Water Resources Research

RESEARCH ARTICLE
10.1029/2022WR033323

Special Section:
Advancing process representation in hydrologic models: Integrating new concepts, knowledge, and data

Interacting Effects of Precipitation and Potential Evapotranspiration Biases on Hydrological Modeling

Jiao Wang[1], Lu Zhuo[1,2], Miguel Angel Rico-Ramirez[1], Ahmed Abdelhalim[1], and Dawei Han[1]

[1]Department of Civil Engineering, University of Bristol, Bristol, UK, [2]School of Earth and Environmental Sciences, Cardiff University, Cardiff, UK

Abstract The quality of precipitation (*P*) and potential evapotranspiration (*PET*) data greatly affects the hydrological modeling performance. Considerable attention has been paid to identifying the influence of biased *P* or *PET* inputs independently. However, few studies have explored the joint interaction of biases in *P* and *PET* inputs on hydrological simulations. Here, we investigate the mutual compensation of *P* and *PET* biases on the performance of two widely used conceptual hydrological models, the Xinanjiang model and the Probability Distributed Model. *P* and *PET* from HYREX (HYdrological Radar EXperiment) and CAMELS-GB (Catchment Attributes and Meteorology for Large-sample Studies in Great Britain) data sets are collected over five catchments with varying characteristics in Great Britain. Different biases are added to these original time series to generate 6560 biased input scenarios. The results suggest that there is a certain compensational relationship between the biases in *P* and *PET* inputs to reproduce desirable streamflow simulations. A new hydrological proxy named Compensational Interaction Angle (CIA) is identified and found to be stationary with various modeling periods, as well as stable with different hydrological models despite model equifinality. Further, the CIA highly relates to the long-term climate aridity ratio. The catchments with greater aridity have larger CIAs. This study offers a fresh perspective to analyze the input errors in hydrological modeling. The results can help to better understand *P* and *PET* interactions in hydrological modeling, and guide the selection/evaluation/bias-correction of *P* and *PET* data sets for hydrological applications.

Plain Language Summary Hydrological modeling helps us understand, predict and manage water resources in the real world. Precipitation (*P*) and potential evapotranspiration (*PET*) are two key inputs. Hence their accuracy has a great impact on the modeling outputs. However, the measurements and estimations of *P* and *PET* are prone to different sources of errors. We explore the joint interaction of their accuracy on hydrological modeling by considering there are mutual compensational effects between them. It is found that there is a stable compensational relationship between their biases in producing desirable hydrological performance. Moreover, the compensation relationship is highly linked to the climate aridity condition of the region. This study provides a new perspective for hydrologists to explore the sensitivity of input errors in hydrological modeling. We strongly encourage the community to apply the proposed method in other regions to further explore and identify the compensational relationship between *P* and *PET* by using different hydrological models. This has the potential to enhance our understanding of the interactions between *P* and *PET* in hydrological modeling.

Key Points:

- Biases in precipitation and potential evapotranspiration data have compensating effects on hydrological modeling performance
- A Compensational Interaction Angle (CIA) is proposed to quantify the compensational relationship in reproducing good streamflow
- The CIA shows stationarity and stability in different hydrological models; the catchments with greater aridity have larger CIAs

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

J. Wang,
jiao.wang@bristol.ac.uk;
jiao.wang_hydro@outlook.com

Citation:

Wang, J., Zhuo, L., Rico-Ramirez, M. A., Abdelhalim, A., & Han, D. (2023). Interacting effects of precipitation and potential evapotranspiration biases on hydrological modeling. *Water Resources Research*, *59*, e2022WR033323. https://doi.org/10.1029/2022WR033323

Received 27 JUL 2022
Accepted 28 FEB 2023

Author Contributions:

Conceptualization: Jiao Wang, Dawei Han
Formal analysis: Jiao Wang, Ahmed Abdelhalim
Funding acquisition: Jiao Wang
Investigation: Jiao Wang, Dawei Han
Methodology: Jiao Wang, Lu Zhuo, Miguel Angel Rico-Ramirez

1. Introduction

A hydrological model normally aims to simulate streamflow (*Q*, runoff or discharge) in a river basin (Ellenburg et al., 2018). Although complexities of conceptualizations and structures vary for hydrological models (Beven, 2012), precipitation (*P*, or rainfall) and potential evapotranspiration (*PET*) are two essential inputs. This is due to the fact that *P* and actual evapotranspiration (*ET*, which depends on *PET*) constitute two major components of the water balance and terrestrial water cycle for a basin (Nonki et al., 2021; Samain & Pauwels, 2013). Generally, *P* is relatively easy to measure while measuring *ET* is challenging, time-consuming and costly (Doorenbos & Pruitt, 1977; Nonki et al., 2021). As a theoretical estimation of *ET*, *PET* is proposed to represent the upper limit of *ET* under the scenario of the unlimited supply of soil water (Allen et al., 1998).

Ideally, we use the "true" value of *P* and *PET* data for driving the hydrological model to reproduce reliable *Q*. However, even for ground-based observations, errors inevitably occur in the input data due to various reasons, such as measurement issues, interpolation methods, wind, obstruction, incorrect equation assumptions, etc

Supervision: Lu Zhuo, Miguel Angel Rico-Ramirez, Dawei Han
Validation: Jiao Wang
Visualization: Jiao Wang
Writing – original draft: Jiao Wang
Writing – review & editing: Jiao Wang, Lu Zhuo, Miguel Angel Rico-Ramirez, Ahmed Abdelhalim, Dawei Han

(Sevruk, 1996; Xu et al., 2006). Consequently, these input errors can introduce significant errors in model calibration and *Q* outputs (Nandakumar & Mein, 1997).

Complicated combinations of different error types would make it hard to interpret the corresponding changes (Oudin et al., 2006), thus most studies have focused on two simple types of errors, namely, systematic and random errors. Paturel et al. (1995) found that errors (both systematic and random) in *P* were amplified on *Q* outputs by the hydrological model, and this amplification behavior varied with the phase of the hydrograph. Xu et al. (2012) also reported that a bias of ±1% in *P* resulted in ±3.35% error in *Q* on average in 193 study catchments across Australia. Xu and Vandewiele (1994) compared the random errors with systematic errors in *P* data and found that systematic errors were less important for estimating monthly *Q*. On the contrary, more studies revealed that the systematic errors in *P* data had the most severe effect on *Q* simulations (Nandakumar & Mein, 1997; Oudin et al., 2006; Xu et al., 2006). This is because a systematic error in *P* causes a bias in the water balance, thereby leading to a systematic error when calibrating the model's parameters (Xu et al., 2006). Wang et al. (2023) also found that biases in rainfall inputs are of great importance for the reliability of *Q* simulations. By comparing two hydrological models, Oudin et al. (2006) suggested that the sensitivity of a hydrological model to errors in rainfall might depend partly on the model structure itself.

Likewise, the influence of different errors in *PET* inputs on hydrological modeling has also been explored. It has been found that poor *PET* inputs could result in poor hydrological model performance (Samain & Pauwels, 2013), while many studies concluded that hydrological models were less sensitive to errors in *PET* than errors in *P* data (Nandakumar & Mein, 1997; Oudin et al., 2005; Paturel et al., 1995). As reported by Nandakumar and Mein (1997), a 10% bias in *P* might cause a bias of up to 35% in simulated *Q*, while the same amount of bias in *PET* might result in up to 10% bias in *Q* in five experimental catchments in Australia. Also, most studies found that systematic errors in *PET* presented a more serious impact than random errors (Oudin et al., 2005, 2006; Parmele, 1972). Jayathilake and Smith (2022) found that the hydrological model was more sensitive to negative *PET* biases than positive ones. Nandakumar and Mein (1997) used a *PET* factor to compensate for the under/over estimation of *PET* data. Similarly, Oudin et al. (2005) introduced a scaling factor to eliminate the systematic biases (systematic difference) in *PET* values. Samain and Pauwels (2013) also suggested that model recalibration was needed when the *PET* input data has systematic errors without rescaling.

Previous studies also indicated that the hydrological model calibration could compensate for inaccurate *P* input data to a certain degree (Beck et al., 2017; Essou et al., 2016; Wang et al., 2023; Xu et al., 2006), as well as for inaccurate *PET* inputs (Andréassian et al., 2004; Bai et al., 2016; Nonki et al., 2021). This can result in good hydrological model performance even though there are biases in *P* and *PET* inputs. However, it is worth pointing out that inaccurate input data would conceivably lead to nonrepresentative parameter optimization to match the recorded *Q* time series (Kabir et al., 2022; Parmele, 1972).

In addition, Hagemann and Jacob (2007) found that the overestimations in modeled *P* and *ET* fluxes from a climate model (known as HadAM3H) compensated for each other, making the simulated runoff in agreement with the observed *Q*. Nandakumar and Mein (1997) also pointed out that errors in *P* and *PET* may mutually offset each other to reproduce satisfactory *Q* simulations. However, previous studies assumed that errors in *P* and *PET* were independent and their effects on hydrological modeling have been well documented. So far, there has been a lack of studies focused on exploring the mutual compensation between *P* and *PET* inputs for hydrological modeling. It remains unknown how, and to what extent the joint interaction of errors in *P* and *PET* affects the hydrological model performance.

The objective of this study is to explore the joint interaction between *P* and *PET* errors in response of hydrological model simulations. In particular, we paid more attention to uncovering the potential compensational relationship between *P* and *PET* errors to reproduce *Q*. This analysis is focused on the biases (systematic errors) in *P* and *PET* inputs since they bring more significant effects than random errors on hydrological model simulations based on previous studies. The specific research questions are: (a) Are there any interacting effects between *P* and *PET*'s biases in the simulated *Q* response? (b) Is there any relationship between the biases in *P* and *PET* data that can still reproduce *Q* well? (c) How does this compensational relationship change with different situations (the length of modeling period, model equifinality, hydrological model, and catchment aridity condition)?

The rest of the paper is organized as follows. Section 2 details the data and models used in the analysis. Section 3 presents the results and discussion; Section 4 provides a summary and concluding remarks.

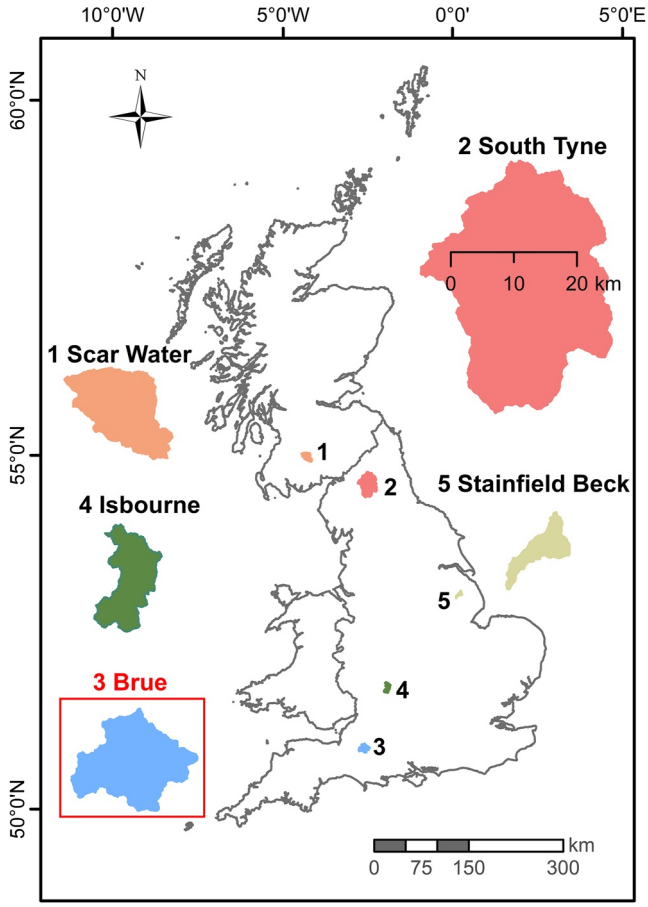

Figure 1. Location map of the catchments used in this study.

2. Data and Methods

2.1. Study Area

To address the above questions, five benchmark (near-natural) catchments including three humid catchments and two arid catchments in Great Britain were selected for the analysis (Figure 1). The Brue catchment has the availability of high-quality observation data. It was first used to investigate the existence, stationarity and stability of the compensational relationship between *P* and *PET* inputs' quality for hydrological modeling. The Brue catchment has been widely used in past hydrological studies (Borga, 2002; Borga et al., 2006; Liu & Han, 2010; Moore et al., 2005; Srivastava et al., 2014). The catchment drains an area of 135 km^2 and had a dense network of 49 tipping-bucket rain gauges (TBRs) operating between 1993 and 1997 (Wood et al., 2000). The TBRs have a resolution of 0.2 mm and the time of the tips is recorded to the nearest 10 s (Wood et al., 2000). The elevation of the catchment ranges between 23 and 253 m, with an average of 104 m (Wood et al., 2000). This catchment is predominantly rural and is dominantly covered by grass (79%) (Roberts et al., 2000). The remaining four catchments were used to investigate how the compensational relationship would change in different regions. The catchments are: Scar Water at Capenoch (No 79004), South Tyne at Haydon Bridge (No 23004), Isbourne at Hinton on the Green (No 54036) and Stainfield Beck at Cream Poke Farm (No 30012). As shown in Table 1, which is derived from the daily CAMELS-GB data set (Coxon et al., 2020a, 2020b), the five catchments vary in climatic and hydrological characteristics, especially the aridity and runoff ratio. The Scar Water and South Tyne catchments have lower aridity ratios than the Brue catchment, whereas the Isbourne and Stainfield Beck catchments have higher aridity ratios than the Brue catchment. The Baseflow index (BFI) represents the contribution of groundwater/stored sources to stream flow and it is given by the ratio between the long-term baseflow and the total stream flow (Bloomfield et al., 2009). The BFI is between 0.43 and 0.59 for those catchments, indicating a small groundwater/stored sources contribution to the river discharge for all five catchments.

Table 1
Basic Information of the Catchments Derived From the Daily CAMELS-GB Data (Coxon et al., 2020a, 2020b)

Catchment ID	Catchment name	Area (km^2)	$\overline{P}$ (mm/day)	$\overline{PET}$ (mm/day)	$\overline{Q}$ (mm/day)	Aridity ratio $\overline{PET}/\overline{P}$	Runoff ratio $\overline{Q}/\overline{P}$	Baseflow index (BFI)	$\overline{T}$ (°C)	Dominate land cover	Percentage of sand, silt, and clay (%)
79004	Scar Water at Capenoch	142.68	4.73	1.25	3.56	0.26	0.75	0.43	7.42	Grass and Pasture	46.84, 26.25, 26.91
23004	South Tyne at Haydon Bridge	749.89	3.18	1.26	2.11	0.4	0.66	0.45	7.29	Grass and Pasture	42.88, 33.73, 23.39
52010	Brue at Lovington	137.81	2.45	1.44	1.21	0.59	0.5	0.54	10.08	Grass and Pasture	37.11, 29.96, 32.93
54036	Isbourne at Hinton on the Green	92.83	1.96	1.39	0.62	0.71	0.32	0.59	9.68	Grass and Pasture	28, 30.59, 41.41
30012	Stainfield Beck at Cream Poke Farm	38.22	1.77	1.42	0.56	0.8	0.32	0.51	9.53	Crops	45.18, 29.59, 25.23

Note. $\overline{P}$, $\overline{PET}$, $\overline{Q}$, and $\overline{T}$ represent annual mean values of precipitation, potential evapotranspiration, streamflow, and temperature respectively averaged over several years.

2.2. HYREX Data

The in-situ data for the Brue catchment were collected from HYREX (HYdrological Radar EXperiment) (Wallingford, 2007). HYREX was a project founded by Natural Environment Research Council and ran from May 1993 to April 1997 in the Brue catchment (data collection was extended to 2000). Hydrological data (P and Q) were collected at 15-min intervals. For this study, the 15-min data were converted to hourly data. The hourly areal P data were calculated by averaging the retrievals of the 49 tipping bucket rain gauges available within the catchment using the Thiessen polygon technique. The River Brue is gauged at the catchment outlet at Lovington. The river level is recorded every 15 min and is converted to flow using a well-established rating curve. The 15-min data were converted to hourly Q data, which were used as the observed Q to calibrate the hydrological models.

The estimates of *PET* are computed from meteorological data. Many formulations have been developed to estimate *PET* (Allen et al., 1994; Ekström et al., 2007; Penman, 1948). The most widely used is the Food and Agricultural Organization (FAO) Penman-Monteith method (Allen et al., 1998), also called grass reference *ET*. The FAO *PET* computation is usually considered to be the most appropriate and physically satisfactory by many hydrologists and meteorologists worldwide (Andréassian et al., 2004; Nonki et al., 2021; Samain & Pauwels, 2013; Shuttleworth, 1993). As such, hourly *PET* values in this study were calculated based on the actual data from meteorological stations installed during HYREX. An automatic weather station and an automatic soil water station were located in the catchment and recorded the net solar radiation, wind speed, wet and dry bulb temperatures, barometric pressure and other atmospheric parameters (Wallingford, 2007).

Due to data discontinuity, we selected a 2-year period from the HYREX data, from June 1994 to May 1996 which includes two high-flow and two low-flow periods for the analysis. These 24 months of hourly data contain various rainfall events with a wide range of flow conditions. They provide sufficient information for calibrating the hydrological models.

2.3. CAMELS-GB Data

The CAMELS-GB (Catchment Attributes and Meteorology for Large-sample Studies in Great Britain) data set contains daily time series of various hydro-meteorological variables encompassing *P, PET*, and Q for 671 catchments across Great Britain (Coxon et al., 2020). However, for this analysis, we needed hourly data (P, *PET* and Q) and thus hourly data were compiled from a range of sources as follows. The hourly P data were derived from the hourly Gridded Estimates of Areal Rainfall (CEH-GEAR1hr) product, developed by the Centre for Ecology & Hydrology (CEH) (Lewis et al., 2019). This data set contains 1 km gridded estimates of hourly P for Great Britain from January 1990 to December 2014. The estimates were derived by applying the nearest neighbor interpolation method to a national database of hourly rain gauge observations (Lewis et al., 2018). Catchment-averaged P time series are then produced by averaging values of all grid squares that lay within the catchment boundary for each catchment. Hourly *PET* data were disaggregated from the daily *PET* data contained within the CAMELS-GB data set. These daily *PET* time series were derived from the Climate Hydrology and Ecology research Support System Potential Evapotranspiration (CHESS-PE) data set, which is a 1 km^2 gridded product of *PET* time series for Great Britain (Robinson et al., 2016). To disaggregate to hourly, the daily *PET* is distributed across daylight hours using a simple sine distribution accounting for the latitude and longitude of the catchment and the length of day throughout the year. Observed hourly Q time series for the period from 1990 to 2014 were retrieved from the Environment Agency (EA), Natural Resources Wales (NRW) and the Scottish Environment Protection Agency (SEPA).

There is a need to validate the robustness of the potential compensational relationship that we aim to investigate. Therefore, we selected a 10-year period (from June 2003 to May 2013) from the CAMELS-GB database (P, *PET* and Q) over the five catchments to further validate the stationarity of the compensational relationship and to explore how the relationship varies in different catchments.

2.4. Rainfall-Runoff Models

We selected the Xinanjiang (XAJ) model for simulating Q due to its efficiency and minimal preparation requirement for input data. The XAJ model has been widely applied to various catchments and has shown good performance worldwide (Sheng et al., 2020; Yang et al., 2020; Zhao, 1992; Zhao et al., 1980; Zhuo, Han, et al., 2015).

The XAJ model consists of four modules: evapotranspiration module, runoff production module, runoff separation module and runoff concentration module. A detailed description of the model is outside the scope of this paper, but can be found in previous studies (Zhao, 1992; Zhao et al., 1980). The ranges for parameter values were determined based on preliminary experiments and related studies (Ye et al., 2014; Zhuo & Han, 2016). The XAJ parameters were calibrated by using a standard gradient-based automatic optimization method (Lagarias et al., 1998).

Further, the Probability Distributed Model (PDM) was employed to explore if the compensational relationship changes with different models. The PDM is a conceptual rainfall-runoff model with analytical and computational simplicity (Beven, 2012; Moore, 2007). Its soil moisture storage capacity is characterized by a simple probability distribution curve (Liu & Han, 2013). This model has been widely applied in the UK and other countries. In particular, many studies on the Brue catchment used the PDM model for hydrological applications (Borga, 2002; Borga et al., 2006; Liu & Han, 2010; Moore et al., 2005; Srivastava et al., 2014). A detailed description of the model can be found in previous studies (Beven, 2012; Liu & Han, 2013; Moore, 2007; Samain & Pauwels, 2013). Both models (XAJ and PDM) use P and PET data as the main inputs.

As concluded by previous studies (Samain & Pauwels, 2013; Wang et al., 2023; Xu et al., 2006), the model recalibrations could partly compensate for the inputs' errors by altering the model parameters. This would cause a misrepresentation of input data accuracy. Thus, the XAJ and PDM models were calibrated by using the original (unbiased) P and PET input data in each experiment. No recalibration was carried out in the biased input scenario runs to avoid the compensational effect from model recalibration.

2.5. Evaluation Criteria

The Nash-Sutcliffe Efficiency (NSE) (Nash & Sutcliffe, 1970) was used as the objective function for model calibration. NSE is a widely used indicator in hydrology (e.g., Beck et al., 2017; Liu & Han, 2010; Wang et al., 2023; Zhuo, Dai, & Han, 2015). In particular, NSE is highly sensitive to peak flows (Krause et al., 2005), as well as long-term biases (Gupta et al., 2009). This makes NSE suitable since we were focused on exploring the joint interaction from the inputs' biases. The NSE is given by:

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{T} (S_i - O_i)^2}{\sum_{i=1}^{T} \left(O_i - \overline{O}\right)^2} \tag{1}$$

where i refers to the time step in hours, T is the total number of hours, S_i and O_i are the simulated and observed Q, and $\overline{O}$ is the mean value of O_i. The closer the NSE is to 1, the better the simulation is.

The NSE can be decomposed into three distinctive components, correlation (r), relative variability (α), and normalized bias (β) according to Gupta et al. (2009). The decomposed components are beneficial to identify which aspects of NSE impact the simulation performance (Lane et al., 2019). They are also helpful in providing insight into the compensational effects between the quality of P and PET inputs in simulating Q. In this study, β was calculated based on the Equations proposed by Gupta et al. (2009), and r and α were computed by using the Equations shown in Lane et al. (2019).

The Percent Bias (PBIAS, %) was used to further evaluate the systematic bias of Q simulations. PBIAS measures the average tendency of the simulated Q to be greater or smaller than the observation. A value of zero suggests a perfect case, while positive or negative values indicate an overestimation or underestimation of the Q simulation respectively. PBIAS is given by:

$$\mathrm{PBIAS} = 100 \times \frac{\sum_{i=1}^{T} (S_i - O_i)}{\sum_{i=1}^{T} O_i} \tag{2}$$

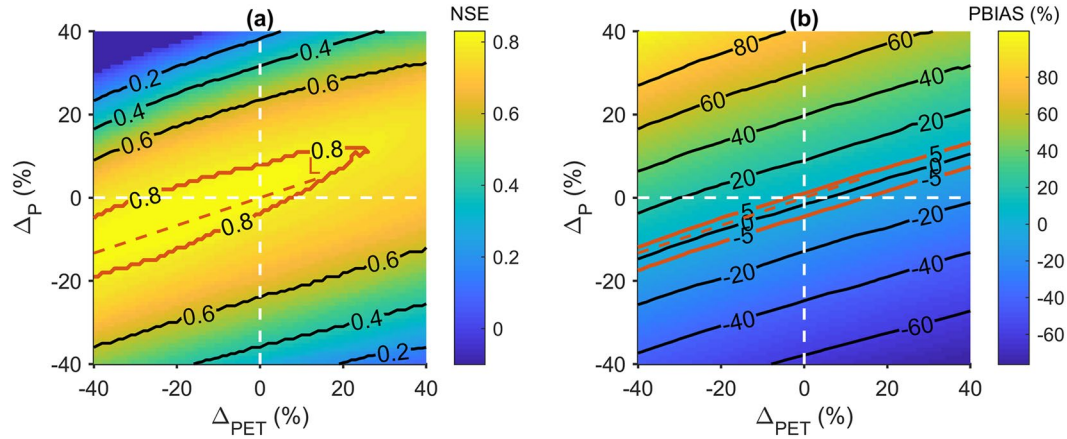

Figure 2. The Xinanjiang (XAJ) model performance results, Nash-Sutcliffe Efficiency (NSE) (a) and Percent Bias (PBIAS) (b) under different scenarios of biased precipitation (*P*) and potential evapotranspiration (*PET*) inputs (Δ_P and Δ_{PET}). The scenarios with the best performance are within the red solid boundary. Within a certain range of biases combination (such as along the dotted line L), *P* and *PET* can compensate for each other, and thereby generate satisfactory streamflow simulations.

2.6. Biased Input Scenarios

In this study, we added various biases (systematic errors) to the original *P* and *PET* data to generate biased input scenarios. This was carried out by introducing two bias coefficients, Δ_P (%) and Δ_{PET} (%), to the original *P* and *PET* time series. The bias coefficients are applied to the whole time series and the new estimates are given by:

$$P_i' = (1 + \Delta_P) \times P_i \tag{3}$$

$$PET_i' = (1 + \Delta_{PET}) \times PET_i \tag{4}$$

where P_i' and PET_i' represent biased *P* and *PET* time series. The bias coefficients indicate how much the original data has been scaled. They are helpful to test the effect of under- or overestimations of *P* and *PET* inputs on simulating *Q*. Based on previous studies dealing with systematic errors to the inputs (Bai et al., 2016; Jayathilake & Smith, 2022; Oudin et al., 2006), a range of −40% to 40% with a step of 1% was applied to Δ_P and Δ_{PET} individually. The biases within ±40% are able to cover the systematic errors that happened in most hydrological applications. A negative bias coefficient means underestimation and a positive one means overestimation, and a value of "0" refers to the original (unbiased) data. There were 6560 biased scenarios in total. These biased *P* and *PET* data were used to drive the hydrological models without recalibration.

Ideally, a good *Q* simulation should meet the requirement of NSE ≥ 0.8 and $|PBIAS| \leq 5\%$. The biased scenarios that meet this requirement would be selected to investigate the specific mutual relationship between the biased *P* and *PET* data in reproducing good *Q*. However, there are large discrepancies in hydrological model performance in different catchments across Great Britain as investigated by Lane et al. (2019). Some of their results showed catchments with NSE values below 0.6 in terms of the best flow simulation performance. Therefore, we expect the NSE performance to vary across the selected catchments and we expect some of the best performing models to show NSE values lower than 0.8 in some cases.

3. Results and Discussion

3.1. Compensational Relationship of *P* and *PET* Inputs' Errors in Simulating *Q*

The simulated *Q* based on the HYREX data in the Brue catchment is shown in Figure S1 in Supporting Information S1, with an NSE of 0.82 and a PBIAS of 3.4%. The XAJ hydrological model was run for the 6560 biased scenarios in order to test how the hydrological model performance changes with biased *P* and *PET* input data. Figure 2 presents the hydrological performance results (NSE and PBIAS) under different combinations of biased *P* and *PET* input data. It can be seen that within a certain range of biased scenarios (i.e., within the red solid boundary), *P* and *PET* can compensate for each other, thereby generating a satisfactory hydrological performance

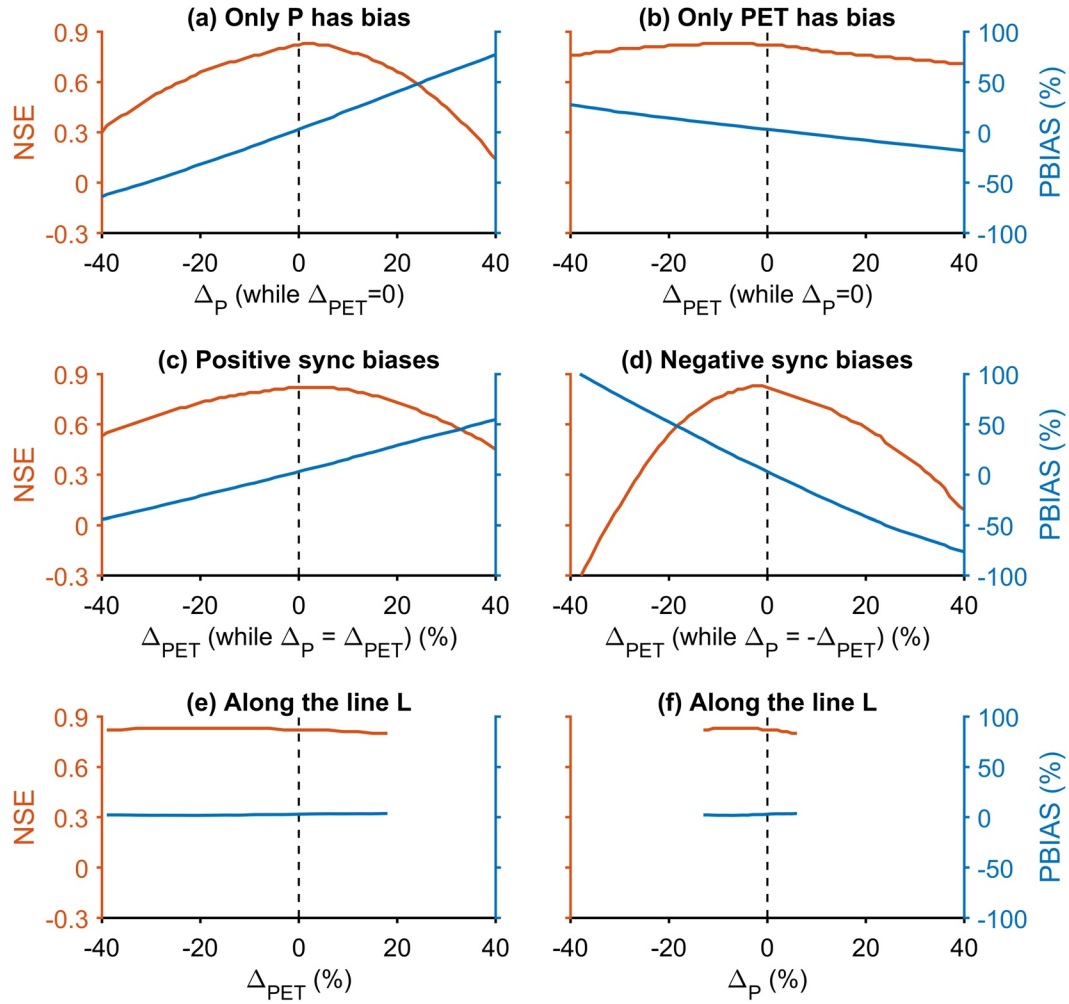

Figure 3. The Xinanjiang (XAJ) model performance under different biased input scenarios: (a) with a bias in precipitation (*P*) only; (b) with a bias in potential evapotranspiration (*PET*) only; (c) with the same bias in *P* and *PET*; (d) with opposite bias in *P* and *PET*; (e, f) with biased scenarios along the line L shown in Figure 2.

(NSE $\geq$ 0.8 in Figure 2a, and $|$PBIAS$|$ $\leq$ 5% in Figure 2b). The bias combinations that appear along the line *L* are selected as a representative example to further analyze the mutual compensation between *P* and *PET* input data's quality in reproducing *Q* simulations.

Figure 3 shows the performance of the hydrological simulation under different biased input scenarios. The accuracy of *P* data is found to be more important than the accuracy of *PET*. By introducing the same magnitude of bias to only one input (Figures 3a and 3b), biased *P* data generates simulations with even worse hydrological performance than using biased *PET*. It can be seen that the model's efficiency significantly decreases when there is a systematic bias of ±20% only in *P* inputs, while insignificantly decreases when the bias only occurs in *PET* inputs (Figures 3a and 3b). This confirms again that the hydrological model is less sensitive to systematic errors in *PET* than in *P* data. These findings are in good agreement with previous studies (Nandakumar & Mein, 1997; Paturel et al., 1995).

Further, *P* and *PET* data with the same biases fail to mutually offset each other to reproduce satisfactory *Q* simulations, which can be found in Figure 3c. The model performance becomes worse if *P* and *PET* have an opposite bias (i.e., *P* with overestimation while *PET* with underestimation) (Figure 3d). For example, when *PET* has a negative bias while *P* has a positive bias with the same magnitude (Figure 3d), the *Q* simulations are found to be poorer than the scenario in which *P* and *PET* have the same negative biases (Figure 3c). Moreover, the overestimation of inputs has a more severe effect on the *Q* outputs, which is in line with the study of Xu et al. (2006). As

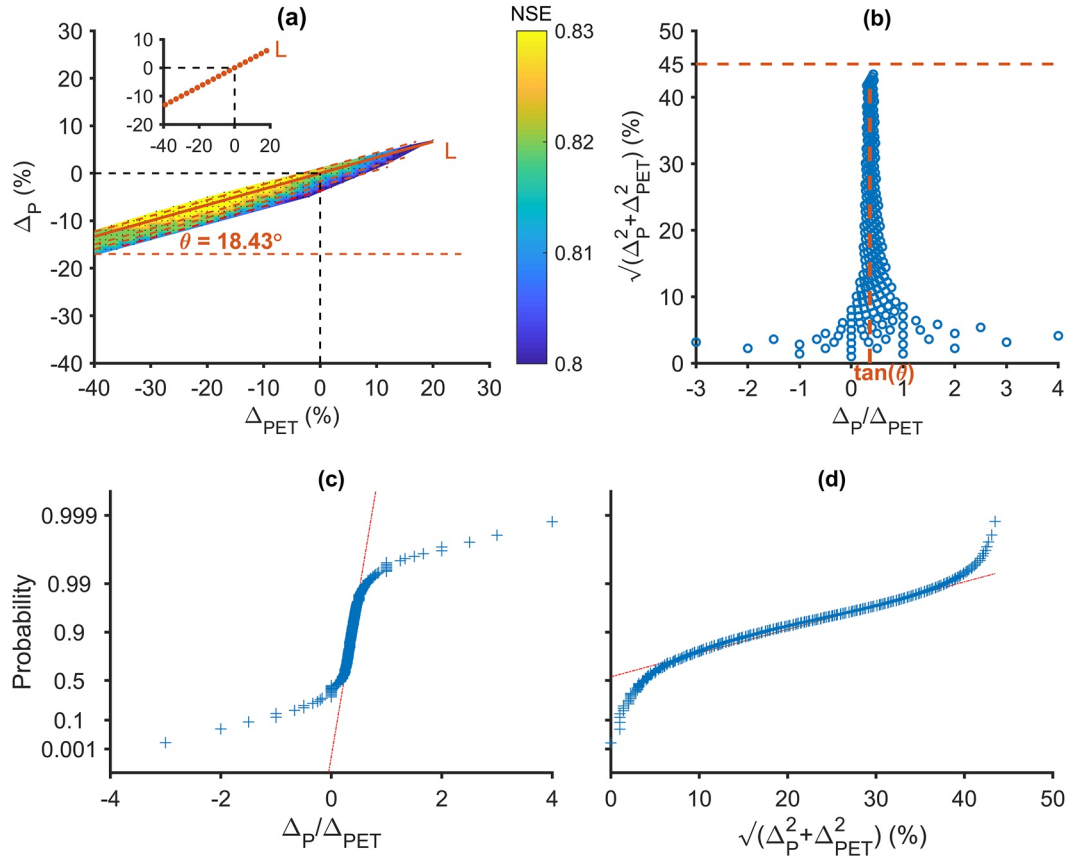

Figure 4. (a) Biased precipitation (P) and potential evapotranspiration (PET) scenarios that can produce good Xinanjiang (XAJ) model simulations based on 2-year hourly data from HYdrological Radar EXperiment. (b) A scatter plot between the bias ratio and the square root of the sum of squared biases for these scenarios. Normal probability plots for the bias ratio (c) and the square root (d), where the data appear along the reference line (in red) indicate that they follow a normal distribution.

shown in Figures 3a–3c, the slopes of the overestimation scenarios (the right side of each plot) are greater than the underestimation scenarios (the left side of each plot).

On the contrary, if the biases in P and PET inputs meet a certain relationship, such as that shown appear along the line L in Figure 2, they can offset each other's biases to produce good Q simulations, which can be found in Figures 3e and 3f. The XAJ NSE scores of these biased scenarios are about 0.82 and the PBIAS scores are about 3%.

3.2. Compensational Relationship Between *P* and *PET* Biases That Produce Good *Q* Simulations

To figure out the specific mutual relationship of biases in P and PET in Q modeling, we selected the biased scenarios that can produce good Q simulations (NSE $\geq$ 0.8 and |PBIAS| $\leq$ 5%) in the Brue catchment. As shown in Figure 4a, a region has been formed and there exists a certain angle and range of values that compensate for the biases in P and PET data. It can also be found that except for a few scenarios, the majority of bias combinations stay along the lines which are parallel to Line L. The gradient of L is tan(18.43°) $\approx$ 0.33, which means the slope $\Delta_P : \Delta_{PET} \approx 1 : 3$. In other words, if the bias in PET input data is three times the bias in P data, then the biases could mutually offset each other to reproduce good Q. Certainly, there are thresholds for the range of P and PET biases.

Figure 4b shows a scatter plot between the bias ratio (Δ_P/Δ_{PET}) and the square root of the sum of squared biases $\left(\sqrt{\Delta_P^2 + \Delta_{PET}^2}\right)$ for the selected biased scenarios that generate satisfactory Q simulations (the region in Figure 4a). A certain pattern can be found in most of those biased scenarios. As the square root becomes greater than 10%, the ratio turns asymptotically closer to tan(18.43°). Although there are a few combinations whose absolute ratio

$|\Delta_P/\Delta_{PET}|$ are more than 1, their square root is less than 5%. In these cases, both Δ_P and Δ_{PET} are quite small. This means P and PET input data are almost unbiased, making them still behave well in terms of Q modeling. Additionally, we used the normal probability plots to statistically test the normality of the distributions. As shown in Figures 4c and 4d, a normal distribution fits well the central part of the quantified distributions (Δ_P/Δ_{PET} and $\sqrt{\Delta_P^2+\Delta_{PET}^2}$), which is approximately symmetrical to the vertical line $x = \tan(18.43°)$, but fails to fit the tails.

On the basis of the above-mentioned, a new hydrological proxy, named Compensational Interaction Angle (CIA, θ) is proposed in this study to quantitatively and intuitively understand the mutual compensational relationship between the biases in P and PET inputs. In specific, the CIA can be extracted from the plot shown in Figure 4a, or be approximately computed as the arctangent of the mean bias ratio (Δ_P/Δ_{PET}) of the central part of the scatter plot (Figure 4b). The maximum square root $\sqrt{\Delta_P^2+\Delta_{PET}^2}$ is about 45% in each case. This is expected since we assume the maximum bias in P or PET is 40%, thus the maximum is $\sqrt{(40\%)^2+(40\%)^2}\approx 45\%$. It can also be derived from Figure 4b that $\sqrt{\Delta_P^2+\Delta_{PET}^2}$ varies from 0% to 45%.

The CIA is found to be 18.43° for the Brue catchment. The next step is to further explore how CIA varies in different situations (the length of modeling period, model equifinality, hydrological model and catchment condition).

3.3. How Does the CIA Vary With Different Modeling Scenarios?

3.3.1. Using Different Modeling Periods

The above analysis was based on the 2-year hourly HYREX rain gauge data. Although HYREX data are of high quality, the short availability is inadequate in investigating the stationarity (i.e., stability with time) of the compensational relationship. In order to test if the CIA changes with the length of the simulation period, we used additional data sets. Thus, two 5-year periods (from June 2003 to May 2008, and from June 2008 to May 2013) and one 10-year period (from June 2003 to May 2013) from the hourly CAMELS-GB data were used to validate the stationarity of the CIA. The corresponding Q simulated by using these 3 additional periods are shown in Figures S2, S3, and S4 in Supporting Information S1. All simulations have reached satisfactory performance with NSE greater than 0.8 and PBIAS within ±5%.

Similar to Figure 4 in Section 3.2, Figure 5 shows the biased input scenarios that reproduce Q well using biased P and PET from different time periods. Notably, the CIA turns out to be the same (19.95°) in these 3 modeling periods (Figures 5a–5c). As shown in Figures 5d–5f, the central parts of the scatter plots between the bias ratio and the square root of the sum of squared biases for these scenarios, also follow a similar normal distribution, which is symmetrical to the vertical line $x = \tan(19.95°)$. This indicates the stationarity of the CIA. Additionally, the angle of 19.95° (found by using the CAMELS data) is very close to the angle 18.43° derived from the HYREX data. The small difference in this angle might be due to the different data sources and length of time periods used in the analysis were used.

Moreover, by comparing the biased input scenarios demonstrated in Figures 5a–5c, a wider range of biased input scenarios and greater NSE scores can be found as the data are more recent and longer. This means the hydrological model appears to be more adaptable to the joint biases when using longer time periods in the input data. Specifically, the bias combinations in P and PET data over the period from June 2008 to May 2013 (Figure 5b) generate better Q simulations than those from June 2003 to May 2008 (Figure 5a). Similarly, Figure 5c (using the 10-year data) has a better NSE performance than Figures 5a and 5b (using the 5-year data). By comparing the results in different periods (Figures 5d–5f), the central parts of the scatter plot become closer to a normal distribution as the data used are more recent and longer. This might be attributed to the better quality of more recent data, which are derived from more accurate and advanced measurements. In this study, the data with the longer time period are more reliable than the data with the shorter time period given that the information content contained in the former is greater than in the latter.

3.3.2. Assuming Hydrological Model Equifinality

It is well known that various model parameter sets could be equally capable of reproducing similar hydrological outputs, which is called equifinality (Beven & Freer, 2001). This means the optimal parameter set for calibrating the XAJ model is not unique. Many other parameter sets could also generate equally good Q simulations as well. Thus, there is a need to further investigate if the compensation relationship would change in terms of model

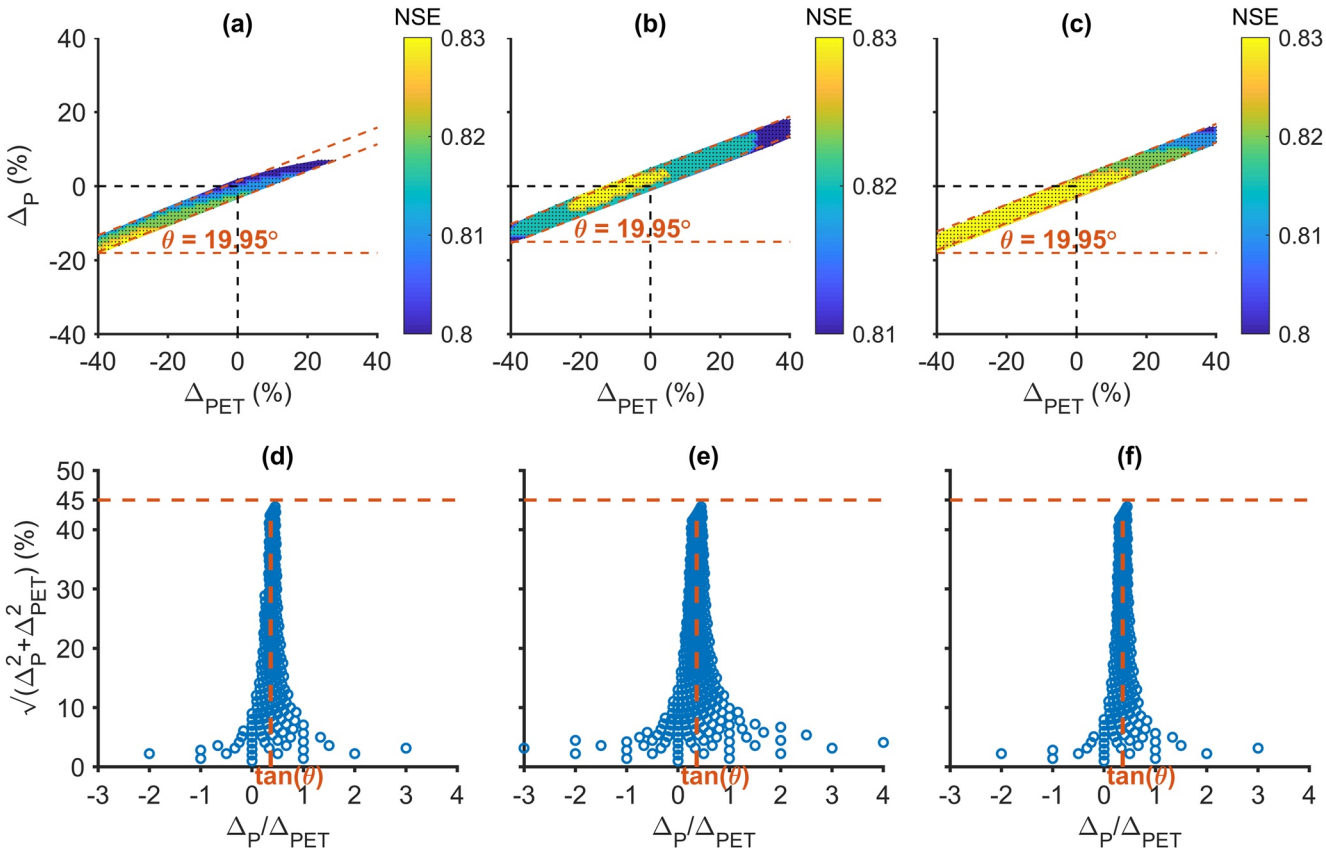

Figure 5. Top row: biased precipitation (*P*) and potential evapotranspiration (*PET*) input scenarios that produce good model results using different time periods from Catchment Attributes and MEteorology for Large-sample Studies in Great Britain data, including from June 2003 to May 2008 (a), from June 2008 to May 2013 (b), and from June 2003 to May 2013 (c). Bottom row: corresponding scatter plots between the bias ratio and the square root of the sum of squared biases for these scenarios.

equifinality. Two different parameter sets were calibrated using the original 2-year HYREX data and the 10-year CAMELS-GB data (Figures S5 and S6 respectively in Supporting Information S1).

Then, the XAJ model was rerun without recalibrations using the 6560 biased input scenarios. Figure 6 shows the biased scenarios that reproduce *Q* well using the HYREX and CAMELS-GB data sets. It can be seen that the CIA remains the same as the one using the previous parameter set, namely, 18.43° for the 2-year HYREX and 19.95° for the 10-year CAMELS-GB data sets (Figures 6a and 6c). A similar normal distribution can also be found in the central part of the scatter plots in Figures 6b and 6d for both model results. This indicates the stability of the compensational relationship under model equifinality.

3.3.3. Using a Different Hydrological Model

Further, as only the XAJ model is considered in this study, there is a question on whether the compensational relationship would change when using a different hydrological model. We tested this by using the 10-year CAMELS-GB data as input data to force the PDM model. The *Q* simulated by the PDM model after calibration can be found in Figure S7 in Supporting Information S1. The NSE and PBIAS scores are 0.84 and −2.2%, respectively. The PDM model was rerun using the 6560 biased input scenarios.

Figure 7 shows the performance results of these 6560 biased input scenarios. It can be seen that the interaction between biased *P* and *PET* inputs is very similar to the one using the XAJ model. The CIA appears to almost remain the same as using the XAJ model, namely, 19.95°. Similarly, the normal distribution fits well with the central part of the scatter plot, which is approximately symmetrical along a vertical line $x = \tan(19.95°)$, but fails to fit the tails (Figures 7e and 7f). These results indicate that the compensational relationship is stable when using a different hydrological model.

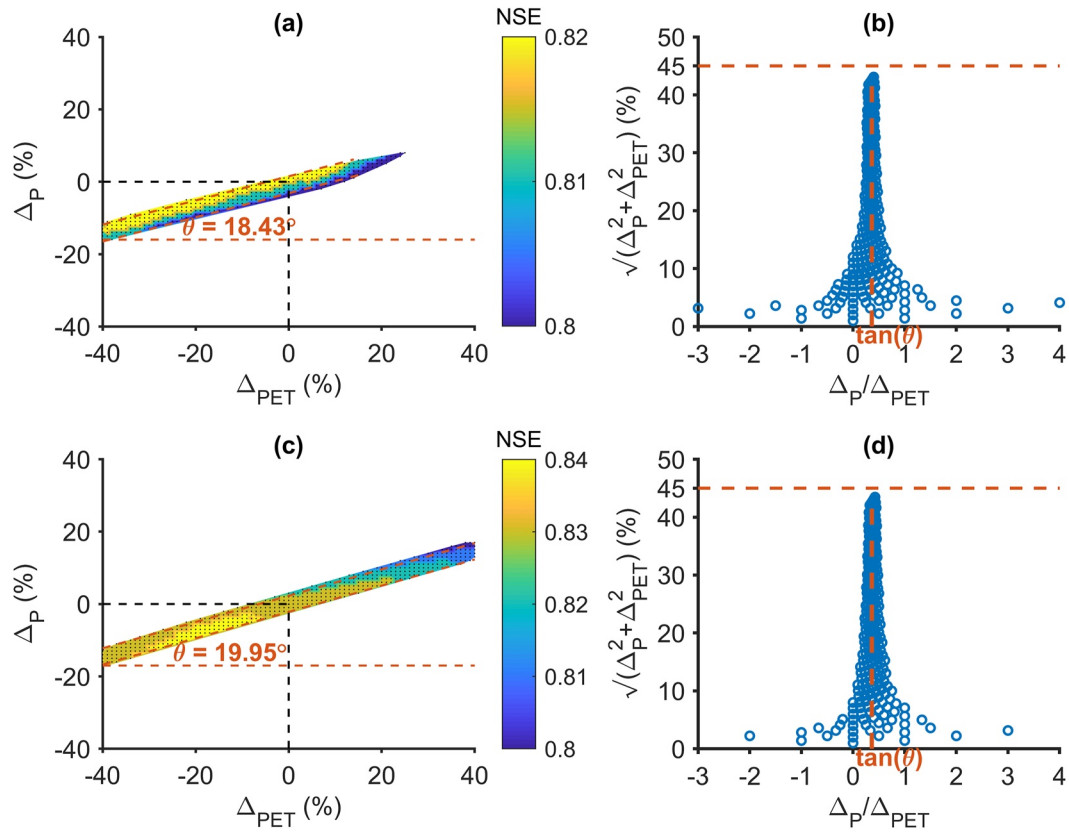

Figure 6. Left column: biased precipitation (*P*) and potential evapotranspiration (*PET*) scenarios that produce good model outputs using a different parameter set for the 2-year HYdrological Radar EXperiment data (a), and the 10-year Catchment Attributes and MEteorology for Large-sample Studies in Great Britain data (c). Right column: corresponding scatter plots between the bias ratio and the square root of the sum of squared biases for these scenarios (b and d).

3.3.4. Using Different Catchments With Varied Aridity Conditions

There is a more interesting question on whether the compensational relationship changes with different aridity conditions. The Scar Water and South Tyne catchments (more humid than the Brue catchment) and Isbourne and Stainfield Beck catchments (arider than the Brue catchment) were selected to explore this question. As found in the previous section, the PDM model shows similar results to the XAJ model. For simplicity and efficiency, the PDM model with fewer parameters was used to simulate *Q* using 10-year hourly data for the 4 catchments independently. The corresponding simulated *Q* for these catchments can be found in Figures S8, S9, S10, and S11 in Supporting Information S1. It is important to highlight that the simulations using the original data exhibited an optimal NSE of 0.76 for the Isbourne catchment and 0.68 for the Stainfield Beck catchment. Therefore, the satisfactory simulations using the biased inputs for the Isbourne and the Stainfield Beck catchments were defined as having an NSE greater than 0.7 and 0.62, respectively, as well as a PBIAS within ±5%. The Scar Water and South Tyne catchments showed better NSE performance (NSE greater than 0.8). Figure 8 shows the biased input scenarios that reproduce *Q* well in different catchments. A similar compensational relationship was identified in these four catchments. The central parts of the scatter plots shown in Figures 8e–8h also follow a normal distribution, which is symmetrical along a vertical line $x = \tan(\theta)$.

Moreover, it can be found from Figures 8a–8d that the catchments with greater aridity have larger CIA (θ) whereas the more humid catchments showed smaller CIAs values. *P* is higher than *ET* in humid regions, thereby the bias in *PET* needs to be higher to accommodate even a small bias in *P* to be able to produce good model results. For example, in the Stainfield Beck catchment (Figure 8d), the CIA is 25.45° and $\tan(25.45°) \approx 0.476$, namely, $\Delta_P : \Delta_{PET} \approx 1 : 2.1$, which indicates that the bias in *PET* is 2.1 times higher than the bias in *P*, then the biases could mutually offset each other to reproduce good *Q*. However, for the more humid Scar Water catchment,

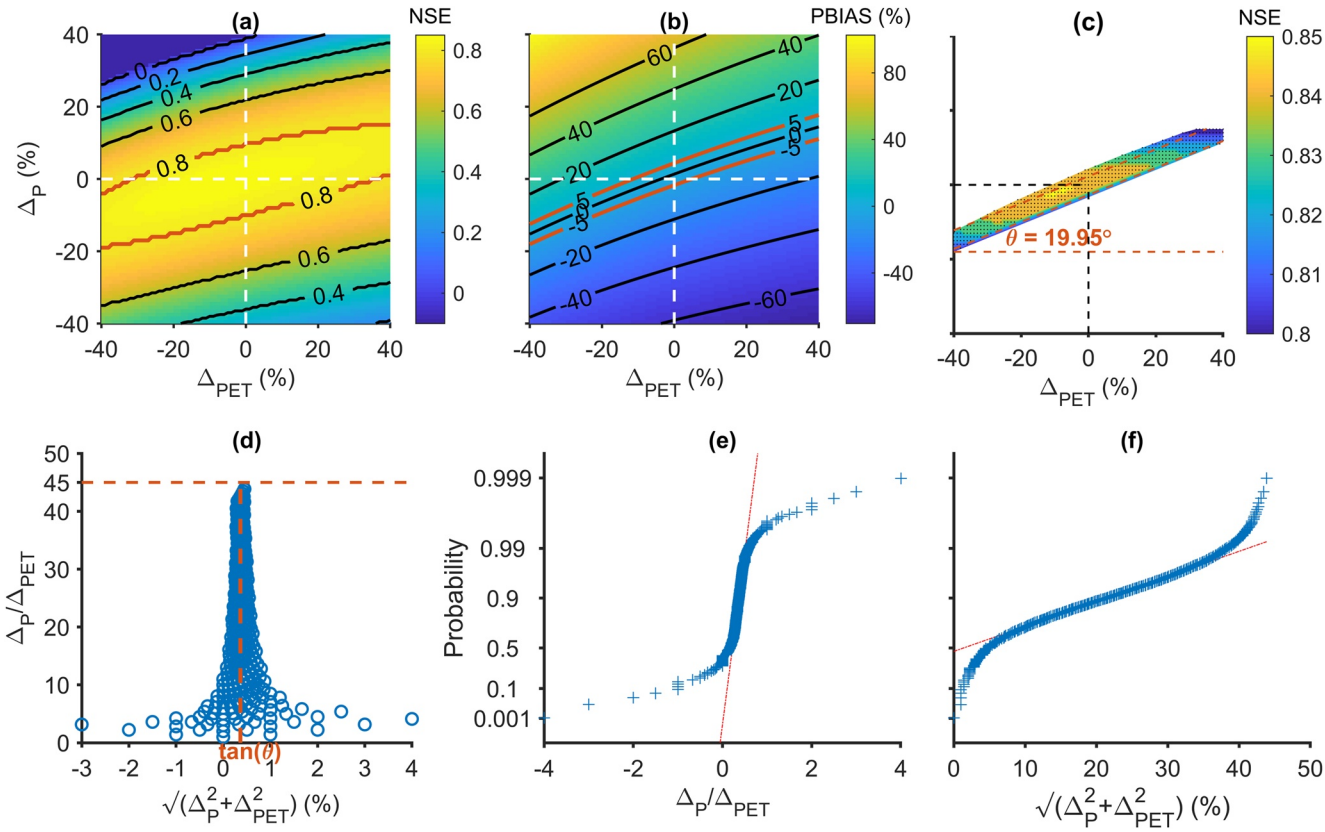

Figure 7. The Probability Distributed Model (PDM) simulation performance, Nash-Sutcliffe Efficiency *(NSE)* (a) and Percent Bias (PBIAS) (b) in 6560 biased Precipitation (*P*) and Potential Evapotranspiration (*PET*) input scenarios. (c) Biased input scenarios that can produce good model performance. (d) A scatter plot between the bias ratio and the square root of the sum of squared biases for these scenarios. Normal probability plots for the bias ratio (e) and the square root of the sum of squared biases (f).

the CIA is 11.25° and tan(11.25°) ≈ 0.2, namely, $\Delta_P : \Delta_{PET} \approx 1 : 5$, which indicates that the bias in *PET* should be five times higher than the bias in *P* for the model to produce acceptable *Q* simulations.

3.4. Which Aspects of NSE Significantly Affect the Compensational Relationship?

The NSE decomposition can give us some insights into how its different components (*r*, α, and β) are related to the compensational relationship. As found in Section 3.3.1, a longer time period gives us a better model performance. Thus, we used the 10-year hourly CAMELS-GB data to drive the XAJ and PDM models for the Brue catchment for further investigation. The three distinct components (*r*, *α*, and *β*) of NSE were calculated for each biased input scenario. As shown in Figure 9, the correlation (*r*) seems to remain at a value of more than 0.9 in most of the biased scenarios (Figures 9a and 9e), implying that the bias added in the inputs does not largely affect the correlation of the model simulations. However, *α* (relative variability) and *β* (normalized bias) vary for most biased scenarios (Figures 9b, 9c, 9f, and 9g).

We selected the biased scenarios that appear along the lines L1 and L2 in Figure 9 as representative examples to intuitively understand the variability of *α* and *β*. The line L1 has been defined previously and it has a slope of tan(19.95°) ≈ 0.363, while the line L2 is perpendicular to L1 and is made up of biased scenarios that do not follow the compensational relationship. It is found that all the biased scenarios along L1 in the XAJ and PDM models (Figures 9i and 9k) can produce good model performance, with NSE, *r*, *α*, and *β* remaining almost unchanged. However, most biased scenarios along L2 fail to produce good model performance, with large discrepancies in *α*, *β*, and NSE, except for *r*, which still seems to show a high correlation ($r > 0.9$). These findings suggest that it is mainly the joint response of the relative variability and normalized bias that contribute to the NSE result in our study, which is in agreement with the study of Wang et al. (2023). In particular, the normalized

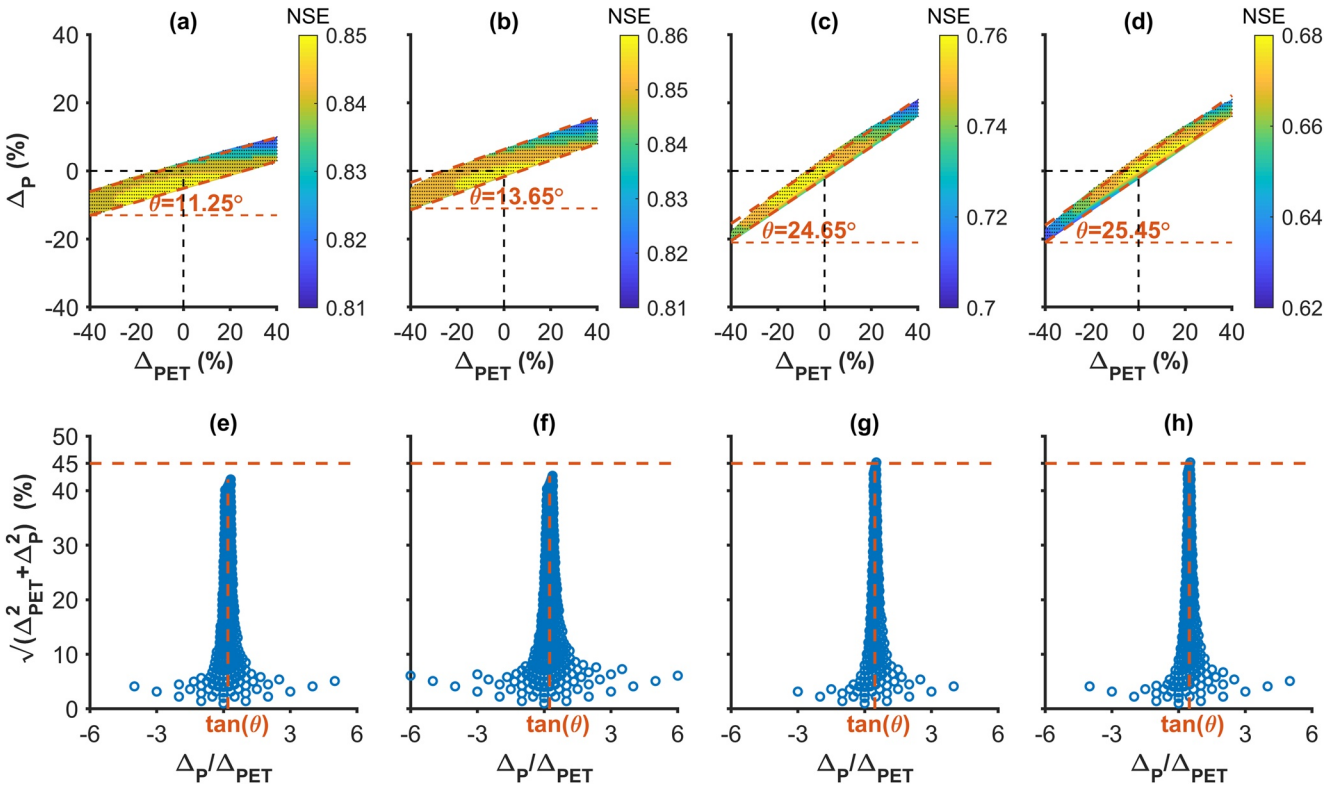

Figure 8. Top row: biased Precipitation (*P*) and Potential Evapotranspiration (*PET*) input scenarios that produce good simulations using the Probability Distributed Model (PDM) and a 10-year hourly data period for the Scar Water (a), South Tyne (b), Isbourne (c) and Stainfield Beck (d) catchments, respectively. Bottom row: scatter plots between the bias ratio and the square root of the sum of squared biases for these scenarios for each catchment respectively (e, f, g, and h). The catchments are presented from left to right in ascending order of their aridity ratios.

bias has the same pattern as the CIA relationship. This is mainly because the CIA identified in the Brue catchment was constrained not only by NSE but also by PBIAS (NSE $\geq$ 0.8 and |PBIAS|$\leq$ 5%).

3.5. Why Does This Compensational Pattern Exist?

Based on the above analysis, *P* and *PET*'s biases have a mutual compensational relationship to reproduce *Q* well. This compensational pattern appears to be stationary with time and stable with different hydrological models, and taking into account model equifinality. Moreover, arid catchments have larger CIAs. The question is, why does this compensational pattern exist? The water balance principle could provide some hints to answer this question since *P*, *PET*, and *Q* are three essential components of the terrestrial water cycle (Lehmann et al., 2022). The long-term water balance equation for a catchment can be simplified as,

$$Q = P - ET - \omega \tag{5}$$

where ω denotes water budget imbalance (including the change in soil water storage, surface water, groundwater, snow, and canopy) (Scanlon et al., 2018). *P*, *ET*, and *Q* are the major components of the terrestrial water balance, and generally, ω is assumed to be zero when assuming a longer time period (for a catchment with negligible groundwater flow such as the Brue catchment), However, ω cannot be ignored over shorter time periods or for catchments where groundwater effects are important. *ET* is considered to have an approximately proportional relationship with *PET*, which can be expressed as:

$$ET = \partial \cdot PET \tag{6}$$

where ∂ refers to the proportional coefficient.

If there are biases in *P* and *PET*, then Equation 5 could be rewritten as,

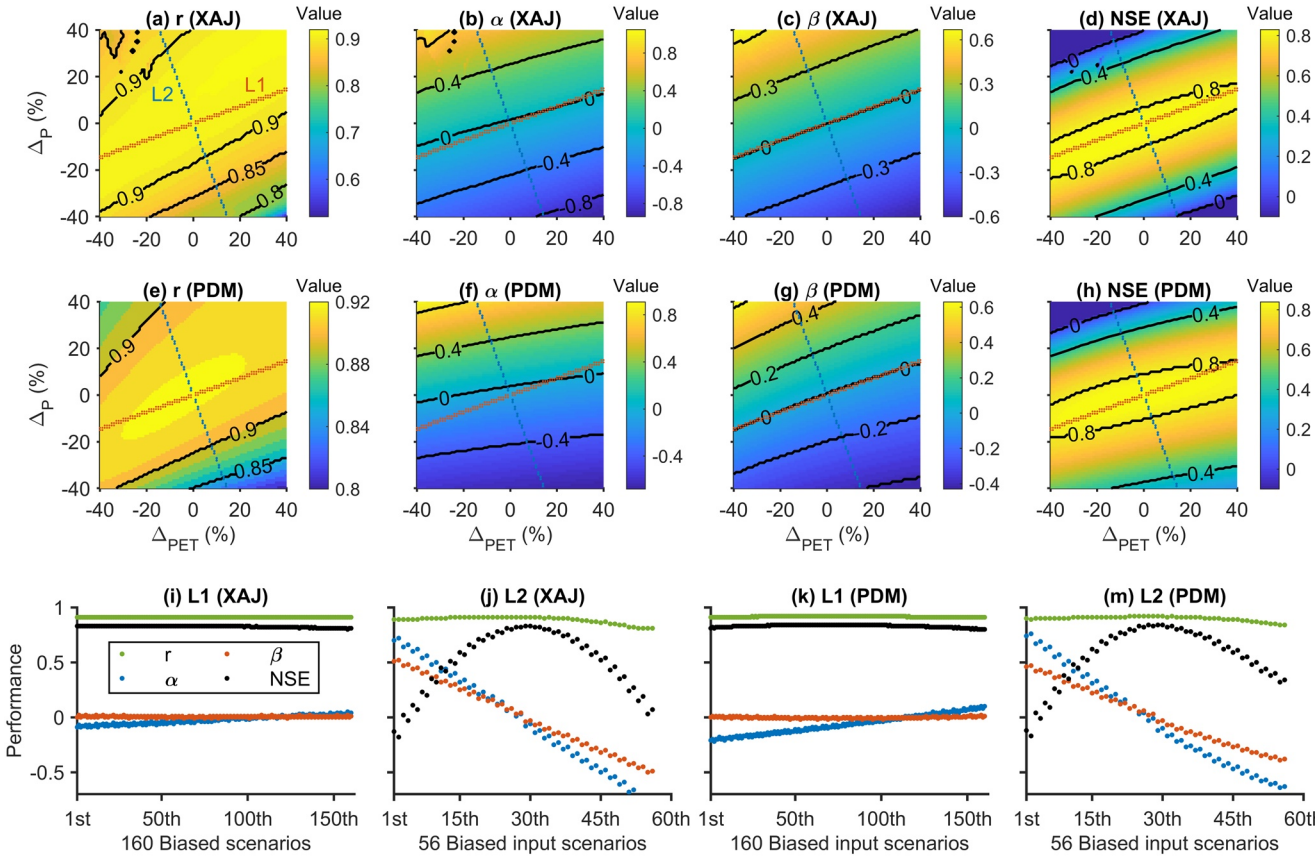

Figure 9. Model performance in terms of Nash-Sutcliffe Efficiency (NSE) decomposition. The top row shows correlation (r), relative variability (α), normalized bias (β), and NSE for the Xinanjiang (XAJ) model, the middle row shows r, α, β, and NSE for the Probability Distributed Model (PDM), and the bottom row demonstrates the performance of the biased scenarios along the lines (L1 and L2).

$$Q = P \cdot (1 + \Delta_P) - \partial \cdot PET \cdot (1 + \Delta_{PET}) - \omega \tag{7}$$

$$Q = P - \partial \cdot PET - \omega + P \cdot \Delta_P - \partial \cdot PET \cdot \Delta_{PET} \tag{8}$$

For a hydrological model to produce a good performance even if there are biases in P and PET, the sum of the last two terms of Equation 8 should be close to zero and therefore:

$$0 = P \cdot \Delta_P - \partial \cdot PET \cdot \Delta_{PET} \tag{9}$$

$$\frac{\Delta_P}{\Delta_{PET}} = \partial \cdot PET \,/\, P = ET/P \tag{10}$$

The mean ET values based on the PDM simulations using the 10-year unbiased CAMELS-GB data for the Scar Water, South Tyne, Brue, Isbourne, and Stainfield Beck catchments are 1.17, 0.86, 1.18, 1.14, and 1.15 mm/day, respectively. Given the mean values of P introduced in Table 1 (4.73, 3.18, 2.45, 1.96, and 1.77 mm/day, respectively for each catchment), we can conclude that Δ_P/Δ_{PET} values are 0.25, 0.27, 0.48, 0.58, and 0.65, respectively for each catchment. The results are a little bit higher than the values of tan(CIA) (0.2, 0.24, 0.36, 0.46, and 0.48, respectively for each catchment) found in Section 3.3, but this might be due to the assumptions used to derive the long-term water balance equation (Equation 5).

Overall, Equation 10 suggests that the ratio of biases in P and PET inputs for producing good Q simulations (Δ_P/Δ_{PET}), namely, tan(CIA), is highly related to the ratio of mean actual ET and P of the catchment. It is also linked to the ratio of mean PET and P, that is, the aridity ratio. Different CIAs found in the five studied catchments and catchments with greater aridity have larger CIAs also support this conclusion (Section 3.3.4.).

Certainly, more investigations in a wider range of catchments are needed to better understand the compensational relationship.

3.6. What Is the Use of the Compensational Relationship?

The water balance-based Budyko hypothesis (Budyko, 1974) is a widely applied empirical "top-down" approach in catchment hydrology (Wang et al., 2016). It hypothesizes that the ratio between mean *ET* and *P* is primarily a function of the ratio between the mean annual *PET* and *P*, which is the long-term climate aridity ratio. The compensational relationship identified in our study is also linked to the Budyko curve. The CIA might be used to elucidate some important relationships of the terrestrial water balance. This compensational relationship also varies with the aridity of the catchments. Additionally, it may also provide useful information for ungauged catchments nearby the gauged ones since it indicates relevant hydrological characteristics in local regions.

However, it is important to highlight that the Brue catchment is a near-natural catchment with insignificant human impact (e.g., no reservoirs or river diversion). This might be the main reason that the compensational pattern is stable using different modeling periods. However, the pattern is likely to be unstable in those regions where noticeable human activities are involved (i.e., reservoir and urbanization).

Since various *P* and *PET* data sets have been developed in recent decades from different sources, these data sets are inevitably prone to different sources of biases (Dai & Han, 2014; Levy et al., 2017), caused by different factors (Sevruk, 1996; Xu et al., 2006), encompassing instrument measurement error, improper placement of measuring equipment, biases in climate model outputs, etc. Consequently, these errors result in different magnitudes of biases with different data sources during different periods (Wang et al., 2023). The compensational pattern identified in the present study could be useful for anticipating how hydrological modeling behaves in such cases. If the biases in *P* and *PET* data sets follow the CIA relationship, then they are highly likely to yield satisfactory hydrological performance.

4. Summary and Conclusions

In this study, we explored the joint interaction of *P* and *PET* input data biases in the response of *Q* simulations. Different biases (systematic errors) were added to the original *P* and *PET* data to produce 6560 biased input scenarios. Those biased inputs were then independently used for driving the XAJ and PDM models to investigate the existence of compensational interaction in simulating *Q*. Two-year HYREX data for the Brue catchment and 10-year CAMELS-GB data over five benchmark catchments in Great Britain were used for the analysis. The compensational relationship was investigated under various situations, including different data period lengths, model equifinality, hydrological models, and catchment conditions. The contributions of the present study are threefold.

First, the results confirm findings reported in previous studies, such as (a) the hydrological model is more sensitive to the bias in *P* than in *PET* data, and (b) overestimation of *P* and *PET* has a greater impact on *Q* simulations than underestimation. Second, the study reveals the following new findings: (a) the biased *P* and *PET* inputs could compensate for each other to some extent in reproducing *Q* well; (b) this compensational relationship is further quantified as the CIA. The normal distribution fits well with the central part of the scatter plot between the bias ratio (Δ_P/Δ_{PET}) and the square root of the sum of squared biases $\left(\sqrt{\Delta_P^2+\Delta_{PET}^2}\right)$, which is approximately symmetrical to the vertical line $x = \tan(\text{CIA})$; (c) further, the CIA appears to be stationary with the length of the modeling period and is stable despite model equifinality (i.e., using a different optimal parameter set). The hydrological model is more adaptable to the joint biases in inputs that contain much more information (i.e., 10-year data provide more information than 2-year data); (d) the CIA is also similar when using different hydrological models; (e) the CIA is highly linked to the long-term climate aridity ratio. The catchments with greater aridity have larger CIAs whereas humid catchments show smaller CIAs. Third, a great number of data sets have emerged in this big data era, making it hard for the research community to choose suitable *P* and *PET* data sets. The compensational pattern diagnosed in our study may provide helpful information in selecting and evaluating suitable *P* and *PET* data sets for hydrological application, as well as provide insights into *P* and *PET* data bias correction and data fusion.

This study aims to provide a new perspective for hydrologists to analyze the input errors and water balance in hydrological modeling. Any generalization would need the use of various input data sets, study areas and hydrological models. The methodology in this study can be easily applied to other catchments and using different hydrological models. Except for the CIA proposed in this study, more perspectives need to be further explored to comprehensively understand the compensational relationship. Additionally, as only five near-natural catchments were analyzed in our study, we recommend using the proposed method in a wider range of catchments (especially the catchments affected by human activities) to further validate our findings and explore the spatial heterogeneity of the compensational relationship. Further investigation is warranted to provide a bigger picture and detailed understanding of the hydroclimatic conditions worldwide, which will be useful for forecasting extreme events (e.g., floods and droughts).

Data Availability Statement

The HYREX rainfall data are publicly available at https://data.ceda.ac.uk/badc/hyrex. The daily CAMELS-GB can be freely downloaded from https://doi.org/10.5285/8344e4f3-d2ea-44f5-8afa-86d2987543a9. The hourly hydro-meteorological data sets (*P*, *PET* and *Q*) are under license and were compiled by the Hydrology group in Geography at the University of Bristol. Please email Dr. Gemma Coxon (gemma.coxon@bristol.ac.uk) to discuss access to the data. The hourly CEH-GEAR1hr *P* time series are available at https://doi.org/10.5285/d4ddc781-25f3-423a-bba0-747cc82dc6fa. The *Q* time series are available upon request from the Environmental Agency (EA), Natural Resources Wales (NRW), and Scottish Environmental Protection Agency (SEPA). Daily CHESS *PET* time series are available at https://doi.org/10.5285/8baf805d-39ce-4dac-b224-c926ada353b7. Additional codes that support the findings of this study are available from the corresponding author upon request.

Acknowledgments
J. Wang is supported by the China Scholarship Council/University of Bristol joint scholarship (No 201906380142). The authors would like to acknowledge Dr. Gemma Coxon and her team from the University of Bristol for sharing the hourly CAMELS-GB data of the studied catchments. This work was carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol—http://www.bris.ac.uk/acrc/.

References

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). *Crop evapotranspiration-Guidelines for computing crop water requirements*. FAO. FAO Irrigation and drainage paper 56.

Allen, R. G., Smith, M., Perrier, A., & Pereira, L. S. (1994). An update for the definition of reference evapotranspiration. *ICID bulletin*, *43*(2), 1–34.

Andréassian, V., Perrin, C., & Michel, C. (2004). Impact of imperfect potential evapotranspiration knowledge on the efficiency and parameters of watershed models. *Journal of Hydrology*, *286*(1–4), 19–35. https://doi.org/10.1016/j.jhydrol.2003.09.030

Bai, P., Liu, X., Yang, T., Li, F., Liang, K., Hu, S., & Liu, C. (2016). Assessment of the influences of different potential evapotranspiration inputs on the performance of monthly hydrological models under different climatic conditions. *Journal of Hydrometeorology*, *17*(8), 2259–2274. https://doi.org/10.1175/jhm-d-15-0202.1

Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., et al. (2017). Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences*, *21*(12), 6201–6217. https://doi.org/10.5194/hess-21-6201-2017

Beven, K. (2012). *Rainfall-runoff modelling: The primer* (2nd ed.). Wiley-Blackwell.

Beven, K., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, *249*(1–4), 11–29. https://doi.org/10.1016/S0022-1694(01)00421-8

Bloomfield, J. P., Allen, D. J., & Griffiths, K. J. (2009). Examining geological controls on baseflow index (BFI) using regression analysis: An illustration from the Thames Basin, UK. *Journal of Hydrology*, *373*(1–2), 164–176. https://doi.org/10.1016/j.jhydrol.2009.04.025

Borga, M. (2002). Accuracy of radar rainfall estimates for streamflow simulation. *Journal of Hydrology*, *267*(1–2), 26–39. https://doi.org/10.1016/S0022-1694(02)00137-3

Borga, M., Degli Esposti, S., & Norbiato, D. (2006). Influence of errors in radar rainfall estimates on hydrological modeling prediction uncertainty. *Water Resources Research*, *42*(8), W08409. https://doi.org/10.1029/2005wr004559

Budyko, M. I. (1974). *Climate and life*. Academic Press.

Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Fry, M., Hannaford, J., et al. (2020a). Catchment attributes and hydro-meteorological time-series for 671 catchments across Great Britain (CAMELS-GB) [Dataset]. NERC EDS Environmental Information Data Centre. https://doi.org/10.5285/8344E4F3-D2EA-44F5-8AFA-86D2987543A9

Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Fry, M., Hannaford, J., et al. (2020b). CAMELS-GB: Hydrometeorological time series and landscape attributes for 671 catchments in Great Britain. *Earth System Science Data*, *12*(4), 2459–2483. https://doi.org/10.5194/essd-12-2459-2020

Dai, Q., & Han, D. (2014). Exploration of discrepancy between radar and gauge rainfall estimates driven by wind fields. *Water Resources Research*, *50*(11), 8571–8588. https://doi.org/10.1002/2014wr015794

Doorenbos, J., & Pruitt, W. O. (1977). *Guidelines for predicting crop water requirement*. FAO Irrigation and Drainage. Paper No. 24.

Ekström, M., Jones, P. D., Fowler, H. J., Lenderink, G., Buishand, T. A., & Conway, D. (2007). Regional climate model data used within the SWURVE project–1: Projected changes in seasonal patterns and estimation of PET. *Hydrology and Earth System Sciences*, *11*(3), 1069–1083. https://doi.org/10.5194/hess-11-1069-2007

Ellenburg, W. L., Cruise, J. F., & Singh, V. P. (2018). The role of evapotranspiration in streamflow modeling—An analysis using entropy. *Journal of Hydrology*, *567*, 290–304. https://doi.org/10.1016/j.jhydrol.2018.09.048

Essou, G. R. C., Arsenault, R., & Brissette, F. P. (2016). Comparison of climate datasets for lumped hydrological modeling over the continental United States. *Journal of Hydrology*, *537*, 334–345. https://doi.org/10.1016/j.jhydrol.2016.03.063

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Hagemann, S., & Jacob, D. (2007). Gradient in the climate change signal of European discharge predicted by a multi-model ensemble. *Climatic Change*, *81*(S1), 309–327. https://doi.org/10.1007/s10584-006-9225-0

Jayathilake, D. I., & Smith, T. (2022). Identifying the influence of systematic errors in potential evapotranspiration on rainfall–runoff models. *Journal of Hydrologic Engineering*, *27*(2). https://doi.org/10.1061/(asce)he.1943-5584.0002157

Kabir, T., Pokhrel, Y., & Felfelani, F. (2022). On the precipitation-induced uncertainties in process-based hydrological modeling in the Mekong river basin. *Water Resources Research*, *58*(2), e2021WR030828. https://doi.org/10.1029/2021wr030828

Krause, P., Boyle, D. P., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, *5*, 89–97. https://doi.org/10.5194/adgeo-5-89-2005

Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence properties of the nelder-mead simplex method in low dimensions. *SIAM Journal on Optimization*, *9*(1), 112–147. https://doi.org/10.1137/S1052623496303470

Lane, R. A., Coxon, G., Freer, J. E., Wagener, T., Johnes, P. J., Bloomfield, J. P., et al. (2019). Benchmarking the predictive capability of hydrological models for river flow and flood peak predictions across over 1000 catchments in Great Britain. *Hydrology and Earth System Sciences*, *23*(10), 4011–4032. https://doi.org/10.5194/hess-23-4011-2019

Lehmann, F., Vishwakarma, B. D., & Bamber, J. (2022). How well are we able to close the water budget at the global scale? *Hydrology and Earth System Sciences*, *26*(1), 35–54. https://doi.org/10.5194/hess-26-35-2022

Levy, M. C., Cohn, A., Lopes, A. V., & Thompson, S. E. (2017). Addressing rainfall data selection uncertainty using connections between rainfall and streamflow. *Scientific Reports*, *7*(1), 219. https://doi.org/10.1038/s41598-017-00128-5

Lewis, E., Quinn, N., Blenkinsop, S., Fowler, H. J., Freer, J., Tanguy, M., et al. (2018). A rule based quality control method for hourly rainfall data and a 1 km resolution gridded hourly rainfall dataset for great Britain: CEH-GEAR1hr. *Journal of Hydrology*, *564*, 930–943. https://doi.org/10.1016/j.jhydrol.2018.07.034

Lewis, E., Quinn, N., Blenkinsop, S., Fowler, H. J., Freer, J., Tanguy, M., et al. (2019). *Gridded estimates of hourly areal rainfall for Great Britain (1990–2014)*. NERC Environmental Information Data Centre. [CEH-GEAR1hr]. https://doi.org/10.5285/d4ddc781-25f3-423a-bba0-747cc82dc6fa

Liu, J., & Han, D. (2010). Indices for calibration data selection of the rainfall-runoff model. *Water Resources Research*, *46*(4), W04512. https://doi.org/10.1029/2009wr008668

Liu, J., & Han, D. (2013). On selection of the optimal data time interval for real-time hydrological forecasting. *Hydrology and Earth System Sciences*, *17*(9), 3639–3659. https://doi.org/10.5194/hess-17-3639-2013

Moore, R. J. (2007). The PDM rainfall-runoff model. *Hydrology and Earth System Sciences*, *11*(1), 483–499. https://doi.org/10.5194/hess-11-483-2007

Moore, R. J., Bell, V. A., & Jones, D. A. (2005). Forecasting for flood warning. *Comptes Rendus Geoscience*, *337*(1–2), 203–217. https://doi.org/10.1016/j.crte.2004.10.017

Nandakumar, N., & Mein, R. G. (1997). Uncertainty in rainfall—Runoff model simulations and the implications for predicting the hydrologic effects of land-use change. *Journal of Hydrology*, *192*(1–4), 211–232. https://doi.org/10.1016/S0022-1694(96)03106-X

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Nonki, R. M., Lenouo, A., Lennard, C. J., Tshimanga, R. M., & Tchawoua, C. (2021). Comparison between dynamic and static sensitivity analysis approaches for impact assessment of different potential evapotranspiration methods on hydrological models performance. *Journal of Hydrometeorology*, *22*, 2713–2730. https://doi.org/10.1175/jhm-d-20-0192.1

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model? Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *Journal of Hydrology*, *303*(1–4), 290–306. https://doi.org/10.1016/j.jhydrol.2004.08.026

Oudin, L., Perrin, C., Mathevet, T., Andréassian, V., & Michel, C. (2006). Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models. *Journal of Hydrology*, *320*(1–2), 62–83. https://doi.org/10.1016/j.jhydrol.2005.07.016

Parmele, L. H. (1972). Errors in output of hydrologic models due to errors in input potential evapotranspiration. *Water Resources Research*, *8*(2), 348–359. https://doi.org/10.1029/WR008i002p00348

Paturel, J. E., Servat, E., & Vassiliadis, A. (1995). Sensitivity of conceptual rainfall-runoff algorithms to errors in input data—Case of the GR2M model. *Journal of Hydrology*, *168*(1–4), 111–125. https://doi.org/10.1016/0022-1694(94)02654-T

Penman, H. L. (1948). Natural evaporation from open water, bare soil and grass. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, *193*(1032), 120–145. https://doi.org/10.1098/rspa.1948.0037

Roberts, A. M., Cluckie, I. D., Gray, L., Griffith, R. J., Lane, A., Moore, R. J., & Pedder, M. A. (2000). Data management and data archive for the HYREX Programme. *Hydrology and Earth System Sciences*, *4*(4), 669–679. https://doi.org/10.5194/hess-4-669-2000

Robinson, E. L., Blyth, E., Clark, D. B., Finch, J., & Rudd, A. C. (2016). Climate hydrology and ecology research support system potential evapotranspiration dataset for Great Britain (1961–2015) [CHESS-PE]. https://doi.org/10.5285/8baf805d-39ce-4dac-b224-c926ada353b7

Samain, B., & Pauwels, V. R. N. (2013). Impact of potential and (scintillometer-based) actual evapotranspiration estimates on the performance of a lumped rainfall–runoff model. *Hydrology and Earth System Sciences*, *17*(11), 4525–4540. https://doi.org/10.5194/hess-17-4525-2013

Scanlon, B. R., Zhang, Z., Save, H., Sun, A. Y., Müller Schmied, H., van Beek, L. P. H., et al. (2018). Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. *Proceedings of the National Academy of Sciences*, *115*(6). https://doi.org/10.1073/pnas.1704665115

Sevruk, B. (1996). Adjustment of tipping-bucket precipitation gauge measurements. *Atmospheric Research*, *42*(1–2), 237–246. https://doi.org/10.1016/0169-8095(95)00066-6

Sheng, S., Chen, H., Guo, F. Q., Chen, J., Xu, C. Y., & Guo, S. L. (2020). Transferability of a conceptual hydrological model across different temporal scales and basin sizes. *Water Resources Management*, *34*(9), 2953–2968. https://doi.org/10.1007/s11269-020-02594-5

Shuttleworth, W. J. (1993). Evaporation. In D. R. Maidment (Ed.), *Handbook of hydrology*. McGraw–Hill.

Srivastava, P. K., Han, D., Rico-Ramirez, M. A., & Islam, T. (2014). Sensitivity and uncertainty analysis of mesoscale model downscaled hydro-meteorological variables for discharge prediction. *Hydrological Processes*, *28*(15), 4419–4432. https://doi.org/10.1002/hyp.9946

Wallingford, C. (2007). HYREX project: Brue River gauging station at Lovington. Retrieved from https://catalogue.ceda.ac.uk/uuid/3a64a223f1f2c3d2009f94e2d145d445

Wang, C., Wang, S., Fu, B., & Zhang, L. (2016). Advances in hydrological modelling with the Budyko framework. *Progress in Physical Geography: Earth and Environment*, *40*(3), 409–430. https://doi.org/10.1177/0309133315620997

Wang, J., Zhuo, L., Han, D., Liu, Y., & Rico-Ramirez, M. A. (2023). Hydrological model adaptability to rainfall inputs of varied quality. *Water Resources Research*, *59*(2), e2022WR032484. https://doi.org/10.1029/2022WR032484

Wood, S. J., Jones, D. A., & Moore, R. J. (2000). Accuracy of rainfall measurement for scales of hydrological interest. *Hydrology and Earth System Sciences*, *4*(4), 531–543. https://doi.org/10.5194/hess-4-531-2000

Xu, C. Y., Tunemar, L., Chen, Y. D., & Singh, V. P. (2006). Evaluation of seasonal and spatial variations of lumped water balance model sensitivity to precipitation data errors. *Journal of Hydrology*, *324*(1–4), 80–93. https://doi.org/10.1016/j.jhydrol.2005.09.019

Xu, C. Y., & Vandewiele, G. L. (1994). Sensitivity of monthly rainfall-runoff models to input errors and data length. *Hydrological Sciences Journal*, *39*(2), 157–176. https://doi.org/10.1080/02626669409492731

Xu, X., Yang, D., & Sivapalan, M. (2012). Assessing the impact of climate variability on catchment water balance and vegetation cover. *Hydrology and Earth System Sciences*, *16*(1), 43–58. https://doi.org/10.5194/hess-16-43-2012

Yang, X., Magnusson, J., Huang, S., Beldring, S., & Xu, C. Y. (2020). Dependence of regionalization methods on the complexity of hydrological models in multiple climatic regions. *Journal of Hydrology*, *582*, 124357. https://doi.org/10.1016/j.jhydrol.2019.124357

Ye, Y., Song, X., Zhang, J., Kong, F., & Ma, G. (2014). Parameter identification and calibration of the Xin'anjiang model using the surrogate modeling approach. *Frontiers of Earth Science*, *8*(2), 264–281. https://doi.org/10.1007/s11707-014-0424-0

Zhao, R. J. (1992). The Xinanjiang model applied in China. *Journal of Hydrology*, *135*(1–4), 371–381. https://doi.org/10.1016/0022-1694(92)90096-E

Zhao, R. J., Zhang, Y. L., & Fang, L. R. (1980). The Xinanjiang model. In *Hydrological forecasting proceedings oxford symposium* (Vol. 129, pp. 351–356). IAHS.

Zhuo, L., Dai, Q., & Han, D. (2015). Meta-analysis of flow modeling performances—To build a matching system between catchment complexity and model types. *Hydrological Processes*, *29*(11), 2463–2477. https://doi.org/10.1002/hyp.10371

Zhuo, L., & Han, D. (2016). Misrepresentation and amendment of soil moisture in conceptual hydrological modelling. *Journal of Hydrology*, *535*, 637–651. https://doi.org/10.1016/j.jhydrol.2016.02.033

Zhuo, L., Han, D., Dai, Q., Islam, T., & Srivastava, P. K. (2015). Appraisal of NLDAS-2 multi-model simulated soil moistures for hydrological modelling. *Water Resources Management*, *29*(10), 3503–3517. https://doi.org/10.1007/s11269-015-1011-1