Project 780118

Horizon 2020

Innovation Action

ICT programme

Project: 780118 FANFAR

Full project title:
Reinforced cooperation to provide operational flood forecasting and alerts in West Africa

Deliverable: D3.2
Report documenting and explaining the hydrological models

FINAL VERSION

Due date of deliverable: 2020-04-30
Date accepted for publication/submission to EU: 2020-04-30
Actual submission date: 2020-04-30

Title	Report documenting and explaining the hydrological models
Creator	Jafet Andersson
Editor	Lorna Little
Brief Description	Report documenting and explaining the hydrological modelling activities in the FANFAR project
Publisher	FANFAR Consortium
Contributors	Léonard Santos, Kristina Isberg, David Gustafsson, Jude Musuuza, Bernard Minoungou, Louise Crochemore
Type (Deliverable/Milestone)	Deliverable
Format	Report
Creation date	2020-04-30
Version number	3
Version date	2020-04-30
Last modified by	Lorna Little
Rights	Copyright “FANFAR Consortium”. During the drafting process, access is generally limited to the FANFAR Partners.
Audience	☐internal ☒ public ☐restricted, access granted to: EU Commission
Action requested	☐to be revised by Partners involved in the preparation of the deliverable ☐for approval of the WP Manager ☐for approval of the Internal Reviewer (if required) ☐for approval of the Project Co-ordinator
Deadline for approval	

Version	Date	Modified by	Comments
V1	15th April 2020	Jafet Andersson	Created document with first draft information
V2	29th April 2020	All contributors	Added results and content to report
V3	30th April 2020	Jafet Andersson	Added all document contributions to final version

Contents

List of Figures

List of tables

Executive Summary

The objective of this deliverable is to describe the hydrological modelling activities undertaken so far in the FANFAR project. Specifically it focuses on selecting, scaling, adapting, and deploying hydrological models on the operational production cloud platform, on summarizing the current forecast skill, and on presenting ongoing developments aimed at further improving the system. The FANFAR system currently deploys two hydrological models in its forecasting chain: the Niger-HYPE model and the World-Wide HYPE model subset to West Africa. HYPE was selected as the first model to be deployed in FANFAR based on a set of criteria: 1) suitable model type, 2) existence of other operational forecast applications, 3) accessible model code and setup, 4) past achievements and competence, 5) model performance, and 6) priorities by the co-design committee. Prior to FANFAR, the HYPE model had already been set up for the Niger River basin. To achieve full coverage in West Africa, we applied the global model World-Wide HYPE and thereby tested the scalability of the model. We found that the model is scalable, but also that the accuracy of the global models need to be improved for the region.

A number of activities have been undertaken to adapt the models to be more useful in the FANFAR context. Primarily, this has been driven by the top priority of FANFAR stakeholder to build a system with high accuracy. Firstly, we have revised the catchment delineation in certain areas in order to better match the location of hydrometric gauges, and in total linked 97 gauges identified in the project to their corresponding catchment in the models. Secondly, we have initiated several activities to improve the meteorological input data. A first test with the 3rd version of SMHI's HydroGFD data set indicates improved performance, and work is therefore underway to deploy this data set in the forecasting chain. A preliminary test has also been made with AGRHYMET's Merged Data product, however before it can be used, the models would have to be recalibrated. Thirdly, the West African subset of the World-Wide HYPE model has been calibrated in several iterations, specifically with respect to floodplains, reservoirs and soil processes. This has increased performance significantly in most parts of West Africa, moving from a median Kling-Gupta Efficiency (KGE) of -0.09 for the global model, to a median KGE of 0.36 for the model calibrated for West Africa (for >100 gauges). The calibration also improved the models' ability to represent high-flows, which is critical for forecasting floods. Aside from HYPE, we are investigating the potential of adding also the SWAT model to the forecasting chain. Preliminary tests with SWAT indicate that additional calibration is needed before it can be deployed. Finally, another way to improve hydrological forecasts is to adjust the models iteratively toward recent observations by assimilating either local hydrometric gauge data or earth observations. This work is currently ongoing, but an initial trial simulating the September 2019 flood in Niamey indicated very promising results reducing the 1-day forecast error from −86% to only −2%. Assimilation will hence be pursued for eventual deployment in the operational forecasting chain.

To better understand the ability of the entire system to forecast floods, a re-forecasting experiment has been carried out with the World-Wide HYPE model version 1.3.6 driven by HydroGFDv.2 and operational deterministic 10-day forecasts of the European Centre for Medium-Range Weather Forecasts (ECMWF), i.e. the currently deployed system. In essence it consisted of 1095 forecasts for each date between 2017-01-01 and 2019-12-31. The forecasted flood severity levels (defined by thresholds determined from return period calculations) were compared with corresponding severity levels derived for streamflow observations at 9 gauged locations. The results display two distinct groups of catchments: a) catchments with high probability of detecting an event but also a high false alarm ratio, and b) catchments with no false alarms but also a low probability of detecting an event.

The experiment also showed that several more years should be investigated to achieve a robust conclusion. We further analysed the potential reasons for the forecast skill. One factor is clearly the latency of meteorological observations: as new observations arrive, the ability of the Niger-HYPE model to forecast the Niamey 2019 flood improved significantly. Another factor is the general model performance. We found that the recalibrated model could forecast the 2019 flood in Niamey much better than the original model. These activities will be extended in order to better understand the current forecasting skill, its driving factors, and the extent to which it can be improved through e.g. new meteorological data, calibration and assimilation.

A core activity has also been to deploy the hydrological models on Hydrology-TEP (the operational production cloud platform). Open-source software has been developed that implements the operational forecasting chain based on Niger-HYPE and World-Wide HYPE (subset to West Africa). During the project, we have worked incrementally to specify the forecasting chain, to implement it in computer code, and to deploy it operationally on Hydrology-TEP using its new scheduling capabilities. This has resulted in a new processing service on the Hydrology-TEP named "FANFAR-forecast". Every day it produces a 10-day hydrological forecast for each deployed model. The forecasting chain has been producing automatic forecasts on Hydrology-TEP since beginning of 2019 for Niger-HYPE, and since beginning of 2020 for World-Wide HYPE. The service has been documented on the FANFAR Knowledge Base. In the future, the service will be extended to include all appropriate features identified as adding value to the forecast (e.g. new meteorological datasets, calibrated models and assimilation of observations).

1. Introduction

In FANFAR, hydrological forecasts and flood risk information are produced using an operational hydrological forecasting and alert pilot system (Figure 1). The overall system has been described in earlier project deliverables (e.g. Deliverable 1.1 and 3.1) and is also available online at the FANFAR knowledge base[1]. The core of the system consists of a hydrological model, indicated by red squares in Figure 1. The main function of the hydrological model is to predict the effects of meteorological dynamics on river flow, water level, soil moisture in rivers, lakes, wetlands, and all land surface areas. Assimilation of past observations is also explored as an option to improve predictions.

The objective of this deliverable is to describe the hydrological modelling activities undertaken so far in the project. Specifically it focuses on selecting, scaling, adapting, and deploying hydrological models on the operational production cloud platform, on summarizing the current forecast skill, and on presenting ongoing developments aimed at further improving the system.

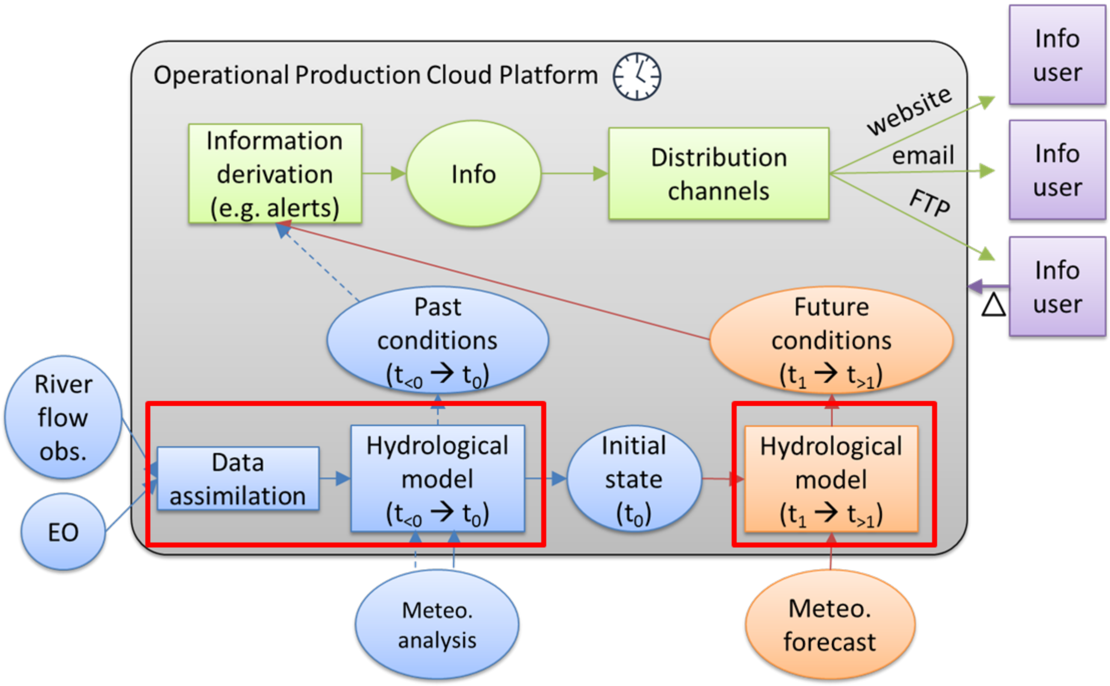

Figure 1. Overview of the essential components and data flow of the operational hydrological forecasting and alert ICT system used in FANFAR. The red squares indicate the hydrological modelling components.

2. Model selection

The FANFAR system currently makes use of two hydrological models: the Niger-HYPE model for the Niger River basin, and the World Wide-HYPE (WWH) model extracted for the West African hydrological domain (Figure 2, Table 1).

[1] https://knowledge.terradue.com/display/FANFAR/EN%3A+FANFAR+system+overview

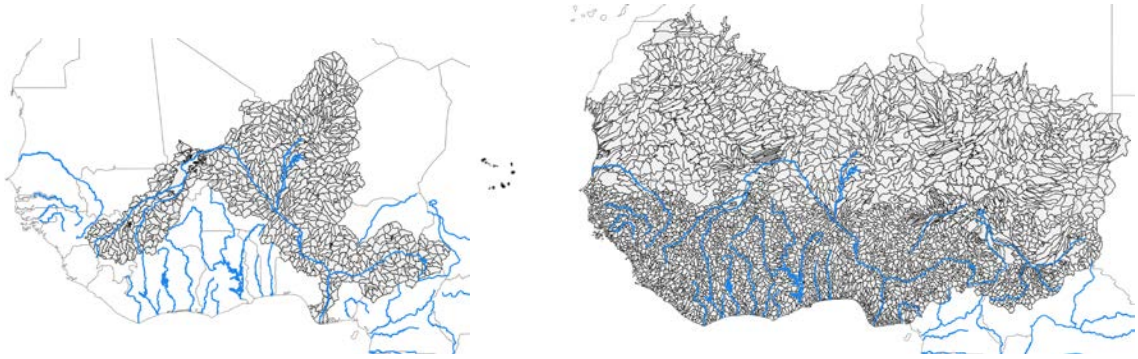

Figure 2. Model domain and catchment delineation of (left) the Niger-HYPE model, and (right) the West Africa subset of the World Wide-HYPE model (i.e. within or draining into a West African country).

Table 1 Key characteristics of the Niger-HYPE model and the West Africa subset of the World Wide-HYPE model

Model	**Niger-HYPE 2.23**	**World Wide HYPE 1.3.6 (West African subset)**
Domain	Niger River basin (2.1 million km^2)	West African hydrological basins (8.6 million km^2)
Number of sub-basins	803	4581
Average sub-basin size	2619 km^2	1870 km^2
Reference	Andersson et al. (2017)	Arheimer et al. (2020)

Before FANFAR; several hydrological models had been set up to simulate one or more basins in West Africa (e.g. HYPE, GeoSFM, SWAT, SOBEK, AFDM, and LSHM). Hence, the first FANFAR activity relating to hydrological modelling was to select model(s) to be applied in the project. A number of criteria influenced our choice, including 1) model type, 2) existence of other operational forecast applications, 3) model accessibility, 4) past achievements and competence, 5) model performance, and 6) priorities by the co-design committee.

Model type (1): We considered only models that fit with the overall objective of FANFAR: to produce operational hydrological forecasts for West Africa. This implies model types that can be applied on large scales, and that can translate meteorological forecasts to hydrological forecasts (e.g. HYPE, SWAT, SOBEK, AFDM and LSHM). In this process, we filtered out two types of models: a) models that depend on upstream hydrological observations (runoff-runoff models like GeoSFM), since they cannot produce forecast for ungauged basins; and b) hydraulic models, since they are too computationally demanding to be applied at the West African scale. Hydraulic model outputs, however, were highly sought-after by stakeholders in West Africa. Therefore, we initiated collaboration with the ANADIA-Niger project to combine FANFAR forecasts with a more detailed hydraulic model in a section of the Sirba River in Niger[2]. Evaluations are currently ongoing, and once results become available we intend to publish the analysis.

Existence of other operational forecasting applications (2): Hydrological models are built for various purposes. Some are built for analysing water resources, some for calculating irrigation water

[2] https://knowledge.terradue.com/display/FANFAR/Second+workshop%2C+9-12+April+2019%2C+Accra%2C+Ghana?preview=/15129551/17629277/4-ANADIA_Sirba_EWS.pdf

requirements, some to aid infrastructure design, some for answering specific research questions, and some for making operational forecasts etc. In FANFAR, we wanted to use a model that is used for operational forecasting elsewhere, in order to benefit from many years of development, practical tailoring and real-world tests that such models undergo. Several of the identified models are used for operational forecasting, including HYPE (e.g. in Sweden, Europe, Arctic, Globe), SWAT (e.g. for Annual Flood Outlooks in Nigeria), SOBEK (e.g. in Ghana), AFDM (in Africa), and LSHM (e.g. in the SATH-NBA system for the Niger River basin).

Model accessibility (3): In the process of creating an operational forecasting system, it is essential to have full control over the core hydrological engine inside the system. This is required for a number of reasons, for example, a) input files have to be updated every day when new forecast arrive, b) the model may need to be calibrated or process representations recoded to better simulate peak flows, c) catchment delineations may need to be adjusted to correspond to local observations. The LSHM, SOBEK and AFDM models were not available to the project partners, therefore could not be deployed. Both HYPE and SWAT are open-source models, and were accessible. A good competence and *modus operandi* for making such modifications is also required in order to be efficient with time. For the project partners, this competence was most developed for the HYPE model.

Past achievements and competence (4): FANFAR is built on a 10-year-long collaboration between European and West African partners. A key purpose of FANFAR is to reinforce this cooperation, to not start all over from scratch but to build on previous achievements. A central activity of past projects has been to develop the HYPE model for the Niger River basin, and to build competence around hydrological modelling using HYPE (Andersson, et al. 2017). Prior to FANFAR, this model had also been integrated with the Hydrology-TEP platform which enabled on-demand simulations. The FANFAR team had thus learned hands-on the pros and cons of the HYPE model in West Africa, and several key partners were already able to apply and refine the HYPE model. Several FANFAR partners also had experience in using the SWAT model. Some use the GeoSFM and AFDM models, but since they were not of the right model type nor accessible, these models were not an option for FANFAR.

Model performance (5): The model employed needs to capture the key processes of the real hydrological system in order for forecasts to be relevant for societies. Hence, the model selected should have an adequate model performance (here interpreted as having a minimum error compared with field measurements of streamflow). There are various ways to increase model performance, which we have used in FANFAR (see section 4). However, for the purposes of selecting a model, we considered only prior published information, focussing on HYPE and SWAT (the two models remaining after the aforementioned criteria). The HYPE model was developed for Sweden, where it displays a very good performance (Lindström et al., 2010; Strömqvist et al., 2012). Prior to FANFAR, HYPE had been applied to the Niger River basin (Andersson et al., 2017), where decent model performance on daily resolution was achieved for most streamflow gauges (Figure 3). There are also indications in the literature of decent SWAT model performance in Africa. For example, Aich et al. (2014) showed that decent performance could be achieved for several African river basins at both monthly and daily resolution. Particularly good performance was achieved for the outlet of the Niger River basin (Figure 4). These studies and evaluations differ in several aspects, but the intention here is not to make a consistent model inter-comparison. Rather, the intention is to investigate past literature for evidence of decent model performance in the region. And our conclusion is that both HYPE and SWAT have the potential to provide decent performance on daily resolution in West Africa.

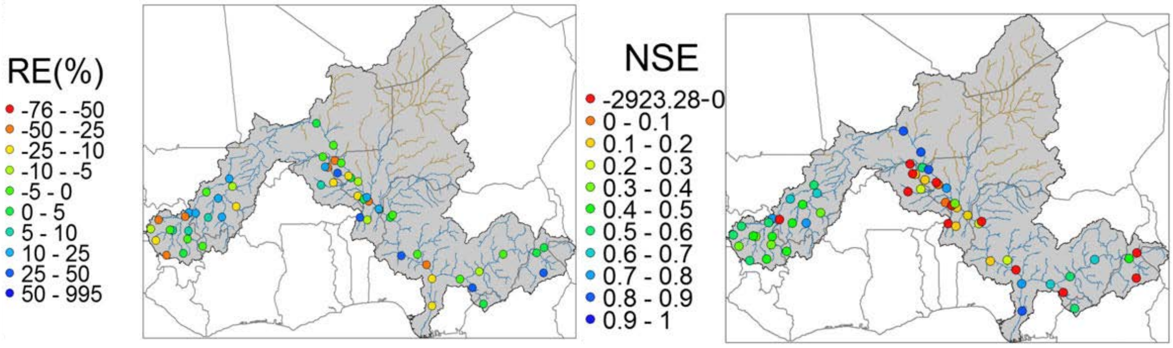

Figure 3. Overview of the performance of the Niger-HYPE 2.0 model at daily resolution for 56 streamflow gauges (1994-2009). The average Nash-Sutcliffe Efficiency (NSE, right) was 0.4 (ranging between 0.9 to -16) and the average relative error (RE%, left) was -1% (ranging between -40% and +38%). The NSE at the outlet was >0.8. Source: Andersson et al. (2017).

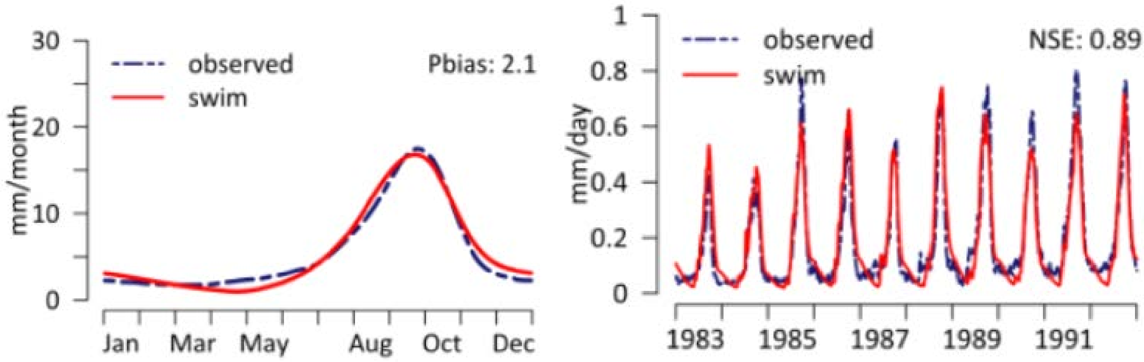

Figure 4. Example comparison of observed and simulated discharge at the outlet of the Niger River basin using the SWIM model (a fork of the SWAT model). (left): Seasonality of monthly runoff and percent bias, (right) daily runoff rate and NSE. Source: Aich et al. (2014).

Priorities by the co-design committee (6): Co-design is an important pillar of FANFAR. One of the topics discussed at the first FANFAR workshop was the choice models in the system[3]. We explored various options in open discussions about any other regional or national model that may exist. We found that the most robust system and the easiest-to-use system would be a system built only around HYPE. However, the most attractive system to the stakeholders would also include other regional and national models. Subsequently, we carried out structured activities to prioritize the different options available, given the resources at hand. We found that the co-design committee put a higher priority on building a functioning end-to-end forecasting chain than on integrating different models. Multiple models were seen as potentially adding value (e.g. through ensemble forecasting), but it was considered even more important to actually have operational outputs even if only from one model.

Taken together, these factors made a strong case for working with the HYPE model to start with, potentially adding other models like SWAT later.

[3] https://fanfar.eu/wp-content/uploads/sites/4/2019/04/190404_FANFAR_ExecReport_WS1_WP2_EN.pdf

3. Model scalability

The intent of FANFAR is to provide information across West Africa. Since the existing Niger-HYPE model only covers part of the area, it was necessary scale up the model to cover the entire domain. We opted for utilizing a parallel initiative applying the HYPE model on global scale (Arheimer et al., 2020, Figure 5). This World-Wide HYPE model (WWH) is an ambitious effort to apply catchment modelling techniques at the global scale, including evaluation against river flow. WWH consists of about 130 000 catchments (average size: 1000km^2), and is based on 20 open databases. It was built with the aim of representing major spatial and temporal variability of the global freshwater resources on monthly resolution. Global parameter values were estimated using a stepwise approach for groups of parameters regulating specific processes and catchment characteristics in representative gauged catchments. Daily and monthly time series (>10 years) from 5338 streamflow gauges were used for model evaluation, resulting in a median monthly Kling-Gupta Efficiency (KGE) of 0.4 (for version 1.3, available in September 2018). However, WWH displays large variation in model performance, and the FANFAR application requires daily resolution. In West Africa, evaluation against 106 streamflow gauges displayed an average daily KGE value of -0.1 (ranging from -19 to 0.7), which is generally not sufficient (Figure 6). Hence FANFAR supported the development of WWH and its further refinement to West Africa (as described in more detail in the next section).

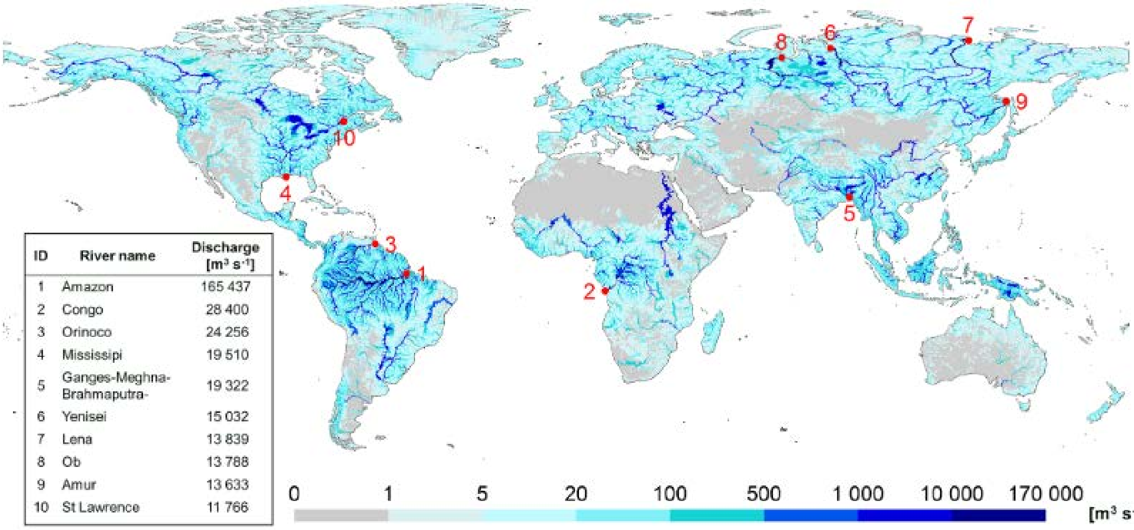

Figure 5. Annual mean of river discharge across the globe for the period 1981–2015 estimated with WWH version 1.3. Source: Arheimer et al. (2020).

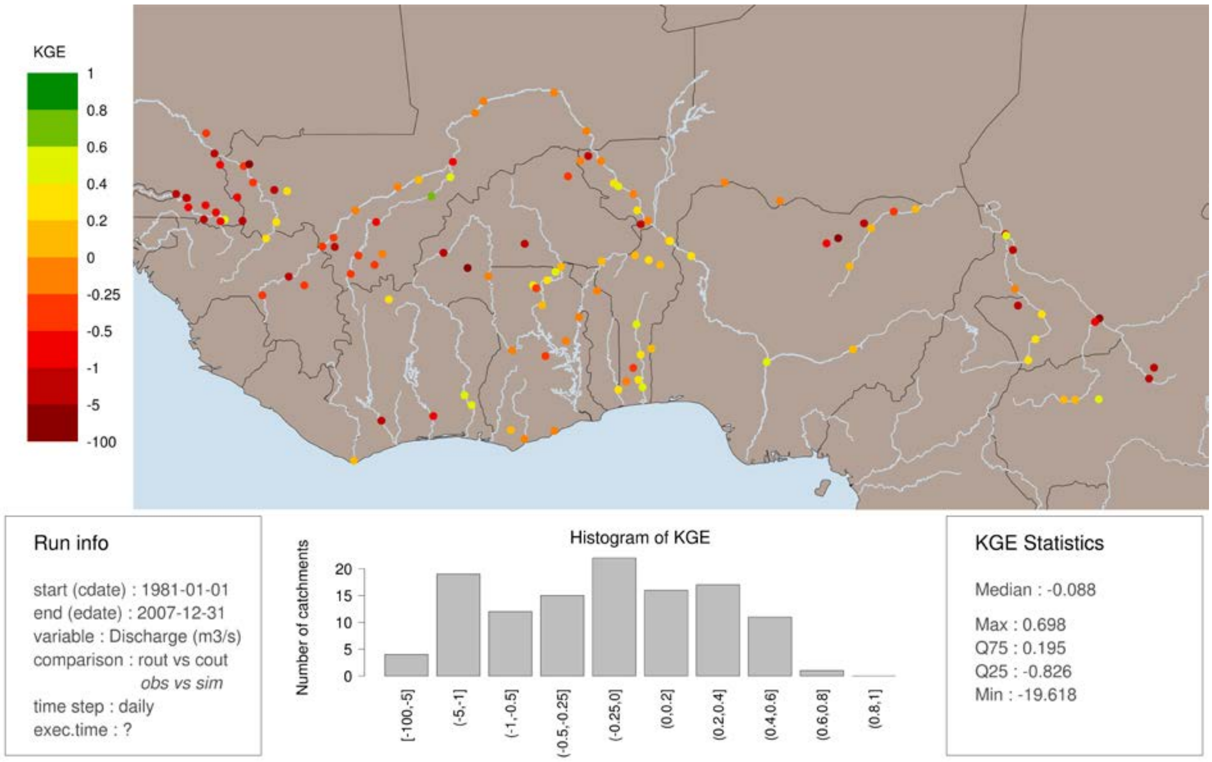

Figure 6. Performance of the World-Wide HYPE model version 1.3.3 in West Africa. The map shows the Kling Gupta Efficiency (KGE) at 106 streamflow gauges evaluated for the period 1981-2007 at daily resolution.

4. Model adaptation

The World-Wide HYPE model provides a good basis to develop a forecast model over West Africa. However, as pointed out in the previous section on model scalability, the World-Wide HYPE model needs to be adapted to the West African region.

At first, a sub-basin delineation step was necessary in order to fit the model to the specifics of the project. It was necessary to add Cape Verde to the model and to perform a delineation adaptation process in order to take into account the new discharge observations (ground or satellite).

The second part of the adaptation procedure was to work on improving the discharge model simulations in all West Africa at the daily time step. Model evaluation based on KGE criteria is shown in Figure 6 in which problems in model performance can be identified all over West African territory. This figure shows that the long term model performance (here illustrated by the KGE score) is not optimal for almost all the available gauged stations in the modelling domain. At first a calibration of model parameters fixed this long term model performance issue. Besides, an underestimation of high flows was identified with this calibration. Feedback from users at the third FANFAR workshop in Abuja (February 2020) confirmed that during the 2018 rainy season, forecasts were underestimated. Feedback from AGRHYMET stated that, even if the timing of the forecast were accurate, the discharge and the severity level were underestimated by the FANFAR forecasting system[4]. These

[4] https://knowledge.terradue.com/pages/viewpage.action?pageId=24674773&preview=/24674773/24675455/AGRHYMET.pdf

feedbacks from regional and national agencies proved that the forecast generation needs to be improved to better capture flood events in the zone. In the context of hydrological modelling, we worked on several ways of improvement. The first step was to recalibrate the model in order to improve the performance especially for high flows. Then this calibration went on with a work in progress that aims to include a new meteorological forcing dataset (HydroGFD3) and to use it for hindcast. The final step (still in progress), aims at including the assimilation of observed discharge data and EO water levels in order to improve forecasts.

4.1 Catchment delineation

4.1.1 Niger-HYPE from version 2.23 to version 2.24

The catchment delineation of the Niger-HYPE model (Andersson et al. 2017) is based on HYDROSHEDS (a digital elevation model at 15 arc-second resolution). The original sub-basin delineation of the model version 2.23 was adjusted for sub-basin outlets at larger lakes, reservoirs, the Inner Niger Delta and, available gauging stations in 2012. In FANFAR, this delineation has now been revised to match the 47 gauges within the Niger River basin that were available on Hydrology-TEP as of 2nd April 2020 (Table 2).

In general, no routing has been affected, but several sub-basins have been adjusted with respect to the outlets. Some sub-basins have been divided into two parts. The number of sub-basins has therefore been increased from 803 to 814 and the average sub-basin size has thus increased from 2619km^2 to 2629km^2.

Table 2. Hydrometric stations listed on Hydrology-TEP (2020-04-02) and their coupling to the World-Wide HYPE and Niger-HYPE models respectively.

Country	Station name	Waterbody	World Wide-HYPE	Niger-HYPE
Bénin	Adjohoun	Oueme	X	
Bénin	Ahlan	Oueme	X	
Bénin	Atchakpa	Oueme	X	
Bénin	Athieme	Mono	X	
Bénin	Beterou	Oueme	X	
Bénin	Bonou	Oueme	X	
Bénin	KOMPONGOU	Mekrou	X	X
Bénin	LANTA	Couffo	X	
Bénin	Malanville	Niger	X	X
Bénin	SAVE	Oueme	X	
Burkina Faso	BOROMO	Mouhoun	X	
Burkina Faso	DAPOLA	Mouhoun	X	
Burkina Faso	FOLONZO	Comoé	X	
Burkina Faso	KORIZIENA	Gorouol	X	X
Burkina Faso	LIPTOUGOU	Faga/Sirba	X	X
Burkina Faso	RAMBO	Nakambé	X	

Burkina Faso	SAMANDENI	Mouhoun	X	
Burkina Faso	wayen	Nakambé	X	
Burkina Faso	YENDERE	Leraba	X	
Côte D'Ivoire	ABOISSO	Bia	X	
Côte D'Ivoire	ABRADINOU	Comoé	X	
Côte D'Ivoire	ANIASSUE	Comoe	X	
Côte D'Ivoire	DIMBOKRO	Bandama	X	
Côte D'Ivoire	KOKPINGUE	Volta	X	
Côte D'Ivoire	KOUTO	Bagoué	X	X
Côte D'Ivoire	SAMANTIGUILA	Baoulé	X	X
Côte D'Ivoire	SAN PEDRO	San Pédro	X	
Côte D'Ivoire	SEREBOU	Comoé	X	
Côte D'Ivoire	TIASSALE	Bandama	X	
Gambie	BANSANG	Gambie	X	
Gambie	GOULOUMBOU	Gambie	X	
Gambie	KUNTAUR	Gambie	X	
Ghana	EKOTSI	OCHI-NAKWO	X	
Ghana	PRESTEA	ANKOBRA	X	
Ghana	SEFWI-WIAWSO	Bia	X	
Ghana	TANOSO	Bia	X	
Guinée	Dialakoro	Niger	X	X
Guinée	Faranah	Niger	X	X
Guinée	Kankan	Milo	X	X
Guinée	KISSIDOUGOU	Niandan	X	X
Guinée	Kouroussa	Tinkisso	X	X
Guinée	Mandiana	Sankarani	X	X
Guinée	Tinkisso	Tinkisso	X	X
Mali	Ansongo	Niger	X	X
Mali	Banankoro	Bagoé	X	X
Mali	Beneny Kegny	Bani	X	
Mali	Bougouni	Baoulé	X	X
Mali	Dioila	Baoulé	X	X
Mali	Diré	Niger	X	
Mali	Douna	Bani	X	X
Mali	Ké-Macina	Niger	X	
Mali	Kirango	Niger	X	X
Mali	Korioume	Niger	X	
Mali	Koulikoro	Niger	X	X
Mali	Mopti	Bani	X	
Mali	PANKOUROU	Bagoé	X	X
Mali	Selinge Aval	Sankarani	X	X

Mali	Taoussa (Tossaye)	Niger	X	X
Niger	Ayorou	Niger	X	X
Niger	BAGARA	Komadougou Yobé	X	
Niger	BAROU	Mékrou	X	X
Niger	Bossey Bangou	Sirba	X	X
Niger	Garbey Kourou	Sirba	X	X
Niger	Kandadji	Niger	X	X
Niger	Niamey	Niger		
Niger	Niamey	Niger	X	X
Nigeria	Baro	Niger	X	X
Nigeria	Jebba downstream dam	Niger	X	X
Nigeria	Jidere Bode	Niger	X	X
Nigeria	Kainji	Niger	X	X
Nigeria	Kainji Dam	Niger		
Nigeria	Kende	Sokoto	X	X
Nigeria	Lokoja	Niger	X	X
Nigeria	Makurdi	Benoué	X	X
Nigeria	ONITSHA	Niger	X	X
Nigeria	Umaisha	Benoué	X	X
Nigeria	WURO BOKKI	Niger	X	X
Sénégal	Bafing Makana	Bafing	X	
Sénégal	Bakel	Senegal	X	
Sénégal	Daka-Saidou	Bafing	X	
Sénégal	Gourbassi	Falémé	X	
Sénégal	KEDOUGOU	Gambie	X	
Sénégal	Kidira	Falémé	X	
Sénégal	Kolda	Casamance	X	
Sénégal	Mako	Gambie	X	
Sénégal	Matam	Sénégal	X	
Sénégal	OUALIA	Bakoye	X	
Sénégal	Simenti	Gambie	X	
Tchad	BONGOR	Logone	X	
Tchad	MOUNDOU	Logone	X	
Tchad	N'DJAMENA TP	Chari	X	
Tchad	PATALAO	Kabia	X	X
Tchad	SARH	Chari	X	
Togo	Kati	Zio	X	
Togo	Kpédji	Lac-Togo	X	
Togo	TETETOU	Mono	X	
Togo	TOGBLECOPE	Lac-Togo	X	

4.1.2 World-Wide HYPE from version 1.3 to version 1.4

The delineation of World-Wide HYPE (and its corresponding subset for West Africa) is based on GWD-LR (a digital elevation model with 3 arc-sec resolution). The original sub-basin delineation was adjusted for sub-basin outlets at larger lakes, reservoirs, the Inner Niger Delta and, gauging stations with position and catchment area information available in 2016. In FANFAR, this delineation has now been revised to match the 97 stations located within the West African region, available on Hydrology-TEP as of 2nd April 2020 (Table 2) and new information about gauging stations within the Sirba river basin (Figure 7). Sub-basins for the islands of Cape Verde were also prepared and added to the model (Figure 8). This was done by preparing flow accumulation and flow direction from another global DEM (SRTM, 3 arc-seconds, since the islands were not present in GWD-LR). A few sub-basins at the main island of Cape Verde were adjusted for the location of gauging stations, unfortunately without any available discharge observations. The number of sub-basins in the West African subset of the World-Wide HYPE model were increased from 4470 (version 1.3.3), to 4581 (version 1.3.6) to 4609 (version 1.4) and the average sub-basin size changed from 1915 km^2 (version 1.3.3), 1870 km^2 (version 1.3.6) to 1858 km^2 (version 1.4). For both models, all the Hydrology-TEP stations with available information about location at April 2, 2020 were used even if no observations were available, in order to facilitate for future inclusion.

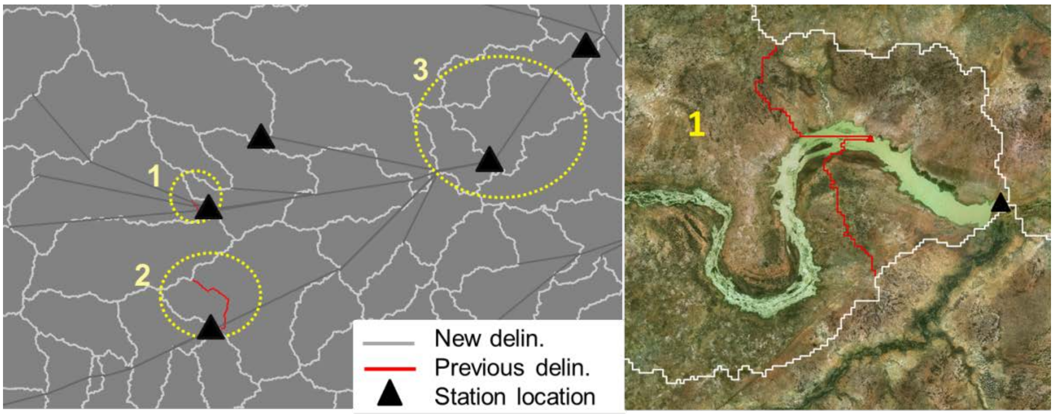

Figure 7. Example of delineation adjustments in the Sirba river basin. 1) minor change of the sub-basin border to more accurately represent the outflow of a reservoir (figure to the right), 2) larger adjustment of the sub-basin border and outlet to better fit to the gauging station, 3) the sub-basin was cut into two new sub-basins in order to be able to simulate a new gauge (the Bossey Bangou gauge in Niger).

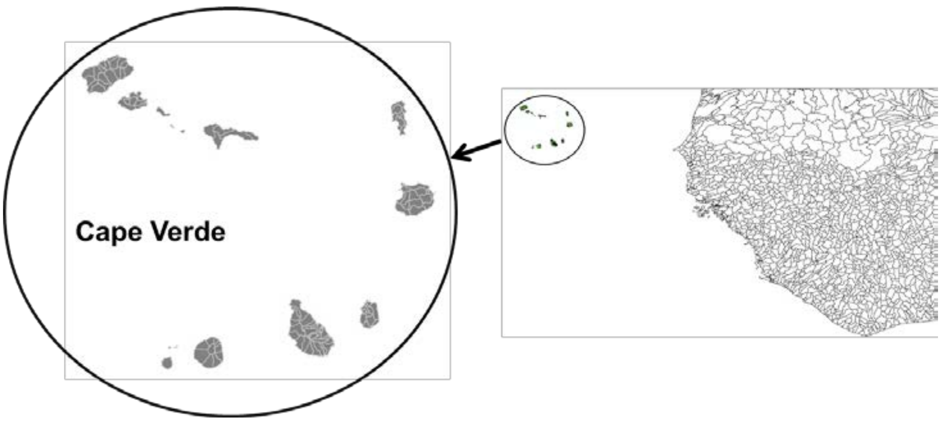

Figure 8. Cape Verde islands were added to the World-Wide HYPE model and the West African subset.

4.2 New global meteorological data: HydroGFD3

FANFAR utilizes the HydroGFD meteorological data set (a.k.a. forcing data, as described in Deliverable 3.1 and 1.1). Operationally, version 1 and version 2 of the data set has been used so far. Work is currently ongoing to use the updated third version of the HydroGFD meteorological data product in the forecasting system. This HydroGFD3 dataset is based on the ERA5 reanalysis produced by ECMWF and replaces ERA-Interim which was used for HydroGFD2. As it was the case for HydroGFD2, the HydroGFD3 dataset uses other climatological data to correct ERA5. The new dataset has a higher spatial resolution than HydroGFD2 (0.25 degrees instead of 0.5). This is supposed to increase the accuracy of the meteorological data at a sub-basin scale.

An initial test was made to evaluate the potential of the product in the context of FANFAR project. The West Africa subset of the World Wide-HYPE model was run using HydroGFD3 as forcing and compared to what was obtained using HydroGFD2. Figure 9 shows the distribution of the model performances (KGE values) for the gauged stations available in West Africa. The performances are clearly better when using HydroGFD3 as forcing, since KGE values are higher and mostly positive, even as the model was calibrated to HydroGFD2. Hence, this new data set has a good potential to improve the simulation in the context of FANFAR project and will be further pursued.

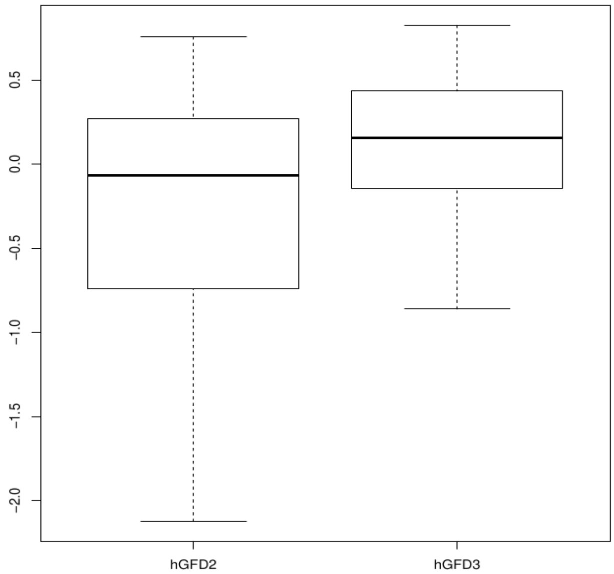

Figure 9. Distribution of KGE values (in the form of boxplots) for the West Africa subset of the World Wide-HYPE model used with HydroGFD2 and HydroGFD3 as input respectively.

4.3 Preliminary results using West African meteorological data

AGRHYMET is collecting meteorological data from West African national meteorological agencies. By combining satellite data and ground data from West African station, AGRHYMET creates a product that can be used as forcing inputs for the hydrological hindcast run or as a correction of the HydroGFD product (the “Merged Data”, described in more detail in Deliverable 3.1 and 1.1). The first version of this “Merged Data” product is here compared to HydroGFD2. The maps in Figure 10 show the average yearly precipitation of the Merged Data and HydroGFD2 respectively in each HYPE sub-basins. This figure shows that the precipitation values of the Merged Data are substantially higher than HydroGFD2 in the northern part of Sahel, but also significantly lower than HydroGFD2 in the rest of the domain (where most runoff is generated).

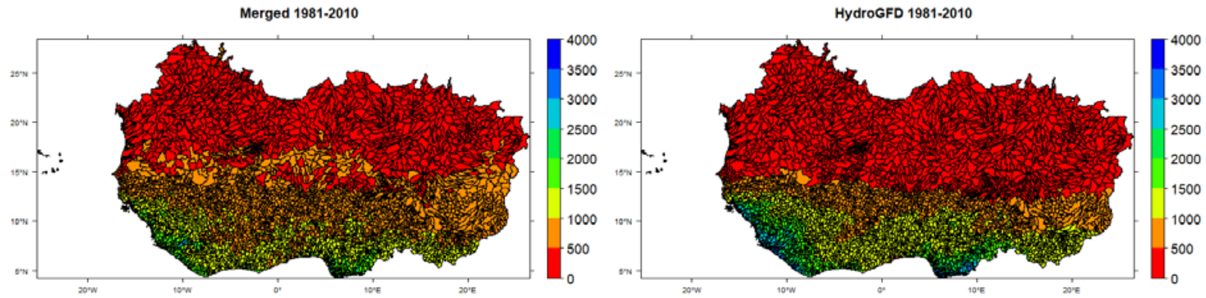

Figure 10. Maps of average yearly rainfall for each sub-basin in West Africa calculated with AGRHYMET's Merged Data (left) and HydroGFD2 (right). Model: World-Wide HYPE subset to West Africa.

We carried out an initial simulation to test the Merged Data with the West African subset of the World Wide HYPE model (that so far has only been calibrated toward HydroGFD2). This resulted in a clear underestimation of discharge values (Figure 11). This systematic underestimation of flow is not observed when HydroGFD2 is used to force the model. It tends to show that even if the Merged Data gives higher values of average rainfall in some catchments, the total amount of rainfall on the domain is significantly lower than for HydroGFD2. The Merged Data does not provide enough rainfall to fill the soil storage of the model which prevents the generation of runoff. However, this underestimation can be explained by the fact that the model was calibrated using HydroGFD2. If the precipitation values are underestimated compared to the precipitation values used for its calibration, the model will consequently simulate underestimated flows. This result does not mean that the Merged Data product is wrong and HydroGFD2 true. These results are only preliminary, and additional adaptations and analyses are required (e.g. calibrating the model toward the Merged Data), in order to understand the potential usefulness of the Merged Data.

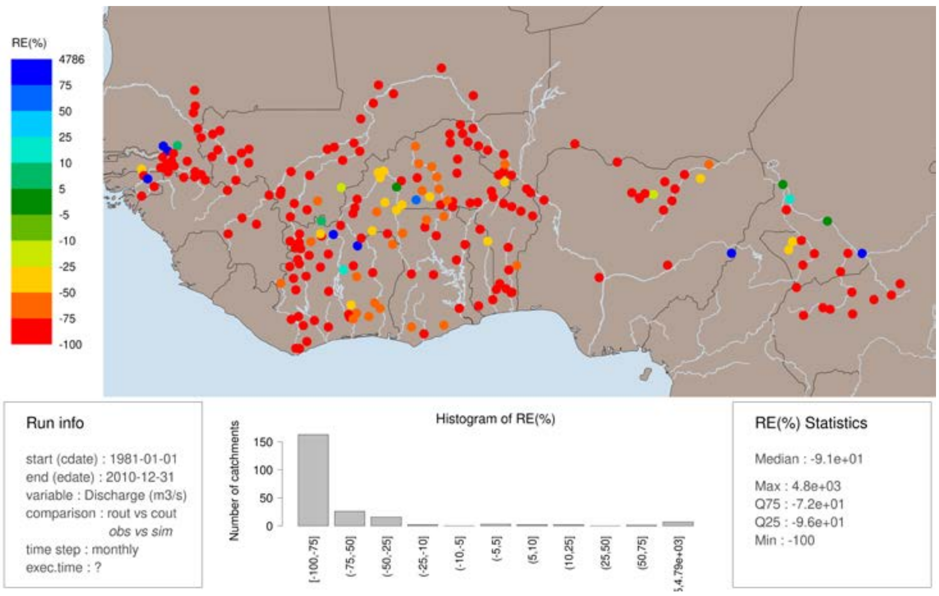

Figure 11. Long-term relative error in accumulated streamflow volumes for 181 West African gauged stations. The model evaluated is the West African subset of the World-Wide HYPE model forced with AGRHYMET merged data ran between 1981 and 2010.

4.4 Calibration of hydrological processes in the World-Wide HYPE model

Model calibration of hydrological processes was carried on for different purposes. The principal efforts have been put on calibrating the World-Wide HYPE model for West Africa. First the global model had to be adapted to the local hydrological context since it was overreacting (runoff was generated just after each rainfall event), especially in the Niger River basin. Then, this process-oriented calibration also aimed at improving simulations of peak flows and at adapting the model to the new meteorological inputs. Two elements of the model were calibrated: floodplains & reservoirs, and processes regulating runoff generation from soils.

4.3.1 Floodplains & reservoirs

The Niger inner delta is an uncommon hydrological feature. It is a wetland area located in an arid zone where the river loses more than half of its discharge through evaporation. Downstream areas along the Niger River are sensitive to flood events, which makes the inner delta a very important zone to calibrate. The HYPE model simulates floodplains as two stores that communicate between them (Figure 12). These stores represent each the river and the floodplain. The transfer of water between the stores is driven by thresholds (green rectangles in Figure 12). The calibration of this simplified floodplain model was made by adapting these thresholds in order to mimic the inner delta behaviour. More details on how this model works are available in Andersson et al. (2017).

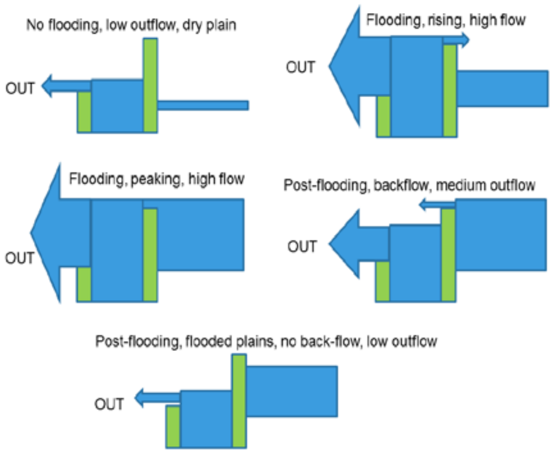

Figure 12. Scheme of floodplain representation based on stores in HYPE model.

When calibrating the floodplain, discharge observations upstream to the inner delta were used to force the model to avoid the influence of other model errors on discharge (such as rainfall generation errors). Figure 13 shows the hydrograph at the outlet of the inner delta, at the Tossayé (Taoussa) station. This figure clearly shows the effect of modifying these thresholds and how this modification improves discharge simulations. The shape of the hydrograph and the timing of high flows are closer to observations when thresholds are adapted.

In addition, the way in which the model takes into account some major reservoirs had to be adapted. The dam module of HYPE model was calibrated using a similar method to the one used for the inner delta to better simulate the hydrological behaviour of the Kainji reservoir on the Niger River and the Volta lake on the Volta River. The thresholds and the recession coefficient of the lakes were calibrated using local hydrometric data.

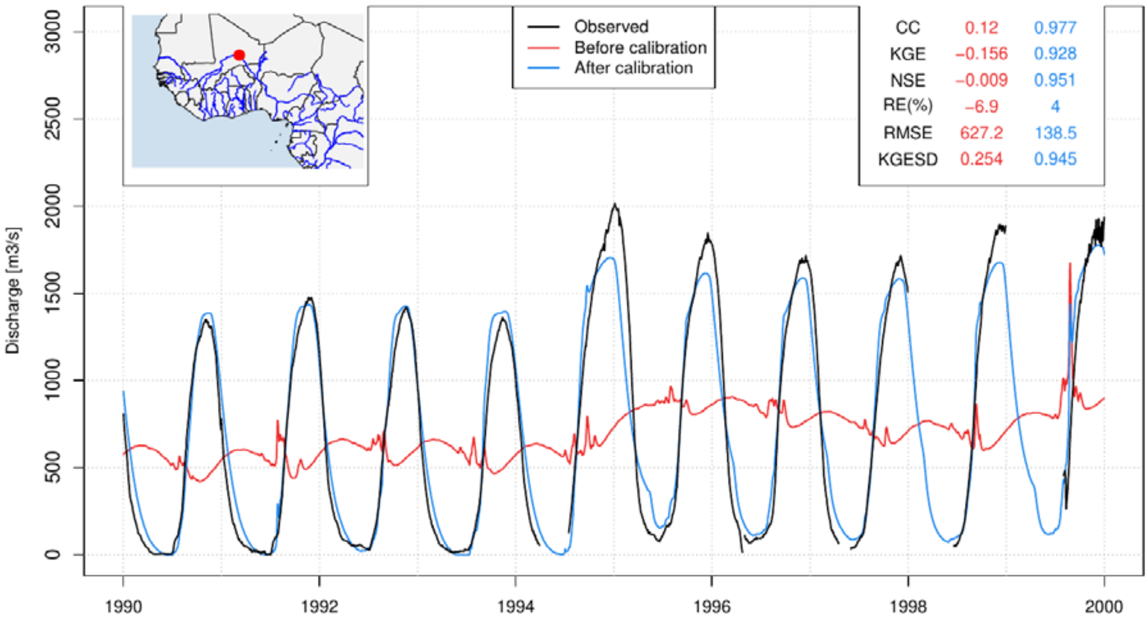

Figure 13. Hydrograph of the observed and simulated discharge before (red) and after (blue) calibration compared with observations (black) at outlet of the Niger inner delta (Tossayé station). Observed discharges upstream to the floodplain were injected in the model (i.e. used to force the model), in order to avoid upstream errors to propagate into the representation of the floodplain processes. Model: World-Wide HYPE subset to West Africa.

4.3.2 Soil processes

The calibration of the floodplains and reservoirs was performed focusing on characteristics of specific locations. However, HYPE was also calibrated all over the West Africa domain by tuning the runoff generation processes. The HYPE model simulates surface runoff based on soil humidity and on rainfall intensity. Surface runoff will be generated if a rainfall event occurs on a saturated soil. It can also be the result of high intensity rainfall, falling into a soil with low infiltration capacity. Subsurface runoff in the soil is also simulated by HYPE, which is mainly regulated by soil saturation and recession coefficients. Figure 14 summarizes in a simplified way the processes involved in runoff generation in the soil. Elements calibrated for West Africa are coloured in green and red depending on the different calibration steps.

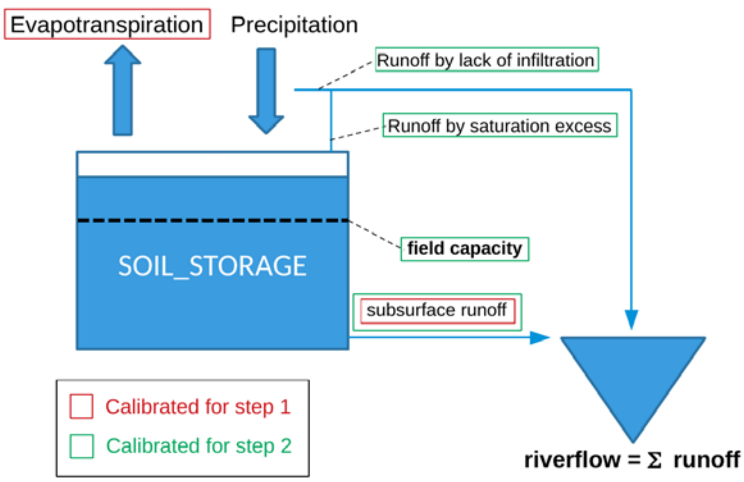

Figure 14. Simplified representation of runoff generation from soil in HYPE model.

A first approach was to calibrate the evapotranspiration and the recession coefficient of the subsurface flow (red squares in Figure 14). Since model response was too reactive due to little memory in the soil (for example, the blue curve in Figure 15) we decided to increase this short memory. To take into account the spatial variability of process in the model domain, four subdomains were created by merging areas with similar hydrological behaviour (left map in Figure 16). In general, the recession of the subsurface flow was slowed down and the potential evapotranspiration was reduced to avoid high water losses. The intensity of these modifications was adjusted for each subdomain. These modifications showed an improvement in terms of long-term bias and KGE values compared to the initial model behaviour (red curve, Figure 15).

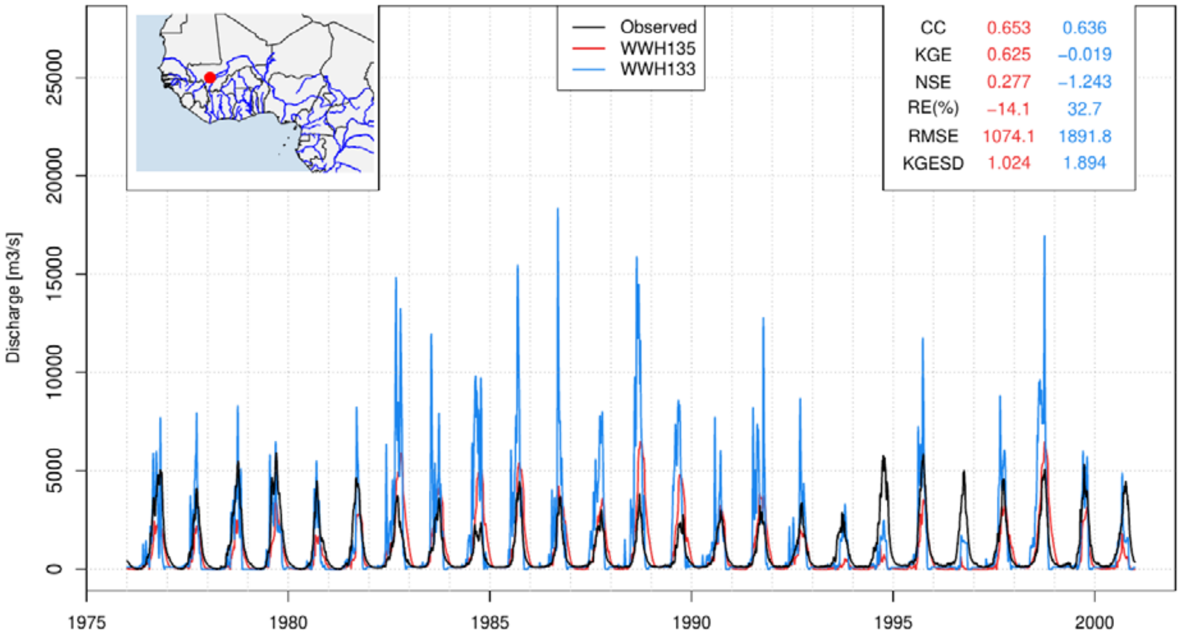

Figure 15. Observed and simulated discharge time series of the Niger River at Koulikoro by World-Wide HYPE 1.3.3 (blue) and the first calibration on West Africa (red).

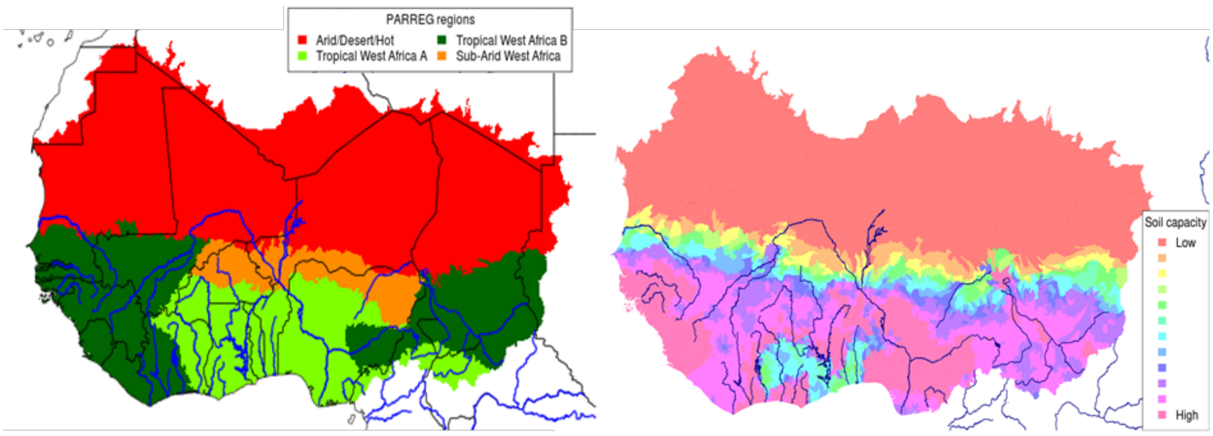

Figure 16. Maps of subdomains with similar hydrological behaviour within the West African territory use for calibration step 1 (left) and step 2(right).

However, even if the overall performances of the calibrated model were far better than the original model, the resulting simulations after the first calibration were too smooth and lead to underestimation of high flows. During the period of use of this first calibrated model (the rainy season 2019) the feedbacks from regional and national agencies have shown that the forecasted flood severity were underestimated[4]. A second calibration was then carried out with the objective to improve specifically high flow simulations. At the same time, this calibration was also designed to adapt discharge simulations to the new version of forcing (HydroGFD3). In order to improve the modelling possibilities, all the processes of runoff generation were calibrated in the model (green squares in Figure 14). Subdomains of identical soil parameters were also refined compared to the ones of step 1 in order to improve the spatial representation of the hydrological processes variability over West Africa (map on the right in Figure 16). This domain division was based on the soil water capacity that was estimated by Wang-Erlandsson et al. (2016). This calibration was more precise for each subdomain in which the field capacity and the occurrence of each runoff process were balanced to better reproduce the available discharge observation. Following the second calibration, the long-term model performance improved for almost all the gauged locations (Figure 17 vs. Figure 6). The performance map of the second calibration (Figure 17) shows that the main rivers are well simulated by the model and that the performance are acceptable in the majority of the gauged station.

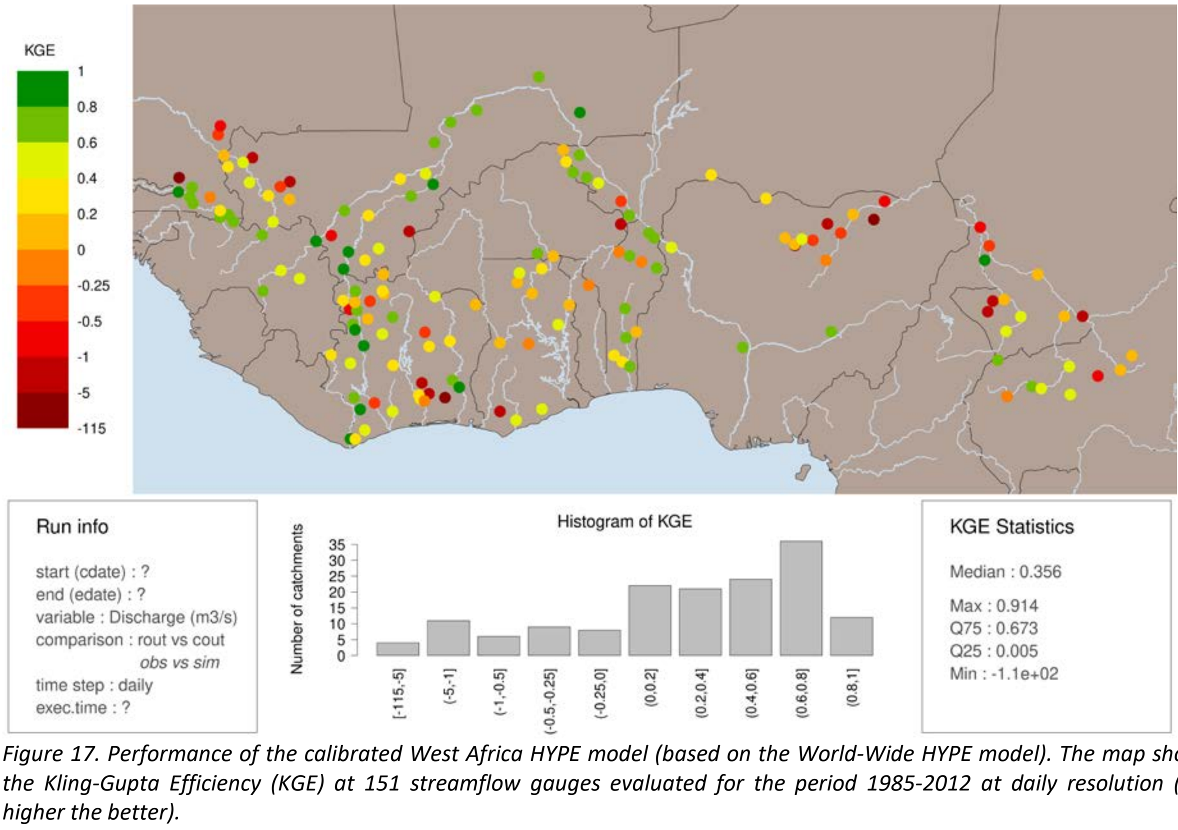

Figure 17. Performance of the calibrated West Africa HYPE model (based on the World-Wide HYPE model). The map shows the Kling-Gupta Efficiency (KGE) at 151 streamflow gauges evaluated for the period 1985-2012 at daily resolution (the higher the better).

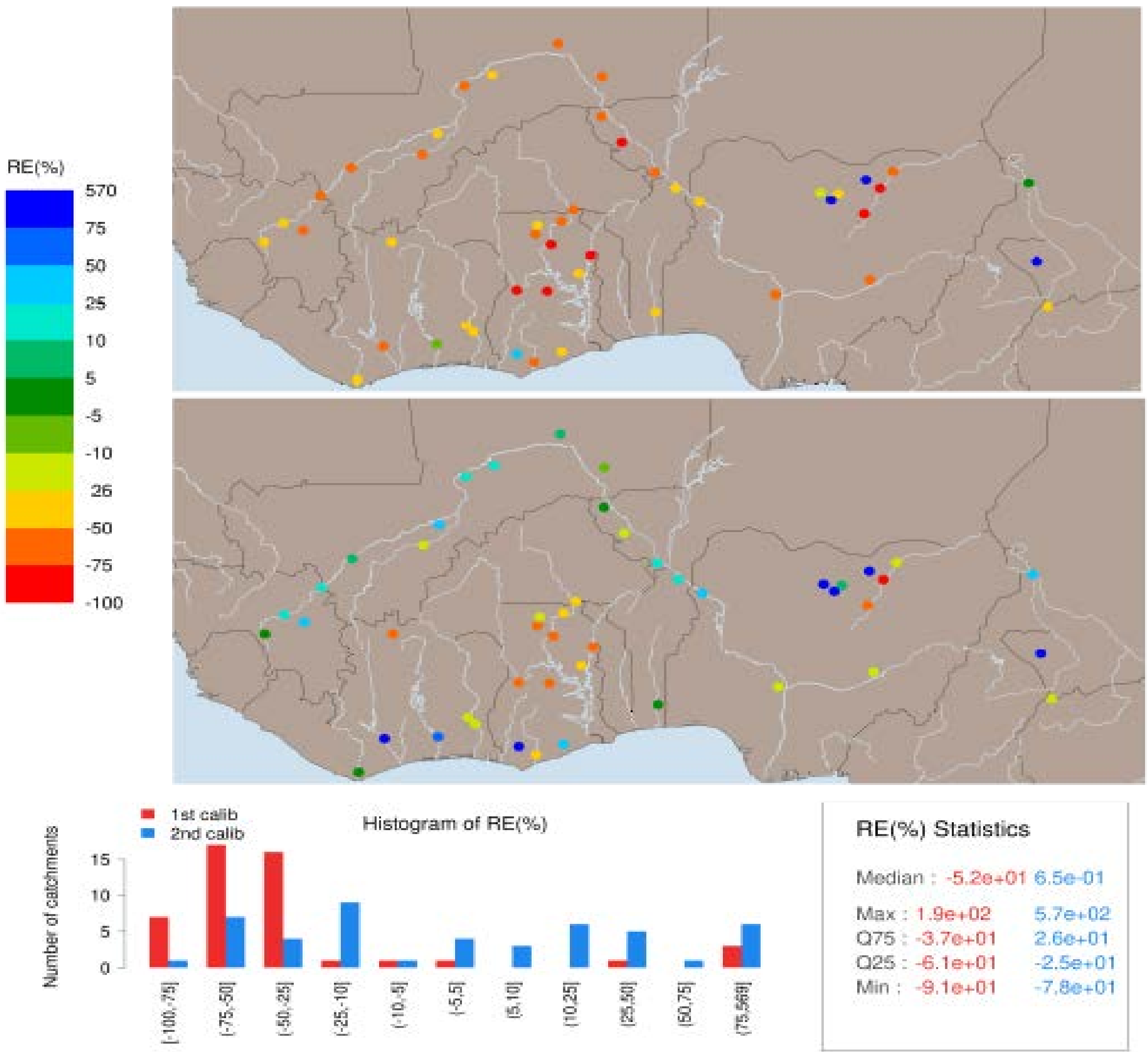

Figure 18. Relative error on the 95th discharge percentile of each year for 47 gauged stations in West Africa. The maps represent the scores of the first step of the calibration (on the top) and the second step of the calibration (middle). The bottom histogram and descriptive statistics represents the distribution of the scores for the two steps, and illustrates that the second step of the calibration is significantly better and more balanced in representing high flows compared with the first calibration. Model: World-Wide HYPE subset to West Africa.

In addition, as flood forecasts were not accurate enough after the first step of calibration, special attention was given to high flows in the project area. Figure 18 shows the difference between the two steps of calibration in terms of relative error for a high-flow indicator. This indicator is calculated as the relative error on the 95th daily discharge percentiles of each year with observed data in the station. It takes into account the high flows of each year, without having to work with the annual maximum (that can be an extreme event or observation error). It has been calculated on the 47 gauged stations with more than 5 years of data. The maps shows that the high flows are underestimated after the first step of the calibration and often better simulated after the second step of the calibration. The balance of the relative error across all available gauges in West Africa is also better after the second calibration step (it no longer has a systematic underestimation as after the first calibration). This shows that this calibration has a good potential on high flow forecasting.

4.5 Preliminary results with the SWAT model

We plan to incorporate other hydrological models available in West Africa, in order to make the FANFAR system more comprehensive and flexible, and to potentially increase forecast accuracy. To this end the AGRHYMET Regional Centre has initiated the process of applying the SWAT model (Arnold et al., 1998) in the Niger River basin. The model was set up using ArcSWAT 2012 with the data listed in Table 3. This resulted in a model with 5534 sub-basins and an average size of 378 km^2.

Table 3. Data used for setting up the SWAT model over the Niger River basin.

Data type	Description	Resolution/Period	Source
Topography	Digital Elevation Model	90 m	USGS HydroSHEDS
Land Cover	GlobCover 2009	1 km	ESA
Soil Data	Soil characteristics database	1 :5000000	FAO
Climate Data	Meteorological re-analysis data	0.5°	ERA-Interim
Discharge	Observed discharge on main stations	1955-2010	AGRHYMET, ABN
Lakes Data	Characteristics of large reservoirs in the basin		AGRHYMET, FAO

Work is currently ongoing to calibrate the model. So far the model has been calibrated using 28 gauging stations on monthly resolution for the period 1990-2012. The model was subsequently evaluated against the available streamflow gauges. On monthly resolution, an acceptable model performance was achieved at 32% of the gauges (NSE >0.5). However, on daily resolution almost all stations have NSE <0.2 (Figure 19), which is not sufficient for daily flood forecasting, and also significantly lower than the Niger-HYPE model (Figure 3). The main reason for this is an overestimation of flow rates in the SWAT model (often >50%), exemplified at the gauges Faranah and Lokoja (Figure 20). This model is still under development, and should be further improved before it can be incorporated in the FANFAR forecasting chain.

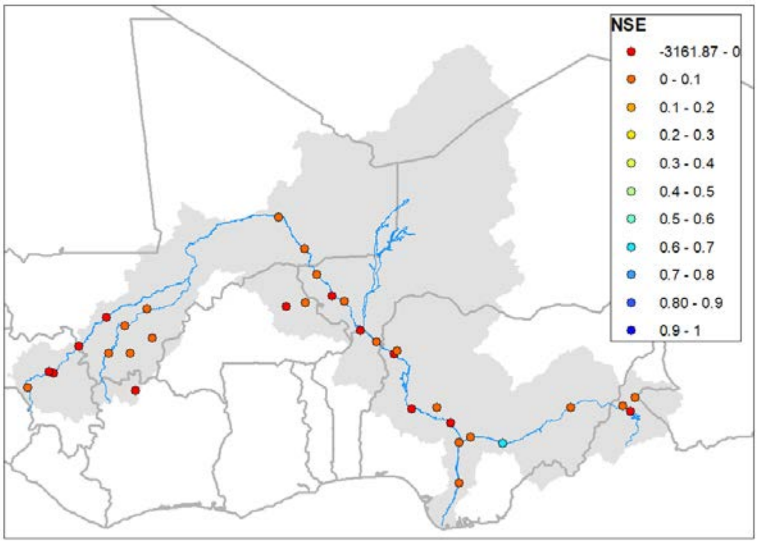

Figure 19. Nash-Sutcliffe Efficiency (NSE) of AGRHYMETs SWAT model in the Niger River basin.

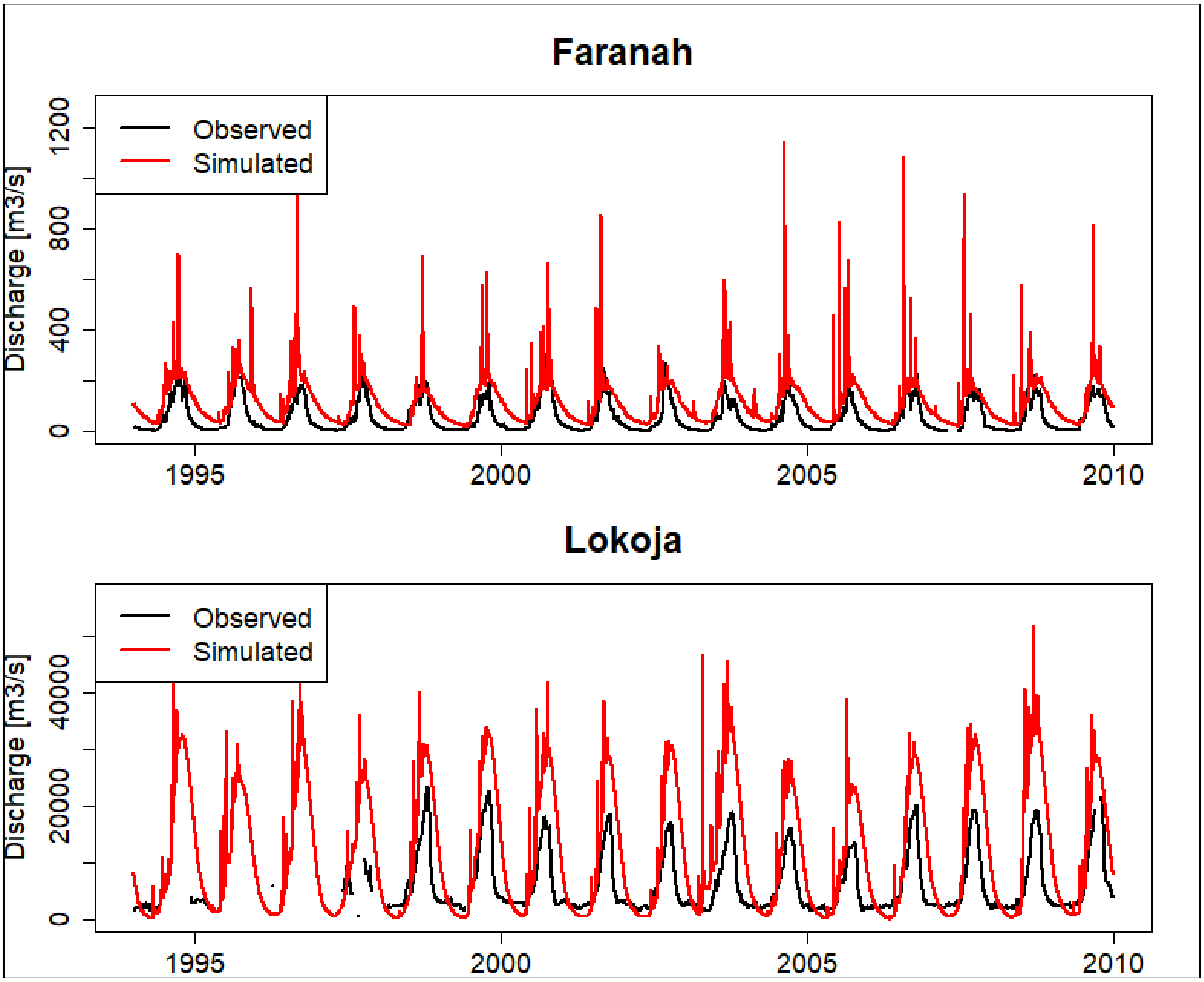

Figure 20. Streamflow simulated with AGRHYMETs SWAT model compared with observations at Faranah and Lokoja. The graphs illustrate the general over-prediction by the model at these and many other locations. Here the model also captures the timing of the rainy season quite well, while often produces too flashy responses at the onset of the rainy season.

4.6 Assimilation of observations

In the FANFAR context, data assimilation is a technique that seeks to optimally combine numerical models with observations in order to improve the representation of the initial conditions for a forecast model. Given that hydrological systems has a memory – i.e. that past conditions influence the state of the system in the coming days – assimilation has the potential to improve hydrological forecasts. Assimilation in FANFAR focus on two types of observations: 1) hydrometric field measurement data (streamflow and water level) collected by national and regional agencies in West Africa, and 2) remotely sensed water levels derived from Earth Observation satellites. Assimilation can be carried out in a number of ways. In this project we focus on two approaches: a simple auto-regressive updating algorithm and a more complex Ensemble Kalman Filter approach.

4.6.1 Auto-regressive updating using local streamflow observations

In the auto-regressive updating method, the simulated discharge at a gauging station is replaced with observations whenever observations are available. On days without observations (e.g. the first day of the forecast), the discharge is estimated from the last known model error multiplied by a scaling factor – the 'auto-regressive' parameter (AR) – that decays exponentially with time[5]. Such days without observations can be the future, in which case the method provides forecasts. The AR parameter varies between 0 and 1, and is a rough indication of the hydrological memory at a given gauge, i.e. the information propagated forward into the future.

To apply the auto-regressive updating method, it is necessary to determine an appropriate AR factor. It was obtained by sampling a range of AR values and calculating the error between observed and simulated discharges using the West Africa domain from the World-Wide HYPE model version 1.3.6 for lead days 1-10. The AR factor that provided the minimum absolute error at each gauge was then selected (Figure 21). The AR values ranged between 0.0008 and 0.9999 with a median of 0.931. The high AR values indicate that there is a significant hydrological memory at most stations, and hence that past observations can be utilized to improve forecasts from one day to the next.

The AR parameter varies in space, but can potentially also vary between seasons and for different lead times. Figure 22 shows the AR factor at the different stations for the wet season (1st June-31st October) and the dry season (1st November-31st May). In general, we found no significant difference between seasons except for at a few stations.

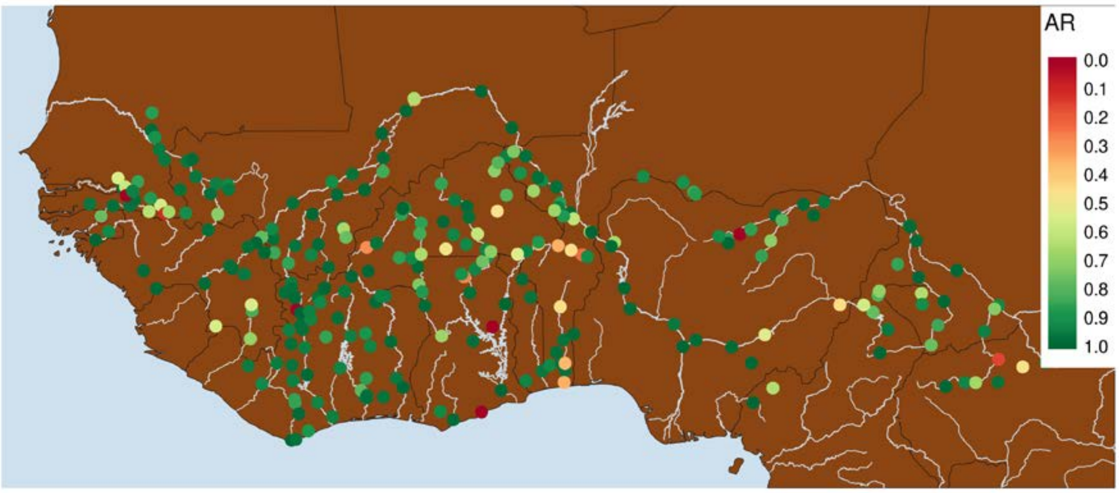

Figure 21. The values of the auto-regressive (AR) parameter providing the minimum absolute error between simulated and observed daily discharge at all available gauged locations in the West Africa domain of the WW-HYPE 3.1.6 for the period 1979-2019.

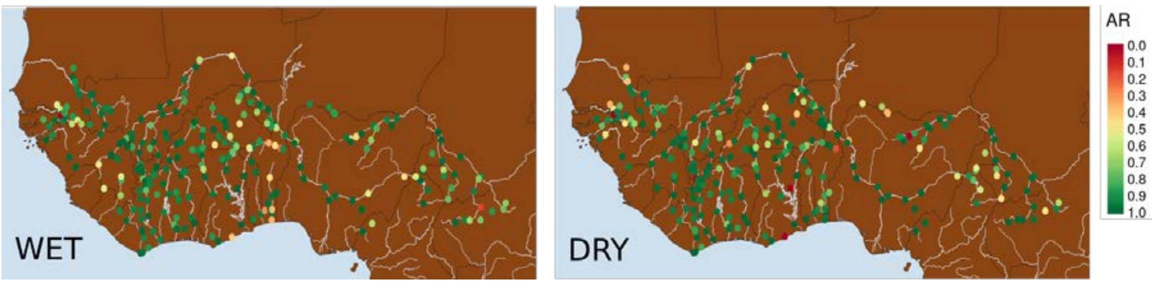

Figure 22. The AR factors for the wet season (left) and dry season (right).

[5] http://www.smhi.net/hype/wiki/doku.php?id=start:hype_tutorials:updating

To test the auto-regressive updating method using local streamflow observations, we carried out a re-forecasting experiment focussing on the 2019 flood in Niamey. Specifically, we re-forecasted every day between 2019-08-15 and 2019-09-15 with the current operational model configuration (without any assimilation, “NO UPDATE”) and with a configuration in which the auto-regressive method was activated (“AR UPDATE”). The results show that the AR-updated model is significantly better than the model without AR-updating (Figure 23). The average error is reduced from −86% to less than 2% for 1-day lead times (Table 4), which is a tremendous improvement.

Forecast accuracy tends to decrease with lead time, hence the 1-day forecast is expected to have highest accuracy. However, longer lead times may still be indicative of events that could take place. One can therefore consider the envelope of all forecasts valid for a given date to be a kind of confidence interval (here represented by the range between the 1-day and 10-day lead times). This is range is shown in Figure 23 on the right. It indicates that observations were within the forecasted confidence interval until 24th August and then slightly below until the peak for the AR UPDATE model, while they were only within the band for a few days for the NO UPDATE model. The figure also reveals another interesting insight. The 10-day lead time displayed a very high peak on 21st August (Figure 23, right), which was no longer present a few days later, when the 1-day forecast was made (Figure 23, left). This trend has also been seen in other events. It indicates that the meteorological forecast model from ECMWF radically changes its precipitation forecast from one day to another. It also underlines the need for practitioners to revisit forecasts day after day, in order to better understand the flood risks ahead.

All in all, the auto-regressive updating method using local streamflow observations has great potential to improve forecasts at and downstream of gauged locations. However, it critically depends on up-to-date observations being continuously delivered to the system, which is currently a challenge in the region.

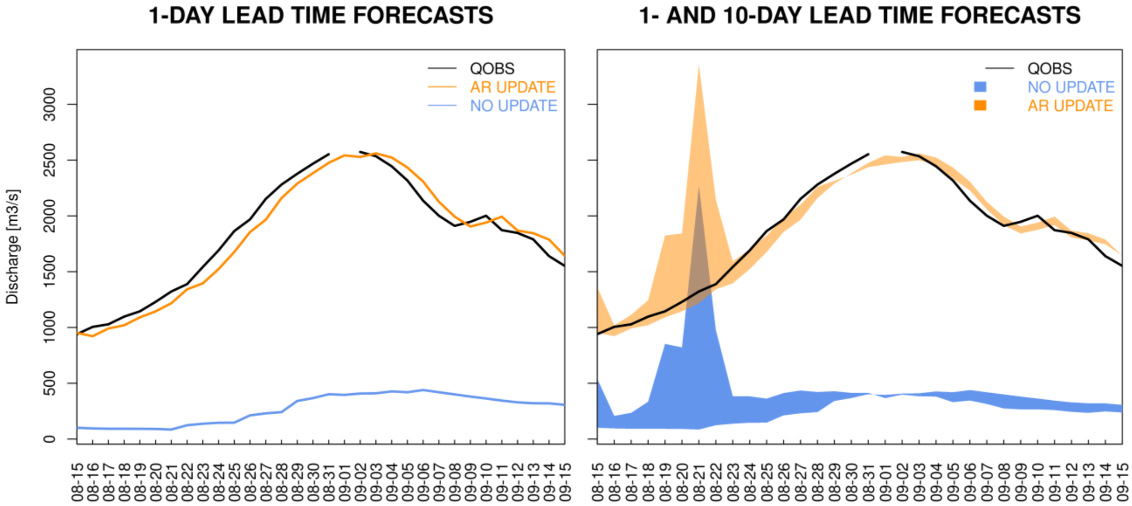

Figure 23. Observed and re-forecasted discharge of the Niger River at Niamey during the high flow event in August-September 2019. Black is observed discharge in both figures. The re-forecasted time series are the results of 32 daily forecasts from 2019-08-15 to 2019-09-15 each providing its forecasted value (left) for the 1-day lead time, and (right) as a range between the 1-day and 10-day lead times. The blue (NO UPDATE) is the currently deployed model configuration (without assimilation), and the orange (AR UPDATE) is the model employing auto-regressive updating with observed discharge.

Table 4. Average deviation between simulated and observed streamflow for the current operational model without assimilation (NO UPDATE), and the model in which the auto-regressive method was activated (AR UPDATE).

Model	1-day-ahead forecast	10-day-ahead forecast
NO UPDATE	−86%	−71%
AR UPDATE	−1.7%	12%

4.6.2 Exploitation of satellite altimetry water level data

One of the main objectives of the FANFAR project is to exploit satellite altimetry-based river water level data to improve the hydrological modelling and forecasting for West Africa. Inland water satellite altimetry has the potential to provide water level data on thousands of unmonitored locations in a river network of the size considered in FANFAR, and could ideally provide data both for model improvements in ungauged basins, as well as input data to real-time assimilation and forecasting – either as backup or complement to data from physical stations that might be available only with some lag time. There is two main problems with the satellite altimetry data: 1) the return period to the same location is often much longer than daily (the Sentinel-3 data used in FANFAR has a return period of 27 days), 2) the footprint of the altimetry sensor requires a certain river width to provide any data (a minimum is about 300 m), and 3) even for rivers considerably wider that this, the data quality is depending much on local adaptation of filters and signal processing. Still, a lot of uncertain observations might be better than no observations at all.

In FANFAR, satellite altimetry data will be used in three steps:

1) To establish rating curves – an empirical relationship between discharge and water level – at remote unmonitored locations (so-called virtual stations); combining discharge simulated by the model and water level from the satellite altimetry.
2) To apply the rating curves for virtual and physical stations to enable prediction of river water level from the river discharge simulated by the models.
3) To apply the virtual station rating curves on the altimetry-based water levels, to generate altimetry-based river discharge estimates, which can be assimilated in the forecast models with the procedure described in the previous section.

For best result in the first step, we use the simulated river discharge based on assimilation of discharge observations from physical stations. The quality of the river discharge will be better in model sub-basins close to upstream physical stations with data, so in theory, the generation of virtual station rating curves should be applied in an iterative procedure. In the first round, rating curves are only generated for virtual stations within a critical range from the physical stations with data in overlapping time periods. In a second round, the simulated discharge data needed for virtual stations further downstream can be improved by discharge data estimated at already established upstream virtual stations.

We use the common exponential rating curve relating river discharge (Q, m3/s) to river water level (w, m above sea level) by:

$$Q = c(w - w_0)^p$$

where c and p are empirical linear and exponential rate coefficients, and w0 is a zero flow water level coefficient that can be based on local knowledge or used as an empirical coefficient as well.

Rating curves are established for physical stations by combining real discharge measurements with water level observations at a range of streamflow conditions. To exemplify this, we estimated a rating curve for the Niger River at Niamey using the official discharge and water level data from the Niamey station (Figure 24, left). This is a theoretical example, since the discharge data have of course been estimated from the observed water levels in the first place with an existing rating curve. However, we are now able to apply this rating curve in the HYPE model to predict water levels also for the time periods without in-situ data (Figure 24, right panel), as well as in the forecasting mode to provide forecasted water levels and not only river discharges.

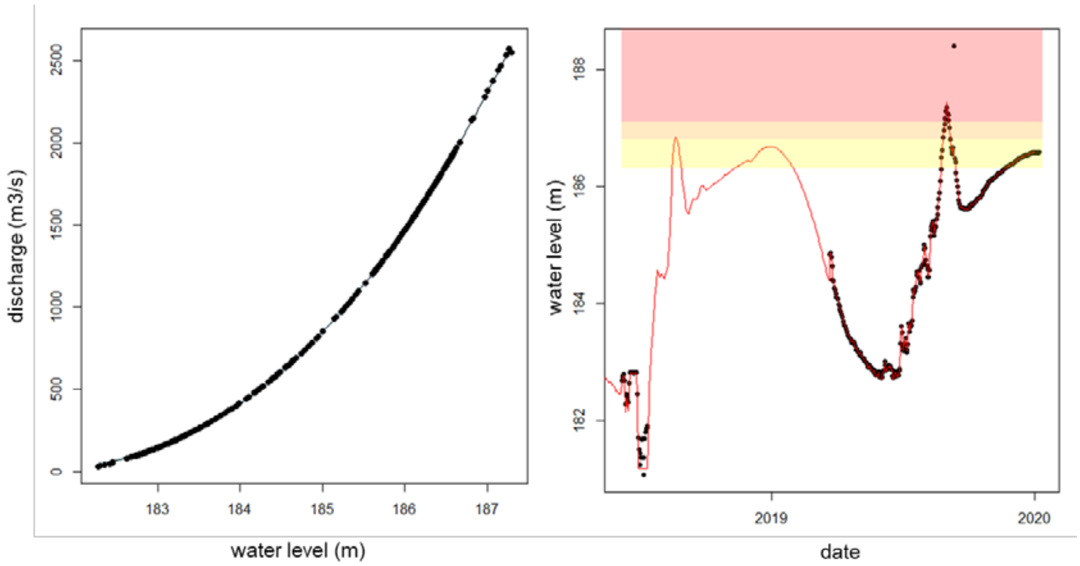

Figure 24. Left: rating curve estimated based on in-situ observations at Niger River at Niamey, and right: water level for Niger River at Niamey calculated from simulated discharge with the World-Wide HYPE model including updating with observed discharge. The black dots are time steps with observations, and the red line is the simulation. Coloured ribbons correspond to the local warning thresholds at 530, 580 and 610 cm above the gauge zero height (181m). The figure illustrates the capacity of the model to provide water levels at key river sections, both for forecasts and for periods without observations.

In a second example, we established a rating curve for a virtual station upstream of Niamey (S3B-200, Figure 25). There was only 11 data points available since 2019, and the uncertainty in the rating curve is much larger than for the in-situ data (right graph of Figure 25). If we apply this rating curve to the simulated discharge in the corresponding location, and compare this to the observed water level in Niamey it still provides relatively good results, especially for flows above the warning thresholds (Figure 26).

This is first results of combining modelled discharges with altimetry based water level data in the FANFAR project. We will continue to establish virtual station rating curves at as many locations as possible, to enable water level forecasting, as well as updating the streamflow simulations with the altimetry based discharge data. However, it is important to remember that the in-situ observations are fundamental to obtain the initial discharge data set for the virtual station rating curves.

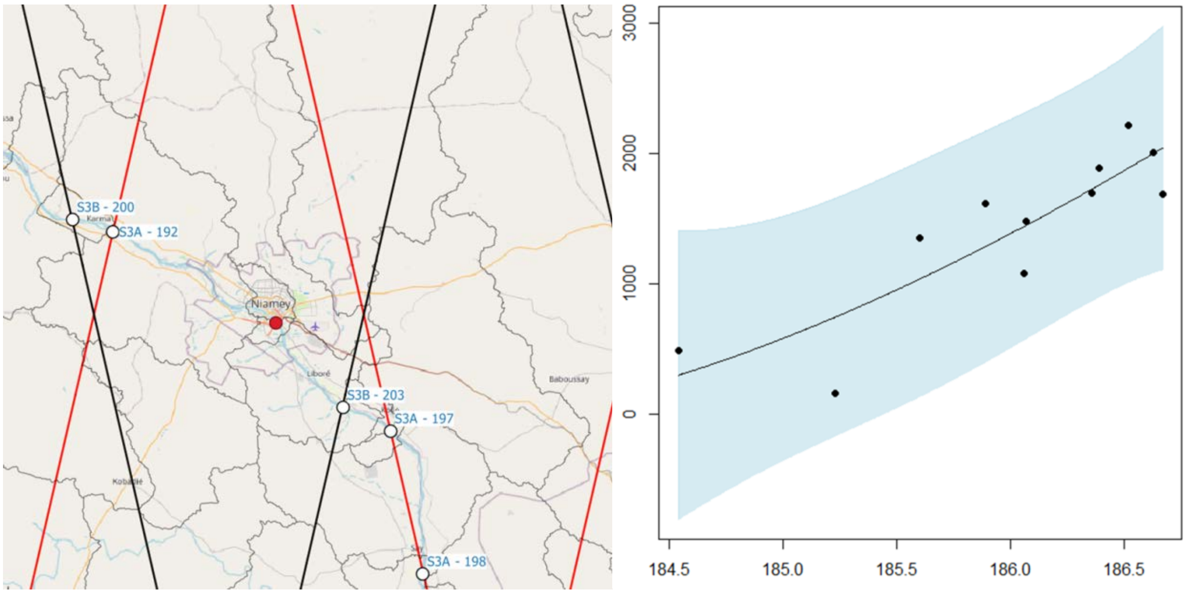

Figure 25. Left: Map of Niger River at Niamey indicating the position of potential virtual stations at the cross-points of the Sentinel-3A and B satellites with the Niger River. Right: rating curve for the virtual station S3B-200 relating water level (X-axis, m) to discharge (Y-axis, m^3/s).

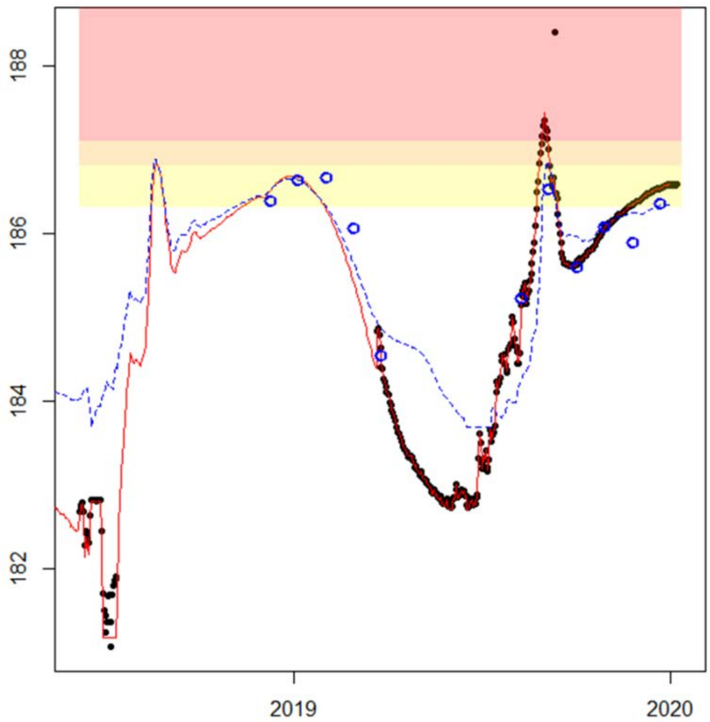

Figure 26. Water level derived from simulated river discharge at Niamey using rating curves based on in-situ water level data (red curve, physical station) and altimetry-based water level data (blue dashed curve, virtual station), compared to observed water levels (black dots). The altimetry-based water level data from S3B-200 is shown by blue circles. Model: World Wide-HYPE including updating with observed discharge.

5. Forecasting skill with respect to high-flows

A core aim of the FANFAR project is to produce forecasts of flood severity for the coming 10 days. The priority of the models is to display forecasts of severity levels that are as close as possible to the actual severity levels. For this purpose, the model does not have to mimic exactly the observed discharge, but rather be able to simulate and predict when the flood hazard thresholds defined for each sub-basin may be exceeded. Therefore, the forecasts are evaluated *with respect* to the flood hazard thresholds (rather than directly toward observed discharge values).

The main approach to represent flood severity levels in FANFAR is to compare forecasted discharge with flood-hazard thresholds based on return periods[6]. A flood with low severity level is forecasted when the discharge is between return periods of 2 and 5 years. The severity level is medium when the discharge is between return periods of 5 and 30 years and it is high when discharge is over the 30-year return period threshold. However, since model simulations are not perfect, the return periods of simulated and observed discharge are different. In this context, a severity level can be well forecasted even if the model discharge simulations are wrong. In contrast, if the model gives a discharge value that corresponds to the observation but not the appropriate severity level the forecast will be wrong (because return periods are different).

The following analysis aims at evaluating the forecast on their capacity to detect severity levels rather than on their capacity to predict absolute discharge values. The analysis was based on the contingency table (Table 5). In this table, an "event" is defined as the discharge being over a given flood hazard threshold. This analysis takes into account the differences between observed and simulated return periods, i.e. observed events are defined with respect to past observations while forecasted events are defined with respect to a long-term historic model simulation. If the measured discharge is higher than the discharge for a given return period (computed based on observations and thus valid for the observations), we consider that the event is observed. In parallel, if the forecasted discharge is higher than a given return period (computed based on simulations and thus valid for the simulations), we consider that the event is forecasted. A good forecasting system would forecast events that are actually observed/simulated (high number of hits) but would not forecast events when they did not actually occur (low number of false alarms).

Table 5. Description of the contingency table (source: National Oceanic and Atmospheric Administration).

2x2 Contingency Table		Event Observed	
		Yes	No
Event Forecast	Yes	a (hits)	b (false alarms)
	No	c (misses)	d (correct negatives)

To evaluate the model performance, two scores derived from the contingency table were then calculated for the available gauged stations:

- The probability of detection (POD) that measures the probability of detecting an observed event. It is calculated using the following equation and the best possible value is 1.

  POD=Hits/(Hits+Misses)
- The false alarm ratio (FAR) which measures the probability that a forecasted event is a false alarm. It is calculated as follows and its best value is 0.

  FAR=False alarms/(False alarms+Hits)

[6] https://en.wikipedia.org/wiki/Return_period

This performance analysis was carried out using the Hydrology-TEP re-forecasting functionality. All days between 2017 and 2019 were re-forecasted using the West Africa subset of the World Wide-HYPE model version 1.3.6 forced with HydroGFD2 and the ECMWF 10-day deterministic forecasts. Then, the forecasts were evaluated for 9 gauged stations. These 9 stations were the only stations with enough available data to calculate all return periods (at least 30 years of available data were necessary) and to calculate scores using the re-forecasts (at least one year of data was necessary between 2017 and 2019).

The results indicate high probabilities of flood event detection (POD) in the upper Niger area but low probabilities in the middle and lower Niger (Figure 27). The false alarm ratio (FAR) has opposite results with low false alarm ratios in the middle and lower Niger but high ratios in the upper Niger. This shows that the model has either a good capacity of detection of events but at the cost of generating too many false alarms or, on the opposite, no false alarms but a low probability of detecting events. However, this conclusion has to be nuanced because the available data was too scarce to make a complete diagnosis of the model forecasting performances.

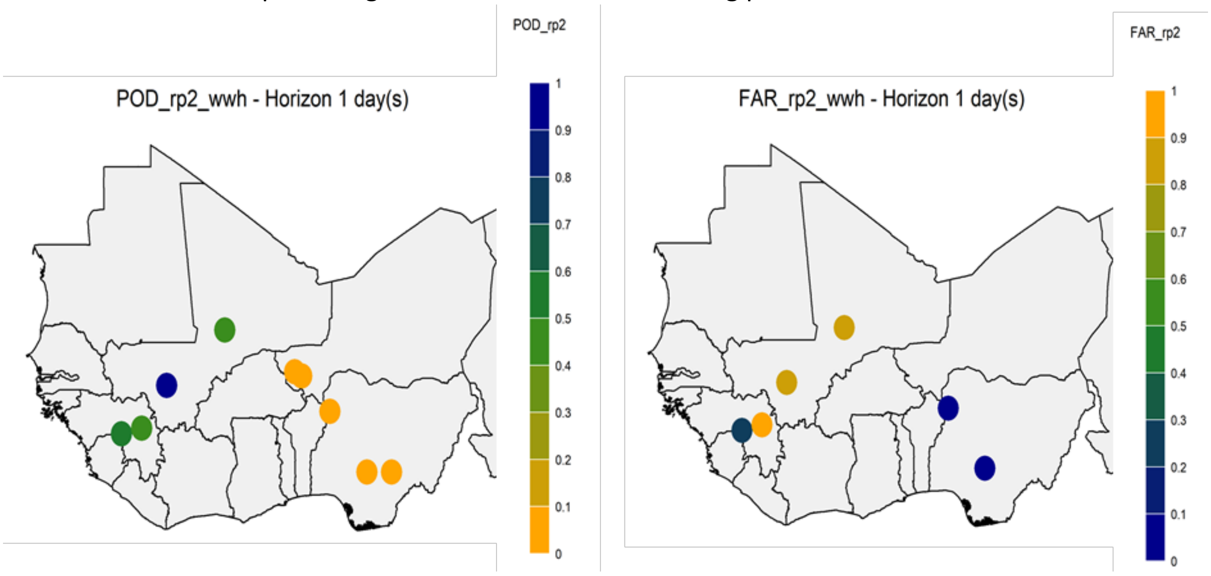

Figure 27. Summary of forecast skills of World Wide-HYPE 1.3.6 forced with HydroGFD2 and the ECMWF deterministic forecasts for the period 2017-2019 in terms of probability-of-detection (POD) and false alarm ratio (FAR) for the 9 gauged stations evaluated using the re-forecast experiment.

The only location where the system performed well for both scores is at the Faranah gauging station in Guinea (the most western station in the maps of Figure 27). The hydrograph in Figure 28 shows this result. It also shows that, even if the discharge is underestimated, the performance of the forecasting system is good, i.e. forecasted streamflow is generally above its simulation-based return-period threshold when the observed streamflow is above its observation-based return-period threshold.

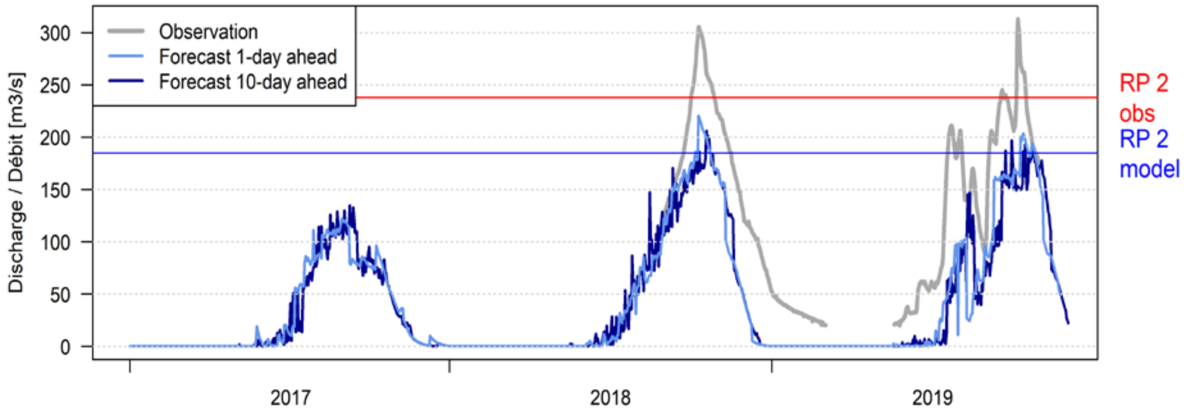

Figure 28. Time series of observed and re-forecasted discharge of the Niger River at Faranah (Guinea) including the observed and simulated 2 years return periods thresholds (RP2 obs and RP2 model). Note that the re-forecasted time series are the result of 1095 forecasts each providing its 1-day and 10-day forecasted value, here stitched together into two time series (i.e. each forecast was limited to its 1-day and 10-day forecast horizon respectively). Model configuration: World Wide-HYPE 1.3.6 forced with HydroGFD2 and the ECMWF deterministic forecasts.

Another way of evaluating the FANFAR forecasting system is to analyse a given flood event as a case study. This kind of case study analysis was done on the flood event that occurred in Niamey, Niger during the first half of September 2019. The Niamey station observed a "high local flood situation" according to the Niger Basin Authority monthly bulletin. At this time, the two FANFAR models (Niger HYPE and West Africa HYPE) did not forecast discharge that would exceed any flood hazard threshold. A further analysis based on the re-forecast provided several indications that could help explain this failure. For the Niger-HYPE model, the failure seems to be the result of a forcing data issue. Indeed, the model gave a correct severity level after the peak, on the 11th of September (Figure 29). This corresponds to the day of the monthly update of the HydroGFD1 meteorological hindcast data. This underlines the critical need for having accurate and up-to-date meteorological data in order to forecast floods in the region.

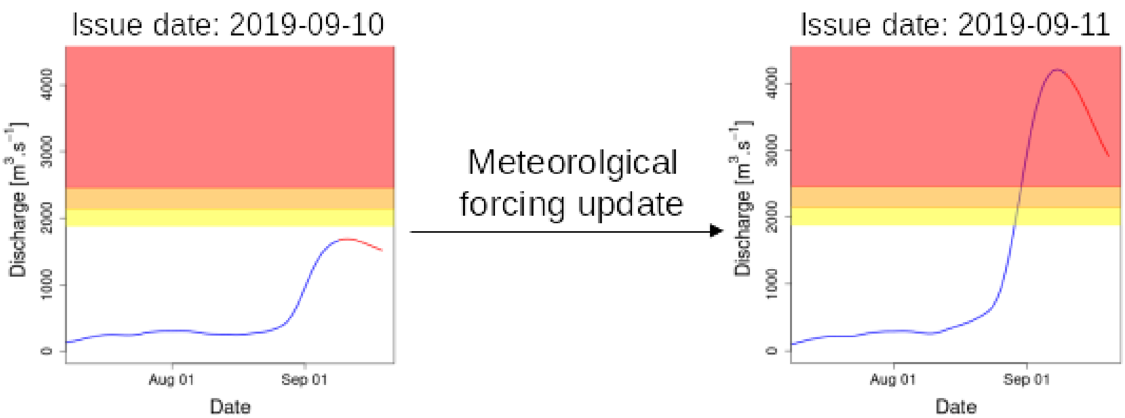

Figure 29. Comparison of the discharge forecast of the Niger River at Niamey made with Niger-HYPE 2.23 on the 10th and the 11th of September 2019. The large difference between the two forecasts comes just after the monthly update of the HydroGFD1 meteorological hindcast data.

The reasons for the forecasting failure of World Wide HYPE v1.3.6 for this event are more complex. The forcing data could be one factor leading to failure, but this is not the only factor since the forecast on the 11th of September shows no indication of surpassing the flood hazard thresholds. As shown in Figure 30 (left), the simulated discharge is more-or-less flat during the entire event. The hypothesis is that the first calibration of the model (step 1 as described in section 4.3.2) was not adapted to simulate this flood. We therefore carried out a new re-forecasting test with the second calibration of the model (step 2 as described in section 4.3.2). This resulted in a much more appropriate representation of the flood (forecasting a flood of high severity).

These two evaluations of forecast skill help to understand the performance of the FANFAR forecasting system. They show that the current system does not always provide accurate forecasts, which corresponds with the user feedback provided during the third FANFAR workshop in Nigeria in February 2020. However, the first tests made with the recalibrated model shows encouraging results (Figure 30, right). The use of new forcing data and the assimilation of observations also have the potential to improve the forecast. Intense efforts are now ongoing to implement all these approaches in the forecasting chain. Subsequently, a new forecast skill experiment will be carried out to better understand the system performance under different configurations.

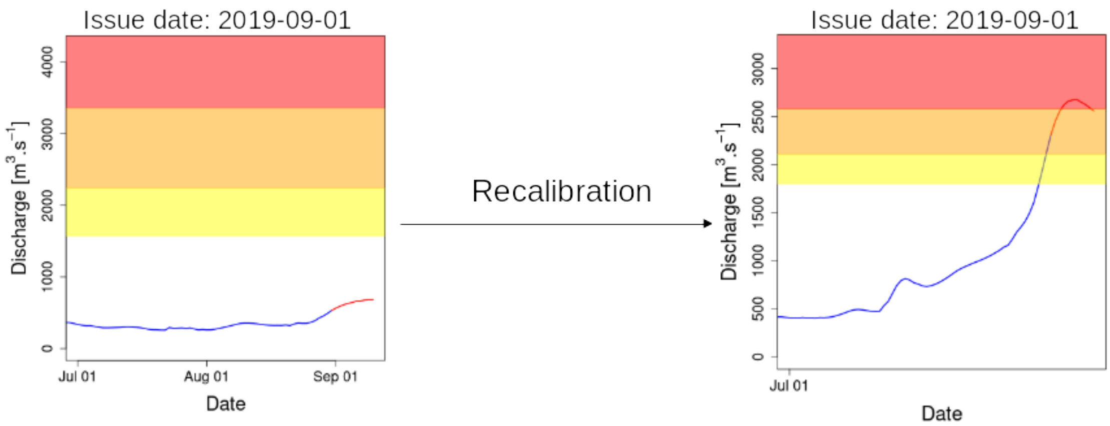

Figure 30. Comparison of the discharge forecast of the Niger River at Niamey made with West Africa subset of the World Wide HYPE model with the first (left) and second calibration (right), forecast start date on the 1st of September 2019.

6. Design and deployment of operational forecasting chain

An operational forecasting chain based on Niger-HYPE and the West African subset of World-Wide HYPE has been designed and deployed in an operational prototype on Hydrology-TEP. Every day it produces 10-day hydrological forecasts for the two model domains. The forecasting chain has been producing automatic forecasts on Hydrology-TEP since beginning of 2019 for Niger-HYPE, and since beginning of 2020 for World-Wide HYPE (but both since September 2018 on SMHI servers).

The basic design of the operational forecasting chain was already defined at the start of the project (Figure 1). In essence it consists of 1) a set of input data (meteorological data, local observations, EO observations), 2) the processing of the input data in two runs with each hydrological model (one to initialize the model for past/present conditions, and one for forecasting the coming 10 days), 3) derivation of flood risk information (e.g. comparing with past flow conditions), 4) information distribution (e.g. on a website), and 5) daily execution and monitoring.

During the project, we have worked incrementally to specify the forecasting chain, to implement it in computer code, and to deploy it operationally on Hydrology-TEP using its new scheduling capabilities. This incremental work is illustrated in Figure 31, Figure 32, and Figure 33. In September 2018 (Figure 31) the system could only produce on-demand/interactive forecasts for Niger-HYPE and interactive EO water levels, no river observations were available, only one meteorological hindcast/analysis dataset was used (HydroGFDv1), and the web visualisation was showing information from SMHI rather than from Hydrology-TEP. Seven months later (April 2019, Figure 32), the automatic scheduling service was available and deployed for both the Niger-HYPE forecasting and EO water level production at Hydrology-TEP (producing updated data every day), and a new version of the meteorological hindcast data was also ingested in the system. By that time we had also developed the connections between fanfar.eu and Hydrology-TEP to display Niger-HYPE forecasts on the interactive visualisation portal, and developed the Email, SMS and API distribution channels (see Deliverable 3.5). By the third FANFAR workshop (February 2020, Figure 33), the system had progressed to include forecasts for the entire West African domain (with the World-Wide HYPE model), ingestion of local streamflow observations, and a set changes to the interactive visualisation portal (including context layers, improved navigation, data download, terminology harmonization). We had also finalized the code for post-processing (Deliverable 3.3), but not yet deployed it operationally on Hydrology-TEP. The plan for the future essentially contains updates and improvements to all components of the chain, but particularly to assimilate local observations and EO-water level data, to deploy a new post-processing service on Hydrology-TEP, to integrate West African meteorological data, to run multiple parallel forecasting chains (with different configurations), and to distribute the data and flood risk information through the visualisation portal, the Email and SMS recipients, as well as the API subscribers (e.g. to the ANADIA-Niger project).

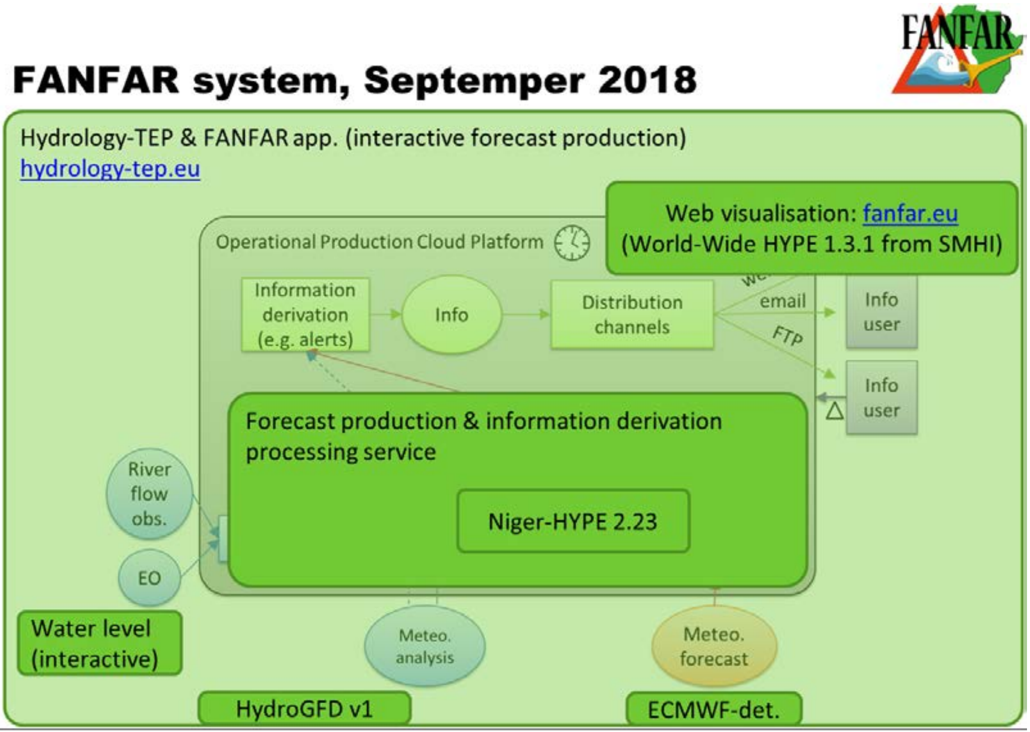

Figure 31. FANFAR system overview, specification and status as of September 2018

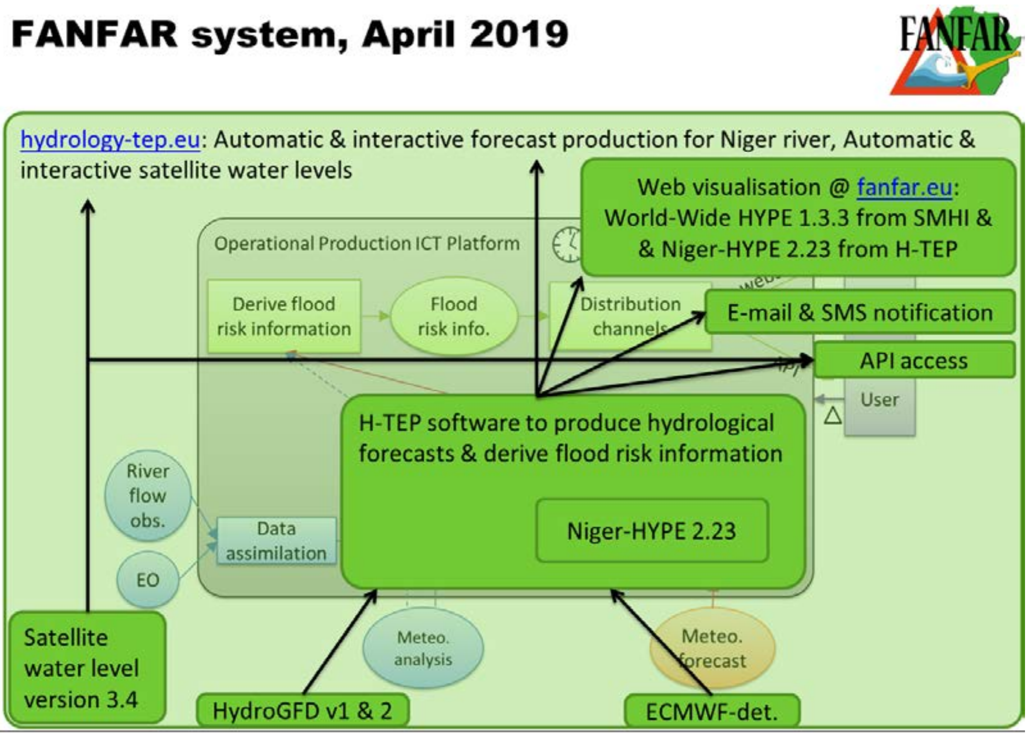

Figure 32. FANFAR system overview, specification and status as of April 2019 (arrows show data flows)

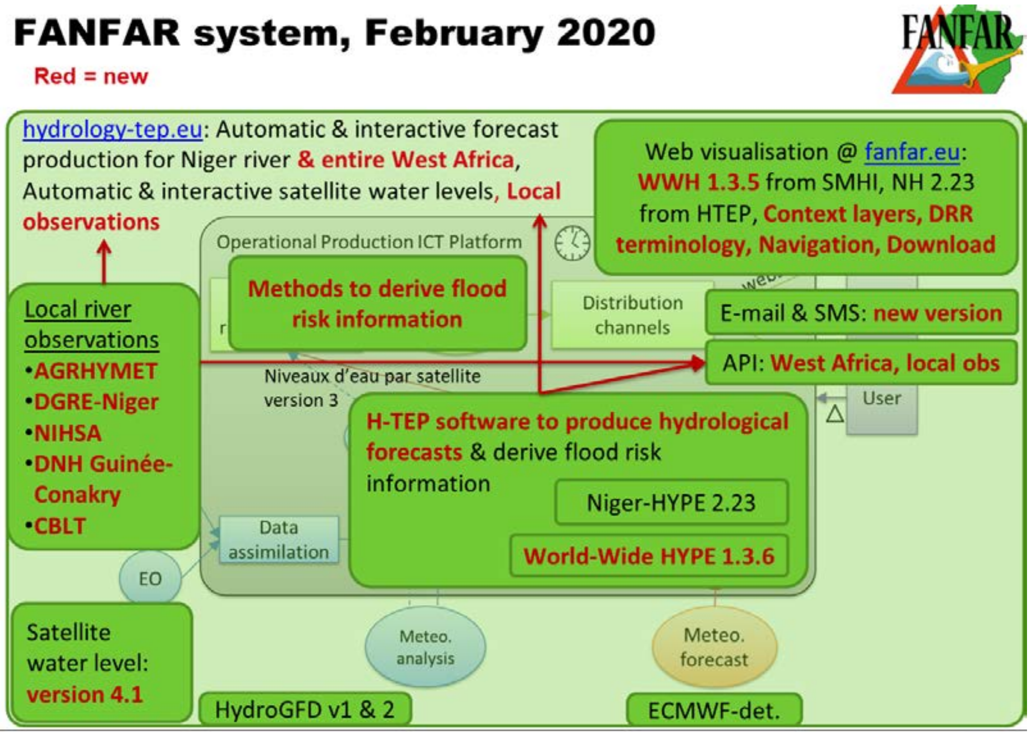

Figure 33. FANFAR system overview, specification and status as of February 2020 (red shows new additions).

A new processing service on Hydrology-TEP has been developed that implements the model runs required to make hydrological forecasts. This service is called "FANFAR-forecast", and it has been documented on the FANFAR knowledge base[7], and its source code is openly available on GitHub[8].

The first date of the forecast is called "Forecast issue date". In this service, it is possible to set the forecast issue date to the current date, or to any other date back to 2017-01-01. The service always makes two simulations: first a warm-up simulation (the hindcast) ending on the day before the forecast issue date, then the actual forecast simulation beginning from the end of the warm-up simulation and running 10 days ahead (the forecast). The service has been implemented in a flexible manner as to allow for multiple model configurations, different run types (e.g. operational for the current day, and re-forecasting of past dates), different initialisation period lengths, different outputs (variables, catchments) etc. The outputs consist of standard HYPE output files[9] for both hindcast and forecast runs, along with supplementary files (e.g. log-files and GIS files). Figure 34 illustrates the workflow of the service. Figure 35 shows the user interface of the processing service, and the output files created. The service has been operational since beginning of 2020 on Hydrology-TEP, but retrospective runs have also been executed to provide an archive of daily forecasts back to 2017-01-01 (currently 1205 forecasts are available with the West African subset of the World-Wide HYPE 1.3.6 model). The next steps in the development of the service consist of assimilating observations, and supporting additional model configurations.

[7] https://knowledge.terradue.com/display/FANFAR/EN%3A+FANFAR+Forecast+service
[8] https://github.com/hydrology-tep/fanfar-forecast
[9] http://www.smhi.net/hype/wiki/doku.php?id=start:hype_file_reference#output_files

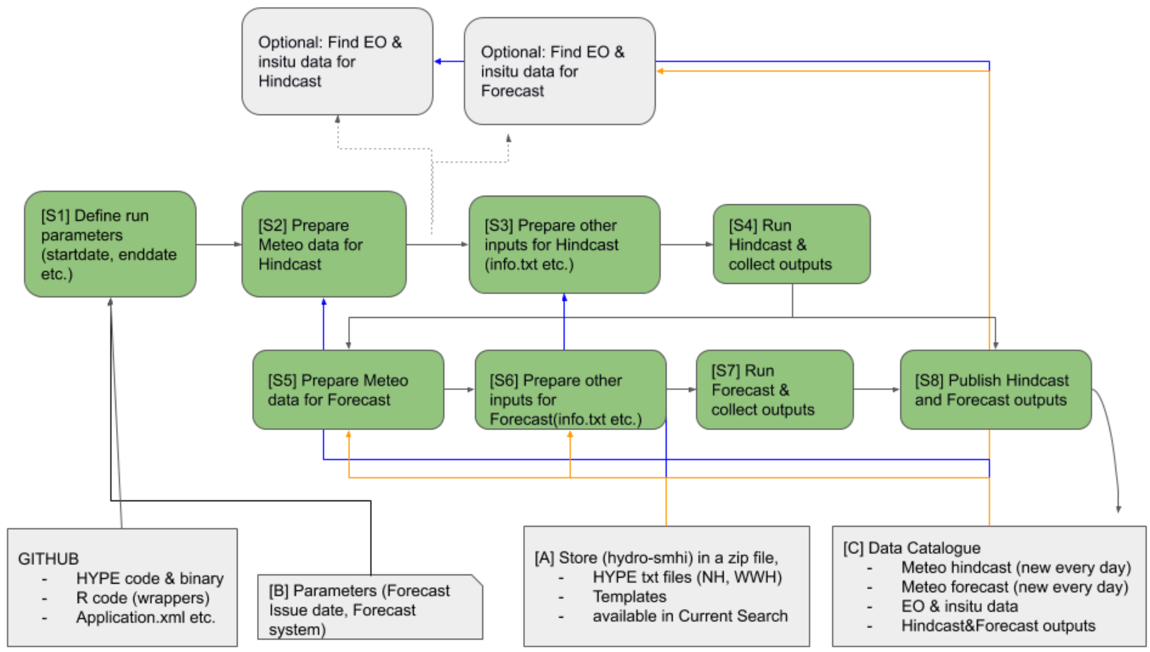

Figure 34. Workflow of the FANFAR-forecast service on Hydrology-TEP.

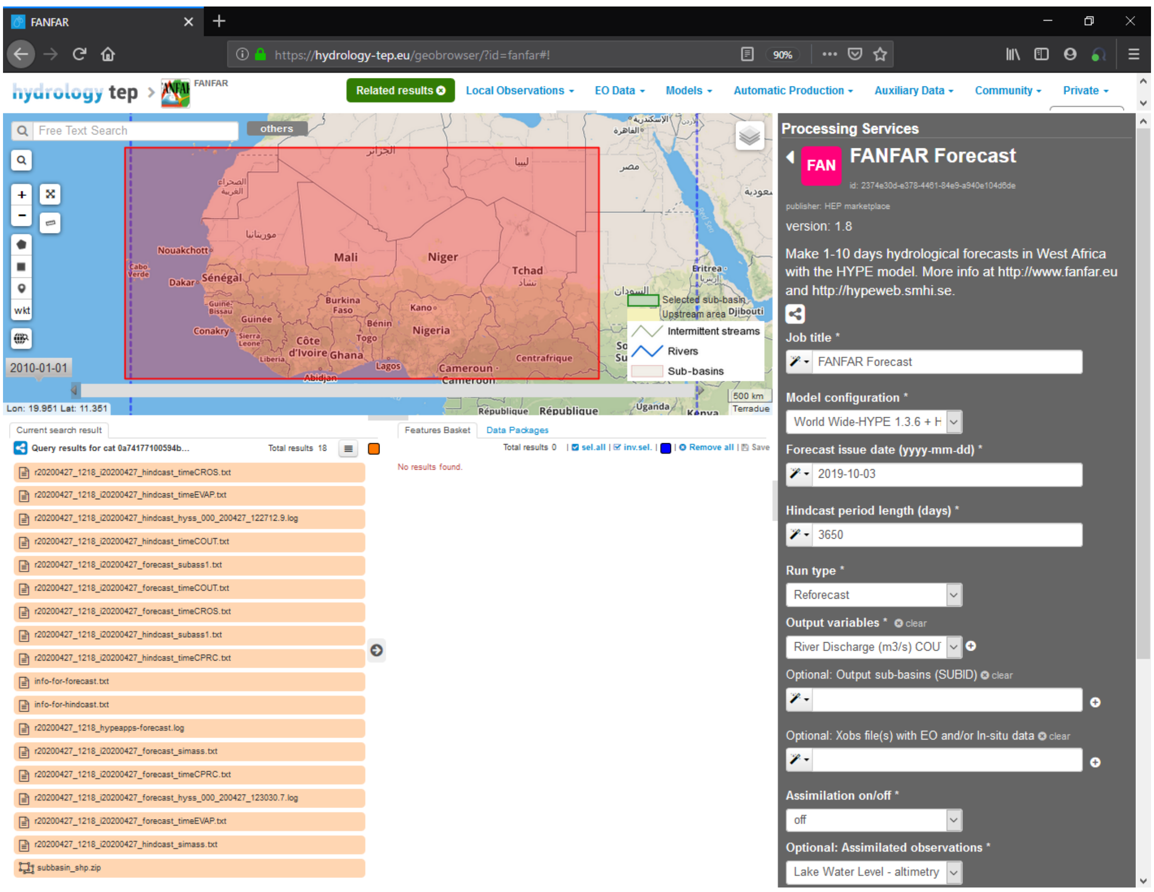

Figure 35. Overview of the FANFAR forecast service components at the Hydrology-TEP. The grey right pane is the user interface of the service, and the orange pane on the bottom left provides the results of the service (the hindcast and forecasts simulation outputs), here shown for the issue date 2020-04-27.

7. References

Aich, V., Liersch, S., Vetter, T., Huang, S., Tecklenburg, J., Hoffmann, P., Koch, H., Fournet, S., Krysanova, V., Müller, E. N., and Hattermann, F. F. (2014): Comparing impacts of climate change on streamflow in four large African river basins, Hydrol. Earth Syst. Sci., 18, 1305–1321, https://doi.org/10.5194/hess-18-1305-2014

Andersson J.C.M., Arheimer B., Traoré F., Gustafsson D., Ali A. (2017): Process refinements improve a hydrological model concept applied to the Niger River basin. Hydrological Processes, 31, 4540–4554, https://doi.org/10.1002/hyp.11376

Arheimer B., Pimentel R., Isberg K., Crochemore L., Andersson J.C.M., Hasan A., and Pineda L. (2020): Global catchment modelling using World-Wide HYPE (WWH), open data, and stepwise parameter estimation, Hydrol. Earth Syst. Sci., 24, 535–559, https://doi.org/10.5194/hess-24-535-2020

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R. (1998): Large area hydrologic modeling and assessment - Part 1: Model development, J. Am. Water Resour. As., 34, 73–89.

Lindström G., Pers C., Rosberg J., Strömqvist J., Arheimer B. (2010) Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. Hydrology Research 41.3–4, 295-319.

Strömqvist, J., Arheimer, B., Dahné, J., Donnelly, C. and Lindström, G. (2012). Water and nutrient predictions in ungauged basins – Set-up and evaluation of a model at the national scale. Hydrological Sciences Journal 57(2):229-247

Wang-Erlandsson, L., Bastiaanssen, W. G. M., Gao, H., Jägermeyr, J., Senay, G. B., van Dijk, A. I. J. M., Guerschman, J. P., Keys, P. W., Gordon, L. J., and Savenije, H. H. G.: Global root zone storage capacity from satellite-based evaporation, Hydrol. Earth Syst. Sci., 20, 1459–1481, https://doi.org/10.5194/hess-20-1459-2016, 2016.