03 – Assessing the validity of scaling catchment parameters from lumped to semi-distributed models

Thibault Hallouin[1], Michael Bruen[1], Eva Mockler[1], Michael Christie[2], Craig Bullock[3], Fiona Kelly[4], Mary Kelly-Quinn[5]

[1] *Dooge Centre for Water Resources Research, University College Dublin, Ireland*
[2] *Blue Island Consulting Limited, Ceredigion, Wales, United Kingdom*
[3] *School of Architecture, Planning and Environmental Policy, University College Dublin, Ireland*
[4] *Inland Fisheries Ireland, Dublin, Ireland*
[5] *School of Biology and Environmental Sciences, University College Dublin, Ireland*

Abstract

The problem of hydrological predictions in ungauged catchments has long been an issue for dynamic hydrological models. This is the case for conceptual rainfall-runoff models that normally require observed rainfall and discharges to parameterise the contribution of the different processes in the model. One approach is the regionalisation of catchment model parameters by developing a relationship between them and the catchment's physical characteristics (descriptors). However, this procedure still requires some gauged catchments with observations, which, in most cases, limits its development to lumped conceptual models for the upstream drainage area of existing hydrometric gauges. For large catchments, and particularly for water quality modelling, it is desirable to have recourse to semi-distributed models to account for the heterogeneity of the soil and the geology, and the spatial variability of the forcing data. Given that the regionalisation relates the parameters to physical descriptors at one scale, the question arises as to whether it is reasonable to expect this relationship to remain valid at different scales. This study looks at the application of the conceptual rainfall-runoff model (SMART) and its regionalisation equations, previously developed from lumped catchments, to a set of fifteen Irish catchments simulated at sub-basin scale (semi-distributed) over the period 2007-2016. The performance of the regionalised catchment model is assessed against observed discharges using the Nash-Sutcliffe efficiency, and the bias. The results show promising capacities in predicting the hydrograph, over a range of scale, e.g. for small catchments such as the Ilen River and large river basins such as the Nore River.

1. INTRODUCTION

In Ireland, the management of freshwater resources is important for water supply but also for a range of freshwater ecosystem services, including water purification and the provision of a habitat for aquatic life. The main quality issue for surface waters in Ireland is eutrophication due to excessive water-borne nutrient losses from land into rivers and lakes. The use of rainfall-runoff catchment models can help in understanding the different pathways responsible for the transport of these nutrients to waterbodies. Such models can then be used to investigate the impacts of human activities, human landscape alterations and mitigation measures in the

catchment on the repartition of water runoff between the different pathways and then on the transport of water-borne pollution.
For large catchments, the heterogeneity of the soil and geology means that different sub-regions will have different hydrological behaviour. For flood forecasting, the approximation of a lumped catchment responding to rainfall events is often sufficient to make satisfactory predictions. However, when water quality is considered, the sources and the flow pathways become crucial to identify the main routes pollution is taking and how it may be attenuated. Beyond the heterogeneity of the hydrological response, land management practices (land use) are not homogeneous either at the catchment scale. The location of the sources of contamination in the catchment will influence how much will be lost to surface and underground waterbodies. Therefore, the use of a finer spatial resolution should be considered. The limited number of hydrometric gauges in Ireland reduces the opportunities for extensive calibration of high resolution rainfall-runoff models. The necessity for hydrometric gauges can be driven mainly by water supply and flood forecasting needs, where flows are relatively large. This means that the application of hydrological model is easier in these regions. In Ireland, more than three quarters of the approximately 74,000 mapped river reaches are first or second order streams. This network of small streams delivers most of the water extracted for drinking purposes. However, the lack of gauges for these small streams limits their hydrological and water quality modelling (Nasr and Hynds, 2014). Therefore, alternatives to calibration are required. Moreover, scenario analyses affecting the landscape, e.g. land cover change, effectively change the hydrological system and its response to rainfall so that calibration with historic data is not suitable where major changes have occurred.

Regionalisation is the most widely used alternative to calibration for predictions in ungauged basins (PUB) and the last IAHS Decade was dedicated to this purpose. These regionalisation procedures infer the parameters of rainfall-runoff models without the need for discharge data for calibration. They rely on the assumption that parameters can be identified from physical characteristics of the catchment. In Ireland, such a regional parameter transfer method has been recently developed for the SMART rainfall-runoff model for use in ungauged Irish catchments (Mockler et al., 2014). This model has demonstrated promising capabilities in predicting the hydrograph, and the partitioning of flows between the different conceptual pathways of the SMART model. However, this regional model has been developed and used only for lumped catchments. Here, we investigate whether the model is suited for use at the river sub-basin scale in order to be used for land use and land cover scenarios analyses.

The aim of this study is to confirm the validity and the performance of the regionalised version of the SMART model on a 10-year period, and to determine the validity of these regression equations at the river sub-basin scale. We compare the model on a set of fifteen catchments in the Republic of Ireland and assess the simulations on two performance criteria.

2. METHODOLOGY

The objective of this study is to assess the performance of a regionalised conceptual rainfall-runoff model for different spatial discretisation schemes: lumped, semi-lumped, and semi-distributed. The rainfall-runoff model used is SMART (Soil Moisture Accounting for Routing and Transport). A regional parameter transfer method allows the SMART model to be used for predictions in ungauged basins. The resulting regionalised model is tested on a set of fifteen Irish catchments. The results at this semi-distributed scale are compared with results at the lumped and semi-lumped scales.

2.1 Rainfall-Runoff Model

The SMART model is a conceptual rainfall-runoff model purposely developed to finely represent the various flow pathways through the catchment to its outlet (Mockler et al., 2016), including a drain flow pathway important for Irish catchments. It combines a soil moisture accounting component dividing the soil horizon into six layers of equal depths and a routing component for the soil moisture outflow distinguishing five pathways including overland flow, drain flow, interflow, shallow groundwater flow, and deep groundwater flow. These five pathways are conceptualised as linear reservoirs and contribute to a final linear reservoir used for channel routing. SMART is characterised with ten parameters: six for the soil moisture accounting component, and four for the routing components.

In order to use SMART in ungauged catchments, a regional parameter transfer method has been developed (Mockler et al., 2014) in order to estimate the parameters of the model from a range of physical catchment descriptors that are described in the data section below. The regionalisation consisted of a multiple linear regression fitted with ordinary least squares that produces, for eight parameters, a relationship between a set of explanatory variables (physical catchment descriptors) and the model parameters. The two remaining parameters were inferred from GIS datasets without calibration. This regionalised version of SMART has been developed for lumped catchments from a dataset including 31 catchments in the Republic of Ireland using rainfall and flow data from the period 1990-2005.

2.2 Spatial Discretisation: Lumped, Semi-Lumped, Semi-Distributed

Since the parameters can be inferred from physical characteristics of the landscape at the lumped scale with satisfactory hydrological predictions (Mockler et al., 2014), it is interesting to investigate the quality of the predictions of the regionalised model at a semi-distributed scale, defined here as a spatial discretisation of the lumped catchment into a collection of lumped river sub-basins, each with its own unique model parameter set, and hydrologically connected through river channels.

An intermediate approach is the semi-lumped approach where the spatial resolution is the same as for the semi-distributed approach but the model parameters are identical across all the river sub-basins, only the forcing inputs, rainfall and potential evapotranspiration, are specific to each river sub-basin.

Here, these three approaches (see Figure 1) are compared using the regionalised SMART model at an hourly time step. These three scales represent three level of inclusion of both the rainfall heterogeneity and the landscape heterogeneity (driving the hydrological response of the catchment).

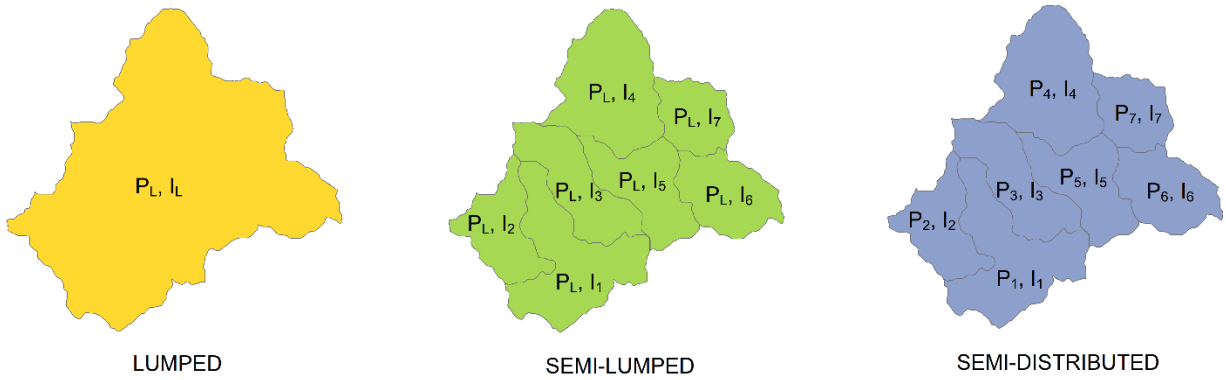

Figure 1: *Conceptual representation of the three spatial discretisation schemes used to assess the performance of the regionalised rainfall-runoff model SMART.* P_X *represents the model parameter sets, where X is a unique identifier for a given set.* I_Y *represents the input forcing datasets, where Y is a unique identifier for a given dataset.*

2.3 Predictions in Ungauged River Sub-Basins

In practice, the SMART model is used to predict streamflow at the outlet of each river sub-basin that is then routed through the channel network. Most of these river sub-basins are ungauged. However, the full catchments selected are all gauged. They are assumed to be ungauged for the simulations (i.e. no calibration is done). The measured discharge data at the outlet of each catchment are just used to verify the performance of the predictions made using the ungauged parameter set(s).

3. CASE STUDY

3.1. Selection of the Catchments

The investigations are carried out on fifteen catchments across the Republic of Ireland, whose surface areas range from 99 up to 2462 square kilometres (see Table 1). They have been selected on the following criteria: absence of large lakes in the catchment, few karst features, and the availability of discharge data for analysis of the performance. The location of the outlet of the catchment has been determined following two criteria: upstream of any tidal influence, and where discharge gauge is present and is in a suitable location (i.e. at or near the downstream end of a drainage area delineated by the Environmental Protection Agency) in order to limit the scaling factor of the discharge data.

Table 1: *Information regarding the catchment delineated areas and the corresponding hydrometric stations whose discharge data were used in this study.*

Catchment Name	Catchment Outlet (EPA WFD identifier)	Delineated Area (km²)	Sub-Basins (units)	Hydrometric Station Name	Gauged Area (km²)	Period
Boyne	IE_EA_07B042100	2461.9	103	Slane Castle	2408.0	2007-2016
Barrow	IE_SE_14B012920	2438.2	116	Royal Oak	2439.5	2007-2016
Nore	IE_SE_15N012200	2221.9	110	Mount Juliet	2225.7	2007-2016
Suir	IE_SE_16S021930	1586.3	83	Cahir Park	1582.7	2007-2016
Slaney	IE_SE_12S022200	1032.0	63	Scarawalsh	1030.8	2007-2016
Maigue	IE_SH_24M010700	770.6	33	Castle Roberts	770.6	2007-2016
Feale	IE_SH_23F010600	647.0	29	Listowel	646.8	2007-2016
Deel	IE_SH_24D021100	438.7	19	Rathkeale	438.8	2007-2016
Bandon	IE_SW_20B020800	421.5	16	Curranure	423.7	2007-2016
Dee	IE_NB_06D011000	376.2	16	Charleville + Coneyburrow	364.4	2007-2016
Avonmore	IE_EA_10A050300	230.4	12	Rathdrum	233.0	2011-2016
Ilen	IE_SW_20I010300	223.2	8	Ballyhilty	236.6	2007-2016
Ryewater	IE_EA_09R010600	211.3	9	Leixlip	209.6	2007-2016
Tolka	IE_EA_09T011100	136.6	8	Botanic Gardens	137.8	2007-2016
Morell	IE_EA_09M010300	98.8	6	Morell Bridge	98.7	2007-2016

It should be noted that the majority of the catchments selected for this study are common with the ones employed in the original regionalisation procedure (see Figure 2). However, some catchments are new inclusions (Ilen, Avonmore, Slaney, Morell, and Tolka) and can be considered completely independent from the regionalisation procedure. The time frame for the regional model development was 1990 to 2005 while the present study uses forcing data for the period 2007-2016 (except for Avonmore).

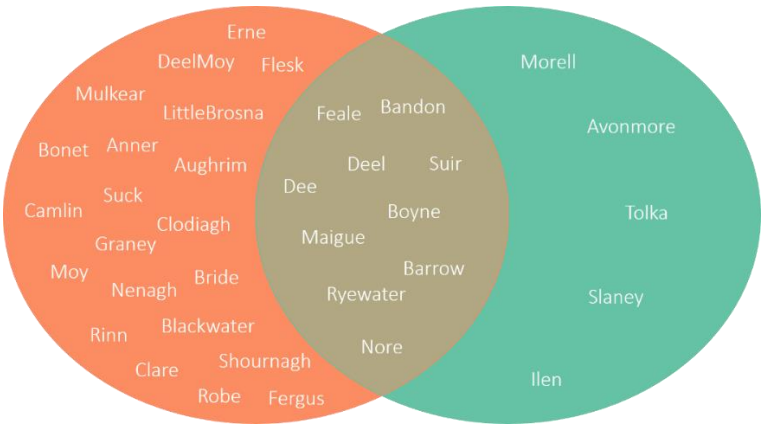

Figure 2: *Comparison of the 31 catchments used to develop the Regional Model (orange ensemble in figure) and the 15 catchments used in this present study to verify the performance of the model (green ensemble in figure).*

3.2. Objective Functions used for Performance Analysis

No calibration is done in this study. However, standard objective functions are used to assess the performance of the simulations against observed data. Because different objective functions give importance to different aspects of the hydrograph (e.g. the NSE given more weight to high flows, the BIAS comparing the overall volume of water), several indicators are used here to compare the simulations against the observations: the Nash-Sutcliffe criterion (Nash and Sutcliffe, 1970), and the bias. These are calculated as follows:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(Q_i^{obs} - Q_i^{mod})^2}{\sum_{i=1}^{n}\left(Q_i^{obs} - \overline{Q^{obs}}\right)^2} \quad ; \quad -\infty < NSE \leq 1$$

$$BIAS = \frac{\sum_{i=1}^{n}(Q_i^{obs} - Q_i^{mod})}{\sum_{i=1}^{n} Q_i^{obs}} \quad ; \quad -\infty < BIAS < +\infty$$

Where n is the total number of observations for the period of simulation; Q_i^{obs} is the ith observed discharge and Q_i^{mod} is the corresponding ith modelled discharge; $\overline{Q^{obs}} = \left(\sum_{i=1}^{n} Q_i^{obs}\right)/n$ is the mean observed discharge over the period of simulation.

4. DATA

4.1. Physical Catchment Descriptors

Various spatial datasets are required to parameterise the regionalised version of the SMART model. Below, only the physical characteristics that demonstrated a sensitive correlation with the hydrological parameters of the model are presented.

The Office of Public Works (OPW) produced a national spatial dataset as part of the Flood Studies update (FSU). Here, we use the meteorological and hydrometrical characteristics it contains such as long-term average rainfall and potential evapotranspiration, and long-term average discharge. Also used here are the drainage density, the base flow index, the slope, the index of arterial drainage, and the index of flood attenuation by lakes and reservoirs. The national Corine Land Cover map produced by the Environmental Protection Agency (EPA) gives the extent of urban and forested areas in the catchment. The national Soil map provided by Teagasc is used to characterise the composition and texture of the upper soil layers important to determine their drainage capacities. The Geological Survey of Ireland (GSI) produced a national map of groundwater recharge that contains much useful information, including, of interest here, the subsoil horizon permeability, the groundwater resources vulnerability, and a fine classification of the aquifers.

4.2. Meteorological Data

The forcing data required by the SMART model are precipitation and potential evapotranspiration (PE hereafter). Daily mean values provided by the Irish national meteorological office Met Éireann are used here. The rainfall average across each river sub-

basin is first determined using the Thiessen polygon method. Then, the average across the lumped catchment is calculated by summation of these semi-distributed mean values in order to guarantee the same volume of rainfall entering the system between lumped, semi-lumped, and semi-distributed simulations.

Due to the lower density of meteorological stations providing PE values, daily PE values for each river sub-basin are determined using the nearest neighbour method.

The daily meteorological values are then used to estimate hourly values using a uniform distribution.

4.3. Hydrometric Data

The data used to assess the performance of the simulations is the daily mean discharge provided by the OPW and the EPA that are the two main operators of hydrometric gauges across the Republic of Ireland. For each of the catchments in this study, the gauged data at or near its outlet is collected. To correct for cases where the gauge is not at the mapped catchment outlet, the discharge data is scaled using the ratio of the delineated area of the catchment over the gauged area at the hydrometric station.

The simulated discharge at the hourly time step is averaged over the day in order to be compared with the daily mean observed discharges.

5. RESULTS AND DISCUSSION

5.1. Performance of the Regionalised Model for Lumped, Semi-Lumped, and Semi-Distributed Catchments

The results of the simulations are shown on Figure 3 for the three performance indicators selected. The catchments are ordered by increasing area from the left to the right. Overall, the results on the NSE are satisfactory for almost all discretisation schemes (lumped, semi-lumped, and semi-distributed) with most simulations above 0.6 (except for the Tolka), 9 out of 15 catchments are even above 0.8. These results demonstrate good prediction capabilities for the regionalised version of the SMART model, which is encouraging for the predictions of flows in ungauged catchments in Ireland.

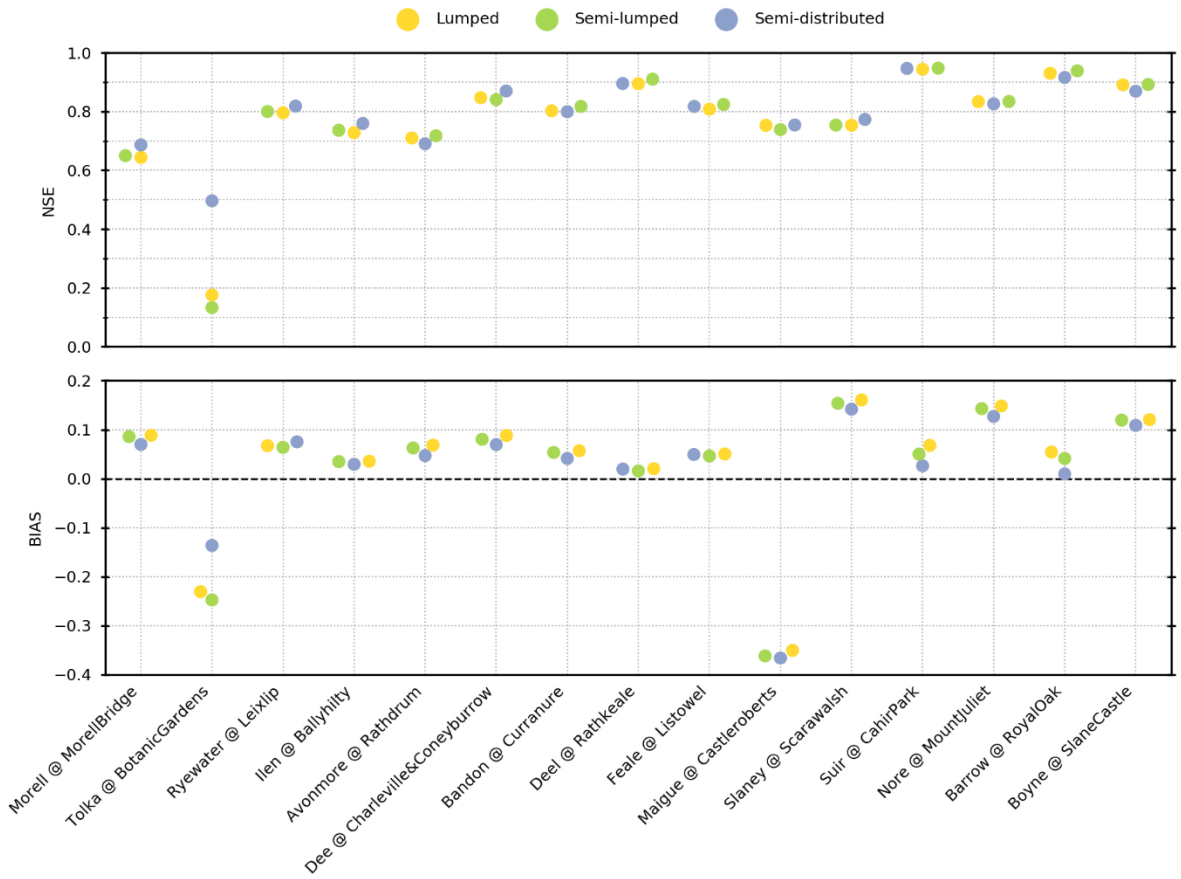

Figure 3: *Comparison of the three discretisation schemes (lumped, semi-lumped, semi-distributed) for the fifteen catchments on NSE, C2M, and BIAS. Catchments are denoted as "River Name @ Gauge Name" and they are sorted by increasing area (from the left to the right).*

The regionalised model performs poorly for the Tolka, where the NSE for both lumped and semi-lumped discretisation schemes are below 0.2, meaning that the model is only slightly better than using the average of the observed discharge for predictions. This demonstrates that either the model structure is inadequate for that catchment, or that appropriate model parameters are not correctly identified by the parameter transfer method. The semi-distributed discretisation for the Tolka yields better but still unsatisfactory predictions (NSE = 0.50). Since the Tolka features a large proportion of urban area compared to the other catchments of this study (see Figure 4), it is likely that the parameter transfer (regionalisation) relationships are not appropriate for catchments with large urban components. Indeed, the catchments used to develop the transfer equations did not contain catchments with large proportions of urban area, which means that the regression equations did not capture the behaviour (processes) of urban catchments. When using the lumped and semi-lumped discretisation, the catchment descriptor for the urban extent (URBEXT) is 27% for the Tolka, whereas it is below 5% for all of the other catchments used to develop the equations.

When using the semi-distributed discretisation, most of the sub-basins have relatively low urban extent, more in line with the other catchments in this set (see Figure 4). This can explain why the semi-distributed discretisation yields better predictions for the Tolka because the behaviour of the more rural sub-basins (with low urban extent) can be accurately represented using the regionalisation procedure (that used rural-dominated catchments as well). Some other catchments in the set feature sub-basins with some significant urban extents (Suir, Moy, Barrow, Boyne), but the total area is relatively small

for the whole catchment, whereas 3 out of 8 sub-basins in the Tolka have urban extents greater than 45%.

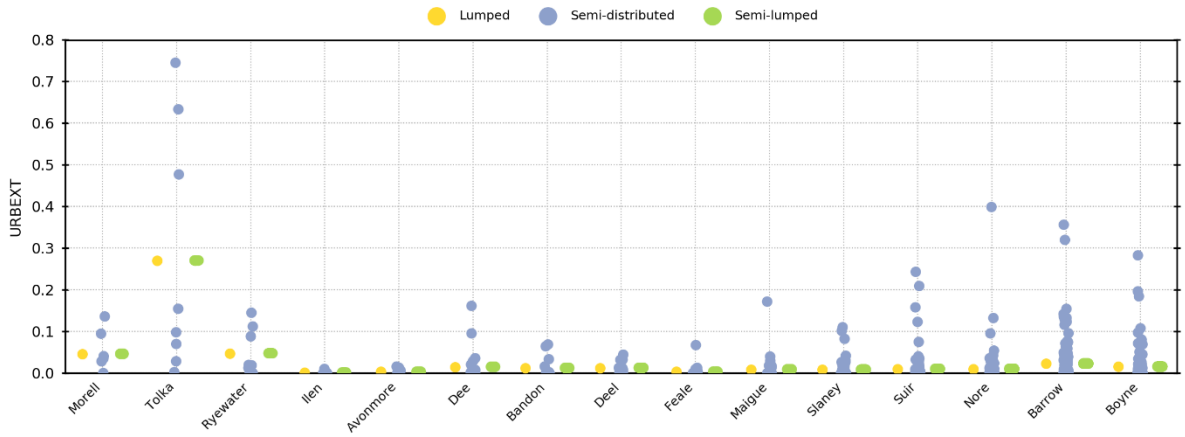

Figure 4: *Comparison of the urban extent (ratio) for the fifteen catchments and for the three discretisation schemes (lumped, semi-lumped, semi-distributed). Catchments are sorted by increasing area (from the left to the right).*

The results of the simulations do not demonstrate any significant improvement or deterioration of the predictions of the regionalised SMART model due to the discretisation scheme used. Indeed, on the Nash-Sutcliffe criterion, lumped, semi-lumped, and semi-distributed schemes perform in a very similar manner for fourteen of the catchments (excluding the Tolka for the reasons suggested above). This is a somewhat unexpected result, since the semi-distributed scheme uses a finer forcing input resolution as well as a finer spatial discretisation scale compared to the lumped scheme. Thus it might be expected that the semi-distributed scheme should yield better predictions than the lumped one because it has more information on the heterogeneity of the hydrological behaviour in the catchment. For example, Andreassian et al. (2004) found better predictions for semi-distributed schemes in French catchments and found that between 66% and 74% of the improvement could be attributed to the improvement spatial distribution of rainfall, with only a minor effect from improving the spatial discretisation of the catchment. However, their study is somewhat different to the present one since they compared calibrated models, whereas here a regionalised parameter transfer method is used.

The fact that the spatial discretisation has little effect on the performance of the regionalised SMART model could mean that the regional model (regression equations) are adapted to a variety of catchment sizes, from largest lumped catchments (≈ 2400 km^2) to the smallest river sub-basins (≈ 5 km^2). However, gauged river sub-basins should be used to confirm this hypothesis, because the present results could be due to a fortunate compensation of over- and underestimating river sub-basins. If the hypothesis is confirmed, this would mean that the dominant processes are the same for this range of catchment sizes.

Concerning the limited effect of the spatial discretisation of the meteorological input data, the results require some explanation. Indeed, if one looks at the variability of the rainfall distribution at the sub-basin scale compared to the average at the catchment scale (see Figure 5), there is a wide range of values for each catchment around the area weighted mean (lumped values). However, this variability does not appear clearly on the results predicting the discharge

at the outlet. In order to increase the performance of the predictions with the semi-distributed model, it might be necessary to significantly increase the temporal resolution for the simulations since the effective catchment scale for river sub-basin is much smaller (from 5 to 90 km^2).

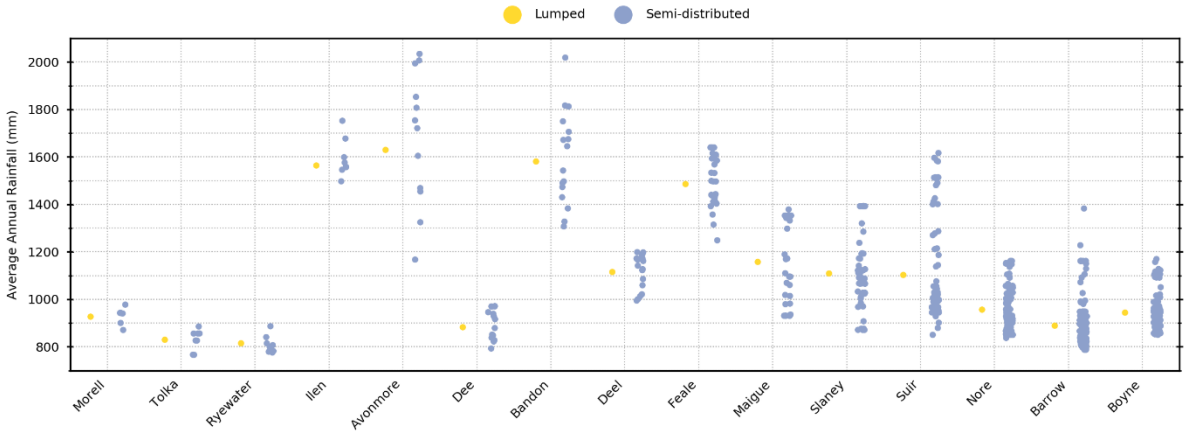

Figure 5: *Comparison of the rainfall distributions (annual average rainfall) for the fifteen catchments between catchment average (lumped) and river sub-basin average (semi-distributed). Catchments are sorted by increasing area (from the left to the right).*

The analysis of the bias index shows that, for most catchments, the regionalised SMART model provides accurate predictions: 10 of the catchments have a bias below 10%, 13 of the catchments below 20%. This means that the model is capable of good estimates of the overall volume of water going through the catchment during the simulated period. For most of the catchments, the model slightly underestimates this volume of water (when bias is positive). Only the Tolka and the Maigue exhibit less accurate model predictions and an overestimation by the model of the volume of water.

6. CONCLUSIONS

The regional parameter transfer method available to parameterise the SMART rainfall-runoff model for ungauged basins has been verified satisfactorily both at the lumped catchment scale and the river sub-basin semi-distributed scale using fifteen Irish catchments. The inclusion of five catchments completely independent from the regionalisation procedure objectively demonstrates the quality of the regionalised SMART model to predict the discharge at the outlet of ungauged Irish catchments. The results also demonstrate that the regional model is not well suited to parameterise the SMART model catchments with large urban areas. We suggest a different set of transfer equations may be appropriate for such cases.

The verification of the performance of the regionalised SMART model at the semi-distributed scale presented here demonstrates promising capabilities of the regionalised SMART model. However, more investigations using gauges within the catchments are required to determine whether underestimated and overestimated river sub-basins compensate each other, or if the satisfactory results at the catchment outlet also means a satisfactory behaviour locally. In other

words, the question remains over whether the dominant hydrological processes captured satisfactorily by the regional model are enough to model the hydrology at finer spatial and temporal scales.

The insensitivity of the spatial rainfall variability exhibited by the comparable performances of the lumped and the semi-lumped discretisation schemes raises questions over the suitability of the simulation time step when moving from lumped catchments (mesoscale) to river sub-basin scale. Investigations using hourly sub-daily rainfall data should be used to investigate this further. The regionalisation procedure only allows for one set of model parameters to be identified per catchment. This is a concern given the well-known problems of model structure and model parameter identifiability, where very often a set of so-called behavioural sets of model parameters is preferred to a single best performing set of parameters to overcome the problem of equifinality. This approach of accounting for uncertainty in model predictions should be transferred to regionalisation procedures (Wagener et al., 2004).

7. ACKNOWLEDGEMENTS

The authors would like to acknowledge funding from the Irish Environmental Protection Agency (EPA) through the Research Projects ESMANAGE (2014-W-LS-5) and DiffuseTools (2016-W-MS-24). Special thanks to Joanne Comer from the Office of Public Works (OPW) and Aidan Duffy from Met Éireann for their assistance in collecting the hydrometric and meteorological data used in this study.

8. REFERENCES

ANDREASSIAN, V., ODDOS, A., MICHEL, C., ANCTIL, F., PERRIN, C. & LOUMAGNE, C. 2004. Impact of spatial aggregation of inputs and parameters on the efficiency of rainfall-runoff models: A theoretical study using chimera watersheds. *Water Resources Research,* 40**,** 9.

MOCKLER, E., BRUEN, M., DESTA, M. & MISSTEAR, B. 2014. Pathways Project Final Report Volume 4: Catchment Modelling Tool (STRIVE Report).

MOCKLER, E. M., O'LOUGHLIN, F. E. & BRUEN, M. 2016. Understanding hydrological flow paths in conceptual catchment models using uncertainty and sensitivity analysis. *Computers & Geosciences,* 90**,** 66-77.

NASH, J. E. & SUTCLIFFE, J. V. 1970. River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology,* 10**,** 282-290.

NASR, A. & HYNDS, P. 2014. Assessment of the Hydrometric Network and Hydrodynamic Behaviour of Small Irish Catchments.

WAGENER, T., WHEATER, H. S. & GUPTA, H. V. 2004. *Rainfall-Runoff Modelling in Gauged and Ungauged Catchments*, Imperial College Press.