



OPW
The Office of Public Works
Oifig na nOibreacha Poiblí

REPORT

Met Éireann

Services for the Review, Development and Trial of a Range of Hydrological Models and Integrator Systems for Use in Operational Fluvial Flood Forecasting in Ireland

Summary Report

10th of June 2020 - version 3.0





International Marine & Dredging Consultants

Colophon

International Marine & Dredging Consultants

Address: Van Immerseelstraat 66, 2018 Antwerp, Belgium

: + 32 3 270 92 95

: + 32 3 235 67 11

Email: info@imdc.be

Website: www.imdc.be

Document Identification

Title	Summary Report
Project	Services for the Review, Development and Trial of a Range of Hydrological Models and Integrator Systems for Use in Operational Fluvial Flood Forecasting in Ireland
Client	Met Éireann - Department
Tender	RFT 125339
Document ref	I/RA/12172/20.029/SME
Document name	K:\PROJECTS\12\12172_P011988 - FFS Ireland\10-Rap\RA20029_summaryreport\RA20029_summaryreport_v3.0.docx

Revision

Version	Date	Description	Authors	Checked	Approved
1.0	30/01/20	Draft Final Report	SME	BPA	SME
2.0	10/4/20	Remarks included	SME	LMA	SME
3.0	10/6/20	Final version	SME	SME	SME

Approval

Authors	SME	
Checked	SME	
Approved	SME	

Contact within IMDC

Name	Steven Smets
Phone number	+32 3 270 92 95
e-mail	Steven.smets@imdc.be

Distribution List

	Hard copy	
x	Pdf	Eoin Sherlock, Sinéad Duffy, Oliver Nicholson, Fasil Gebre

Executive summary

In this project, candidate hydrological models and integrator systems for a future Irish National Flood Forecasting and Warning Service have been reviewed.

In Stage 1 of the project an extensive literature review was carried out to look into potential hydrological models and integrator systems and an assessment was made of the data input sources to be used in the models and systems. A longlist of fourteen hydrological models were evaluated in the literature review. Based on a set of requirements each model was scored, and then weighted using a multi criteria analysis. This led to the selection of four hydrological models for further testing in Stage 2 and 3 of the project. The selected models are WRF-HYDRO, WFlow, HYPE, URBS. These models cover a range of modelling approaches, from distributed and physically-based models to purely conceptual models.

The same approach was used for evaluating the integrator systems, leading to a longlist of five integrator systems. After scoring and evaluation, two integrator systems were selected for further testing in Stage 4 of the project: Delft FEWS and Mike Operations. With respect to data inputs, the main focus was given to the meteorological input data as they are the main driver for the hydrological models. A mix of observed data (gridded daily rainfall, weather station data) and forecasted data (HARMONIE) is proposed for further testing in the project.

In Stage 2 of the project, four hydrological models were built for each of the 5 catchments (Nore, Slaney, Barrow, Moy & Killala, Shannon). The resulting models were calibrated and validated. To make this process as objective as possible, calibration metrics and calibration time periods, etc., were kept identical. Seven different calibration metrics were taken into account. The HYPE models achieved on average the highest scores, followed by URBS, WFlow and WRF-HYDRO. Additionally, the model building process resulted in some recommendations for further improvements regarding additional gauging stations and additional observed rainfall gauges.

In Stage 3 of the project, the calibrated and validated models were tested using historical forecasted data (HARMONIE). Ten historical storm periods/events between 2013 and 2018 were simulated with a forecast lead time of upto 54 hours and a model run frequency of 6 hours. The results of these simulations were compared with the observed discharges using four different metrics: Relative Mean Absolute Error (RMAE), Nash Sutcliffe Efficiency (NSE), Peak Weighted Root Mean Square Error (PWRMSE) and peak timing. The outcome of this analysis showed URBS to be performing best for most of the metrics, whilst HYPE is a close second. WFlow and WRF-Hydro perform worse, consistent with their model performance in Stage 2. These differences can be explained both by some minor adaptations of the methodology between the calibration phase and the testing phase, and more in general by the focus of the testing phase on metrics taking into account the full shape of the hydrograph rather than only the peak values.

An important conclusion of Stage 3 is that the required target accuracy is in general not met: *“For this Contract the target forecast accuracy will be for a model that predicts the actual peak flow to within +/- 10% and 6 hours of the actual peak at the gauged forecast points, and with a lead time of greater than 24 hours.”*. For some station/storm combinations good results are achieved, but for many other combinations this is not the case.

In Stage 4 the two selected integrator systems were setup and configured for the Barrow catchment for the 4 different models. The aim of this Stage was to test configurability and robustness of the integrator systems, rather than for performance of the forecasts (which is related to the models). This configuration exercise brought up interesting differences in user friendliness, which were reported. After configuration, both integrator systems were run with real-time data during a trial period of eight weeks. In general, both systems (integrator systems and models) worked well during the trial period. The issues that were encountered were more related to missing real-time input data rather than with the systems themselves.

Based on the hardware setups that was used for the testing of the two integrator systems and the results of the trial period, an assessment was carried out on some ICT considerations for the future production setup of the National Flood Forecasting and Warning System.

Based on the results of the previous Stages, HYPE is proposed to be the selected hydrological model and Delft FEWS is the preferred integrator system for the National Flood Forecasting and Warning System.

To finish the report, a chapter with recommendations regarding additional data, dataflow improvements, hydrological model/integrator system improvements is included.

Table of Contents

1. INTRODUCTION	8
2. COMPREHENSIVE LITERATURE REVIEW	9
3. HYDROLOGICAL MODEL BUILDING	13
4. TRIALLING OF HYDROLOGICAL MODELS FOR FLUVIAL FLOOD FORECASTING 20	
5. TRIALLING OF INTEGRATOR SYSTEMS FOR FLUVIAL FLOOD FORECASTING	23
6. SELECTION OF THE HYDROLOGICAL MODEL AND INTEGRATOR SYSTEM	24
6.1 HYDROLOGICAL MODEL.....	24
6.2 INTEGRATOR SYSTEM.....	25
7. RECOMMENDATIONS	27

List of Tables

TABLE 2-1: REQUIREMENTS FOR THE FUTURE HYDROLOGICAL MODEL	9
TABLE 2-2: PROPOSAL OF METEOROLOGICAL MODEL INPUT	11
TABLE 3-1: COMPARISON OF PERFORMANCE METRICS FOR NORE	15
TABLE 3-2: COMPARISON OF PERFORMANCE METRICS FOR SLANEY	15
TABLE 3-3: COMPARISON OF PERFORMANCE METRICS FOR BARROW	15
TABLE 3-4: COMPARISON OF PERFORMANCE METRICS FOR MOY AND KILLALA.....	16
TABLE 3-5: COMPARISON OF PERFORMANCE METRICS FOR SHANNON.....	16
TABLE 4-1: OVERVIEW OF THE METRICS FOR EACH CATCHMENT AND MODEL. COLOUR CODES ARE SET UP FOR EACH METRIC SEPARATELY (GREEN = BETTER, RED = WORSE).....	21

List of Figures

FIGURE 2-1: FINAL SCORES FOR THE DIFFERENT MODELS USING DIFFERENTIATED WEIGHTS FOR EACH REQUIREMENT	10
FIGURE 2-2: FINAL SCORES FOR THE DIFFERENT INTEGRATOR SYSTEMS USING EQUAL WEIGHTS FOR EACH REQUIREMENT	11
FIGURE 3-1: TYPICAL MODELLING RESULTS FOR STATION 14018 IN THE BARROW CATCHMENT, TOP FIGURE FOR 1 YEAR AND 1 MODEL (HYPE), AND BOTTOM FIGURE IS A DETAIL OF THE PEAK SIMULATION FOR THE 4 MODELS	14
FIGURE 3-2: COMPARISON OF NSE VALUES FOR ALL MODELLED GAUGED FORECASTING POINTS	18
FIGURE 3-3: COMPARISON OF R2 ON THE PEAKS FOR ALL MODELLED GAUGED FORECASTING POINTS	18
FIGURE 3-4: COMPARISON OF NSE VALUES FOR ALL MODELLED GAUGED FORECASTING POINTS, WITHOUT THE SHANNON CATCHMENT.....	19

FIGURE 4-1: TYPICAL RESULTS, BOTH SIMULATION RESULTS (RED LINE) BASED ON OBSERVATIONS AND FORECASTING RESULTS BASED ON HISTORICAL HARMONIE FORECASTS (GREY DASHED LINES) FOR THE SAME PERIOD AS WHICH WAS USED FOR FIGURE 3-1.....21

FIGURE 6-1: FINAL SCORES AND WEIGHTS FOR THE FOUR TESTED HYDROLOGICAL MODELS (MARKED IN RED ARE CHANGES COMPARED TO THE MCA OF THE LITERATURE REVIEW)25

FIGURE 6-2: FINAL SCORES AND WEIGHTS FOR THE 2 TESTED INTEGRATOR SYSTEMS (MARKED IN RED ARE CHANGES COMPARED TO THE MCA OF THE LITERATURE REVIEW)26

Abbreviation list

API	Application Programmer Interface
CAMP	Climate Automatisatation & Modernisation Project
DEM	Digital Elevation model
EFAS	European Flood Alert System
HRLDAS-HR	High resolution land-surface data assimilation – Hourly
KGE	Modified Kling-Gupta efficiency criteria
MAE	Mean absolute error
MCA	Multi Criteria Analysis
MO	Mike Operations
NFFWS	National Flood Forecasting and Warning Service
NSE	Nash–Sutcliffe model efficiency coefficient
OPW	Office of Public Works
PET	Potential evapotranspiration
POT	Peak over threshold
PWRMSE	Peak Weighted Root Mean Square Error
RFT	Request for tendering
RMAE	Relative mean absolute error

1. INTRODUCTION

Met Éireann, in collaboration with the Office of Public Works (OPW), is currently engaged in the establishment of a National Flood Forecasting and Warning Service (NFFWS) to forecast for fluvial and coastal floods, following a Government decision in January 2016. To inform and support this task, a range of existing available hydrological models and integrator systems are to be reviewed, developed and trialled.

The study contains the following Stages:

- Stage 1: Comprehensive literature review of existing hydrological models, existing integrator systems and model data input options;
- Stage 2: Model development for five representative catchments: Shannon, Barrow, Nore, Slaney and Moy & Killala;
- Stage 3: Trialling of hydrological models for fluvial flood forecasting. Models are tested on five representative catchments;
- Stage 4: Trialling of integrator systems for fluvial flood forecasting, including a 2-month pre-operational test and an ICT assessment.
- Stage 5: Final reporting. The final report presents in detail all of the findings and outcomes of Stages 3, 4 and 5.

Based on the literature review of Stage 1, a set of hydrological models and integrator systems are selected for further development. These models are then trialled on the five representative catchments. Each model is first calibrated and validated, and then tested both with historic data (Stage 2) and with real-time forecasted data (Stage 3) from the Harmonie model. The selected integrator systems were configured to work for one catchment and the selected hydrological models were tested for eight weeks with real time data (Stage 4).

2. COMPREHENSIVE LITERATURE REVIEW

In Stage 1 of the project an extensive literature review was made with the aim to assess the data input sources used for the various possible hydrological models and integrator systems. The scope of the forecasting system was described in the RFT, and further discussed during the kick-off meeting. During this meeting, OPW also clarified the intention to build the future NFFWS as a multi-level system, with four modelling levels ranging from national to local scale. The four levels of flood forecasting proposed are as follows:

- Level 1 – National scale forecasts;
- Level 2 – Catchment (hydrometric area) scale forecast;
- Level 3 – Sub-catchment scale forecasts;
- Level 4 – Individual street / property scale forecast.

Each of these levels of forecasting has different model requirements. This literature review focuses mainly on models that are most suitable for Level 1 and Level 2 forecasting systems.

To evaluate the hydrological models and integrator systems, it was proposed to use a Multi Criteria Analysis, in which scores are given for the different requirements, after which a weighting method is used to calculate an overall score per hydrological model/integrator system.

For the hydrological models, a set of ten requirements (Table 2-1) was defined regarding model structure, outputs, inputs, calibration, accuracy, maturity, computational speed, easy setup, code availability and coupling.

Table 2-1: Requirements for the future hydrological model

Requirement	Explanation
REQ 1 - model structure	The model structure is suitable for Irish catchment types, allowing integration of typical hydrological processes: reservoirs, lakes, control structures, drainage.
REQ 2 - output	The model can calculate flows at all modelling points. If a gridded structure is used, the resolution is preferably 2x2km or higher.
REQ 3 - input	The model can handle the available input data, as well as meteorological (deterministic and probabilistic), static input data such as land use, soil types and flow data for data assimilation.
REQ 4 - calibration	The model can be calibrated with suitable tools, or does not require much calibration.
REQ 5 - accuracy	The selected model improves on the EFAS-accuracy or reaches the required target accuracy.
REQ 6 - maturity	The model has a proven stability and maturity and has been used for similar applications in Ireland and internationally.

Requirement	Explanation
REQ 7 - computational speed	The model runs within an acceptable time span, twice daily at national level (level 1), or four times daily at catchment/sub-catchment level (level 2).
REQ 8 - easy set-up	The model can be easily set up by a trained user, is documented well, and does not require too much user experience.
REQ 9 - code availability	The code is readily available for the intended usage.
REQ 10 - coupling:	The model can be easily coupled with other components.

For a longlist of fourteen models (HYPE, WFlow, WRF-HYDRO, URBS, HEC-HMS, PDM, NAM, SMART, JULES, TOPLATS, mHM, iFramework, LISFLOOD, G2G), a qualitative evaluation was done based on available literature (e.g. manuals, webpages). This document was sent to the model developers for review.

The scores and weights were discussed and agreed upon with the steering committee of the project. It was agreed to exclude seven models (SMART, JULES, TOPLATS, mHM, iFramework, LISFLOOD, G2G), in the further analysis of the MCA, as they scored insufficiently for some of the requirements, and were therefore deemed unsuitable for Levels 1 and 2 of the NFFWS. Figure 2-1 shows the result of the MCA-analysis. Based on the MCA-results and on further discussion with the steering group, four hydrological models were selected for further testing in Stages 2, 3 and 4 of the project: WRF-HYDRO, WFlow, HYPE, URBS. These models cover a broad spectrum of possibilities from distributed and physically-based models to purely conceptual models.

	Requirements										Total score
	Model structure	Input	Output	Calibration	Accuracy	Maturity	Computational speed	Easy set-up	Code availability	Coupling	
HYPE	10	10	10	9	10	10	10	9	10	9	89.0
wFlow	10	10	10	9	10	10	9	9	10	10	89.0
WRF-HYDRO	9	9	10	7	10	10	8	8	10	10	66.3
URBS	8	10	10	6	10	10	10	10	8	10	66.0
HEC-HMS	8	8	8	7	10	10	10	10	9	10	64.7
PDM	7	10	10	6	10	10	10	10	8	10	61.0
NAM	7	8	8	6	10	10	10	10	8	10	52.7
Min score	4	7	8	6	9	8	8	6	6	8	
Max score	10	10	10	10	10	10	10	10	10	10	
Weight	30	5	5	20	0	8	8	8	8	8	

Figure 2-1: Final scores for the different models using differentiated weights for each requirement

The same approach was used for evaluating the integrator systems, leading to a longlist of six integrator systems (Delft FEWS, Mike Operations, EFAS, ICM Live, HEC-RTS and Kalypso), which were evaluated for 4 main requirements: modularity, inputs/outputs, efficiency and maturity.

The integrator system EFAS was on the longlist, but is excluded from further analysis because the system is considered too fragmented and insufficiently generic to serve as a candidate for this study.

The MCA analysis leads to the scores as presented in Figure 2-2. This clearly shows that two of the integrator systems (Delft FEWS and Mike Operations) are much more generic than the other ones. Therefore, it was decided to only include Delft FEWS and Mike Operations in the trials of Stage 4.

	modularity	input/output	efficiency	maturity	Total score
Delft-FEWS	5	5	5	5	100.0
MIKE OPERATIONS	4	5	4	4	81.3
Kalypso	3	3	3	2	43.8
ICMLive	1	3	4	3	43.8
HEC-RTS	2	3	2	3	37.5
Min score	1	1	1	1	
Max score	5	5	5	5	
Weight	25	25	25	25	

Figure 2-2: Final scores for the different integrator systems using equal weights for each requirement

With respect to data inputs, the main focus was given to the meteorological input data as they are the main driver for the hydrological models. A mix of observed data (gridded daily rainfall, weather station data) and forecasted data (mainly HARMONIE) is proposed for further testing in the project, which is summarized in Table 2-2.

Table 2-2: Proposal of meteorological model input

Type of the run	Period	Data product for rainfall and ET0
Stage 2: Model calibration	1981 - recent	Rainfall hourly profiles: synoptic (+ weather) stations, Rainfall volumes: daily grids ET0: daily grids
Stage 2: Model validation	1981 - recent	Rainfall hourly profiles: synoptic (+ weather) stations, Rainfall volumes: daily grids ET0: daily grids
	1981 - recent	Rainfall + ET0: MÉRA (if needed)
Stage 3: Trialling of models	start of simulation - t0	Same as Operational state runs

Type of the run	Period	Data product for rainfall and ET0
	t0 - end of simulation	After 2012: HARMONIE (+ MÉRA) Before 2012: MÉRA (if relevant)
Stage 4: Operational forecast	start of simulation - t0	Rainfall radar 15 min or hourly values ET0: near real-time synoptic stations
	t0 - end of simulation	Rainfall + ET0: Harmonie + ECMWF
Stage 4: Operational state run	all state runs	Rainfall hourly profiles: synoptic (+ weather) stations, Rainfall volumes: daily grids (70-90 stations are available daily) ET0: daily grids

3. HYDROLOGICAL MODEL BUILDING

In Stage 2 of the project, hydrological models were built for the four selected model softwares for each of the five catchment areas (Nore, Slaney, Barrow, Moy & Killala, Shannon). This included the delineation of sub-catchments, preprocessing of all meteorological data, preprocessing of flow data and preparation of all required input files. An assessment of the building process of each model can be found below:

- For HYPE, the main difficulty was to find an efficient method to create the model through the existing preprocessing tools, and to get acquainted with the different files and formats. Simulation times were short and the number of parameters to be calibrated were quite limited, which made calibration efficient. Lakes were also easily implemented.
- The URBS manual lacks a clear structure and consistency, forcing the user to advance through trial and error. Nonetheless, the very simple structure and limited amount of parameters made it relatively simple to calibrate, but the parameters are hard to extrapolate to ungauged basins. As URBS (almost) does not include a concept of varying baseflow, parametrizing and simulating lake behaviour properly is difficult.
- In WFlow many modelled processes are spatially distributed, though the basic model structure and equations are kept quite simple. After consultation with Deltares it was chosen to work with the SBM model, which only has a relatively limited number of parameters to be calibrated. The code of the model is not mature yet and some important bugs were encountered; because of that the calibration process was not very efficient.
- Discovering and setting up a WRF HYDRO model was a real challenge, because there is a very steep learning curve to go through and requires an IT skill set rather than knowledge on hydrology. As some documentation is missing and stored in different files, it is a difficult model to understand. As it is spatially distributed and some processes are modelled in a physically based way, simulation times are longer, making it more difficult to calibrate. Within the timeframe of the project it was impossible to create a working routing grid (resolution 250m) for the Shannon catchment. Due to these obstacles, no simulations were undertaken for the Shannon WRF HYDRO model.

After the models were built, they were calibrated and validated using seven calibration metrics. The following metrics were taken into account: Nash–Sutcliffe model efficiency coefficient (NSE), Modified Kling-Gupta efficiency criteria (KGE), Mean absolute error (MAE), Relative mean absolute error (RMAE), Coefficient of determination of the maximum values of each storm event (R^2 peak), Coefficient of determination of the volume of each storm event above 40% of average yearly maximum (R^2 volume), Median peak timing as the median difference between observed and simulated peak timing.

As an example some modelling results can be found in Figure 3-1. The upper frame of the figure shows the results of a year of simulation results and observations for one of the four model software packages. Overall it shows a quite a good match. The lower frame of the figure shows the peak period simulation results for the 4 hydrological

modelling software packages. In this figure larger variations are evident, but overall there is quite a good match.

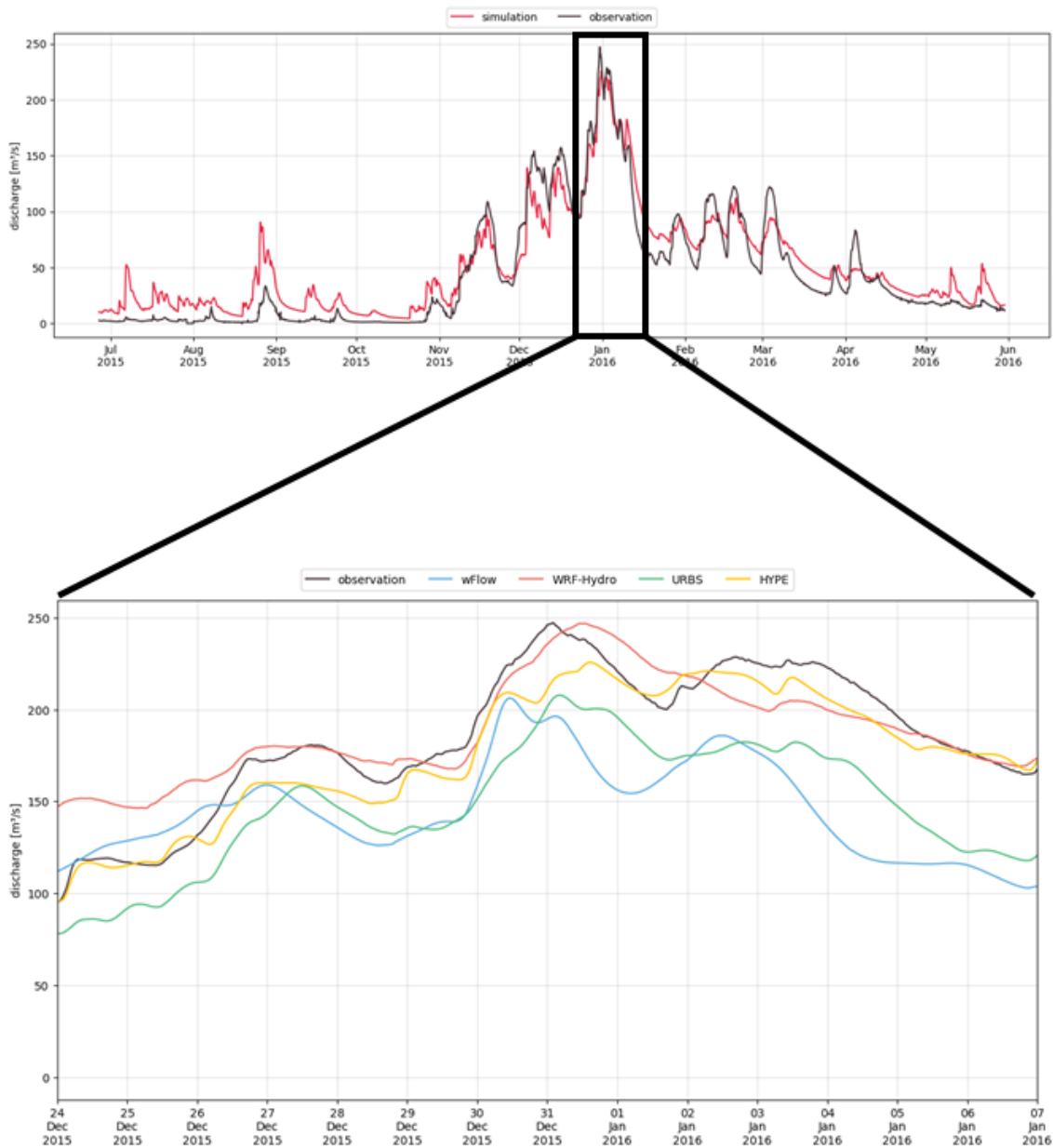


Figure 3-1: Typical modelling results for station 14018 in the Barrow catchment, top figure for 1 year and 1 model (HYPE), and bottom figure is a detail of the peak simulation for the 4 models

Table 3-1 to Table 3-5 give an overview of the calibration/validation results per catchment. The tables also indicate the metric used for each gauge and which of the four models performed best. From these tables, it is clear that HYPE is outperforming the other models.

Table 3-1: Comparison of performance metrics for Nore

Gauge id/metric	KGE	NSE	MAE	RMAE	R2 volume	R2 peak	Peak timing
15001	HYPE	URBS	URBS	URBS	HYPE	WFlow	URBS
15002	HYPE	HYPE	HYPE	HYPE	WFlow	WFlow	HYPE
15004	HYPE	URBS	URBS	URBS	URBS	URBS	URBS
15005	HYPE	HYPE	HYPE	HYPE	HYPE	WFlow	WFlow
15006	HYPE	URBS	URBS	URBS	URBS	WFlow	HYPE

Table 3-2: Comparison of performance metrics for Slaney

Gauge id/metric	KGE	NSE	MAE	RMAE	R2 volume	R2 peak	Peak timing
12001	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE	WFlow
12013	URBS	URBS	HYPE	HYPE	HYPE	URBS	HYPE
12016	HYPE	HYPE	HYPE	HYPE	HYPE	WRF-HYDRO	WRF-HYDRO
12037	WRF-HYDRO	WRF-HYDRO	WRF-HYDRO	WRF-HYDRO	WFlow	HYPE	WRF-HYDRO

Table 3-3: Comparison of performance metrics for Barrow

Gauge id/metric	KGE	NSE	MAE	RMAE	R2 volume	R2 peak	Peak timing
14004	HYPE	HYPE	HYPE	HYPE	URBS	URBS	HYPE
14005	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE
14006	HYPE	URBS	URBS	URBS	URBS	HYPE	HYPE
14007	HYPE	HYPE	HYPE	HYPE	HYPE	URBS	WRF-HYDRO
14009	URBS	URBS	URBS	URBS	HYPE	URBS	HYPE
14011	HYPE	URBS	HYPE	HYPE	URBS	WFlow	WFlow
14013	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE
14018	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE
14022	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE
14019	HYPE	HYPE	HYPE	HYPE	URBS	HYPE	URBS
14029	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE	URBS

Table 3-4: Comparison of performance metrics for Moy and Killala

Gauge id/metric	KEG	NSE	MAE	RMAE	R2 volume	R2 peak	Peak timing
34001	HYPE	HYPE	HYPE	HYPE	WFlow	HYPE	HYPE
34003	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE
34005	URBS	URBS	HYPE	URBS	URBS	HYPE	URBS
34007	WRF-HYDRO	WRF-HYDRO	URBS	URBS	HYPE	WRF-HYDRO	HYPE
34009	URBS	URBS	URBS	URBS	WFlow	HYPE	URBS
34010	WRF-HYDRO	WRF-HYDRO	WRF-HYDRO	WRF-HYDRO	WRF-HYDRO	WRF-HYDRO	HYPE
34011	URBS	URBS	URBS	URBS	URBS	HYPE	WFlow
34018	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE
34024	URBS	URBS	HYPE	URBS	URBS	WRF-HYDRO	URBS

Table 3-5: Comparison of performance metrics for Shannon

Gauge id/metric	KEG	NSE	MAE	RMAE	R2 volume	R2 peak	Peak timing
25001	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE
25002	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE
25003	URBS	HYPE	URBS	URBS	URBS	HYPE	URBS
25005	WFlow	WFlow	WFlow	WFlow	WFlow	HYPE	HYPE
25006	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE
25011	HYPE	HYPE	HYPE	HYPE	URBS	URBS	WFlow
25014	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE
25017	WFlow	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE
25020	HYPE	URBS	HYPE	HYPE	HYPE	HYPE	HYPE
25021	URBS	URBS	URBS	URBS	URBS	URBS	URBS
25022	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE
25023	HYPE	WFlow	HYPE	HYPE	HYPE	WFlow	HYPE
25025	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE	URBS
25027	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE
25029	URBS	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE
25030	HYPE	URBS	URBS	URBS	URBS	HYPE	URBS
25034	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE
25040	URBS	URBS	HYPE	HYPE	HYPE	URBS	HYPE
25044	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE
25124	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE
26002	HYPE	HYPE	HYPE	HYPE	WFlow	HYPE	HYPE
26005	HYPE	HYPE	HYPE	HYPE	WFlow	HYPE	HYPE
26006	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE
26007	HYPE	HYPE	HYPE	HYPE	WFlow	HYPE	WFlow

Gauge id/metric	KGE	NSE	MAE	RMAE	R2 volume	R2 peak	Peak timing
26008	HYPE	HYPE	URBS	URBS	WFlow	HYPE	URBS
26009	HYPE	URBS	URBS	HYPE	HYPE	URBS	HYPE
26010	URBS	URBS	URBS	URBS	HYPE	HYPE	WFlow
26012	HYPE	HYPE	HYPE	HYPE	URBS	HYPE	HYPE
26018	HYPE	HYPE	HYPE	HYPE	URBS	HYPE	URBS
26019	HYPE	HYPE	URBS	URBS	HYPE	HYPE	URBS
26021	HYPE	HYPE	HYPE	HYPE	WFlow	HYPE	HYPE
26022	HYPE	HYPE	WFlow	WFlow	HYPE	HYPE	WFlow
26108	HYPE	HYPE	HYPE	HYPE	HYPE	HYPE	URBS

Figure 3-2 and Figure 3-3 show the empirical cumulative density function of the performance metrics (NSE and R² peak) integrated over all modelled gauged forecasting points. The dotted lines in these figures show the empirical cumulative density function when only the validation events are considered. From the figures, we can see that an NSE value of 0.6 is exceeded in 75% of the catchments for HYPE, in 50% of the catchments for URBS and in about 20% of the catchments for WFlow and WRF-Hydro.

HYPE also scores the best at the R² peak CDF curve, followed by URBS and WFlow. It is however interesting to see that WFlow performs better at the R² peak score compared to the NSE score.

The difference between URBS and HYPE as well as the difference between the calibration and validation results is mainly due to the calibration results in the Shannon catchment, as can be seen by comparing Figure 3-2 (with the Shannon included) and Figure 3-4 (without the Shannon). For the four other catchments, there are no significant differences in the performance between both models. Given the size of the Shannon catchment, it was opted to calibrate URBS for only a subset of the total storm events. Based on the current results, we can conclude that this subset was not representative of the entire catchment and that some overfitting was introduced by limiting the number of storm events.

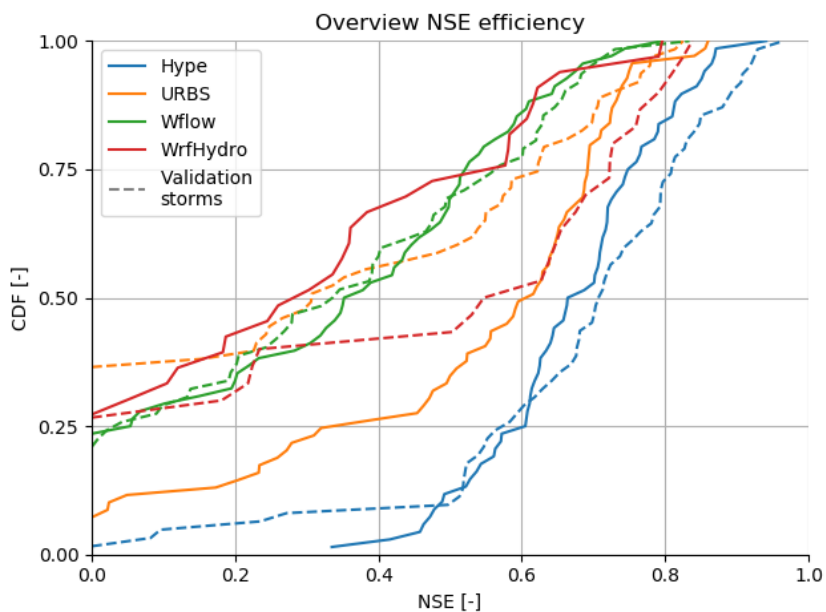


Figure 3-2: Comparison of NSE values for all modelled gauged forecasting points

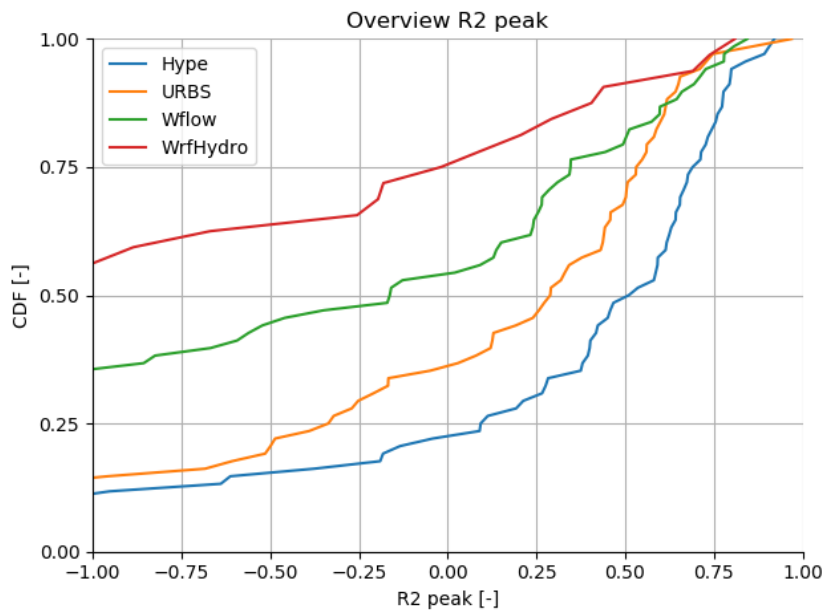


Figure 3-3: Comparison of R2 on the peaks for all modelled gauged forecasting points

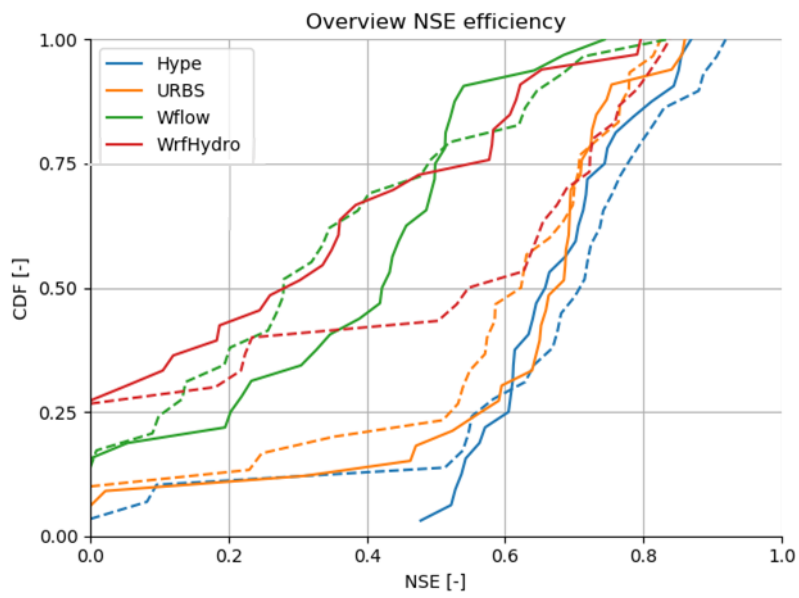


Figure 3-4: Comparison of NSE values for all modelled gauged forecasting points, without the Shannon catchment

4. TRIALLING OF HYDROLOGICAL MODELS FOR FLUVIAL FLOOD FORECASTING

In Stage 3, the calibrated and validated models were tested using historical forecasted data, mainly HARMONIE. For 10 historical storm periods between 2013 and 2018, simulations were done with lead time of up to 54 hours and at 6-hour intervals. The results of these simulations were compared with the observed discharges using 4 different metrics: Relative Mean Absolute Error, Nash Sutcliffe Efficiency, Peak Weighted Root Mean Square Error and the peak timing.

Table 4-1 gives an overview of the results of Stage 3 with respect to the different metrics.

Overall the required target accuracy (*“the actual peak flow needs to be predicted to within +/- 10% and 6 hours of the actual peak at the gauged forecast points, and with a lead time of greater than 24 hours”*) is not met for any of the model software packages. However, for some model/station/storm combinations the target accuracy is met, typically for the stations which also gave reasonably good results in Stage 2.

WFlow and WRF-HYDRO give results which are in line with their performance in Stage 2 and which are significantly worse than the results of HYPE and URBS.

Though overall still close together, URBS now performs better than HYPE for some of the metrics (PWRMSE, NSE and timing), mainly due to small differences in methodologies between Stages 2 and 3: less storms (10) but more recent events (from 2013 onwards) were used in Stage 3. Fewer metrics were used in Stage 3 but in Stage 2 no aggregation of metrics was done across all of the considered gauging stations.

It is also not a coincidence that URBS performs better than HYPE on exactly the PWRMSE and NSE, because these metrics take into account the complete shape of the hydrograph rather than just being scored based on the peak value alone. URBS has a specific parameter with which the shape of the hydrograph can be controlled, while in HYPE this has to be achieved with a combination of parameters.

Figure 4-1 gives an example of typical results from the model trialling stage. The dashed grey lines each represent individual historical forecast simulations based on consecutive HARMONIE forecasts. As can be seen from these figures, the closer to the peak the HARMONIE forecasts start, the better the alignment between the simulation and observed measurements.

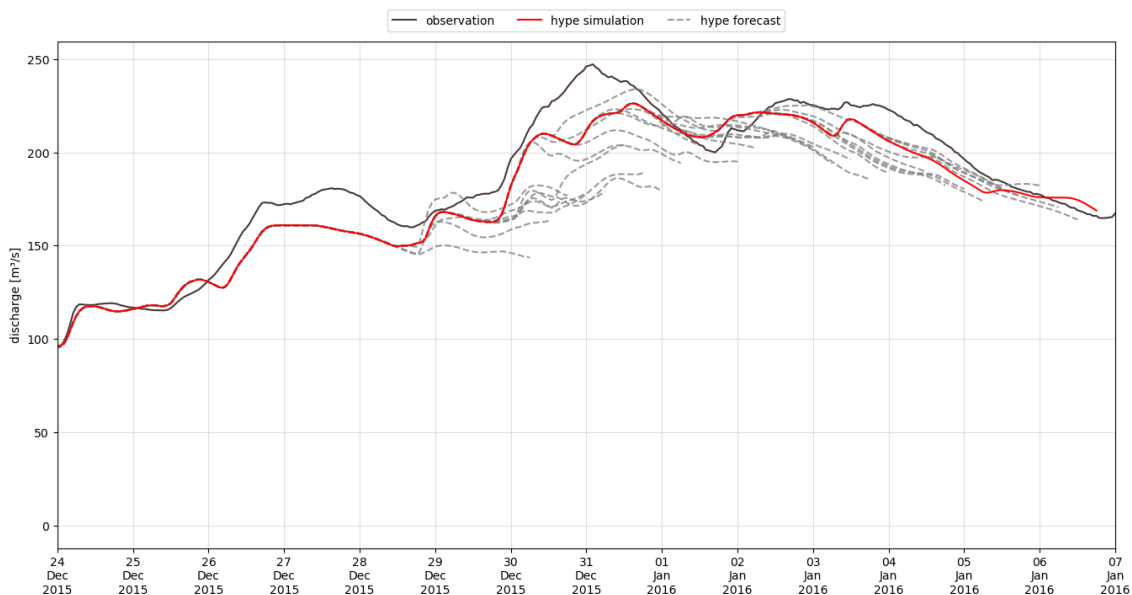


Figure 4-1: Typical results, both simulation results (red line) based on observations and forecasting results based on historical HARMONIE forecasts (grey dashed lines) for the same period as which was used for Figure 3-1

Table 4-1: Overview of the metrics for each catchment and model. Colour codes are set up for each metric separately (green = better, red = worse).

Metric	Catchment	HYPE	URBS	WFlow	WrfHydro
RMAE	Barrow	0.13	0.18	0.22	0.49
	Moy & Killala	0.19	0.26	0.37	0.49
	Nore	0.3	0.3	0.35	0.49
	Shannon	0.2	0.2	0.4	-
	Slaney	0.51	0.58	0.7	0.71
Timing (h)	Barrow	13	9.5	13	9
	Moy & Killala	11	10	12	19
	Nore	11	12	12.5	28
	Shannon	15	13	11	-
	Slaney	11.5	17.5	9	10
NSE	Barrow	0.32	0.17	-11.18	-0.25
	Moy & Killala	0.38	0.49	-3.5	-0.94
	Nore	-0.13	0.03	-2.17	-0.8
	Shannon	-0.05	0.29	-6.91	-
	Slaney	0.02	-0.03	-0.66	-0.11
PW RM SE	Barrow	9.11	9.1	11.22	11.89

Metric	Catchment	HYPE	URBS	WFlow	WrfHydro
	Moy & Killala	12,07	10,48	17,98	19.29
	Nore	29.69	27.87	37.87	33.27
	Shannon	7.72	6.57	14,6	-
	Slaney	23.91	23.68	32.97	26.24

5. TRIALLING OF INTEGRATOR SYSTEMS FOR FLUVIAL FLOOD FORECASTING

In Stage 4 of the project, the two selected integrator systems were setup and configured for the Barrow catchment for the four different models, based on manuals, expert knowledge and for Mike Operations some limited support by DHI. After this configuration, both integrator systems were trialled for eight weeks with real-time data.

Though there were big differences in the development of the configurations, generally both integrator systems can be configured (or programmed) with the basic functionality needed to perform flood forecasts, such as for example the preprocessing of the real-time incoming data and forecasts, the scheduling and execution of simulations and the postprocessing of the forecast results

Overall the configuration of Delft FEWS was much easier than Mike Operations. The main reasons for this were:

- Lack of a proper manual in Mike Operations versus fully-documented manuals in Delft FEWS;
- 1 frequently occurring licensing issue, slowing down the configuration process in Mike Operations;
- Mike Operations requires much more dedicated programming (Python or Iron Python), while FEWS has many out of the box configuration options all accessible through xml.

The results of the 8 week testing showed that both integrator systems can perform the tasks they are supposed to do but that significant differences in efficiency occur. The time series preprocessing within Delft FEWS is much more efficient (i.e. uses less time) than the preprocessing within Mike Operations.

Data not arriving in the integrator system in time or at all resulted in crashes of the models in the integrator system. In a production setup of the integrator systems, additional backups for these data sources need to be foreseen to overcome this problem.

Based on these setups and tests an ICT assessment was made. The used hardware setups were evaluated and improvements were suggested. Additionally, some general IT considerations were discussed in the report, allowing the client to make the best choices. Based on the simulation times during the 8-week trial period an estimate was made for the extension to 30 catchments for all of Ireland. For most of the models the impact of this on the required calculation capacity is rather limited.

6. SELECTION OF THE HYDROLOGICAL MODEL AND INTEGRATOR SYSTEM

6.1 HYDROLOGICAL MODEL

Based on the outcomes of Stages 2 and 3, an update is proposed to the scores and weights used in the MCA for the hydrological systems compared to the scores calculated during the literature review.

- Requirement Calibration:
 - Additional points for URBS since it proved to be fast to calibrate;
 - Penalty points for WFlow since the calibration process was less straightforward than anticipated;
 - The score for WRF-Hydro remains the same since the encountered difficulties were foreseen and, once setup, the model calibration is not significantly more difficult than WFlow;
- Requirement Accuracy:
 - Penalty points for WFlow and WRF-Hydro based on the current results although we acknowledge that additional gain in the results is most likely possible with more time investment (section 3.2);
- Requirement Maturity:
 - Penalty points for WFlow as a significant bug was discovered in the model code during the project which caused us to restart the calibration process. Manuals were also available but for some items these were not up to date.
- Requirement Easy setup:
 - Penalty points for URBS as it does require some processing and scripting to setup a model and is certainly not easier than setting up a HYPE model
 - Bonus points for WFlow as it was straightforward and fast to setup the model if the modelbuilder can be used;
 - The score for WRF-Hydro remains the same despite the difficulties in setting up the model. Those difficulties were anticipated prior to the start of this Stage and hence WRF-Hydro already has the lowest score of all models.

As accuracy was something difficult to assess purely based on literature, no weight was given to it for the MCA of the literature review. Based on the results of Stage 2 and 3 it is now possible to do this. It is proposed to equally divide the weight attributed to the Calibration criterion between Accuracy and Calibration. The min/max scores are kept identical to those used for the literature review for consistency reasons.

Figure 6-1 shows the updated result of the Multi Criteria Analysis, in which HYPE is still ranked highest however, due to it's performance in Stage 3 URBS is now scoring higher than WFlow. WRF-HYDRO is still in fourth place.

Based on the results of our testing and this updated MCA, HYPE is proposed to be the selected hydrological model for the Fluvial Flood Forecasting System on a national or catchment scale level.

Compared to the other models, HYPE has the following advantages:

- It provides a good balance between physically based concepts and faster calculating conceptual aspects, which makes it fast to calculate for many catchments on a national level, and also shows the most potential to perform well for ungauged catchments;
- It outperforms the other selected models on the more downstream gauges;
- It outperforms the other selected models on gauges downstream of lakes, which are present quite frequently in Ireland;
- Though not tested explicitly, it has some data assimilation options already included, which are expected to further improve the results.
- Due to the open source character (compared to URBS), adaptations and extensions are feasible.

	Model structure		Calibration	Accuracy	Maturity	Computational speed	Easy set-up	Code availability	Coupling	Total score
	Input	Output								
HYPE	10	10	9	10	10	10	9	10	9	91.5
URBS	8	10	9	10	10	10	9	8	10	81.5
wFlow	10	10	7	9	9	9	10	10	10	77.5
WRF-HYDRO	9	9	7	8	10	8	8	10	10	63.8
Min score	4	7	6	9	8	8	6	6	8	
Max score	10	10	10	10	10	10	10	10	10	
Weight	30	5	10	10	8	8	8	8	8	

Figure 6-1: Final scores and weights for the four tested hydrological models (marked in red are changes compared to the MCA of the literature review)

6.2 INTEGRATOR SYSTEM

Delft FEWS was the highest scoring integrator system during the literature review. Though both integrator systems in essence managed to do all of the basic tasks successfully during the eight-week trial, significant differences in efficiency and user friendliness were noticed. Mike Operations proved to be less efficient than expected, particularly on input/output (because of the lack of proper manuals), based on the literature review.

This is reflected in Figure 6-2, in which the scores for input/output and efficiency are lowered for Mike Operations. Based on the results of the testing during this project, Delft FEWS is proposed as the selected integrator system for use in Ireland on the national or catchment scale level.

Compared to Mike Operations, Delft FEWS has the following main advantages:

- It is currently much more efficient for processing of data from all kinds of (non-DHI) models, and therefore is much more flexible to use;
- It is much more mature, with a much larger user community;
- It includes much more out of the box functionalities, such as a big data archive.

	modularity	input/output	efficiency	maturity	Total score
Delft-FEWS	5	5	5	5	100
MIKE OPERATIONS	4	4	3	4	62.5
Score min	1	1	1	1	
Score max	5	5	5	5	
Weight	25	25	25	25	

Figure 6-2: Final scores and weights for the 2 tested integrator systems (marked in red are changes compared to the MCA of the literature review)

7. RECOMMENDATIONS

Based on the experiences gathered during the different Stages and the results of the trialing, recommendations regarding the data, models and integrator systems are listed below:

1. Density of the rain gauge network: in Stage 2, observed daily gridded rainfall and hourly point data from the synoptic stations were combined into a (sub)catchment average rainfall. For each grid cell, data of the nearest available synoptic station were used to give a sub-daily temporal profile to the daily gridded amounts. This in practice meant that only between 2 (Nore) and 3 (other catchments) synoptic stations were being used for this, which is far too limited in our experience to properly model and calibrate the timing of the peaks well. If the newly installed, and yet to be installed, 60-80 sub-daily CAMP (Climate Automatisatation & Modernisation Project) stations are also included in the calculations, between 5 and 10 stations would be available per catchment, which seems more appropriate.
2. Density of flow gauging network: based on the calibration results, some recommendations can be made for future additional flow gauging stations.
 - a. **Nore**: Though in theory quite a lot of gauged data were available in this catchment, some of them, especially on the tributaries, were judged to be unreliable. Therefore, it is recommended to make sure more reliable data are available for all of the existing gauges, rather than installing additional gauges.
 - b. **Slaney**: For the Slaney catchment only 4 gauging stations were available. It is recommended to have more gauging stations on the faster reacting tributaries to the main river, for example on the downstream parts of the Derreen, Derry, Bann and Urin tributaries.
 - c. **Barrow**: The number of water level/discharge gauging stations in the catchment were spread relatively well between upstream and downstream locations, and between small and large catchments. No obvious need for improvements could be detected compared to the other catchments.
 - d. **Moy & Killala**: Though results are varying greatly between the different models, it is difficult to give real recommendations with respect to adding more reliable gauging stations. Some upstream tributaries like the Trimoge and the Glore are ungauged and could benefit from having a flow gauges installed.
 - e. **Shannon**: Most of the gauges are situated on upstream tributaries. The main downstream river with some large regulated lakes on it, could use more gauges, for example downstream of Loughs; Allen, Ree and Derg.
3. Measurements of lake levels: The HYPE model can output water levels on the simulated lakes. Observed data on lake levels could immediately lead to significant improvements in model calibration when lakes are present. These can then also be used to test the effectiveness of including more detailed operating rules within the hydrological models.
4. Quality control of the gauged flow data was not part of the project. Because of that, it is quite difficult to assess the quality of the flow data and the rating curves. For this

project, only rating curves, which were deemed to be very reliable, were used. For future model calibrations, it is recommended to undertake a more detailed review before using them for model calibration.

5. Timeliness of data flow between systems: The observed flow data were also configured in the integrator systems, so they could be visualized in the user interfaces. For the Barrow catchment, however, many of the available gauges were not available during the 8-week trial and also the delay in arrival (more than 1 day) in the integrator system was significant. If in the future data assimilation is used for correcting the forecasted flows, it is very important that the observed flow data arrive in the integrator system in a timely and reliable way. If this cannot be achieved for flow data but for water level data, it could be an option to use the water levels and configure rating curves in the integrator system to convert to flows.
6. Expansion of setup to all of Ireland: The integrator systems were tested for only one catchment. To expand this to all of Ireland (assumed to be 30 catchments), it is recommended to model and configure more separate (up to 30) catchments rather than grouping everything into one very large model. The integrator system can group all of these catchments to behave as one big model to run at the same time, which allows for much more flexibility in run scheduling.
7. Data assimilation: Data assimilation to improve the modelled and forecasted flows was not used in this project, because the use of data assimilation could have resulted in some confusion over which hydrological models were considered to be more suitable (with and without data assimilation). However, most operational flood forecasting systems in Europe use one or more forms of data assimilation, from simple error corrections (e.g. ARMA) on the forecasted flows to more complicated state corrections on the modelled soil moisture content. Data assimilation should be considered in future model improvements.
8. Incoming rainfall/PET data: Regarding the incoming data of daily rainfall/PET in the integrator systems, it is recommended to try and speed up delivery, so they are available to integrator system with less delay.
9. Improvements to the hydrological models: One of the main purposes of this project was to select the most suitable hydrological modeling software for a national fluvial flood forecasting system and not necessarily to calibrate all these models to their maximum potential. Below some suggestions are formulated for improvements for the different models, per model listed in decreasing priority:
 - a. **HYPE**: Use more/better input hydrometeorological data, use of water levels for calibration of lakes, testing of different soil/land use combinations, more detailed routing methods.
 - b. **URBS**: Use more/better input hydrometeorological data, try to improve the modeling of lakes.
 - c. **WFlow**: Check and revise the used modeling approach, i.e. model structure, data calibration strategy, to be done in cooperation with Deltares. Use a more detailed spatially varying approach for the overland and river roughness to improve the shape and peaks of the modelled flows.

- d. **WRF-HYDRO:** Check and revise the used modeling approach, use more/better input hydrometeorological data, speed up calculations by using parallel computing abilities thus enabling a faster calibration.
10. Recalibration of the hydrological models: Recalibration of the hydrological models is recommended if:
- a significant revision of the rating curves is envisaged;
 - a significant event has occurred;
 - a significant bug is reported and solved in the hydrological model software;
 - in case sufficiently long (5-10 years) new hydrometric data is newly available;
11. Alternative data inputs: To overcome issues with missing data, it is recommended to configure alternative options, which are automatically used in case a data source is missing (redundancy options).
12. Regular model verifications: In an operational flood forecasting system, forecasts are continuously being produced and more observed data are being gathered. To be able to improve the quality of these forecasts, it is therefore recommended to include a regular (monthly, quarterly) model verification into the operational procedures. As a consequence, all forecasts and probably also the data used for producing the forecasts need to be properly archived.
13. Use of ensemble forecasts: the results of Stage 3 of this project show that despite good results at some locations, there still is a significant uncertainty in the forecasts, which, amongst other things, is mostly caused by the uncertainty in the input rainfall forecasts. To address these uncertainties it is recommended to use ensemble meteorological forecasts as input into the models.