearly resolution are arguably sufficient, and thus catchments included in Set A are recommended. For studies focusing on more detailed aspects of the hydrologic regime, high-quality data at the seasonal and monthly resolutions are needed, and thus stricter criteria should be considered, leading to a smaller subset included in Set B. Although the number of catchments available in the proposed networks is relatively small (22 for set A and 11 for Set B), they generally provide good spatial coverage and are still useful to provide a large picture of any changes in the hydrologic regime across Vietnam.

4. Conclusions

This study reports the first systematic effort to identify the network of stations that are relatively unaffected by on-ground human activities. Using a data-driven approach, we have identified two subsets of stations that could be used as the initial search for catchments that are suitable for future investigations into natural changes in hydrologic regimes and water accounting across Vietnam in a warming climate. Our investigation also shows that some important regions are not adequately covered by the proposed “reference stations”, including the Central Highland, the Red River, and the South. This finding indicates an urgent demand for establishing new stream gauges across Vietnam to support water decision-making processes.

We note that our ad hoc effort by no means provides the definition of the RHN for Vietnam due to the limited resources available for this investigation. For instance, our dam database was limited to those with a total storage capacity of at least 10 million m^3 and thus might not represent the actual influence of dams and reservoirs on streamflow regimes. Another limitation lies in our simple method to quantify land-use change, as only seven land cover types were considered.

These limitations do not overshadow our intention, which was to initiate the discussion from water agencies as well as the Vietnam hydrology community about an RHN for Vietnam. To establish the RHN, ad hoc efforts such as that described in this article are not (and never be) sustainable. Indeed, an orchestrated effort and substantial investments coordinated by water authorities are required to identify, maintain suitable stations as well as to establish new stations that meet international standards. Such a hydrologic reference network, should it be officially established, will be a crucial asset for Vietnam to better understand the uncertain future facing us in a warming climate.

To achieve this important objective, collective efforts should be implemented by the authorities as well as Vietnamese hydrologists. From our perspective, water agencies remain the key players and should invest more efforts to make streamflow records become more FAIR (Findable, Accessible, Interoperable and Reusable; Wilkinson et al., 2016) to the public at large. A data portal similar to that of the Mekong River Commission (https://portal.mrcmekong.org/; last access: 21 Sep 2021) would be a good starting point to make streamflow data more accessible, even with or without a payable fee. Other important databases, such as dams and reservoirs operational rules (currently coordinated by the Ministry of Industry and Trade), should also be synthesized and made accessible through the streamflow data portal. This initiative will make it is possible to assess human interventions to streamflow regimes more comprehensively in the future. Vietnamese hydrologists should also pay more attention to archiving the data underlying independent investigations (at the local scale) using data repositories that are accessible to the larger community. Finally, investment from the government into the water sector should consider a larger share into establishing new stream gauges as well as maintaining active stations. This effort is particularly important for future research, as our findings have demonstrated a very low density of stations that are relatively free from human interventions.

Acknowledgments

Hong Xuan Do and Tu Hoang Le are funded by a grant for basic research (No. CS-CB21-MTTN-07) from Nong Lam University - Ho Chi Minh City to conduct this project. We thank Vietnam Meteorological and Hydrological Administration for providing hydrological records. Due to data policy from water agencies of Vietnam, streamflow data underlying this investigation cannot be redistributed. The authors also wish to thank two anonymous reviewers for their constructive comments and suggestions.

References

Addor N., Do H.X., Alvarez-Garreton C., Coxon G., Fowler K., Mendoza P.A., 2020. Large-sample hydrology: recent progress, guidelines for new datasets and grand challenges. Hydrological Sciences Journal, 65(5), 712-725.

Alexandersson H., 1986. A homogeneity test applied to precipitation data, J. Climatol., 6, 661-675.

Archfield S.A., Hirsch R.M., Viglione A., Blöschl G., 2016. Fragmented patterns of flood change across the United States. Geophysical Research Letters, 43(19), 10-232.

Beck H.E., van Dijk A.I., Roo A.D., Dutra E., Fink G., Orth R., Schellekens J., 2017. Global evaluation of runoff from 10 state-of-the-art hydrological models. Hydrology and Earth System Sciences, 21(6), 2881-2903.

Blöschl G., Kiss A., Viglione A., Barriendos M., Böhm O., Brázdil R., Coeur D., Demarée G., Llasat M.C., Macdonald N., Retsö D., 2020. Current European flood-rich period exceptional compared with past 500 years. Nature, 583(7817), 560-566.

Buishand T.A., 1982. Some methods for testing the homogeneity of rainfall records, Journal of Hydrology, 58, 11-27

Burn D.H., Hannaford J., Hodgkins G.A., Whitfield P.H., Thorne R., Marsh T., 2012. Reference hydrologic networks II. Using reference hydrologic networks to assess climate-driven changes in streamflow. Hydrological Sciences Journal, 57(8), 1580-1593.

Burn D.H., Whitfield P.H., Sharif M., 2016. Identification of changes in floods and flood regimes in Canada using a peaks over threshold approach. Hydrological Processes, 30(18), 3303-3314.

Decision No. 90/QD-TTg dated January 12nd, 2016 of the Prime Minister approving the National Plan of Monitoring network of Natural Resources and Environment in the period of 2016-2025 with a vision to 2030. Available at: https://vanban.chinhphu.vn/default.aspx?pageid=27160&docid=183027 (Accessed: 16 September 2021).

Do H.X., Westra S., Leonard M., 2017. A global-scale investigation of trends in annual maximum streamflow, Journal of Hydrology, 552, 28-43. https://doi.org/10.1016/j.jhydrol.2017.06.015.

Do H.X., Gudmundsson L., Leonard M., Westra S., 2018. The Global Streamflow Indices and Metadata Archive (GSIM) - Part 1: The production of daily streamflow archive and metadata, Earth Syst. Sci. Data, 10, 765-785. https://doi.org/10.5194/essd-10-765-2018.

Do H.X., Zhao F., Westra S., Leonard M., Gudmundsson L., Boulange J.E.S., Chang J., Ciais P., Gerten D., Gosling S.N., Müller Schmied H., Stacke T., Telteu C.-E., Wada Y., 2020a. Historical and future changes in global flood magnitude- evidence from a model-observation investigation, Hydrology and Earth System Sciences, 24, 1543-1564.

Do H.X., Mei Y., Gronewold A.D., 2020b. To what extent are changes in flood magnitude related to changes in precipitation extremes? Geophysical Research Letters, 47, e2020GL088684. https://doi.org/10.1029/2020GL088684.

ECA & D Project Team and Royal Netherlands Meteorological Institute, 2013. Algorithm Theoretical Basis Document (ATBD), available at: https://www.ecad.eu/documents/atbd.pdf.

ESA. Land Cover CCI Product User Guide Version 2. Tech. Rep., 2017. Available at: maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf.

Fowler H.J., Lenderink G., Prein A., Westra S., Allan R., Ban N., Barbero R., Berg P., Blenkinshop S., Do H.X., Guerreiro S., Haerter J., Kendon E., Lewiz E., Schaer C., Sharma A., Wasko C., Villarini G., Zhang X., 2021. Understanding the intensification of short-duration rainfall extremes with global warming and the implications for flood hazard. Nat Rev Earth Environ., 2, 107-122. https://doi.org/10.1038/s43017-020-00128-6.

Gudmundsson L., Do H.X., Leonard M., Westra S., 2018. The Global Streamflow Indices and Metadata Archive (GSIM) - Part 2: Quality Control, Timeseries Indices and Homogeneity Assessment, Earth Syst. Sci. Data, 10, 787-804. https://doi.org/10.5194/essd-10-787-2018.

Gudmundsson L., Boulange J.E.S., Do H.X., Gosling S.M., Grillakis M.G., Koutroulis A.G., Leonard M., Lui J., Papadimitriou L., Pokhrel Y., Schewe J., Zhang X., Schmied H.M., Seneviratne S.I., Thiery W., Westra S., Zhao F., 2021. Globally observed trends in mean and extreme river flow attributed to man-made climate change. Science, 371(6534), 1159-1162. https://doi.org/10.1126/science.aba3996.

Harrigan S., Hannaford J., Muchan K., Marsh T.J., 2018. Designation and trend analysis of the updated UK Benchmark Network of river flow stations: the UKBN2 dataset. Hydrology Research, 49(2), 552-567.

Hodgkins G.A., Whitfield P.H., Burn D.H., Hannaford J., Renard B., Stahl K., Fleig A.K., Madsen H., Mediero L., Korhonen J., Murphy C., 2017. Climate-driven variability in the occurrence of major floods across North America and Europe. Journal of Hydrology, 552, 704-717.

IPCC, 2021. Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu and B. Zhou (eds.)]. Cambridge University Press. In Press.

Langbein W.B., Hoyt W.G., 1959. Water facts for the nation's future: Uses and benefit of hydrologic data programs. New York: The Ronald Press Company.

Le H.M., Corzo G., Medina V., Mercado V.D., Nguyen B.L., Solomatine D.P., 2019. A Comparison of Spatio-Temporal Scale Between Multiscalar Drought Indices in the South Central Region of Vietnam. Spatiotemporal Analysis of Extreme Hydrological Events, 143-169.

Lehner B., Liermann C.R., Revenga C., Vörösmarty C., Fekete B., Crouzet P., Döll P., Endejan M., Frenken K., Magome J., Nilsson C., 2011. Global reservoir and dam (grand) database. Technical Documentation, Version, 1, 1-14002E.

Lins H.F., 2012. USGS hydro-climatic data network 2009 (HCDN-2009). US Geological Survey Fact Sheet, 3047(4).

Mallakpour I., Villarini G., 2015. The changing nature of flooding across the central United States. Nature Climate Change, 5(3), 250-254.

Mast M.A., Clark M.L., Eddy-Miller C.A., Clow D.W., Turk J.T., 1984. Environmental Characteristics and Water Quality of Hydrologic Benchmark Network Stations in the West-central United States, 1963-95. US Geological Survey Report, 2000 (1173).

Marengo J.A., Espinoza J.C., 2016. Extreme seasonal droughts and floods in Amazonia: causes, trends and impacts. International Journal of Climatology, 36(3), 1033-1050.

Marsh T.J., Hannaford J., 2008. High-flow and flood trends in a network of undisturbed catchments in the UK. International Journal of Climatology, 28, 1325-1338.

Tran T.X., 2006. Hydrology and Water Resources Characteristics of Vietnam (in Vietnamese), Hanoi, Agricultural Publisher, pp.426.

Tran T.X., Hoang M.T., Tran H.T., Nguyen K.D., 2012. Water resources on Vietnam's river system (in Vietnamese), Hanoi, Science and Technology Publisher, pp.618.

Pettitt A.N., 1979. A Non-Parametric Approach to the Change-Point Problem, J. Roy. Stat. Soc. C-App., 28, 126-135.

Slack J.R., Landwehr J.M., 1992. Hydro-climatic data network: a US Geological Survey streamflow data set for the United States for the study of climatic variations, 1874-1988. US Geological Survey Open-File Report 92-129.

Truong N.C.Q., Nguyen H.Q., Kondoh A., 2018. Land use and land cover changes and their effect on the flow regime in the upstream Dong Nai River Basin, Vietnam. Water, 10(9), 1206.

Turner M., Bari M., Amirthanathan G., Ahmad Z., 2012. Australian network of hydrologic reference stations-advances in design, development and implementation. In Hydrology and Water Resources Symposium 2012 Engineers Australia, pp.1555.

von Neumann J., 1941. Distribution of the Ratio of the Mean Square Successive Difference to the Variance, Ann. Math. Stat., 12, 367-395.

Whitfield P.H., Burn D.H., Hannaford J., Higgins H., Hodgkins G.A., Marsh T., Looser U., 2012. Reference hydrologic networks I. The status and potential future directions of national reference hydrologic networks for detecting trends. Hydrological Sciences Journal, 57(8), 1562-1579.

Wasko C., Nathan R., 2019. Influence of changes in rainfall and soil moisture on trends in flooding. Journal of Hydrology, 575, 432-441.

WMO (World Meteorological Organization), 2005. WHYCOS Guidelines - Hydrological Information Systems for Integrated Water Resources Management, WMO/TD-No. 1282, Geneva.

World Bank Group, 2011. Vietnam Urbanization Review Technical Assistance Report. Available at: https://documents1.worldbank.org/curated/en/225041468177548577/pdf/669160ESW0P1130Review000Full0report.pdf (Accessed: 16 September 2021).

Wilkinson M.D., Dumontier M., Aalbersberg I.J., Appleton G., Axton M., Baak A., Blomberg N., Boiten J.W., da Silva Santos L.B., Bourne P.E., Bouwman J., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1), 1-9.

Appendix 1. Key metadata of all 68 streamflow stations assessed in this study

Q_BA_0001	AnKhe	1440	108.652343	13.958335	7.132	0.233	2	2	1	2	2
Q_BA_0002	AyunPa	2940	108.45327	13.397588	7.518	0.19	1	1	2	2	2
Q_BA_0003	CungSon	12800	108.999982	13.04301	6.238	0.147	1	2	1	2	2
Q_BA_0004	PoMoRe	312	108.350936	14.032699	5.259	0	1	1	2	2	1
Q_SD_0001	CanDang	617	105.99963	11.542277	1.934	0.067	1	1	1	1	1
Q_SD_0002	DakNong	300	107.687968	12.004217	12.883	0.237	2	1	1	2	1
Q_SD_0003	DaiNga	367.5	107.877855	11.534809	12.168	0.19	2	2	2	2	1
Q_SD_0005	PhuocHoa	5240	106.766946	11.244803	8.699	0.139	2	2	1	2	2
Q_SD_0006	TaLai	10170	107.361247	11.376967	8.702	0.637	1	1	1	1	2
Q_SD_0007	TaPao	2012	107.729436	11.134845	7.102	0.396	2	2	1	2	2
Q_SD_0008	ThanhBinh	290	108.285296	11.776028	3.893	7.424	1	1	1	1	1
Q_MK_0001	BanDon	10600	107.783103	12.897855	9.971	0.732	2	2	1	2	2
Q_MK_0002	BanYen	638	103.014142	21.274199	1.125	0.028	1	1	1	1	1
Q_MK_0003	Cau14	8640	107.928854	12.613317	10.203	0.464	2	2	1	2	2
Q_MK_0004	DakMot	752	107.77078	14.656968	7.141	0.052	1	1	1	1	1
Q_MK_0005	GiangSon	3100	108.183431	12.510001	8.779	0.73	1	1	1	1	1
Q_MK_0006	Konplong	963	108.180321	14.470389	5.328	0.01	1	1	2	2	2
Q_MK_0007	KonTum	3030	108.034111	14.34711	7.342	0.642	1	1	1	1	1
Q_MK_0008	KrongBuk	503	108.37559	12.771826	12.591	1.995	1	1	1	1	2
Q_SK_0001	AnChi	834	108.80721	14.985987	11.812	0.024	2	2	1	2	1
Q_SK_0002	AnHoa	383	108.909297	14.577917	8.182	0.023	1	1	1	1	1
Q_SK_0003	BinhLieu	505	107.387022	21.488299	4.663	0	2	2	1	2	1
Q_SK_0004	DongTam	1150	106.020672	17.904059	3.46	0	1	1	1	1	1
Q_SK_0005	DongTrang	1244	109.012254	12.280838	7.219	0	1	2	1	2	1
Q_SK_0006	GiaVong	267	106.950942	16.956599	3.681	0	2	2	1	2	1
Q_SK_0007	SonGiang	2440	108.566667	15.033333	17.792	0.063	1	1	1	1	1
Q_SK_0008	SongLuy	964	108.346016	11.195616	3.362	0	1	1	1	1	1
Q_SK_0009	ThuongNhat	208	107.684719	16.128361	3.841	0	2	2	1	2	1

Station ID	Station Name	Area	Longitude	Latitude	LU change (%)	Urban (%)	usefulMon	usefulSea	usefulYear	usefulAll	UpstreamDams
Q_HT_0002	BanCung	2620	103.831333	21.825913	6.087	0.438	2	1	2	2	2
Q_HT_0003	BaoYen	4960	104.525036	22.211149	3.991	0.353	2	2	1	2	1
Q_HT_0004	CauPha	363	105.83775	22.153239	1.446	0.44	1	1	2	2	1
Q_HT_0005	ChiemHoa	16500	105.274337	22.142407	5.126	0.017	1	1	1	1	2
Q_HT_0006	Chu	2090	106.590319	21.362772	1.356	0.034	1	1	1	1	1
Q_HT_0007	DaoDuc	8298	105.00551	22.768776	4.88	0.908	1	1	1	1	1
Q_HT_0008	DauDang	1890	105.557872	22.448166	1.701	0	1	1	2	2	1
Q_HT_0009	GhenhGa	29600	105.203372	21.881692	4.291	0.292	2	2	1	2	1
Q_HT_0010	GiaBay	2760	105.8347	21.5993	1.16	0.425	2	2	1	2	1
Q_HT_0011	HamYen	11900	105.089663	22.054125	3.672	0.65	1	1	1	1	1
Q_HT_0012	HoaBinh	51800	105.356606	20.843148	4.642	0.202	2	2	1	2	2
Q_HT_0013	KheLech	503	104.281144	22.145064	0.802	0.071	2	2	1	2	1
Q_HT_0014	LaiChau	33800	103.17194	22.077226	5.047	0.087	2	1	1	2	1
Q_HT_0015	LamSon	33.1	105.495937	20.882525	2.546	0	1	1	1	1	1
Q_HT_0016	LaoCai	41000	103.97013	22.49796	3.514	0.31	2	2	2	2	1
Q_HT_0017	MuCangChai	230	104.059433	21.855945	12.704	0	1	1	2	2	1
Q_HT_0018	NaHu	155	102.869502	22.373179	4.898	0	1	1	2	2	1
Q_HT_0019	NamGiang	6740	103.161135	22.277709	4.497	0.063	1	1	1	1	1
Q_HT_0020	NamMuc	2680	103.291854	21.88467	1.724	0.022	2	2	1	2	1
Q_HT_0021	NgoiHut	602	104.549636	21.95343	0.924	0.014	2	2	1	2	1
Q_HT_0022	QuangCu	1190	105.498638	21.518397	0.441	0.023	1	1	2	2	1
Q_HT_0023	SonTay	143600	105.511923	21.153991	3.878	0.293	1	1	1	1	2
Q_HT_0024	TaBu	49500	104.056368	21.448679	4.929	0.215	1	1	1	1	2
Q_HT_0025	ThanhSon	1590	105.152134	21.191483	1.666	0	2	2	1	2	1
Q_HT_0027	VinhYen	138	104.472919	22.377588	2.256	0.13	2	2	1	2	1
Q_HT_0028	VuQuang	36790	105.25266	21.576135	3.997	0.29	2	2	1	2	2
Q_HT_0029	YenBai	48000	104.869147	21.700615	3.192	0.326	1	1	1	1	1
Q_BK_0001	LangSon	1560	106.759609	21.841984	10.594	0.272	2	1	1	2	1
Q_MA_0001	CamThuy	17500	105.483215	20.209717	1.732	0.038	1	2	1	2	1
Q_MA_0002	CuaDat	6000	105.306721	19.881466	2.288	0.001	1	1	1	1	2
Q_MA_0004	XaLa	6430	103.924276	20.938545	1.756	0.104	1	1	1	1	1
Q_CA_0001	Dua	20800	105.0364	18.994201	2.419	0.006	1	1	1	1	2
Q_CA_0002	HoaDuyet	1880	105.598643	18.383378	7.122	0.077	1	1	1	1	1
Q_CA_0003	MuongXen	2620	104.11804	19.413919	3.367	0	1	1	1	1	1
Q_CA_0004	NghiaKhanh	4024	105.413132	19.222488	2.537	0.023	1	1	1	1	1
Q_CA_0005	QuyChau	1500	105.149316	19.556999	2.427	0	2	2	1	2	1
Q_CA_0006	SonDiem	599	105.356568	18.507271	3.39	0.015	1	1	1	1	1
Q_CA_0007	YenThuong	23000	105.441818	18.684929	2.473	0.02	1	1	1	1	2
Q_VT_0001	NongSon	3155	108.033986	15.702739	14.232	0	2	1	1	2	2
Q_VT_0002	ThanhMy	1850	107.830234	15.766938	8.448	0	2	2	1	2	1

Notes:

- For columns 8 to 11 (“useful” time series indicators): a value of 1 indicates “Yes” (i.e., no abrupt change detected using the statistical tests described in section 2.2.1 for all indices listed in Appendix 2); a value of 2 indicates “No” (i.e., at least one index shows statistical evidence toward an abrupt change in more than one homogeneity test described in section 2.2.1).
- For columns 12 (dam presence): a value of 1 indicates that the total capacity of all upstream dams is less than or equal to 10% of the average annual flow of the corresponding catchment (i.e., dams have a low influence on streamflow regime); a value of 2 indicates the opposite (i.e., dams imply a substantial influence on streamflow regime)

Appendix 2. Definition of streamflow indices calculated in this investigation (reproduced from Gudmundsson et al., 2018)

Title	Abbrev.	Units	Resol.	Definition
Mean daily streamflow	MEAN	$(\mathrm{m}^3\,\mathrm{s}^{-1})$	Y, S, M	Arithmetic mean of daily streamflow.
Standard deviation of daily streamflow	SD	$(\mathrm{m}^3\,\mathrm{s}^{-1})$	Y, S, M	Standard deviation of daily streamflow.
Coefficient of variation of daily streamflow	CV	(–)	Y, S, M	Standard deviation of daily streamflow divided by the mean daily streamflow (SD/MEAN).
Interquartile range of daily streamflow	IQR	$(\mathrm{m}^3\,\mathrm{s}^{-1})$	Y, S, M	75th–25th percentile of daily streamflow.
Minimum daily streamflow	MIN	$(\mathrm{m}^3\,\mathrm{s}^{-1})$	Y, S, M	Minimum value of daily streamflow.
Maximum daily streamflow	MAX	$(\mathrm{m}^3\,\mathrm{s}^{-1})$	Y, S, M	Maximum value of daily streamflow.
Minimum 7-day mean streamflow	MIN7	$(\mathrm{m}^3\,\mathrm{s}^{-1})$	Y, S, M	Minimum 7-day arithmetic mean streamflow. For computation, the complete daily time series are first smoothed with a backward looking moving average with a 7-day window. Subsequently, the minimum value for each yearly, seasonal or monthly period is determined.
Maximum 7-day mean streamflow	MAX7	$(\mathrm{m}^3\,\mathrm{s}^{-1})$	Y, S, M	Maximum 7-day arithmetic mean streamflow. For computation, the complete daily time series are first smoothed with a backward looking moving average with a 7-day window. Subsequently, the maximum value for each yearly, seasonal or monthly period is determined.
10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th and 90th percentiles of daily streamflow	P10, P20, P30, P40, P50, P60, P70, P80, P90	$(\mathrm{m}^3\,\mathrm{s}^{-1})$	Y, S	Percentile values of daily streamflow computed for each yearly and seasonal period, where low percentiles (e.g. 10th percentile) correspond to low flows.
Centre timing	CT	(doy)	Y	The day of the year (doy) at which 50 % of the annual flow is reached. The index is computed for calendar years, where 1 denotes 1 January.
Day of minimum streamflow	DOYMIN	(doy)	Y	The day of the year (doy) at which the minimum flow occurred, where 1 denotes 1 January. The maximum value is 365 for normal years and 366 for leap years.
Day of maximum streamflow	DOYMAX	(doy)	Y	The day of the year (doy) at which the maximum flow occurred, where 1 denotes 1 January. The maximum value is 365 for normal years and 366 for leap years.
Day of minimum 7-day mean streamflow	DOYMIN7	(doy)	Y	Day of the year (doy) at which the minimum 7-day arithmetic mean streamflow occurred, where 1 denotes 1 January. The maximum value is 365 for normal years and 366 for leap years. For computation, the daily time series is first smoothed using a backward looking moving average with a 7-day window length. Subsequently, the day of the minimum of each year is determined.
Day of maximum 7-day mean streamflow	DOYMAX7	(doy)	Y	Day of the year (doy) at which the maximum 7-day arithmetic mean streamflow occurred, where 1 denotes 1 January. The maximum value is 365 for normal years and 366 for leap years. For computation, the daily time series is first smoothed using a backward looking moving average with a 7-day window length. Subsequently, the Julian day of the maximum of each year is determined.
Gini coefficient	GINI	(–)	Y	For daily runoff values q of each year, that are sorted with index i in increasing order such that $q_i \leq q_{i+1}$ GINI is defined as $\frac{1}{n}\left(n+1-2\left(\frac{\sum_{i=1}^{n}(n+1-1)q_i}{\sum_{i=1}^{n}q_i}\right)\right)$, where n is the number data points available for that year. The Gini coefficient ranges from 0 to 1. Values of 0 indicate uniform distribution of flows throughout the time period (i.e. year), whereas values close to 1 indicate that all the flows occur on a single day.