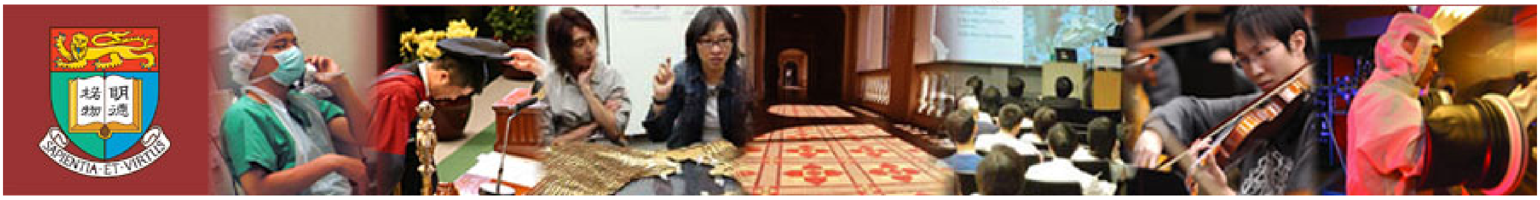
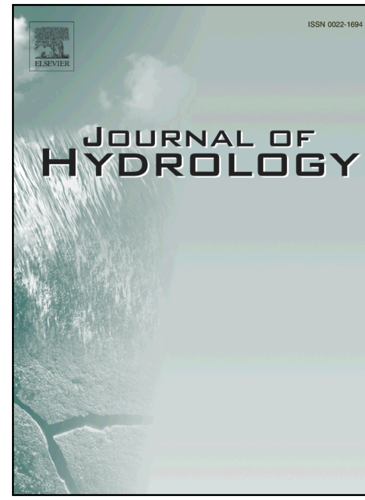


Title	A service-oriented architecture for ensemble flood forecast from numerical weather prediction
Author(s)	Shi, H; Li, T; Liu, R; Chen, J; Li, J; Zhang, A; Wang, G
Citation	Journal of Hydrology, 2015, v. 527, p. 933-942
Issued Date	2015
URL	http://hdl.handle.net/10722/215180
Rights	Copyright © 2015 Elsevier B.V.; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.



A service-oriented architecture for ensemble flood forecast from numerical weather prediction

Haiyun Shi, Tiejian Li, Ronghua Liu, Ji Chen, Jiaye Li, Ang Zhang, Guangqian Wang

PII: S0022-1694(15)00406-0
DOI: http://dx.doi.org/10.1016/j.jhydrol.2015.05.056
Reference: HYDROL 20491

To appear in: *Journal of Hydrology*

Received Date: 4 February 2015
Revised Date: 15 May 2015
Accepted Date: 28 May 2015

Please cite this article as: Shi, H., Li, T., Liu, R., Chen, J., Li, J., Zhang, A., Wang, G., A service-oriented architecture for ensemble flood forecast from numerical weather prediction, *Journal of Hydrology* (2015), doi: http://dx.doi.org/10.1016/j.jhydrol.2015.05.056

A service-oriented architecture for ensemble flood forecast from numerical weather prediction

Haiyun SHI[a,b,*], Tiejian LI[a,*], Ronghua LIU[c], Ji CHEN[b], Jiaye LI[a], Ang ZHANG[a], Guangqian WANG[a]

[a]*State Key Laboratory of Hydroscience and Engineering, Tsinghua University, Beijing, China*

[b]*Department of Civil Engineering, The University of Hong Kong, Pokfulam, Hong Kong, China*

[c]*China Institute of Water Resources and Hydropower Research, Beijing, China*

E-mail addresses:

Haiyun SHI: shihaiyun@mail.tsinghua.edu.cn

Tiejian LI: litiejian@tsinghua.edu.cn

Ronghua LIU: liurh@iwhr.com

Ji CHEN: jichen@hku.hk

Jiaye LI: thulijy@gmail.com

Ang ZHANG: zhanga11@mails.tsinghua.edu.cn

Guangqian WANG: dhhwgq@tsinghua.edu.cn

Revised manuscript for the *Journal of Hydrology*

May 2015

SUMMARY:

Floods in mountainous river basins are generally highly destructive, usually causing enormous losses of lives and property. It is important and necessary to develop an effective flood forecast method to prevent people from suffering flood disasters. This paper proposed a general framework for a service-oriented architecture (SOA) for ensemble flood forecast based on numerical weather prediction (NWP), taking advantage of state-of-the-art technologies, e.g., high-accuracy NWP, high-capacity cloud computing, and an interactive web service. With the predicted rainfall data derived from the NWP, which are automatically downloaded, hydrological models will be driven to run on the cloud. Judging from the simulation results and flood control requirements offered by users, warning information about possible floods will be generated for potential sufferers and then sent to them as soon as possible if needed. Moreover, by using web service in a social network, users can also acquire such information on the clients and make decisions about whether to prepare for possible floods. Along with the real-time updates of the NWP, simulation results will be refreshed in a timely manner, and the latest warning information will always be available to users. From the sample demonstrations, it is concluded that the SOA is a feasible way to develop an effective ensemble flood forecast method. After being put into practice, it would be valuable for preventing or reducing the losses caused by floods in mountainous river basins.

Keywords: Ensemble flood forecast; Numerical weather prediction; Service-oriented architecture; Cloud computing; Web service

1. Introduction

Mountainous regions, where the natural conditions are extremely complicated, account for nearly two-thirds of the total land area in China. In such regions, high-intensity rainstorms occur frequently during the flood period, which can lead to serious flood disasters and cause enormous losses of lives and property to the inhabitants living in the villages at the riversides or near the outlets of rivers (Caruso et al., 2013; Mazzorana et al., 2013; Ruiz-Villanueva et al., 2013; Shi and Wang, 2015). According to the statistic, there are 29 provinces, 274 prefecture-level cities and 1,836 county-level cities that suffer from flood disasters in China, covering an area of 4.63 million km^2 and involving 0.56 billion people. The key prevention and control area is 0.97 million km^2, involving 0.13 billion people, among which 74 million people suffer a direct threat (Chen, 2010). Specifically, the death toll caused by flash flood disasters accounted for two-thirds of the total death toll caused by all flood disasters every year in China before the 1990s; the percentage has risen to 80% since 2000. Approximately 4,000 people, accounting for 90% of the death toll caused by all flood disasters, were dead or missing in the flash flood disasters in 2010. This indicates that the situation of flood prevention and control will still be severe in the future; more technical and financial support should be provided for these flood-prone regions. Consequently, it is important and necessary to develop an effective flood forecast method to prevent people from having to suffer flood disasters.

Generally, traditional methods for flood forecast include the following two types. The first type comprises those methods based on critical rainfall, including the static and dynamic critical rainfall methods (Carpentera et al., 1999; Georgakakos, 2006; Liu et al., 2010). It is supposed that floods and some secondary disasters (e.g., debris flow and landslide) may occur in a river basin when the rainfall during a certain time interval reaches a certain amount or intensity (i.e., the static critical rainfall). Because the saturated degree of soil or the antecedent precipitation index has an important effect on the formation of floods, the critical rainfall should not be constant; thus, the dynamic critical rainfall method has been developed. Overall, the critical rainfall methods are easy to use, with no need for rainfall-runoff calculation; however, they cannot reflect the spatial variation of rainfall, making their applications in flood forecast limited. In contrast, the second type, which is applicable for river basins with sufficient, observed hydrological data, comprises those methods based on critical streamflow computed by using an empirical approach or hydrological models (Liu et al., 2005; Nayak et al., 2005; Cane et al., 2013; Moreno et al., 2013). Comparing simulation results against the flood control requirements, warnings concerning floods can be made early, if needed. Due to its high forecast accuracy, this type of method has been widely used; however, limited by the demand for massive observed data, it seems to be useless for river basins with poor data quality, especially for ungauged river basins.

China has dealt with the task of flood prevention and control for a long time. Moreover, more attention will be paid to floods in the future, and the construction of automatic weather stations, county-level data processing centers and early warning systems will be carried out in the near future. A number of provinces in China have set up their own flood warning systems; however, there are still gaps between the requirements for flood warning and reality. For example, a typical process of a flood warning system below the county level includes several levels (e.g., county, town, village, group and family). The rainfall and hydrological regime is reported to the superior level by level, and the warning information is sent to the inferior level by level as well. Although such a process is in accord with the status quo in China, the lengthy information transmission may considerably affect the efficiency. Thus, it is important and necessary to develop a high-efficiency method for flood forecast.

To this end, this paper aims to propose such a flood forecast method. Currently, numerical weather prediction (noted as NWP hereafter), at the global scale, has developed rapidly with the development of science and technology (Demeritt et al., 2007; Pappenberger et al., 2008). Furthermore, service-oriented architecture (noted as SOA hereafter) has been successfully applied in a wide variety of fields (Erl, 2005; Bell, 2008; Linthicum, 2009); however, the application of SOA for ensemble flood forecast cannot be found in the literature. As a result, this paper proposes the general framework of an SOA for ensemble flood forecast based on the NWP for the first time, taking advantage of the high-accuracy NWP, high-capacity cloud computing and an interactive web service. On the one hand, NWP is introduced to increase the forecast lead time; on the other hand, SOA is introduced to improve the forecast efficiency. In this study, the major challenges in developing such a flood forecast method are i) automatically downloading and updating the predicted rainfall from the NWP in real time, ii) implementing multiple scenarios for flood forecast at the same time by using high performance computing (noted as HPC hereafter) job scheduling, and iii) transferring warning information efficiently by using an interactive web service. Due to these state-of-the-art technologies, this method would be useful for preventing or reducing the losses caused by flood disasters in mountainous river basins after being put into practice.

2. Methodology

A conceptual framework of the SOA for ensemble flood forecast based on the NWP, combining the advantages of the high-accuracy NWP, high-capacity cloud computing and an interactive web service, is proposed in this paper (Fig. 1). In the following, the basic structures of the SOA, NWP, hydrological model used for ensemble flood forecast, HPC job scheduling used for multiple scenarios, and interactive web service used for information transfer will be introduced in detail.

2.1. Service-oriented architecture

Service-oriented architecture (SOA) is essentially a software design methodology based on structured collections of discrete services that collectively provide the complete functionality of a complex application (Erl, 2005). Each service is a well-defined, self-contained set of functions and built as a discrete piece of code, which makes it possible to reuse the code by changing only the interactions of a certain service with other ones rather than the code of the service. Moreover, services communicate with each other closely, involving either simple data passing or complex coordination (Bell, 2008, 2010). Hence, SOA is considered as the infrastructure supporting communications between services, and some connecting services are required. Currently, web service, a set of protocols enabling services to be published, discovered and used in a technology neutral form, seems to be the most feasible way for developing the SOA (Benslimane et al., 2008). By using a web service, a service consumer (e.g., the user client) can send a request message to a service provider (e.g., the cloud server), and then the service provider can return a response message to the service consumer as soon as possible (Linthicum, 2009). It is worth noting that the SOA is an architecture not only of services, as seen from a technology perspective, but also of policies, practices and frameworks, by which we can ensure that the right services are provided and consumed.

As shown in Fig. 1, the cloud server and user client are regarded as the two primary systems in the SOA. Data collection and management, hydrological simulation, flood forecast and early warning are achieved on the cloud server; meanwhile, messages of flood control requirements from different users and flood early warnings from the cloud server are transferred between the cloud server and user client by using a web service in a social network. Fig. 2 presents the flowchart of the SOA for ensemble flood forecast based on the NWP, and the details are introduced as follows.

First, the NWP data are downloaded automatically, in real time, from websites that provide relevant data and then stored in the database on the cloud server (see Section 2.2 for details); moreover, flood control requirements are provided by users on the clients and stored in the database as well. Second, a physically based hydrological model, the Digital Yellow River Integrated Model (noted as DYRIM hereafter) (Wang et al., 2007, 2015; Li et al., 2009a, 2009b), is adopted to compute the streamflow by using the NWP data; the simulation results are also stored in the database. Third, judging from the comparison of the simulation results against the flood control requirements from different users, flood early warnings are generated, if necessary; in addition, along with the real-time updates of the NWP data, the simulation results are refreshed in a timely manner so that the latest early warnings are always available. Finally, by using a web service, the flood forecast and early warning system on the cloud server can send warning information to the potential sufferers. Moreover, users on the clients can also run the DYRIM on the cloud server by themselves at any time to acquire the simulation results within the forecast lead time of the NWP data so that better preparation can be made for the possible flood disasters. Overall, it is clear that such a process of flood forecast and early warning is quite different from the current process and is useful to the potential sufferers, affording them much more response time for flood disasters.

2.2. Numerical weather prediction

Normally, numerical weather prediction (NWP) can be divided into two categories: single NWP and ensemble NWP. The single NWP is usually insufficient for flood forecast because it involves considerable uncertainties that may lead to lots of errors (Demeritt et al., 2007); meanwhile, the ensemble NWP may provide an opportunity to significantly improve the quality of flood forecast, including not only accuracy but also lead time (Pappenberger et al., 2008). It is believable that a more accurate prediction for atmospheric conditions and meteorological phenomena (e.g., rainfall) in the near future can be obtained from the ensemble NWP.

The THORPEX (i.e., The Observing System Research and Predictability Experiment) Interactive Grand Global Ensemble (noted as TIGGE hereafter) dataset is one of the acknowledged NWP datasets available at present (Richardson, 2005; Park et al., 2008). This dataset has been available since 2006 from ten institutions worldwide, including the Bureau of Meteorology (BoM, Australia), the China Meteorological Administration (CMA), the Canadian Meteorological Center (CMC), the Center for Weather Forecast and Climate Studies (CPTEC, Brazil), the European Centre for Medium-Range Weather Forecasts (ECMWF), the Japan Meteorological Agency (JMA), the Korea Meteorological

Administration (KMA), the MeteoFrance (MF), the US NCEP, and the United Kingdom Meteorological Office (UKMO). All of these datasets have the same temporal resolution of 6 hours, while their horizontal resolutions are quite different (e.g., 9/16° for the CMA and 1° for the NCEP and CMC). Moreover, the forecast lead time of the TIGGE data can be 1-16 days, which indicates that the TIGGE data can be a promising tool for short- or medium-term flood forecast.

Due to the uncertainty of the NWP, this indicates that the TIGGE datasets from different institutions can provide a variety of alternatives and seem to be more suitable for this study. Generally, all of these TIGGE datasets can be downloaded for free from their respective official websites (e.g., *http://tigge.ecmwf.int/*). To acquire the newest rainfall data in time for flood forecast, a method to automatically download and manage the NWP data is proposed, taking advantage of the Htmlunit (2013) and Apache CFX web services (2014). Three steps are introduced in detail as follows.

Step 1: Download the NWP data from the websites. In this step, operations on the web browser are simulated and realized by using the Htmlunit; two tasks, including login authentication and data downloading, are realized by using the source codes given in Appendix A.

Step 2: Manage the NWP data on the cloud server. In this step, the downloaded NWP data are interpreted first and then stored in a specific database. Moreover, the NWP data are converted into visual images in the TIFF format (see Appendix B for the source codes).

Step 3: Make the above two steps autorun. In this step, the above two functions (i.e., data downloading and management) are packaged into a one by using the Apache CFX web service. The program for downloading the NWP data is encapsulated as a service component, which runs automatically at a fixed time interval in the background of a server; moreover, this service is configured into the web server (i.e., Tomcat).

2.3. Digital Yellow River Integrated Model

Digital Yellow River Integrated Model (DYRIM) is a distributed model platform developed by Tsinghua University for hydrological and sediment simulations in river basins (Wang et al., 2007, 2015; Li et al., 2009a, 2009b). The DYRIM uses a high-resolution digital drainage network that is extracted from a digital elevation model (noted as DEM hereafter) (Bai et al., 2015) and coded using a modified binary tree method (Li et al., 2010) to simulate runoff yield and flow routing on each hillslope-channel unit. Moreover, dynamic parallel computing technology based on sub-basin decomposition has been developed to speed up the simulation (Li et al., 2011; Wang et al., 2011, 2012; Wu et al., 2013).

The DYRIM hydrological model is a physically based distributed model that represents the infiltration-excess runoff yield mechanism. This model uses a hillslope-channel as a basic hydrological unit because of the different hydrological response mechanisms of hillslopes and channels. The runoff-yield model is established based on the hillslope unit, where the soil mass is divided into topsoil and subsoil layers. A variety of hydrological processes are simulated, including vegetation interception, evapotranspiration, infiltration-excess runoff on the surface, subsurface flow in these two layers, and water exchange between these two layers. In the DYRIM hydrological model, the temporal resolution is six minutes and the rainfall data are uniformly assigned to each time step;

moreover, the rainfall data are spatially interpolated by using the inverse distance weighted method. It is worth noting that parameters in the DYRIM hydrological model can be divided into two types: (i) invariant parameters used for describing the properties of land use and soil type, influenced by the basic features of the river basin and determined from the literature, fieldwork and prior studies; and (ii) adjustable parameters that are calibrated and verified with the observed data.

2.4. HPC job scheduling

HPC is an important branch of computer science that focuses on the development of high performance computers and relevant software. It is a technology that can improve the capability of scientific computing through organizing a number of processors or computers as members of a cluster; it is based on parallel computing technology, a way of enabling an application to be divided into multiple parts that can be executed in parallel multiple processors. There are several types of HPC systems (e.g., large clusters and highly specialized hardware), most of which are based on clusters and interconnect with each other by using a high performance network, e.g., the Quad Data Rate (QDR) InfiniBand network. HPC allows scientists and engineers to solve complex scientific, engineering and business problems by using applications that require high bandwidth, low latency networking, and very high computing capability. In the future, HPC will be more networked, open, standard, structured and diversified in application. For example, in the field of hydrological simulation, HPC can be used when a parallel hydrological model (e.g., the DYRIM in this study) is applied in large-scale river basins.

The framework of the SOA for ensemble flood forecast based on the NWP in this study tries to provide a two-layer parallelism (Fig. 3). The lower layer is the parallelism in the DYRIM hydrological model; the upper layer is the parallelism in the hydrological simulations with the NWP data from different institutions, which is realized by using a job scheduling function. Moreover, the Windows HPC Server 2012 used in this study has 24 compute nodes with 20 processor cores on each of them, i.e., 480 processor cores in total. Generally, one processor core can execute only one process each time; using more processor cores at one time means less time consumption. The number of processor cores used for hydrological simulation at one time can significantly affect the efficiency of the lower-layer parallelism and further affect the efficiency of the upper-layer parallelism. Namely, if N processor cores are used for hydrological simulation with the NWP data from one institution, then hydrological simulations with the NWP data from INT(480/N) (note: the symbolic function INT(X) means the integer part of a real number X) institutions can be carried out at the same time by using the HPC job scheduling.

2.5. Web service

An interactive web service is developed to receive the flood control requirements from the users and send early warnings to the users. Moreover, it is also used for queries on a variety of hydrological information (e.g., the digital drainage network, historical and predicted rainfall data, and streamflow predictions). All of the data are stored in the databases on the cloud server and can be inquired by the user clients at any time. For example, based on the global drainage network (Bai et al., 2015) extracted from the 30-m-resolution Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)

Global DEM dataset (ASTER GDEM Validation Team, 2009, 2011), users can define a watershed by specifying the location of the watershed outlet (i.e., longitude and latitude) and the corresponding river reach. Then, the drainage network of the entire watershed will be selected for hydrological simulation (i.e., flood forecast).

3. Results and discussions

In this study, two river basins in China, including the Juma River basin in the southwest suburb of Beijing and the upper Baishui River basin in the north of Sichuan province, are regarded as the study areas for the application of the SOA for ensemble flood forecast based on the NWP. In the following, the available research data used for each case are introduced, and the results as well as discussions are presented.

3.1. Case study of the Juma River basin

The Juma River basin is located in the southwest of Beijing (114°27'-115°47' E, 39°12'-40°04' N). As shown in Fig. 4, there is only one hydrological station (i.e., the Zhangfang hydrological station) in this river basin; the drainage area in the upstream of this station is over 3,800 km^2. The high-resolution digital drainage network is also shown in Fig. 4; there are 25,833 river reaches and nearly 65,000 hillslopes in total in the extracted digital drainage network.

The Juma River basin was severely affected by the notorious rainstorm on July 21, 2012, in Beijing. This rainstorm was characterized by a large rainfall depth, long duration and high intensity. According to the information from the Beijing Water Authority, this rainstorm lasted for nearly 16 hours, and the mean rainfall depth of the whole city was 170 mm, with a significantly uneven spatial distribution. For example, the Fangshan District, in the southwest of Beijing, had the maximum rainfall depth of 301 mm, while Yanqing County, in the northwest, had the minimum rainfall depth of 69 mm. The area with a rainfall depth over 200 mm was approximately 6,000 km^2, covering 36% of the total area of Beijing, and the largest point rainfall (i.e., 460 mm, with a return period of 500 years) occurred in the Fangshan District. As a result, approximately 1.9 million people had property loss, and among them, 0.8 million were in the Fangshan District; furthermore, there were 79 persons killed due to this rainstorm.

To accurately forecast floods due to severe rainfall, the reliable estimation of rainfall is paramount; then, hydrological models can be used to forecast streamflow with more accuracy. In this case, the TIGGE data derived from six of the above-mentioned ten institutions (i.e., the CMA, CMC, CPTEC, ECMWF, NCEP and UKMO) are selected as the research data (i.e., the predicted rainfall data) because there was no data available from the other four institutions for the period of this rainstorm. After being downloaded automatically in real time from the official websites (see Section 2.2 for details), the NWP data are used as the basic input data for the follow-up streamflow simulation and flood forecast. The NWP data (i.e., the TIGGE data in this study) are considered to be an important factor that can affect the result of streamflow simulation; thus, it is possible that using the NWP data derived from different sources may lead to different flood forecast results.

To investigate the features of the various NWP data, the spatial distributions of total rainfall depth in Beijing and the surrounding area during the time period of 0:00-24:00 UTC, July 21, 2012, which were described by the TIGGE data derived from six institutions (i.e., the CMA, CMC, CPTEC, ECMWF, NCEP and UKMO) released at 0:00 UTC, July 21, 2012, were shown in Fig. 5. Overall, the six TIGGE datasets have significantly different features for describing the spatial distribution of rainfall depth. On the one hand, the maximum values of total rainfall depth inside the Juma River basin during this period varied widely, e.g., 58.5 mm for the CMA data, 68.9 mm for the CMC data, 61.2 mm for the CPTEC data, 92.5 mm for the ECMWF data, 134.8 mm for the NCEP data and 151.9 mm for the UKMO data; the highest value was nearly three times as much as the lowest one. On the other hand, the rainfall centers described by these six TIGGE datasets during this period appeared in different locations. For example, the rainfall center described by the ECMWF data appeared in the downstream of the Juma River basin (Fig. 5d) but that described by the UKMO data appeared in the upstream of this river basin (Fig. 5f); moreover, for the NCEP data, heavy rainfall almost covered the entire river basin (Fig. 5e).

Furthermore, to evaluate the performances of the various NWP data in flood forecast, streamflow processes of the reach where the Zhangfang hydrological station is located were computed with these six TIGGE datasets by using the DYRIM hydrological model and HPC job scheduling. Because the features of these six TIGGE datasets for describing the spatial distribution of rainfall depth were significantly different, it is presumable that the simulation results would be different (see Fig. 6). In this study, the parameters derived from the previous study (Shi, 2013) were directly used, and Fig. 6 shows the comparisons of streamflows computed with the six TIGGE datasets against the observed data recorded at the Zhangfang hydrological station. Overall, all of the simulated values computed with these six TIGGE datasets were not close to the observed ones. Only by using the NCEP data or UKMO data could the peak flow be simulated; however, both the peak value and appearance time were not accurate when they were compared with those presented by the observed data (approximately 2,500 m^3/s appeared at 23:00 UTC, July 21, 2012). By using the NCEP data, the computed peak value was approximately 3,800 m^3/s, 52% larger than the observed peak value, and the appearance time was six hours in advance; by using the UKMO data, the computed peak value was approximately 2,600 m^3/s, only 4% larger than the observed peak value, and the appearance time was only four hours in advance. It is inferred that the UKMO data showed much better performance than the NCEP data for this case. In addition, no peak flows could be simulated by using the other four TIGGE datasets (i.e., the CMA, CMC, CPTEC and ECMWF).

Furthermore, Table 1 lists the results of streamflow simulation by using these six TIGGE datasets. It is observed that the simulation results were markedly different as a whole. The values of flood volume were 3.92×10^6 m^3 (-95.27%) for the CMA data, 3.98×10^6 m^3 (-95.20%) for the CMC data, 3.94×10^6 m^3 (-95.26%) for the CPTEC data, and 3.92×10^6 m^3 (-95.27%) for the ECMWF data. For these four TIGGE datasets, the intensity of the predicted rainfall was not high enough for the runoff yield; the computed streamflow was actually the base flow, resulting in no peak flows appearing. If so, no floods would occur in the Juma River basin, which indicated that people living in this river basin would be safe. In contrast, the high intensity of the predicted rainfall for the other two datasets led to extremely large values of flood volume, e.g., 156.76×10^6 m^3 (88.98%) for the NCEP data and 112.17×10^6 m^3 (35.23%) for the UKMO data; moreover, as mentioned above, the peak values were very large as well (i.e., 3,800 m^3/s for the NCEP data and 2,600 m^3/s for the UKMO data); enormous losses of lives and property would be caused by such large floods if they came true.

3.2. Case study of the upper Baishui River basin

The upper Baishui River basin is located in the north of Sichuan province (103°22'-103°47' E, 33°06'-33°40' N). The region in the upstream of the Batun hydrological station with an area of 1,198 km^2 is considered in this study (Fig. 7). The four rainfall stations with hourly observed data are also shown in Fig. 7. Moreover, there are 7,019 river reaches and nearly 17,500 hillslopes in total in the extracted digital drainage network.

For this river basin, the flood occurred during July 16-23, 2010 is regarded as the study case. The TIGGE data derived from four of the above-mentioned ten institutions (i.e., the CMA, CPTEC, ECMWF and UKMO) are selected as the research data (i.e., the predicted rainfall data), as the data from the other six institutions are not available during this period. Fig. 8 shows the comparison of the simulated streamflows calibrated with the observed station rainfall against the observed streamflows recorded at the Batun station, and the results were generally satisfactory when they were compared with those presented by the observed data (approximately 75 m^3/s appeared at 3:00 UTC, July 17, 2010), with peak value error of -17% and peak time error of six hours delay (Table 2). Fig. 8 also shows the comparisons of streamflows computed with the four TIGGE datasets against the observed data recorded at the Batun station. Overall, the results computed with these four TIGGE datasets were not so close to the observed data. Only by using the ECMWF data or CMA data could the peak flow be simulated; however, both the peak value and appearance time were not accurate. Table 2 also lists the comparisons of streamflows computed with the observed station rainfall and the four TIGGE predicted rainfall inputs against the observed data recorded at the Batun station. Using the predicted rainfall, the simulated peak value and time were 36.0 m^3/s (-52%) and six hours later for the ECMWF data and

30.4 m^3/s (-56%) and 7 hours later for the CMA data; no peak flow could be forecasted by using the CPTEC and UKMO data.

3.3. Discussions

From the case studies, it can be seen that the results of flood forecast obtained by using different NWP data can be markedly different, even completely opposite. To this end, it is important and necessary to provide these various results of streamflow computation by using different NWP data for users at the same time. In this study, implementing multiple scenarios of flood forecast (i.e., streamflow simulations by using different NWP data) at the same time can be realized by using the two-layer paralleled HPC job scheduling on the cloud server (see Section 2.4 for details). Generally, these simulations can be completed within a few minutes (e.g., 3 minutes for the first case and 2 minutes for the second case). Thereafter, all of the simulation results will be compared with the flood control requirements offered by users, and the probability of flood can be described by the percentage of possible floods that are simulated with different NWP data (i.e., 2 in 6 for the first case and 2 in 4 for the second case). If needed, relevant warning information of floods will be generated and sent to potential sufferers immediately. In addition, users on the clients can also acquire such warning information by using the web service in a social network at any time.

Generally, methods for flood risk and vulnerability analyses have been proposed for ensemble flood forecast (UNDRO, 1991; Willows and Connell, 2003; Wu et al., 2012). For river basins with sufficient historical hydrological data, the frequency of the predicted peak flow from each NWP data can be obtained from the probability distribution function derived from the observed streamflow series. Meanwhile, the critical frequency of flood to cause potential disaster can be determined for each NWP agency. Thereafter, through comparing the forecasted frequency of peak flow from each NWP agency against its critical frequency, flood risk degree can be evaluated separately. Finally, a method to generate a synthetic warning from the separate risks, considering different weights according to the historical performance of each NWP agency, is needed. This method is hoped to be adopted in the proposed system in future work. However, for river basins with no historical hydrological data, this method is still tough to succeed. Therefore, the proposed system can only provide the various flood forecast results for users at this stage. After being put into practice for years, accumulation of hydrological data may make a result interpretation method more applicable to the proposed system, which is much more useful to provide the decision-support for users.

Furthermore, it can be inferred that the discrepancies in peak values and times are mainly caused by the low temporal-spatial resolution of the predicted rainfall data. For flood forecast in any given river basin, the globally available predicted rainfall from the NWP (e.g., TIGGE) data is not accurate enough. Nevertheless, the proposed system has made the techniques and the platform ready for better flood forecast, when better predicted rainfall data can be obtained from much finer national and regional NWP data.

4. Conclusions

This paper proposed a conceptual framework for the SOA for ensemble flood forecast from the NWP, combining the advantages of state-of-the-art technologies, e.g., high-accuracy NWP, high-capacity cloud computing and an interactive web service. The significance of this paper can be concluded as follows: first, a method to automatically download and update the predicted rainfall derived from the NWP (e.g., the TIGGE data) in real time was developed. Second, HPC job scheduling was adopted to implement multiple scenarios of flood forecast at the same time; accordingly, various results of streamflow simulation could be provided, and the latest warning information of floods could be generated for potential sufferers. Third, by using the interactive web service in a social network, users can either acquire such warning information on the clients at any time or be informed to prepare for possible floods. It is concluded that the SOA will be a feasible way for ensemble flood forecast based on the NWP, affording potential sufferers much more response time when confronted with possible floods. After being put into practice, the proposed system would be useful for preventing or reducing the losses caused by flood disasters in mountainous river basins.

Acknowledgements

This study was supported by the National Science & Technology Pillar Program in the Twelfth Five-year Plan Period (Grant No. 2013BAB05B03, 2013BAB05B05), the China Postdoctoral Science Foundation funded project (Grant No. 2014M550069) and the Hong Kong Scholars Program project (Grant No. XJ2014059). We are also grateful to the two anonymous reviewers who offered insightful comments leading to the improvement of this paper.

Appendix A

Source codes for downloading the NWP data from websites are given as follows:

Login authentication:

```
WebClient client = new WebClient(BrowserVersion.FIREFOX_10);
HtmlPage  homePage = client.getPage("URL");
HtmlInput name = homePage.getInputByName("Name");
name.setValueAttribute("Value");
HtmlPage loginPage = homePage.getAnchorByText("Name of Link").click();
```

Data downloading:

```
URL url = new URL(downloadurl);
URLConnection conn = url.openConnection();
InputStream inStream = conn.getInputStream();
filestream = new FileOutputStream("@Path"+filename+".grib");
datewriter.write(filename);
datewriter.flush();
byte[] buffer = new byte[1204];
while ((byteread = inStream.read(buffer)) != -1){
        bytesum += byteread;
        filestream.write(buffer, 0, byteread);
}
```

Appendix B

Source codes for the NWP data interpretation are given as follows:

Data interpretation:

```
public class GridData; //the class of rainfall data
String filepath = "F:\\data.grib";
File fileptr = new File(filepath);
FileInputStream filestream = new FileInputStream(fileptr);
BufferedInputStream bufferstream = new BufferedInputStream(filestream);
private void GridSection(BufferedInputStream bufferstream, List<GridData> Points);
private void ProductSection(BufferedInputStream bufferstream);
private void DataSection(BufferedInputStream bufferstream, List<Double> PointsValue) throws
IOException {
        int Length = ConvertInt(4, bufferstream, "the total length");
        ConvertInt(1, bufferstream, "the serial number");
        bufferstream.mark(Integer.MAX_VALUE);
        for (int i = 0; i < TotalNumberofPoints; i++) {
                double PValue = ConvertPoint(24, bufferstream, "第" + (i + 1) + "点： ");
                PointsValue.add(PValue);
        }
        bufferstream.reset();
        bufferstream.skip(Length - 5);
}
```

References

Apache CFX, 2014. <http://cxf.apache.org/>.

ASTER GDEM Validation Team. 2009. ASTER global DEM validation summary report. METI & NASA.

ASTER GDEM Validation Team. 2011. ASTER global DEM version 2 – summary of validation results. METI & NASA.

Bai, R., Li, T.J., Huang, Y.F., Li, J.Y., Wang, G.Q., 2015. An efficient and comprehensive method for drainage network extraction from DEM with billions of pixels using a size-balanced binary search tree. Geomorphology, DOI: 10.1016/j.geomorph.2015.02.028.

Bell, M., 2008. Service-Oriented Modeling: Service Analysis, Design, and Architecture. Wiley, New Jersey.

Bell, M., 2010. SOA Modeling Patterns for Service Oriented Discovery and Analysis. Wiley, New Jersey.

Benslimane, D., Dustdar, S., Sheth, A., 2008. Services Mashups: The New Generation of Web Applications. IEEE Internet Computing, 10(5), 13-15.

Cane, D., Ghigo, S., Rabuffetti, D., Milelli, M., 2013. Real-time flood forecasting coupling different postprocessing techniques of precipitation forecast ensembles with a distributed hydrological model. The case study of May 2008 flood in western Piemonte, Italy. Natural Hazards and Earth System Sciences, 13(2), 211-220.

Carpentera, T.M., Sperfslage, J.A., Georgakakos, K.P., Sweeney, T., Fread, D.L., 1999. National threshold runoff estimation utilizing GIS in support of operational flash flood warning systems. Journal of Hydrology, 224, 21-44.

Caruso, B.S., Rademaker, M., Balme, A., Cochrane, T.A., 2013. Flood modelling in a high country mountain catchment, New Zealand: comparing statistical and deterministic model estimates for ecological flows. Hydrological Sciences Journal, 58(2), 328-341.

Chen, L., 2010. The speech in the start video conference of the construction of the non-engineering measures in the mountain flood prevention at the county level. Beijing. [In Chinese]

Demeritt, D., Cloke, H., Pappenberger, F., Thielen, J., Bartholmes, J., Ramos, M.-H., 2007. Ensemble predictions and perceptions of risk, uncertainty, and error in flood forecasting. Environmental Hazards, 7(2), 115-127.

Erl, T., 2005. Service-Oriented Architecture: Concepts, Technology, and Design. Prentice Hall, New Jersey.

Georgakakos, K.P., 2006. Analytical results for operational flash flood guidance. Journal of Hydrology, 317, 81-103.

HtmlUnit, 2013. <http://htmlunit.sourceforge.net/>.

Li, T.J., Wang, G.Q., Chen, J., 2010. A modified binary tree codification of drainage networks to support complex hydrological models. Computers & Geosciences, 36(11), 1427-1435.

Li, T.J., Wang, G.Q., Chen, J., Wang, H., 2011. Dynamic parallelization of hydrological model simulations. Environmental Modelling & Software, 26, 1736-1746.

Li, T.J., Wang, G.Q., Huang, Y.F., Fu, X.D., 2009a. Modeling the Process of Hillslope Soil Erosion in the Loess Plateau. Journal of Environmental Informatics, 14(1), 1-10.

Li, T.J., Wang, G.Q., Xue, H., Wang, K., 2009b. Soil erosion and sediment transport in the gullied Loess Plateau: Scale effects and their mechanisms. Science in China Series E - Technological Sciences, 52(5), 1283-1292.

Linthicum, D.S., 2009. Cloud Computing and SOA Convergence in Your Enterprise: A Step-by-Step Guide. Pearson Education, New Jersey.

Liu, Z.Y., Martina, M.L., Todini, E., 2005. Flood forecasting using a fully distributed model: application of the TOPKAPI model to the Upper Xixian Catchment. Hydrology and Earth System Sciences, 9(4), 347-364.

Liu, Z.Y., Yang, D.W., Hu, J.W., 2010. Dynamic critical rainfall-based torrential flood early warning for medium-small rivers. Journal of Beijing Normal University, 46(3), 317-318. [In Chinese]

Mazzorana, B., Comiti, F., Fuchs, S., 2013. A structured approach to enhance flood hazard assessment in mountain streams. Natural Hazards, 67(3), 991-1009.

Moreno, H.A., Vivoni, E.R., Gochis, D.J., 2013. Limits to flood forecasting in the Colorado Front Range for two summer convection periods using radar nowcasting and a distributed hydrologic model. Journal of Hydrometeorology, 14(4): 1075-1097.

Nayak, P.C., Sudheer, K.P., Ramasastri, K.S., 2005. Fuzzy computing based rainfall-runoff model for real time flood forecasting. Hydrological Processes, 19(4), 955-968.

Pappenberger, F., Bartholmes, J., Thielen, J., Cloke, H.L., Buizza, R., Roo, A., 2008. New dimensions in early flood warning across the globe using grand-ensemble weather predictions. Geophysical Research Letters, 35(10), L10404.

Park, Y.-Y., Buizza, R., Leutbecher, M., 2008. TIGGE: Preliminary results on comparing and combining ensembles. European Centre for Medium-Range Weather Forecasts, Reading.

Richardson, D., 2005. The THORPEX Interactive Grand Global Ensemble (TIGGE). Geophysical Research Abstracts, 7, Abstract EGU05-A-02815.

Ruiz-Villanueva, V., Bodoque, J.M., Diez-Herrero, A., Eguibar, M.A., Pardo-Iguzquiza, E., 2013. Reconstruction of a flash flood with large wood transport and its influence on hazard patterns in an ungauged mountain basin. Hydrological Processes, 27(24), 3424-3437.

Shi, H.Y., 2013. Computation of spatially distributed rainfall by merging raingauge measurements, satellite observations and topographic information: A case study of the 21 July 2012 rainstorm in Beijing, China. The 35th IAHR World Congress, Chengdu, China.

Shi, H.Y., Wang, G.Q., 2015. Impacts of climate change and hydraulic structures on runoff and sediment discharge in the middle Yellow River. Hydrological Processes, DOI: 10.1002/hyp.10439.

United Nations Disaster Relief Organization (UNDRO), 1991. Mitigating Natural Disasters: Phenomena, Effects and Options: A Manual for Policy makers and Planners. New York: United Nations, 1-164.

Wang, G.Q., Fu, X.D., Shi, H.Y., Li, T.J., 2015. Watershed Sediment Dynamics and Modeling: A Watershed Modeling System for Yellow River. In Yang C.T. and Wang L.K. (eds), Advances in Water Resources Engineering, Handbook of Environmental Engineering, Volume 14. Springer International Publishing.

Wang, G.Q., Wu, B.S., Li, T.J., 2007. Digital Yellow River model. Journal of Hydro-Environment Research, 1, 1-11.

Wang, H., Fu, X.D., Wang, G.Q., Li, T.J., Gao, J., 2011. A common parallel computing framework for modeling hydrological processes of river basins. Parallel Computing, 37, 302-315.

Wang, H., Zhou, Y., Fu, X.D., Gao, J., Wang, G.Q., 2012. Maximum speedup ratio curve (MSC) in parallel computing of the binary-tree-based drainage network. Computers & Geosciences, 38, 127-135.

Willows, R.I., Connell, R.K., 2003. Climate adaptation: Risk, uncertainty and decision-making. United Kingdom: UK Climate Impacts Programme, 39(9): 829-840.

Wu, H., Adler, R.F., Hong, Y., Tian, Y.D., Policelli, F., 2012. Evaluation of Global Flood Detection Using Satellite-Based Rainfall and a Hydrologic Model. Journal of Hydrometeorology, 13, 1268-1284.

Wu, Y.P., Li, T.J., Sun, L.Q., Chen, J., 2013. Parallelization of a hydrological model using the message passing interface. Environmental Modelling & Software, 43, 124-132.

ACCEPTED MANUSCRIPT

List of figure captions

Fig. 1. The framework of the service-oriented architecture (SOA) for ensemble flood forecast from numerical weather prediction (NWP).

Fig. 2. The flowchart of the SOA for ensemble flood forecast from the NWP.

Fig. 3. The HPC job scheduling used in this study.

Fig. 4. The location and the digital drainage network of the Juma River basin.

Fig. 5. The spatial distributions of rainfall depth in Beijing and the surrounding area during 0:00-24:00 UTC, July 21, 2012, which were described by six TIGGE datasets (i.e., the CMA, CMC, CPTEC, ECMWF, NCEP and UKMO) released at 0:00 UTC, July 21, 2012.

Fig. 6. Comparison of the streamflows computed with the six TIGGE datasets against the observed streamflow data recorded at the Zhangfang hydrological station.

Fig. 7. The location and the digital drainage network of the upper Baishui River basin.

Fig. 8. Comparison of the streamflows computed with the observed station rainfall and the four TIGGE datasets against the observed streamflow data recorded at the Batun hydrological station.

Figure 1

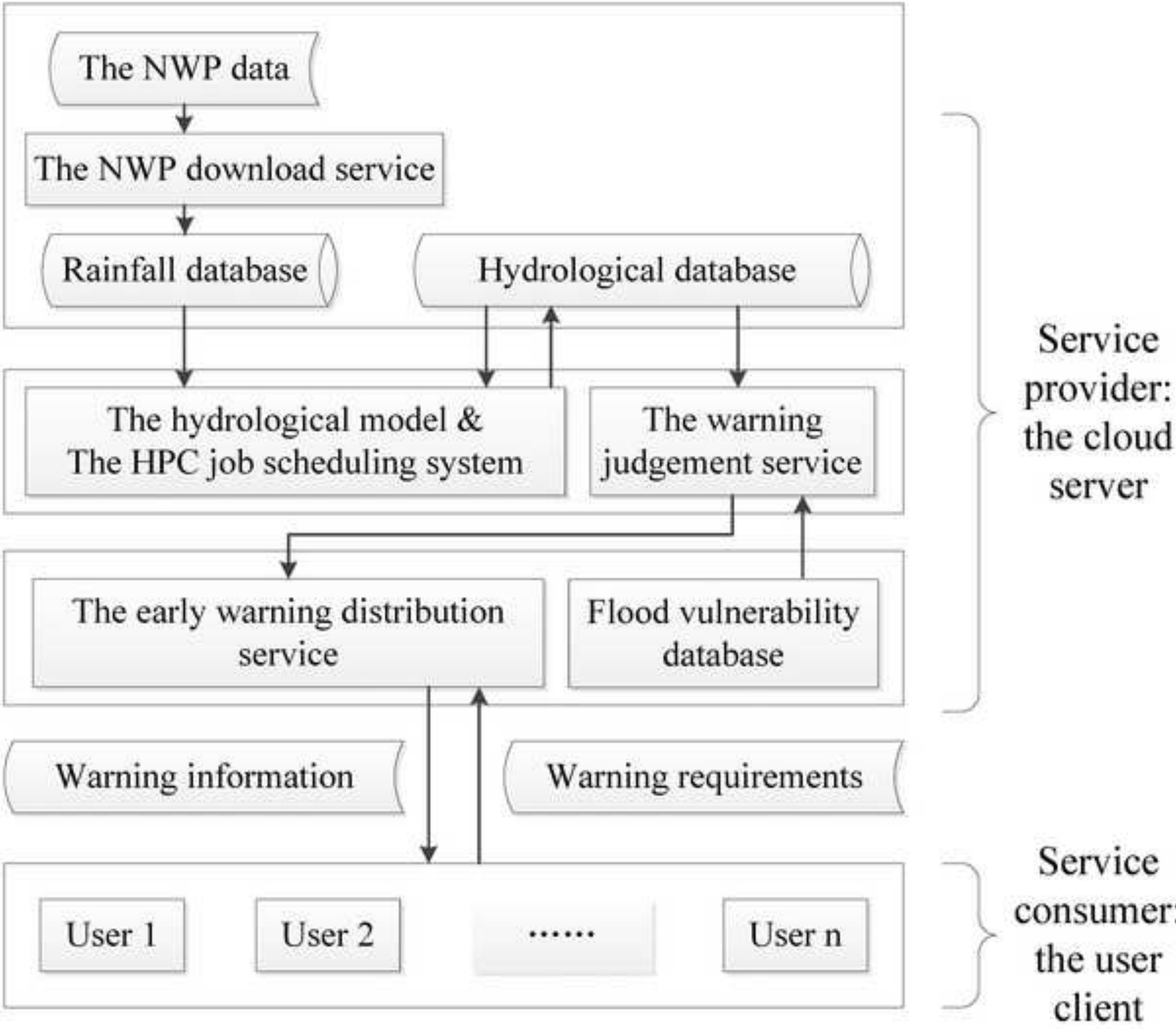

The NWP data
The NWP download service
Rainfall database
Hydrological database
The hydrological model &
The HPC job scheduling system
The warning judgement service
The early warning distribution service
Flood vulnerability database
Warning information
Warning requirements
User 1
User 2
......
User n
Service provider: the cloud server
Service consumer: the user client

Figure 2

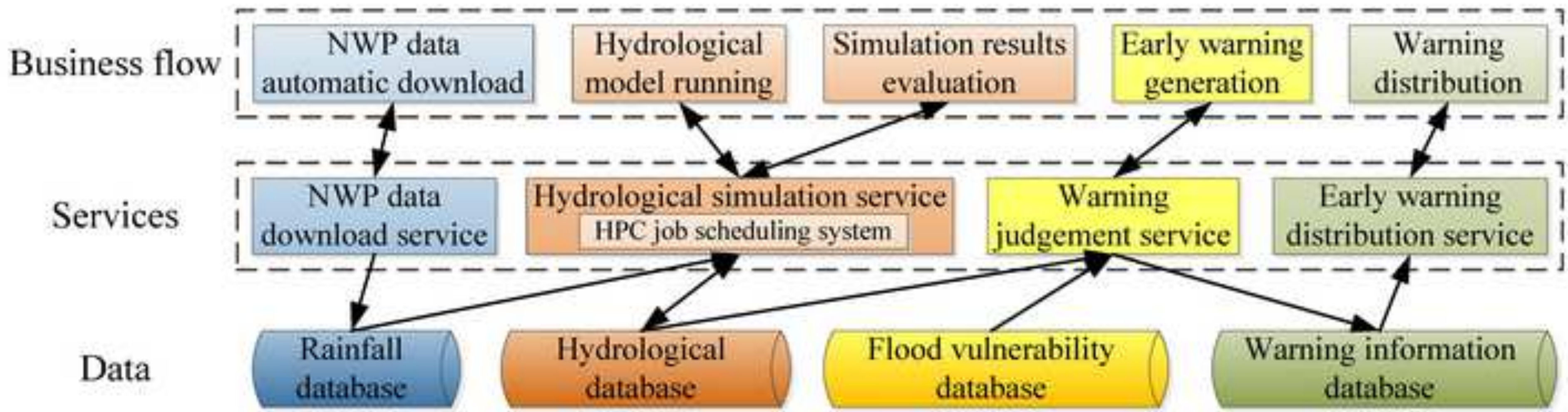

Business flow
NWP data automatic download
Hydrological model running
Simulation results evaluation
Early warning generation
Warning distribution
Services
NWP data download service
Hydrological simulation service
HPC job scheduling system
Warning judgement service
Early warning distribution service
Data
Rainfall database
Hydrological database
Flood vulnerability database
Warning information database

Figure 3

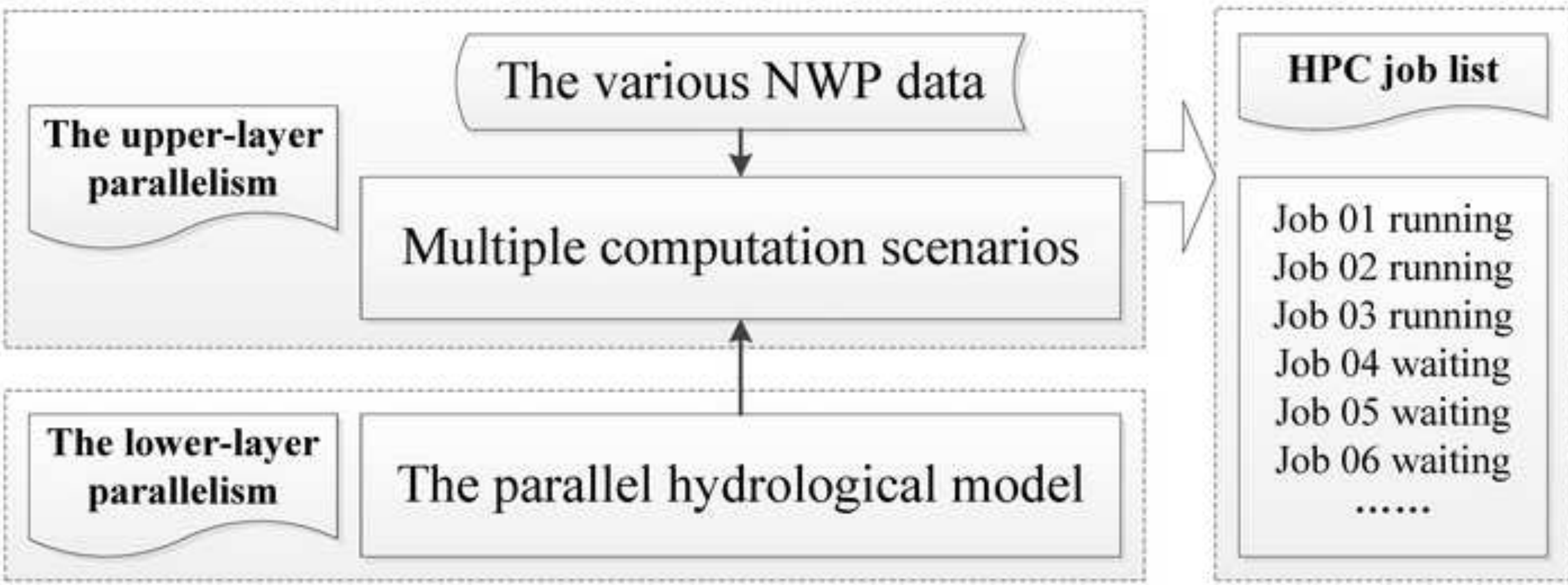
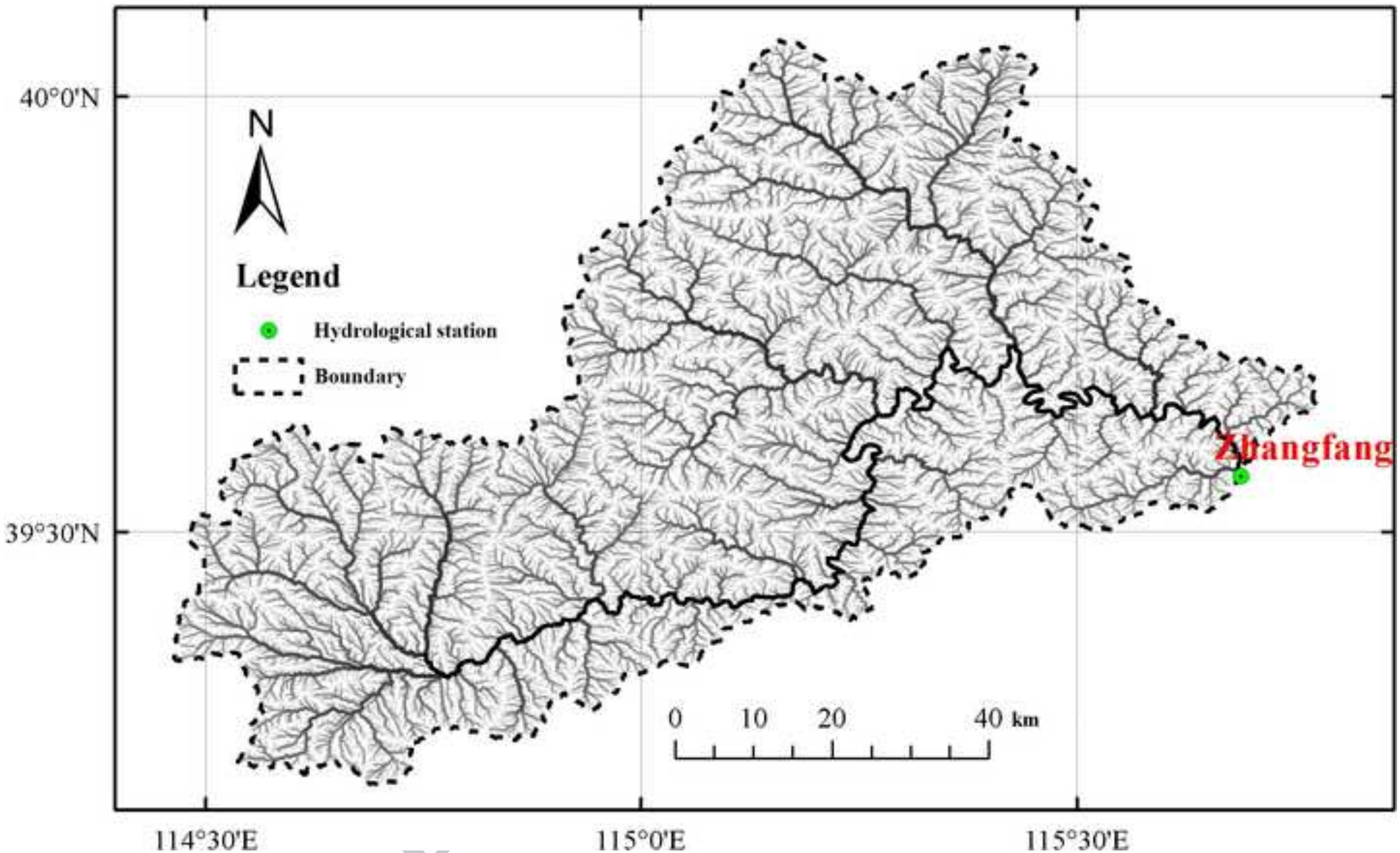
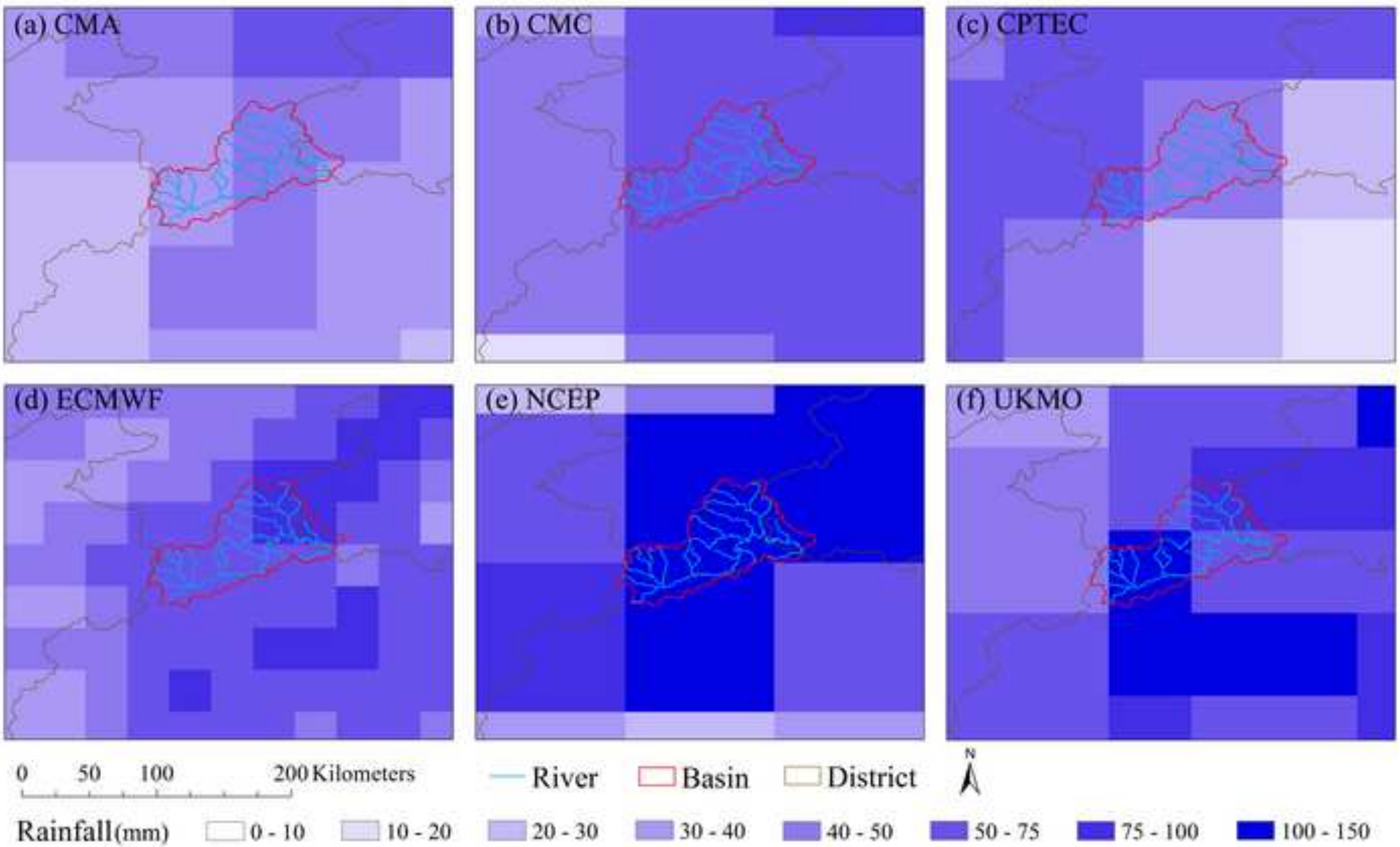

The upper-layer parallelism
The various NWP data
Multiple computation scenarios
The lower-layer parallelism
The parallel hydrological model
HPC job list
Job 01 running
Job 02 running
Job 03 running
Job 04 waiting
Job 05 waiting
Job 06 waiting
......

Figure 4



40°0'N
39°30'N
N
Legend
Hydrological station
Boundary
Zhangfang
0
10
20
40 km
114°30'E
115°0'E
115°30'E

Figure 5



(a) CMA
(b) CMC
(c) CPTEC
(d) ECMWF
(e) NCEP
(f) UKMO
0 50 100 200 Kilometers
River
Basin
District
N
Rainfall(mm)
0 - 10
10 - 20
20 - 30
30 - 40
40 - 50
50 - 75
75 - 100
100 - 150

Figure 6

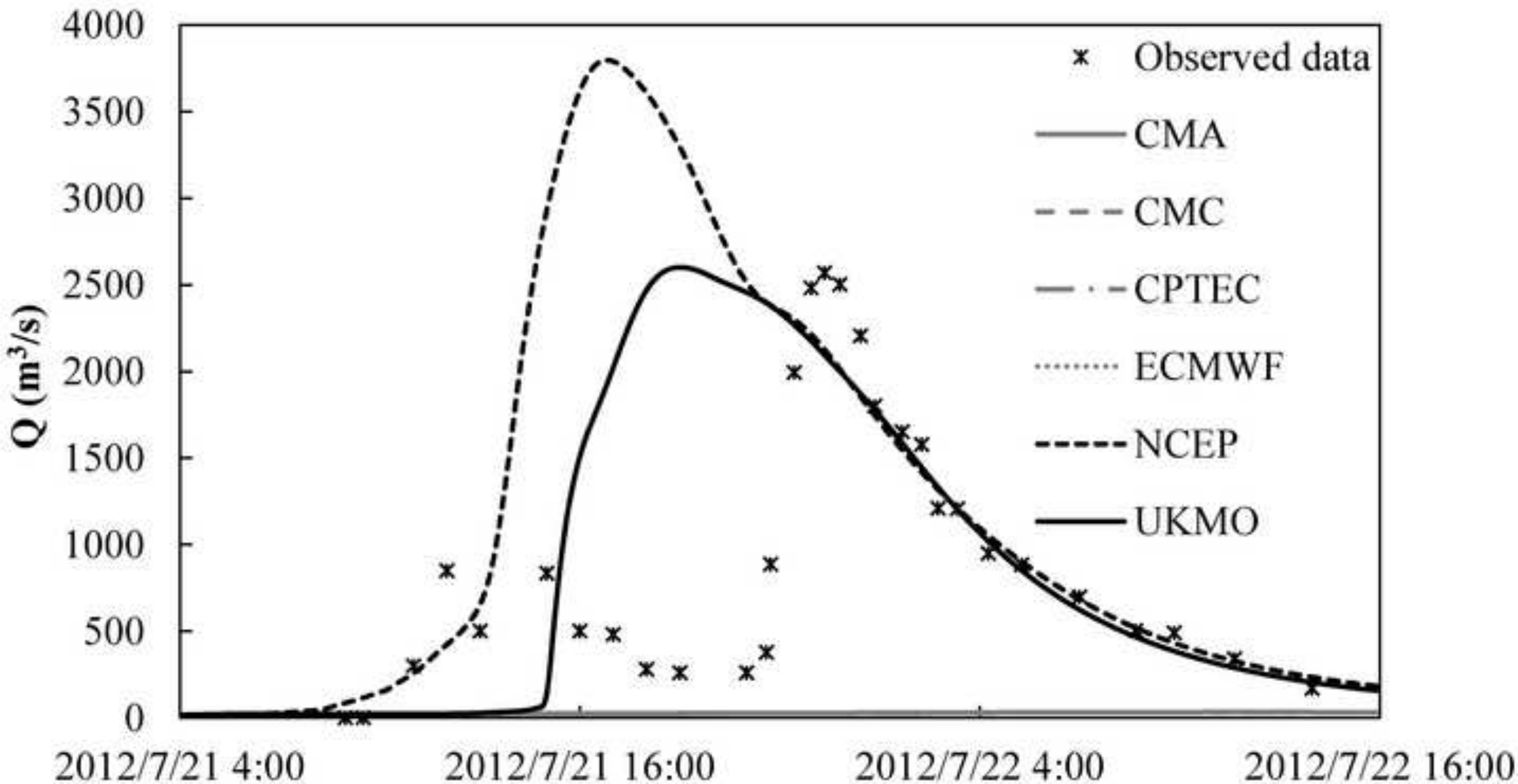

Q (m³/s)
4000
3500
3000
2500
2000
1500
1000
500
0
2012/7/21 4:00
2012/7/21 16:00
2012/7/22 4:00
2012/7/22 16:00
Observed data
CMA
CMC
CPTEC
ECMWF
NCEP
UKMO

Figure 7

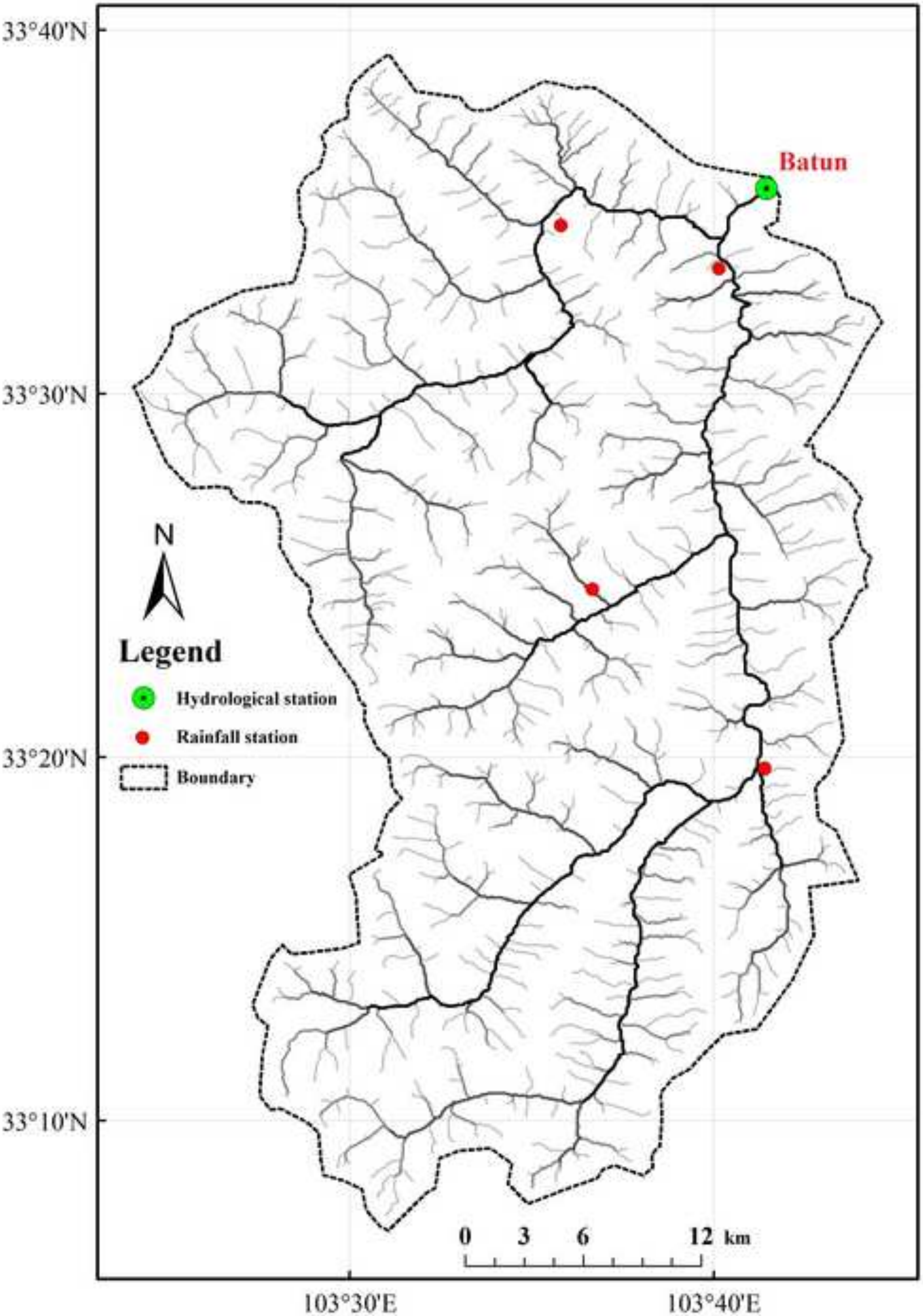
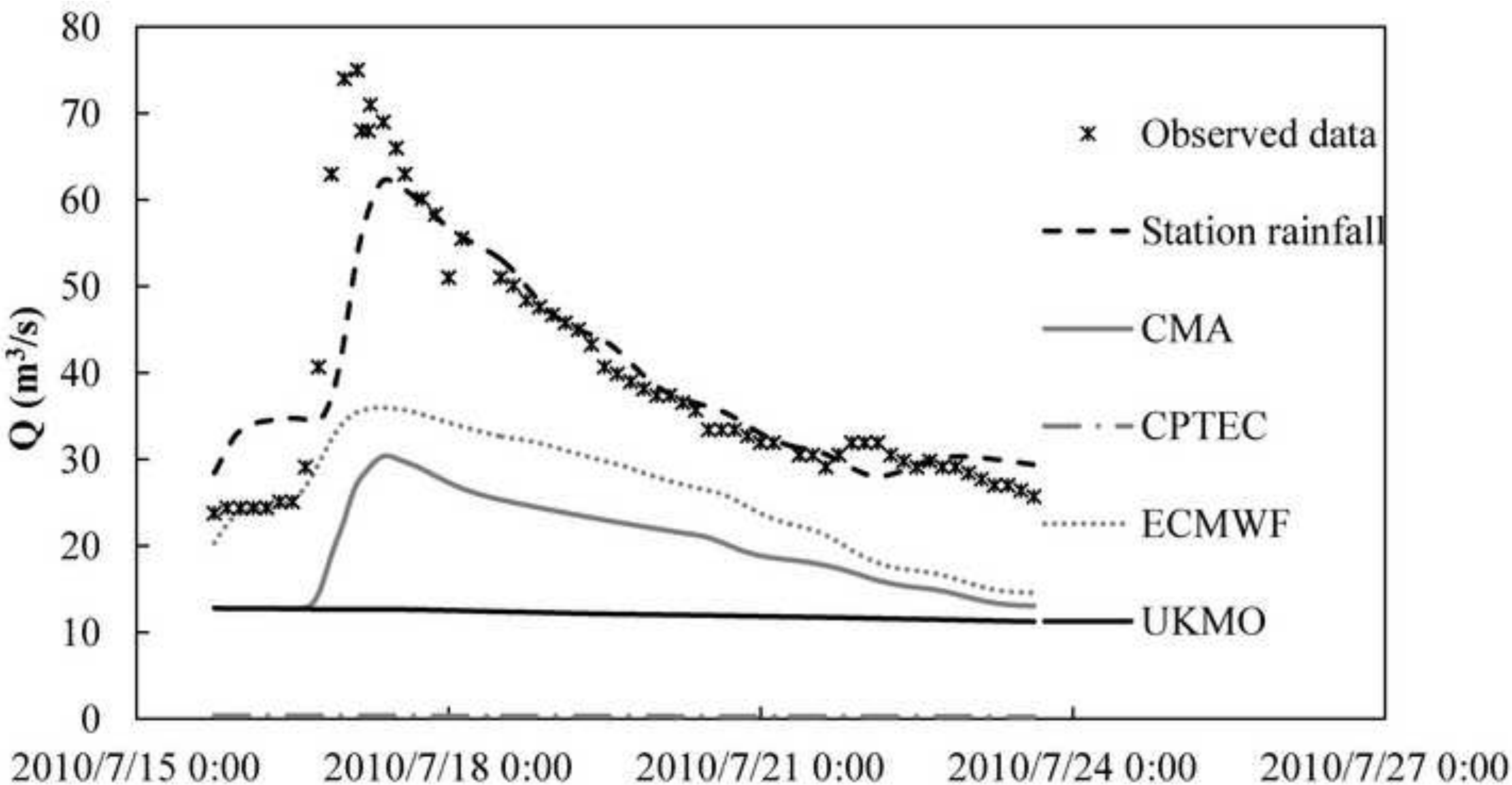

33°40'N
33°30'N
33°20'N
33°10'N
103°30'E
103°40'E
Batun
N
Legend
Hydrological station
Rainfall station
Boundary
0 3 6 12 km

Figure 8



80
70
60
50
40
30
20
10
0
Q (m³/s)
2010/7/15 0:00
2010/7/18 0:00
2010/7/21 0:00
2010/7/24 0:00
2010/7/27 0:00
Observed data
Station rainfall
CMA
CPTEC
ECMWF
UKMO

Table 1 Results of streamflow computation by using six different TIGGE data in the Juma River basin.

Data	Flood volume (10^6 m^3)	Relative error (%)	Peak value (m^3/s)	Relative error (%)	Peak time (UTC)	Peak time error
Observation	82.95	/	2,500	/	2012/7/21 23:00	/
CMA	3.92	-95.27	/	/	/	/
CMC	3.98	-95.20	/	/	/	/
CPTEC	3.94	-95.26	/	/	/	/
ECMWF	3.92	-95.27	/	/	/	/
NCEP	156.76	+88.98	3,800	+52	2012/7/21 17:00	-6 hours
UKMO	112.17	+35.23	2,600	+4	2012/7/21 19:00	-4 hours

Table 2 Results of streamflow computation by using six different TIGGE data in the upper Baishui River basin.

Data	Flood volume (10^6 m^3)	Relative error (%)	Peak value (m^3/s)	Relative error (%)	Peak time (UTC)	Peak time error
Observation	26.94	/	75.0	/	2010/7/17 3:00	/
Station rainfall (Calibration)	26.92	-0.06	62.3	-17	2010/7/17 9:00	+6 hours
CMA	13.55	-49.72	30.4	-56	2010/7/17 10:00	+7 hours
CPTEC	0.22	-99.17	/	/	/	/
ECMWF	17.68	-34.35	36.0	-52	2010/7/17 9:00	+6 hours
UKMO	8.25	-69.36	/	/	/	/

Research Highlights

1. Development of the framework of SOA for ensemble flood forecast from NWP
2. Development of a method to automatically download and update the NWP
3. Realization of implementing multiple scenarios of flood forecast at the same time
4. Validation of the new method through simulating flood flow at two river basins