TẠP CHÍ
KHÍ TƯỢNG THỦY VĂN

Research paper

The impact of digital elevation model data sources on identifying catchment boundary in Vietnam

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

[1] Faculty of Environment and Natural Resources, Nong Lam University – Ho Chi Minh City, Ho Chi Minh City, Vietnam; doxuanhong@hcmuaf.edu.vn
[2] Research Center for Climate Change, Nong Lam University – Ho Chi Minh City, Ho Chi Minh City, Vietnam; tu.lehoang@hcmuaf.edu.vn
[3] Faculty of Water Resources Engineering, The University of Danang - University of Science and Technology, Danang City, Vietnam; nqbinh@dut.udn.vn; pthung@dut.udn.vn
[4] National Center for Water Resources Planning and Investigation, Hanoi, Vietnam; manhhung.le510@gmail.com
[5] Department of Engineering Systems and Environment, University of Virginia, Charlottesville VA, USA

*Corresponding author: doxuanhong@hcmuaf.edu.vn; Tel.: +84–907433031
Received: 5 February 2022; Accepted: 28 March 2022; Published: 25 April 2022

Abstract: In recent years, new advances in remote sensing techniques have made Digital Elevation Models (DEMs) become popular elevation data sources for delineating catchment boundaries. This application of DEMs is particularly useful in water accounting and river basin management for Vietnam, of which the river network has very high drainage density and has been facing many pressures arising from recent economic advances. However, catchment delineated from DEMs is highly dependable to the quality of original data sources, leading to potential discrepancy in the shape as well as catchment area of the boundaries delineated from different DEMs over specific locations in Vietnam. This study comprehensively investigates this issue by analyzing the differences across catchment boundaries delineated from the most popular DEMs (i.e., HydroSHEDS, MERIT, and TanDEM–X). The impacts of these discrepancies (due to using different DEMs) on identifying areal rainfall from a gridded data product are assessed to highlight the importance of selecting DEM data sources that are suitable for specific study area.

Keywords: DEM; River basin; Water resources; Remote sensing; Catchment delineation.

1. Introduction

The rapid development of remote sensing technology in recent decades has enabled new advances in many scientific disciplines, including hydrology [1]. Among remote–sensing based data products, Digital Elevation Model (DEM) data is perhaps the most popular asset for hydrologists [2–3]. DEM data provides abundant information which is very useful for hydrological investigations as well as the development of better physically based distributed models [4–5]. Catchment boundary delineation is the most basic (and arguably most important) applications of DEM in hydrology, as the boundary will set the extent for any hydrological investigations, or model developments [6]. Catchment boundaries (derived from DEM) has also been used extensively to extract catchment information such as the areal rainfall across the catchment, or land surface attributes that are available from gridded data products [7–8].

Recent advances in remote sensing techniques have provided new options for delineating catchment boundary as several global datasets – each derived from an independent satellite mission – are now publicly accessible to support hydrological research [9]. However, catchment delineated from DEMs is highly dependable to the quality of original data sources, and the uncertainty of the delineated boundary can introduce substantial errors [10]. As a result, it is essential to review the quality of DEM data product(s) prior to using it to support any hydrological research to reduce the uncertainty introduced by DEM.

This article showcases an example of this procedure by identifying the uncertainty in catchment boundary delineated from different DEMs. Specifically, we assessed the quality of catchment boundaries derived from three popular DEMs (i.e., HydroSHEDS, MERIT, and TanDEM–X) using two delineation scenarios. In the first scenario, a reference river network is not available to support hydrologists in determining catchment boundary while the second scenario comes with a reference river network. To obtain generalized conclusion from our investigation, we implemented the assessment procedure over eleven locations across Vietnam. The propagated impacts of uncertainty in delineated boundaries to catchment attributes estimated across these eleven catchments were also assessed to highlight the importance of choosing the suitable DEM data sources for specific study area. Specifically, we investigated the discrepancies of monthly areal rainfall over the 1980–2005 period when different catchment boundaries were used.

2. Data and Methodology

2.1. Data

We obtained metadata information (i.e., geographic locations, reference river networks) for eleven stream gauges (see Appendix) from the Vietnam Meteorological and Hydrological Administration (VMHA, http://kttvqg.gov.vn/). These stations spread across Vietnam and have catchment area ranging from 138 to 4024 km^2 (see Figure 1 for a summary of station locations). Among collected stream gauges, three gauges belong to the Hong–Thai Binh river system, two gauges from the Ba river system, the rest from the Ca, Dong Nai, Gianh, Sesan, An Lao, and Cai (Nha Trang) river systems, respectively. We also collected reference river networks from the Department of Water Resources Management (DWRM, http://dwrm.gov.vn/) to support catchment boundary delineation in the scenario when a reference river network is accessible.

Three available global DEM products were used to delineate catchment boundary associated with the selected streamflow stations (Table 1), including HydroSHEDS (Hydrological Data and Maps Based on SHuttle Elevation Derivatives at Multiple Scales), MERIT (Multi–Error–Removed Improved–Terrain), and TanDEM–X (TerraSAR–X add–on for Digital Elevation Measurements). The first DEM product used in this study was obtained from the HydroSHEDS website (https://www.hydrosheds.org/downloads) [11]. This DEM product is based on elevation data obtained in 2000 by NASA's Shuttle Radar Topography Mission (SRTM) [12]. Since there is no underlying SRTM elevation data available for regions above 60 degrees northern latitude, the quality of HydroSHEDS DEM is significantly lower for those regions. HydroSHEDS provides two DEM formats including void–filled and hydrologically conditioned. In the void–filled DEMs, no–data voids are filled in and the main elevation inconsistencies have been removed. Moreover, hydrologically conditioned DEMs are available for hydrological applications and are future conditioned to produce an actual river network. The void–filled DEM was used in this study. MERIT DEM is a 90m resolutions DEM product that is freely available to the public (http://hydro.iis.u–tokyo.ac.jp/~yamadai/MERIT_DEM/) [13–14]. This DEM was made by processing the NASA SRTM3 DEM v2.1, JAXA AW3D–30m DEM v1 and Viewfinder Panoramas DEM products as baseline data [13]. Thus, it could eliminate significant errors from the baseline data [15]. In addition to the baseline DEMs, several supplementary data (e.g., NASA–NSIDC ICESat/GLAS

GLA14 data, NASA Global Forest height data) were also used to develop MERIT DEM [13]. The third DEM data product is TanDEM–X, which was developed through a collaboration between the German Aerospace Center (DLR) and Airbus Defense & Space (ADS). The dataset was built using interferometric synthetic aperture radar (InSAR; accessible at https://download.geoservice.dlr.de/TDM90). The global product of TanDEM–X (v1.0; 90m spatial resolution) was made freely available in late 2018 [16]. It covers all Earth's landmasses from pole to pole with World Geodetic System 1984 (WGS84) ellipsoid datum.

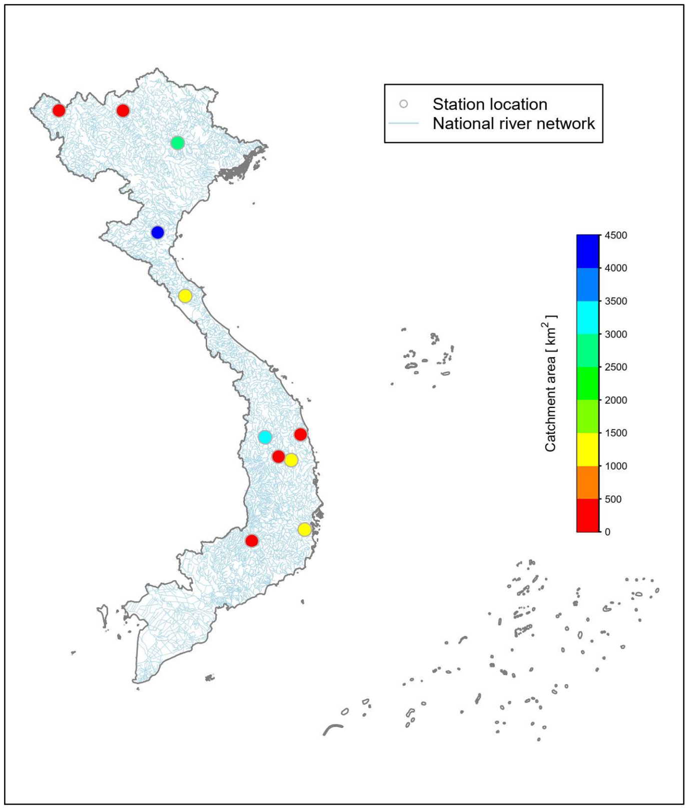

Figure 1. Location of hydrological stations selected for this investigation (colors illustrate the catchment area). Vietnam's river network was shown for reference (blue lines).

Table 1. The information of DEM products used in this study.

DEMs products	Resolution (m)	Developed Agencies (Download website)	Collection period
HydroSHEDS	90	The World Wildlife Fund (WWF) Conservation Science Program, the U.S. Geological Survey, the International Centre for Tropical Agriculture, The Nature Conservancy, and the Center for Environmental Systems Research of the University of Kassel, Germany. (https://www.hydrosheds.org/downloads)	2000
MERIT	90	The Tokyo University (http://hydro.iis.u–tokyo.ac.jp/~yamadai/MERIT_DEM/)	2000–2017
TanDEM–X	90	The German Aerospace Center (DLR) and Airbus Defense & Space (ADS) (https://download.geoservice.dlr.de/TDM90/)	2010–2015

The Vietnam Gridded Precipitation dataset (VnGP) was also used in this study to represent rainfall observation for Vietnam. Our objective was to see whether rainfall information extracted using different catchment boundaries (each shapefile was delineated from an independent DEM data product) have a large discrepancy – which will potentially lead to uncertainty in water accounting. The VnGP was developed using the Spheremap interpolation technique based on rainfall records across 481 rain gauges across Vietnam [17]. The VnGP data is currently available at the Data Integration and Analysis System (DIAS) managed by the University of Tokyo, Japan. In this study, we used the VnGP dataset at the resolution of 0.1° which provides daily gridded rainfall from 1980 to 2005.

2.2. Catchment delineation

To keep the consistency of catchment boundary delineated across all locations, we used the algorithms readily available in the ArcGIS interface of the SWAT model for this task [18]. We used the geographical coordinates (reported in Table S1) of the stations to represent the outlets of the catchments. It is important to note that the delineation algorithm of SWAT–ArcGIS has the capacity to adjust the outlet locations to match the river networks derived from DEMs.

To analyze the uncertainty in catchment boundaries, the delineation procedure was adopted under two scenarios: whether a "reference" river network is available or not. The first scenario assumes that end–users do not have any reference river network to support the delineation process, thus the quality of catchment boundary will depend solely on the quality of chosen DEM data product (denotes as the "no–burn" procedure hereafter). The catchment boundary delineated through the no–burn procedure is arguably prone to errors, as most available DEM are global data products, and likely contain local errors. Specifically, the DEM data product could yield incorrect river network over regions with complex topography (e.g., mountainous areas) and ultimately lead to incorrect catchment boundary.

The second scenario assumes that end–users would be able to consult a reference river network that is available to fix errors introduced by DEM in delineating catchment boundaries (denotes as the "burn" procedure hereafter). This burning technique was shown to improve sub–watershed boundary delineation substantially [19] as they will leverage the information of the river network from local data sources (assumed to better reflect the correct topography) to offset errors associated with DEM data products. To execute the burn procedure, DEM data was first loaded into the SWAT model, then using the "Burn–in" function and imported stream network data. The stream network is superimposed onto the DEM to define the location of the stream network. After "Burn–in" is completed, the SWAT model will pre–process the DEM by filling sinks and calculating the flow direction, flow accumulation and delineating catchment boundary. The DWRM's river network was chosen as the reference to implement the burn procedure.

2.3. Identifying uncertainty in catchment attributes attributing to DEM data sources

To illustrate the potential uncertainty in catchment boundaries derived from different DEM data sources, this study assessed three aspects: (i) the shape of the delineated catchment boundaries, (ii) the delineated catchment area, and (iii) areal rainfall across the catchments.

We first visualized the catchment boundaries derived from three DEM data sources using the two delineation procedures described in section 2.2 (leading to a total number of six boundaries for each station) to identify possible errors in the shape of the catchment. It is important to note that there is no official catchment boundary upstream of each selected stream gauge that could be used as a reference, making it impossible to evaluate the "true" errors associated with the shape of each delineated catchment. It is also unreasonable to assume that a specific DEM data source (e.g., TanDEM–X) could outperform the others (e.g., HydroSHEDS and MERIT) as the quality of each DEM data product could vary substantially across a region with complex topography such as Vietnam. We therefore only assessed whether (i) there are substantial variations regarding the shape in the catchment boundaries and (ii) the delineated catchments are consistent to the reference river network or not.

We then calculated the area of each delineated boundary and compared it to the metadata area reported in station document (obtained from the VHMA) to assess the potential errors of this important attribute. To represent errors in delineated catchment area, we calculated the area discrepancy metric (δ; in %) for each catchment as described in Equation (1):

$$\delta_{j,i} = \frac{(\beta_{j,i} - \alpha_i)}{\alpha_i} * 100 \quad (1)$$

where $\beta_{j,i}$ is the area estimated from the j delineated catchment boundary (j in [1, 6]) associated with station i (i in [1, 6]); $\alpha_{j,i}$ is the catchment area reported in the document of station i (i in [1, 6]).

We also calculated the areal rainfall from a national gridded data product and assessed the range of extracted information. The rationale of this analysis was to investigate the uncertainty of information that was extracted using catchment boundaries generated from different DEM data sources. We specifically calculated the standard deviation (SD) and the Coefficient of Variation (CV) of areal rainfall estimated using different catchment boundaries at each catchment. The SD and CV was calculated separately for each month over the 1980–2005 period (e.g., SD and CV of areal rainfall for March 2005). The range of all calculated SD and CV values was then analyzed to evaluate this uncertainty.

3. Results and Discussion

3.1. Many DEMs incorrectly identify catchment boundary at specific locations

Among the assessed catchments, several locations show a substantial discrepancy between the shape of the catchment boundaries delineated under two scenarios. In some instance, specific DEM data does not correctly reflect the local topography, leading to incorrect river network was obtained and ultimately high uncertainty in delineated catchment boundaries. Figure 1 provides an example of this issue, showing the shape of the catchment boundary derived using different DEM data sources (three columns) for station Gia Bay at the Hong – Thai Binh river system (station ID: Q_HT_0010). Under the no–burn scenario (panel a, b, c), both TanDEM and HydroSHEDS have incorrectly reflected the spatial variation of the local topography, leading to the removal of a large tributary over the southeast of the basin. MERIT DEM, on the other hand, has captured variation in local topography better, thus showing a large discrepancy in shape of the catchment relative to the results obtained from other DEM data products. Burning the reference river network (DWRM river network) into the DEMs (panel d, e, f) have addressed this issue and thus the shapes of catchment boundary obtained under the burn scenario do not show any apparent variation.

To further investigate the uncertainty in catchment boundary derived from different DEM data sources and delineation procedures, we also analyzed the discrepancy (in percentage) between the delineated catchment area relative to that reported in the metadata. Figure 3 shows the results of this assessment, indicating the errors might be relatively high in some instances. The scatter plot between the delineated catchment and reported catchment area (Figure 3a) suggests that the error is generally independent to catchment area. The presence of local errors in DEM information, therefore, are likely more important when deciding which DEM data source is appropriate for a specific investigation. This comparison also shows the value of burning the reference river network into the DEM, as all burn boundary has substantially smaller errors relatively to the no–burn counterparts (Figure 3b). It is also apparent from Figure 3b that there is one instance where all DEMs have severely flaws, leading to a discrepancy of more than 50% of the catchment area when the no–burn delineation technique was used (this issue was solved under the burn scenario). This result further highlights the importance of local perspectives in any applications of remotely–sensed datasets and also proved the usefulness of the river network developed by the national authority in Vietnam.

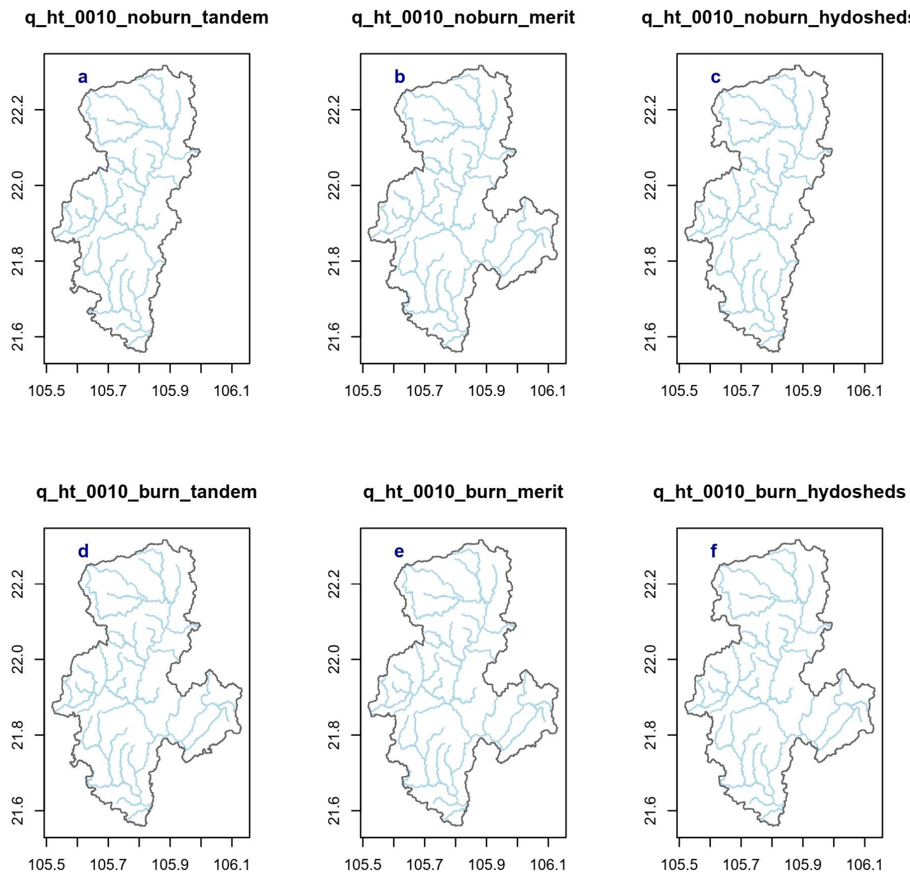

Figure 2. Uncertainty in catchment shapefile derived from different DEM sources at Gia Bay station (station ID: Q_HT_0010) on the Hong–Thai Binh river system: (a), (b), (c) the catchment boundary was delineated without a reference river network; (d), (e), (f) the catchment boundary was delineated with a reference river network. River network obtained from DWRM was shown for reference.

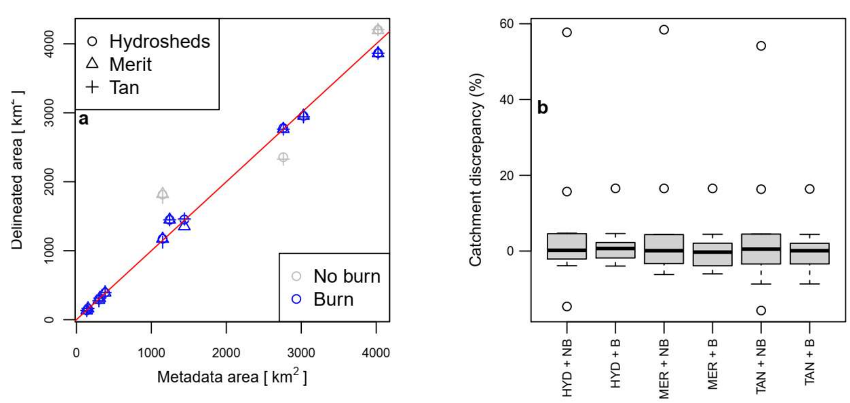

Figure 3. Discrepancy in estimated catchment area (relative to area documented in station reports) using different DEM and delineation technique: (a) the relationship between catchment area documented and area estimated from the delineated boundary; (b) the full range of catchment discrepancy (in %) between delineated catchment area and reported catchment area (HYD: HydroSHEDS, MER: Merit, TAN: TanDEM–X, B: burn, NB: no–burn).

3.2. The influence of delineated boundaries on information extracted from gridded products

This section reports the range of variation in monthly rainfall extracted from a common gridded rainfall dataset using different DEM data sources and boundary delineation techniques. We note that the standard deviation (SD) and the coefficient of variation (CV; the ratio between the standard deviation and the mean in the form of a percentage) was first computed separately for each catchment at each timestep (i.e., each month over the 1980–2005 period). The calculated SD and CV were then classified into twelve groups (based on calendar months) to take into account the seasonal variation of monthly rainfall.

Among the 11 stations, the choice of DEMs and delineation scenarios did not create any changes in the extracted information over four (i.e., the calculated SD and CV are equal to zero across all months) catchments. These four catchments all have relatively small size (catchment area is less than 400 km^2), making it not possible to observe any differences in the areal rainfall obtained from a gridded product that has a relatively coarse resolution (0.1×0.1 arc–degree longitude–latitude resolution, corresponding to grid cell area of about 100 km^2). However, the SD (CV) can be as high as 46.4 mm (28.3%) in some instance, demonstrating that the discrepancy in catchment boundary delineated using different DEMs and delineation techniques could lead to a large uncertainty in information extracted from gridded data products.

We note that the magnitude of variation in extracted rainfall is generally higher from May to November relative to the other months (Figure 4a). This result is expected as these months represent the rainy season that is characterized with convective storms. The magnitude of monthly rainfall during this season, therefore, is generally higher relative to the rest of the year, making any discrepancies easier to be detected.

The relative variation (i.e., the CV) of rainfall extracted using different DEMs, on the other hand, has an opposing pattern: higher over the dry season relative to the wet season (Figure 4b). During the dry season, rainfall events usually occur over small area due to the absence of largescale convection activities. As a result, rainfall during the dry season generally has a stronger spatial pattern. This result indicates that the influence of uncertainty in delineated catchment boundary could be substantial when extracting data associated with a high spatial gradient.

4. Conclusion

This study investigated the potential influence of DEMs on identifying catchment boundary. Using a relatively large sample of Vietnamese catchments, we showed that the uncertainty underlying catchment boundary due to using different DEMs is not trivia. Specifically, using different DEMs and delineation techniques could lead to substantial errors in the shape and area of delineated catchment (a discrepancy of more than 50% relative to the documented catchment area in the most severe case). We also showed the value of the national river network developed by Vietnam water agencies, as burning this "reference" river system into the DEMs during the delineation process has substantially reduced the discrepancy in both the shape of the boundary and the delineated catchment area. The study also found that uncertainty in delineated catchment boundaries could propagate to a substantial discrepancy of information extracted from rasterized data products. Using catchment boundaries delineated from different DEMs and delineation techniques have introduced some variations in the monthly rainfall extracted from the VnGP gridded rainfall dataset, with the standard deviation and coefficient of variations could be as high as 46.4 mm and 28.3% respectively.

Our study suggests the importance of carefully investigating the quality of DEM data products prior to any hydrological applications that need DEM–based topography information. The simple procedure and datasets described in this article could be used in future studies to identify the product that is most appropriate for specific locations within Vietnam.

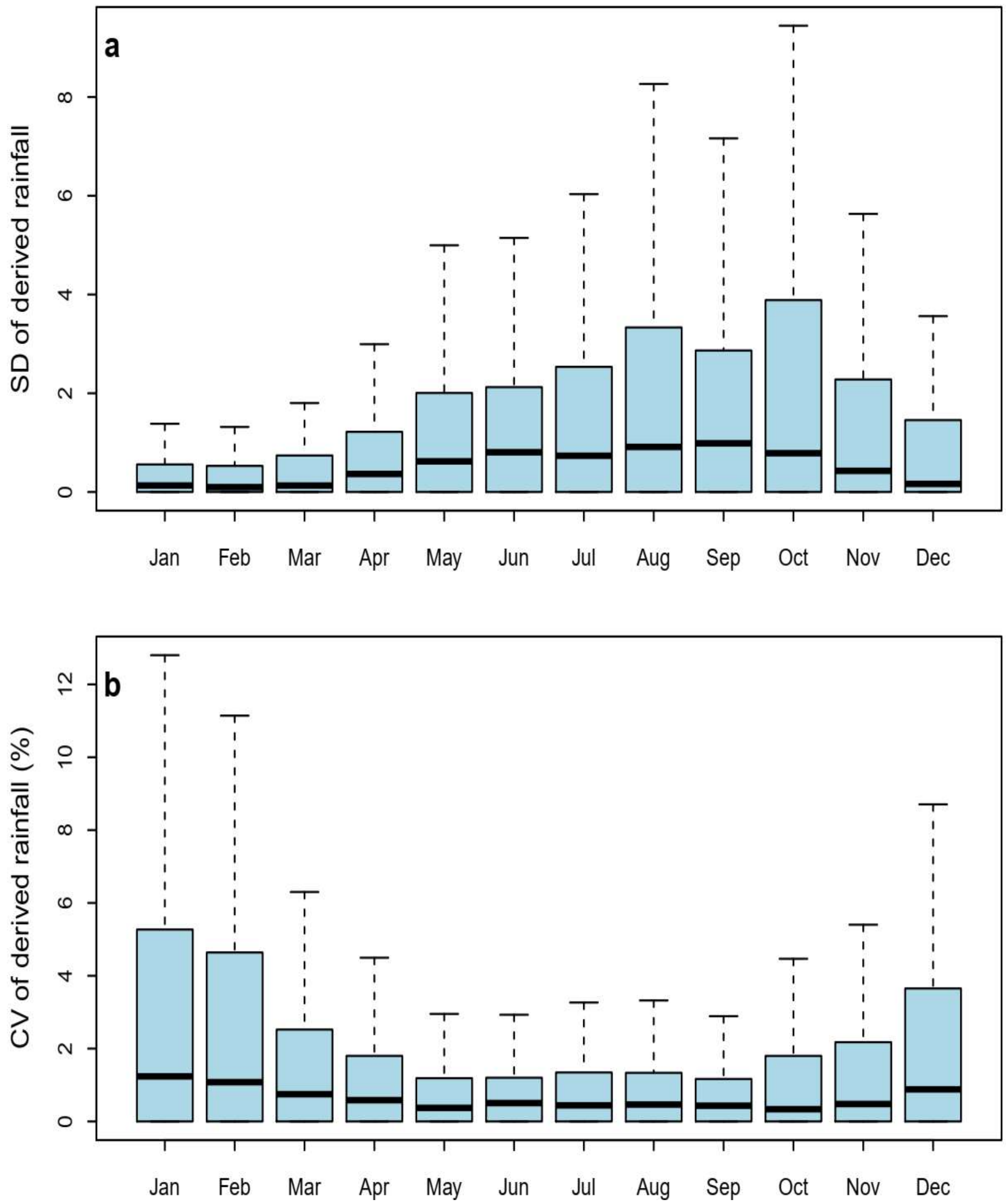

Figure 4. The distribution of (a) the standard deviation and (b) the coefficient of variation (i.e., the ratio between the standard deviation and the mean) of the areal rainfall derived from a common gridded rainfall dataset and different DEM data sources. Note that the metrics were first calculated for individual month at each catchment then grouped into twelve calendar–month groups. A common national data product (daily rainfall over the 1980–2005 period available through the VnGP gridded product) was used in this analysis.

Author contribution statement: Conceived and designed the experiments: D.X.H. and all co–authors; Contributed data: P.T.H., L.M.H., L.H.T., N.Q.B.; Analyzed and interpreted the data: D.X.H. and all co–authors; Prepare the draft manuscript: D.X.H.; Manuscript editing and revising: D.X.H. and all co–authors.

Acknowledgement: D.X.H. and L.H.T. are supported by a grant for basic research (No. CS–CB21–MTTN–07) from Nong Lam University – Ho Chi Minh City to conduct this project.

Competing interest statement: The authors declare no conflict of interest.

Appendix

Table S1. Description of stream gauges used in this study.

No.	Station ID	Station Name	River system	Longitude (°)	Latitude (°)	Area (km²)
1	Q_HT_0010	Gia Bay	Hong – Thai Binh	105.83	21.59	2,760
2	Q_HT_0018	Na Hu	Hong – Thai Binh	102.87	22.37	155
3	Q_HT_0027	Vinh Yen	Hong – Thai Binh	104.47	22.36	138
4	Q_CA_0004	Nghia Khanh	Ca	105.33	19.43	4,024
5	Q_BA_0001	An Khe	Ba	108.67	13.95	1,440
6	Q_BA_0004	PoMoRe	Ba	108.35	14.03	312
7	Q_SD_0002	Dac Nong	Dong Nai	107.68	12.00	300
8	Q_MK_0007	Kon Tum	Sesan	108.02	14.50	3,030
9	Q_SK_0002	An Hoa	An Lao	108.90	14.57	383
10	Q_SK_0005	Dong Trang	Cai Nha Trang	109.00	12.28	1,244
11	Q_RN_0009	Dong Tam	Gianh	106.02	17.90	1,150

References

1. Campbell, J.B.; Wynne, R.H. Introduction to remote sensing. Guilford Press, 2011.
2. Zhang, H.; Li, Z.; Saifullah, M.; Li, Q.; Li, X. Impact of DEM resolution and spatial scale: analysis of influence factors and parameters on physically based distributed model. *Adv. Met.* **2016**, 8582041. https://doi.org/10.1155/2016/8582041.
3. Pham, H.T.; Marshall, L.; Johnson, F.; Sharma, A. A method for combining SRTM DEM and ASTER GDEM2 to improve topography estimation in regions without reference data. *Remote Sens. Environ*. **2018**, *210*, 229–241.
4. Smith, M.B.; Seo, D.J.; Koren, V.I. The distributed model intercomparison project (DMIP): motivation and experiment design. *J. Hydro*. **2004**, *298(1–4)*, 4–26.
5. Beven, K. Changing ideas in hydrology–the case of physically–based models. *J. Hydro.* **1989**, *105(1–2)*, 157–172.
6. Freeman, T.G. Calculating catchment area with divergent flow based on a regular grid. *Comput. Geosci*., **1991**, *17(3)*, 413–422.
7. Le, H.M.; Corzo, G.; Medina, V.; Mercado, V.D.; Nguyen, B.L.; Solomatine, D.P. A Comparison of Spatio–Temporal Scale Between Multiscalar Drought Indices in the South Central Region of Vietnam. Spatiotemporal Analysis of Extreme Hydrological Events, **2019**, 143–169.
8. Do, H.X.; Gudmundsson, L.; Leonard, M.; Westra, S. The Global Streamflow Indices and Metadata Archive (GSIM) – Part 1: The production of daily streamflow archive and metadata, *Earth Syst. Sci. Data* **2018**, *10*, 765–785. https://doi.org/10.5194/essd-10-765-2018.
9. Garbrecht, J.; Martz, L.W. Digital elevation model issues in water resources modeling. In Hydrologic and hydraulic modeling support with geographic information systems, ESRI Press, Redlands, **2000**, 1–28.
10. Li, J.; Wong, D.W. Effects of DEM sources on hydrologic applications. *Comput. Environ. Urban Syst*. **2010**, *34(3)*, 251–261.
11. Lehner, B.; Verdin, K.; Jarvis, A. HydroSHEDS Technical Documentation, version 1.0.

World Wildlife Fund US, Washington, DC, 2006, 1–27.

12. Farr, T.G.; Kobrick, M. Shuttle radar Topography mission produces a wealth of data eos. *Trans. Am. Geophys. Union.* **2000**, *81*, 583–585.
13. Yamazaki, D.; Ikeshima, Tawatari, D.; Yamaguchi, R.; O'Loughlin, T.; Neal, F.; Sampson, J.C.; Kanae, C.C.; Bates, S.P.D. A high accuracy map of global terrain elevations. *Geophys. Res. Lett.* **2017**, *44*, 5844–5853. doi: 10.1002/2017GL072874.
14. Yamazaki, D.; Ikeshima, D.; Sosa, J.; Bates, P.D.; Allen, G.H.; Pavelsky, T.M. MERIT Hydro: A High–Resolution Global Hydrography Map Based on Latest Topography Dataset. *Water Resour. Res*. **2019**, *55(6)*, 5053–5073. https://doi.org/10.1029/2019WR024873.
15. Xu, F.; Dong, G.; Wang, Q.; Liu, L.; Yu, W.; Men, C.; Liu, R. Impacts of DEM uncertainties on critical source areas identification for non–point source pollution control based on SWAT model. *J. Hydrol*. **2016**, *540*, 355–367. https://doi.org/10.1016/j.jhydrol.2016.06.019.
16. Hawker, L.; Neal, J.; Bates, P. Accuracy assessment of the TanDEM–X 90 Digital Elevation Model for selected floodplain sites. *Remote Sens. Environ.,* **2019**, *232*. https://doi.org/10.1016/j.rse.2019.111319
17. Nguyen–Xuan, T.; Ngo Duc, T.; Kamimera, H.; Trinh–Tuan, L.; Matsumoto, J.; Inoue, T.; Phan–Van, T. The Vietnam Gridded Precipitation (VnGP) dataset: Construction and validation. *SOLA* **2016**, *12*, 291−296. https://doi.org/10.2151/sola.2016-057.
18. Winchell, M.; Srinivasan, R.; Di Luzio, M.; Arnold, J. ARCSWAT 2.3. 4 interface for SWAT2005: User's guide. Blackland Research Center: Texas Agricultural Experiment Station, Temple, **2009**, pp. 465.
19. William, K.; Saunders, B.; David, R. A GIS assessment of nonpoint source Pollution in the San Antonio–Nueces Coastal Basin. Center for Research in Water Resources, The University of Texas at Austin, 1996.