Automating Urban Hydrology Models: What Do Hydrologists expect From Model Developers?

RMM Pradeep[1#], NTS Wijesekera [2]

[1] General Sir John Kothelawala Defence University, Sri Lanka
[2]University of Moratuwa ,Sri Lanka
#Corresponding author; <rayantha@gmail.com>

Outline - *Availability of domain knowledge is significant for successful system automation. Therefore the input and output of the processes should be clearly understood. In the case of urban flood management, complex, specific and automated hydrologic models are required. For conceptual automated models, input and output needs must be clearly understood. The present work is to identify such needs through consultation of local and international hydrologic model developers. A comprehensive questionnaire was developed for consultation. The 3rd version of the questionnaire was answered by 58 hydrologists from Sri Lanka, India, Nepal, Nigeria, UK and Japan. Hydrologists show less interest on input data format because they are highly competent in data conversions using different techniques. Nevertheless, a majority stresses the need of to correctly interpret the outputs. Hence the work identified that, when developing a tool, system developers need to pay more attention to the conversion of the hydro model output to end user understandable format specially in the case of urban flood management decision makers.*

Keywords— **Model Automation, Hydro models, Input-process-output, Hydrologists' Requirements**

I. Introduction

A. *Multi-stakeholder Urban Flood Management*

Urban flood is one of the frequently occurring natural disasters, which can manage through engineering options. (Carver, 2016; Hellmers, Manojlovic, Palmaricciotti, & Fröhle, 2014; Konrad, 2003; Xia, Falconer, Lin, & Tan, 2011; Yu, Yin, & Liu, 2016) Unlike other flood analysis models, the hydrological models behind the urban flood management are sensitive due to the small size of; watershed and effective area. Same time the urban flood management process involves number of stakeholders varying from highly technical hydrology model developers to citizens / general public. However the model developers are developed the model and can handover the procedures to be followed by flood management decision makers of the urban areas. Then decision makers can run the model and reach to decisions which should be followed and applied by general public/citizens (Eger, Chandler, & Driscoll, 2017; Fatichi et al., 2016; Gray, Paolisso, Jordan, & Gray, 2017, p. 304; Gupta, 2012; Weiler & Beven, 2015).

B. *Automation of Hydrology Model*

The urban decision makers are not highly professional in hydrology and can be considered as non-technical. But the flood management decisions have to be taken on daily basis as the land enhancement of the urban area is a common phenomenon which contributes to increase surface runoff, a.k.a. flood generation. Therefore, the decision makers have to run the hydro model for new land modifications and make the decision to reduce the effect on flood generation. When running a hydro model, it needs to input various spatial and non-spatial data, manipulate them to; find the effect and optimized solution. This is a difficult task to non-technical person, hence required to be automated (Assaf et al., 2008).

C. *Input, Process and Output*

The automation of any process can be divided to three categories, inputs, process and outputs(Waring, 1996). Then each category should be clearly understood through proper requirement engineering. In the urban flood management scenario, the model is developed by hydrology professionals and to be used by non-technical decision makers. Therefore rather than the inside process, non-technical decision makers handle the input and output to and from process. Hence the input and output of hydrology model should be aligning to hydro modellers' aspirations whilst it should be understood by non-technical decision makers.

D. *The Problem*

Then the hydrologists' requirements on input and output of the process are become one of the important considerations when facilitate the non-technical person to handle and understand the model outputs. To automate, the system developers need to find out the flexibilities of inputs and outputs.

E. *Aim*

Therefore the aim of the present work is to identify hydrologists' view on inputs and outputs of urban flood management hydro model.

II. Methodology

A. Research Methodology

The present work initially identified the basic requirements through a literature survey. Then outlined a questioner and evaluated with 2 hydrologists. With their observation the second questionnaire was developed and evaluated with selected hydrologists from the cross section of the target sample. With the observation, the final questionnaire was developed and share to all possible hydrologists via email. Then acquired the views and analysed the result to reach conclusions.

B. Literature survey to identify basic concerns

Referable to the hardbound relationship between water and mankind, human perpetually curious about the behaviour of water. This curiosity lead mankind to distinguish the water cycle process as, long ago, 1200 B.C. Around A.D. 1800's the water experiments started to flourish and since then hydrologist used to utilize such experiment-resulted hydrologic models' output to water resource management actions such as flood management. Now hydrological models considered as sufficiently matured to provide flood management information.(Chow, Maidment, & Mays, 1988).

However, when utilise such models for practical applications, it need to modify the model to match with the local scenario as no hydrological model is universally applicable (Kavetski & Fenicia, 2011; Siderius, Biemans, Kashaigili, & Conway, 2018). Hence the initial requirement in hydrology modelling is handling the spatially distributed large amount of data to be input, then predict and simulate the nature accurately. Then this refers to not only the amount but also resolution of the input data. (Ogden, Garbrecht, Debarry, & Johnson, 2001). As well when hydro modelling, the user friendliness has highlighted as one of the main draw backs. Due to vague descriptions about the limitation of the models, the hydrologists need to clearly understand the applicability of the models and demarcate the model boundaries when creating of model (Devi, Ganasri, & Dwarakish, 2015).

Then when reviewing these literatures, it understood that the inputs to the hydro model are independent, but outputs are dependent on hydro model. As well creation and testing of such model are a responsibility of hydro modellers. Therefore, it needs to clarify the resolution and format requirements of inputs and outputs.

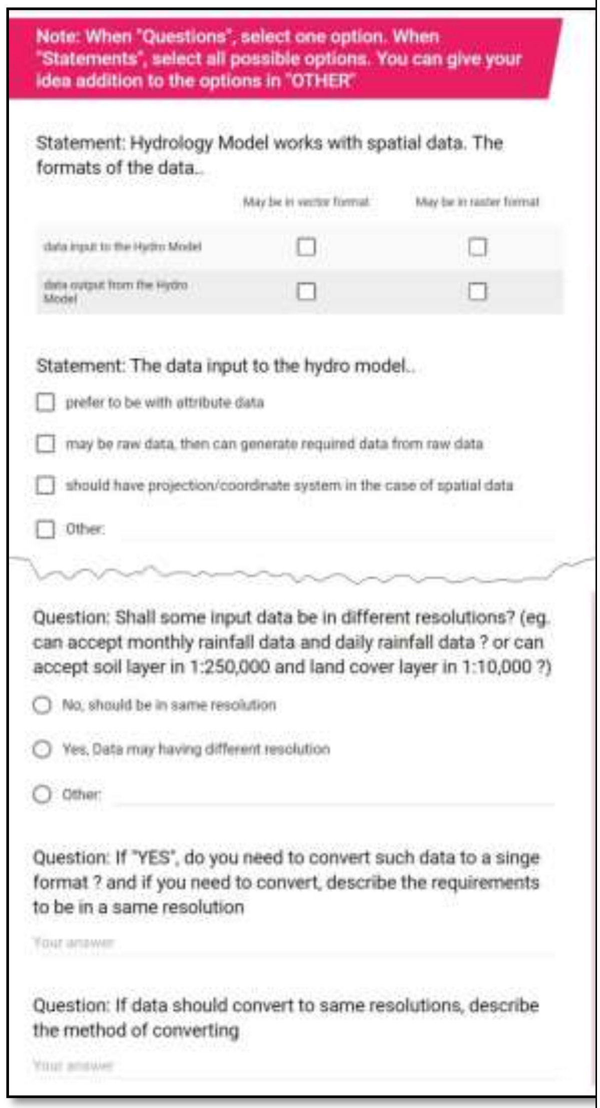
C. Questionnaire Development

With these references it prepared an interview questions to get the hydrologists' requirements. A single question was developed to acquire the format preferences on vector and raster formats based on the input or output situation. Other questions based on the resolution of the temporal and spatial resolution of the data.

After a discussion with a hydrologist, questions were modified and 1st version of the questionnaire was developed. It contains questions about; converting input data resolution to required resolution, output data (resolution and converting), and lookup data (requirement and classification).

The 1st version of questionnaire was created using Google form and evaluated with two other hydrologists and got the comments on the questions. The main observation was on questioning method as it consists of questions as well as statements. Further the observers requested to allowing a space to give ideas of the hydrologist about the requirement.

With these comments the questions were rearranged and created the 2nd version of the questionnaire. Special note was included in the beginning of the questionnaire to educate how to answer the questions. In there the instruction was given to select one or multiple options when a "statement" is given and select only one option or fill the blank when "question" is asked (Figure 1.0).


Note: When "Questions", select one option. When "Statements", select all possible options. You can give your idea addition to the options in "OTHER"

Statement: Hydrology Model works with spatial data. The formats of the data..

	May be in vector format	May be in raster format
data input to the Hydro Model	☐	☐
data output from the Hydro Model	☐	☐

Statement: The data input to the hydro model..

☐ prefer to be with attribute data

☐ may be raw data, then can generate required data from raw data

☐ should have projection/coordinate system in the case of spatial data

☐ Other:

Question: Shall some input data be in different resolutions? (eg. can accept monthly rainfall data and daily rainfall data ? or can accept soil layer in 1:250,000 and land cover layer in 1:10,000 ?)

○ No, should be in same resolution

○ Yes, Data may having different resolution

○ Other:

Question: If "YES", do you need to convert such data to a singe format ? and if you need to convert, describe the requirements to be in a same resolution

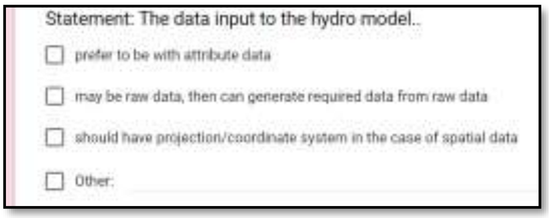
Your answer

Question: If data should convert to same resolutions, describe the method of converting

Your answer

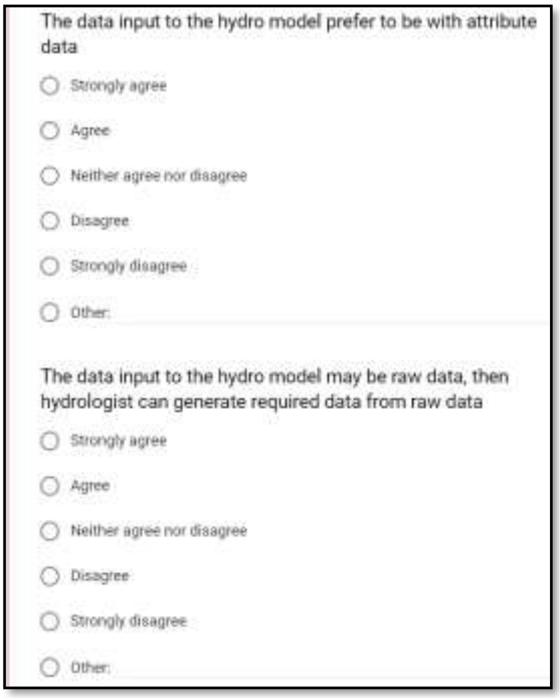
Figure 1.0 Questions and Statements of Questionnaire Version 2.0

The questionnaire was distributed among 10 hydrologists, but replied only by 4. Then personally contact them and find the reasons to not to reply. Apart from the personal reason, the main obstacle with the questionnaire was, asking the methods (related to hydrological model) in questions. Due to the difficulty of remembrance of the appropriate and suitable method or fear of missing most important ones, few hydrologists refused to answer the questionnaire. The same comment was given by the answered-hydrologist about inquiring methods. As well the comments were requested to make the questions / statements simpler and allow select the suitable option/s from a list. With these observations, the questionnaire was rearranged and created additional 2-3 questions from a single question in version 2.0 to get the clear view of the hydrologist (Figure 2.0). Then it created version 3.0 of the questionnaire. It was discussed and adjusted with the answered-hydrologists and confirmed whether all their needs had being incorporated.


Statement: The data input to the hydro model..

- ☐ prefer to be with attribute data
- ☐ may be raw data, then can generate required data from raw data
- ☐ should have projection/coordinate system in the case of spatial data
- ☐ Other:

Sample questions in questionnaire version 2.0


The data input to the hydro model prefer to be with attribute data

- ○ Strongly agree
- ○ Agree
- ○ Neither agree nor disagree
- ○ Disagree
- ○ Strongly disagree
- ○ Other:

The data input to the hydro model may be raw data, then hydrologist can generate required data from raw data

- ○ Strongly agree
- ○ Agree
- ○ Neither agree nor disagree
- ○ Disagree
- ○ Strongly disagree
- ○ Other:

Expansion of same question in questionnaire version 3.0

Figure 2.0 Expansion of question to multiple questions

Then the final questionnaire (version 3.0) was distributed among more than 100 professionals worldwide using emails and researchgate.com and was replied by 58 hydrologists.

III. Analysis and Results

A. *Respondents*

All most the Sri Lankan and Indian hydrologists who received the questionnaire have answered. Then, it can be considered that the result is bias towards the Asian philosophy. The respond participation is shown in the Figure 3.0

As well, the majority of participants (41%) are field engineers who play the hydrologist role in government and non-government organization. Their day to day task is to provide hydrological options to nation-level decision making in flood management activities. 75% of the academics are professors/doctors who are in the civil engineering field. The "Students" category in the survey includes only the participants of MSc and PhD programs in water resource management. The researchers are the academics who are retired or independent consultants in water resource management. (Figure 4.0)

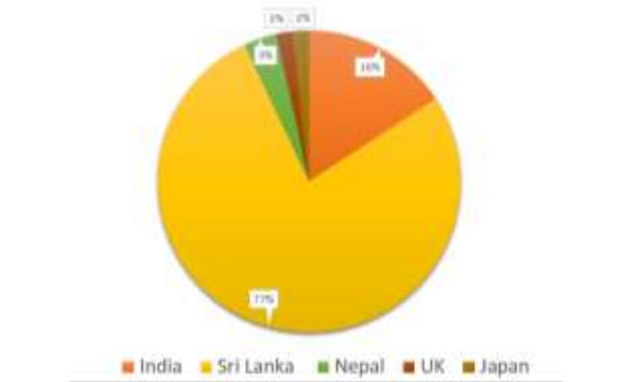


Figure 3.0 Geographical Distribution of Participants

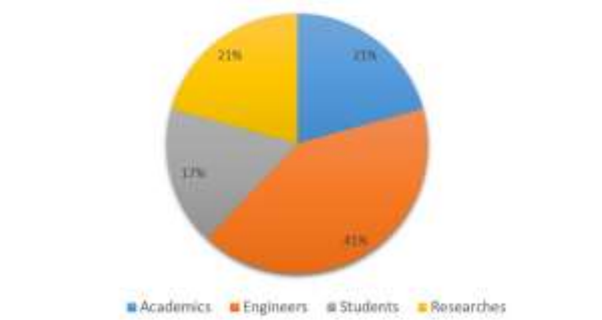


Figure 4.0 Profession of Participants

B. *Data format of input and output data*

The data format of the inputs and outputs is one of the interested components to automation due to the hydrology models are bound to utilise the spatial data. The question was asked about the preferred data formats for used in the models. The options were given to select multiple options out of 4; (1) Any data format (spatial and non-spatial), (2) Spatial data only, (3) Spatial data, but vector format only, (4) Spatial data, but raster format only.

The preference shows that a minority (12%) preferred to work with non-spatial data whilst majority (55%) needs to work with the all types of formats such as non-spatial, vector and raster. However, 27% preferred to work with spatial data, either in vector or raster (Figure 5.).

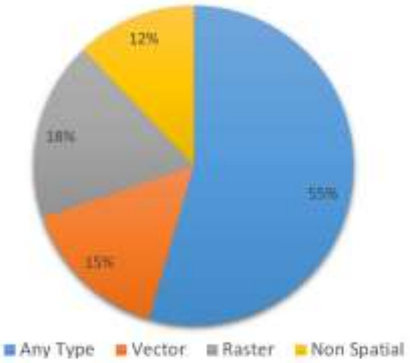


Figure 5.0 Preference of working data format

C. Input Data and data conversion

The majority of the hydrologists (58%) preferred to input the spatial data with the attribute data whilst 34% considered it as a sine-qua-non. However, it was no negative feedback about this requirement whilst 8% does not care about the attribute data (Figure 6.0).

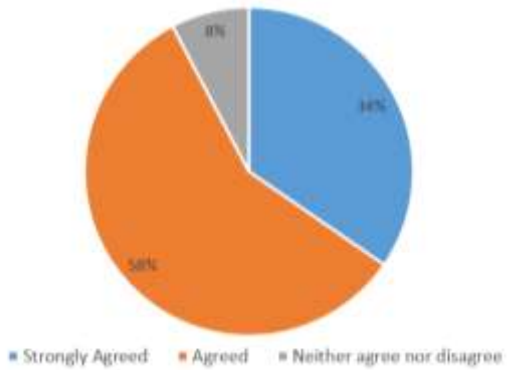


Figure 6.0 Agreement level about the inclusion of attribute data to spatial data

As well majority (73%) of the hydrologist shows that the capability of converting the input data to the format they required whilst 19% does not care. However, 8% comment differently about the conversion requirements. They stated that it should purify the data for remove errors and complexities before input(Figure 7.0).

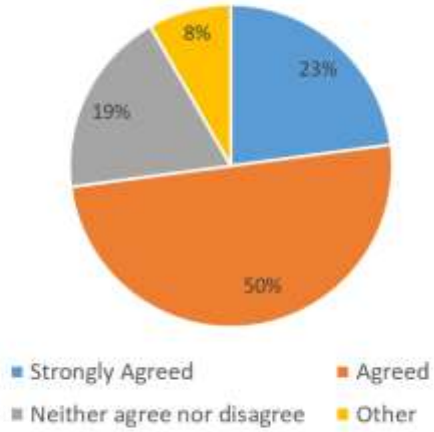


Figure 7.0 Agreement level about the hydrologists' capability of data conversion

Further, in the case of spatial data, 87% of the hydrologist required the input spatial data to be provided with a projection /coordination system details which is a standard requirement of spatial data. Nevertheless 9.3% does not bother about such whilst 3.7% feels it is not a requirement even it is a standard.

As well, due to multiple data used for hydrological analysis, it is required to get to know the compatibility of data resolution among the different sources. On this question, hydrologist shows divided preferences. The majority (51%) needs inputs must be in the same resolution, but considerable (26%) amount disagrees and hopes to accept different resolutions. However, 22% does not prescribe any.

Further around 50% hydrologists need to convert the data as it needs to fulfil the requirements of standardization of the developed model, portability of the developed model across organizations and standard requirement of input data of the particular hydrological process. 35% of hydrologists state that the conversion is required due to the availability of data from verified sources such as survey department. They have stated the cost of development may increase due to requirement of developing multiple process paths for different resolution data. However, one hydrologist highlighted the generation of the required layers from satellite images which required extensive conversions.

When considering the data conversion methods, the most popular one is Re-sampling. The other preferences are shown in the Table 1.0.

Table 1.0 Preferred data conversion methods of hydrologists

Preferred Method	Preferences
Re-sampling	35%
Aggregation	19%
Multidimensional scaling (MDS) techniques	16%
Factor analysis and Kohonen nets	10%
Clustering and geometric triangulation	10%
Self-organizing map (SOM) or self-organizing feature map (SOFM)	6%
Anything Appropriate	3%

D. Output Data and data conversion

When the questioning about the outputs of the hydro models the hydrologists preferences show a distinct diversification (Table 2.0).

Table 2.0

Output Format	Preferences
Vector, Raster and Non-spatial	37%
Vector and raster only	19%
Non-spatial summaries only	13%
Only vector	10%
Only Raster	8%
Non-spatial summaries with Raster	8%
Non-spatial summaries with Vector	4%
Allow customizing	2%

They have a question about the users of the output. Then their answers started to vary from allowing users to change format, to, use the output without changing. However, 37% of hydrologists allow having vector, raster and non-spatial outputs. As some hydrologist (19%) believes map outputs are much more convenient for non-technical users. Then they need to output to be in spatial format, vector or/and raster. 18% (10% for vector and 8% for raster) also supported this preference, but they have bound to specific data formats. In the same way some hydrologists (13%) believe the outputs to be simpler to match with reference values, such as thresholds, preferred non-spatial outputs. Very few (2%) hydrologists allow to users to play with the outputs.

Then conversion of the output to match with the end user requirement creates a problem among the hydrologist. The majority (65%) wanted to consult them when such conversion. Meantime another 22% does not care about the situation. However, only very few (13%) allow customizing the outputs without consulting the hydrologists (Figure 8).

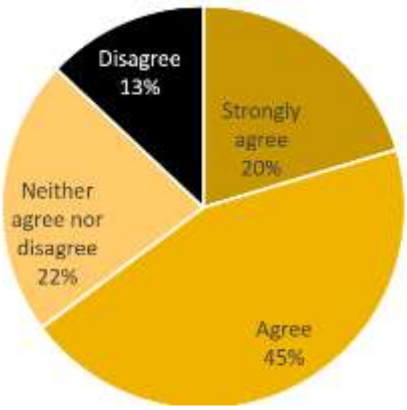


Figure 8.0 Agreement level about the consulting the hydrologists for customizing the output

Then the next question was inquiring the reason for consulting the hydrologist for output conversions. Then it could able to find five reasons as shown in the table 3.0.

Table 3.0 Reasons - why consult hydrologist?

Reason	Indicated by
The output result may give a wrong interpretation due to the customized change	81.8%
Result's resolution/format based on Input data resolution/format. Hence need to consult	52.3%
The output is accurate only on given format/resolution	29.5%
Before generating output model must be calibrated and validated as per site conditions	2.3 %
The user shall validate the output and calculate the performance indices	2.3 %
There could be unseen errors	2.3 %

The "Indicated by" column describes the preference percentage of the total population. Then the majority of hydrologist prescribed to consult them when customized the output as the output may give the wrong interpretation. As well, 52% further indicated, consultation of the hydrologist to be sought due to the output format is based on the input format. A same kind reason is highlighted by 29.5%, which they indicate the accuracy level is only correct in output resolution/ format. Hence, when make changes, the accuracy to be maintained and need to consult hydrologist for accuracy reasons. The other minor reasons for consultation are applicable changes due local site/ geographical differences, requirement to align with performance indices and unseen errors in the output may magnitude with modifications.

E. The reference data to the Hydrology Model

Apart from the input and output data which subjected to utilise in the hydrological processes, it may require reference data, such as runoff coefficients etc. to re-classify input or output data. Then through the questionnaire it evaluates the real requirement of such lookup data for attribute classifications. The results shows that 80 % required to refer but very few 4% does not.

When it inquired about the classification schema preferences, (whether need to be qualitative or quantitative) it was not shown considerable variation; both the methods received same preferences. However, when it needs to convert qualitative answers to quantitative answer, the preference is to provide higher number for most positive answer (eg. Very High is 5) and lowest number to most negative answer (eg. Very Low is 1).

F. Other requirements of hydrologist

Apart from the input, output and reference data, the hydrologist has commented about various other requirements when automating models. The following list shows a summary;

a) The automated model should be simple like mobile application

b) Graphical Interface should be user-friendly for handling the model

c) Strong calibration and verification option should be available.

d) Automated tool should prevent and solve the possible errors.

e) Automation should provide a method to prevent input incorrect / invalid data

V. DISCUSSION

The present work required to get the requirements and restrictions when a hydro model is automated. As the hydro modellers are the creators of the models, the questionnaire was targeted to get their views on inputs, outputs, reference information and additional views on automation.

The intended questionnaire was developed systematically, which, gather outlined data using literature survey, frame the questions and get the view of hydrology modellers, develop the 1st version, get the views about the questionnaire using hydrologist, modify and create 2nd version, evaluate the applicability of the 2nd version to distribute among the community with few other hydrologists, get their views and modify to create a final version, version 3.

The questionnaire was distributed around the world-wide hydrology modellers, but the most of the replies got through Sri Lanka and India. Then the results are having a bias to Asian philosophy on hydro modelling.

As well as, around 62% of the views are from the field engineers and researchers who developed and practised hydrology models for day to day decision making activities. Hence, it can be considered that the results denote the practical requirement, rather than theoretical requirements.

When consider the input data formats, the hydrologist tends to use any available format and showed the capability of converting such data to the format and resolution they want. Hydrologists have diversified ideas about the resolution and conversion of such, but the answers show that, the data conversion should follow only under the guidelines of hydrology modellers.

The work found that the Re-sampling, Aggregation and Multidimensional scaling (MDS) techniques are popular among hydrologist as data conversion methods.

However, in the case of output customization (this is needed as the output should be understood by end users) hydrologist demonstrated much responsibility. Due to the possibility of wrongly interpret the output, the accuracy depends on the input data and the results' accuracy depends on the resulting format, hydrologist urges to consult them for interpreting or customizing the hydrological model output.

Hydrologists need the intermediate referencing data based on the situations, however, no vast deviations could be observed. In classification schemas, the hydrologists' philosophy is to provide higher values to positive attitudes and lower numerical values to negative attitude.

Finally hydrologist required the automated tools for hydrology models which are with user friendly graphical interfaces that provide easy calibration and verification options and user guidance to prevent from errors.

VI. CONCLUSION

The present work able to utilise systematically developed questionnaire to be distributed via Google form and collected the hydrologists' views on model automation successfully.

The data analysis shows the views and requirements of the Asian hydrologists. The major view is the hydrologists ready to use any data for their models by adjusting and modifications. But, they are reluctant to allow customising the outputs for the purpose of easy understands to end users without their consultation.

Therefore, when automating a hydrology model to be used by the non-hydrological decision makers in urban flood management, a special attention and limitation have to be incorporated when automatically converting the output to understandable versions.

Acknowledgment

Author pays his honour to Dr.R.L.H.L.Rajapakse of University of Moratuwa for providing opportunity to link most senior hydrologists in Sri Lanka. Further author would like to offer his special thanks to Prof. K.D.W Nandalal and Dr.(Mrs). H.K. Nandalal of University of Peradeniya and Dr. T. M. N. Wijayaratna of University of Moratuwa for their valuable views and ideas.

Finally author's special thanks are extended to the Ms. E.A.G.Chandaramali and Ms Vinu of UNESCO - Madanjeet Singh centre for South Asia water management, University of Moratuwa for remarkable assistance whilst data collection.

References

Assaf, H., Beek, E. van, Gijsbers, P., Jolma, A., Kaden, S., Kaltofen, M., … Borden, C. (2008). Generic Simulation Models for Facilitating Stakeholder Involvement in Water Resources Planning and Management : A Comparison , Evaluation , and Identification of Future Needs . *Developments in Integrated Environmental Assessment, Elsevier*, *3*, 229–246.

Carver, S. (2016). Flood management and nature – can rewilding help? *ECOS: A Review of Conservation*, *37*(1), 32–42. https://doi.org/ISSN 0143-9073

Chow, V. Te, Maidment, D. R., & Mays, L. W. (1988). *Applied Hydrology*. (B. J. Clark & J. Morriss, Eds.) (Internatio). Singapore: McGraw-Hill.

Devi, G. K., Ganasri, B. P., & Dwarakish, G. S. (2015). A Review on Hydrological Models. *Aquatic Procedia, Elsevier*, *4*, 1001–1007. https://doi.org/10.1016/j.aqpro.2015.02.126

Eger, C. G., Chandler, D. G., & Driscoll, C. T. (2017). Hydrologic processes that govern stormwater infrastructure behaviour. *Hydrological Processes*, *31*(25), 4492–4506. https://doi.org/10.1002/hyp.11353

Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., … Tarboton, D. (2016). An overview of current applications , challenges , and future trends in distributed process-based models in hydrology. *Journal of Hydrology*, *537*, 45–60. https://doi.org/http://dx.doi.org/10.1016/j.jhydrol.2016.03.026

Gray, S., Paolisso, M., Jordan, R., & Gray, S. (Eds.). (2017). *Environmental Modeling with Stakeholders*. Springer. https://doi.org/10.1007/978-3-319-25053-3

Gupta, P. K. (2012). User Friendly Open GIS Tool for Large Scale Data Assimilation – A Case Study of Hydrological Modelling. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XXXIX-B4*, 427–430. https://doi.org/10.5194/isprsarchives-XXXIX-B4-427-2012

Hellmers, S., Manojlovic, N., Palmaricciotti, G., & Fröhle, P. (2014). Modelling Decentralised Systems for Urban Drainage and Flood Mitigation. In Kopmann & Lehfeldt (Eds.), *ICHE 2014* (pp. 375–382). Hamburg.

Kavetski, D., & Fenicia, F. (2011). Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights. *Water Resources Research*, *47*(11), 1–19. https://doi.org/10.1029/2011WR010748

Konrad, C. P. (2003). Effects of Urban Development on Floods : USGS Fact Sheet FS-076-03. *U.S. Geological Survey*, (November), 1–4. https://doi.org/USGS Fact Sheet FS-076-03

Ogden, F. L., Garbrecht, J., Debarry, P. A., & Johnson, L. E. (2001). DIS and Distributed Watershed Models I. *Journal of Hydrologic Engineering*, *6*(6), 515–523.

Siderius, C., Biemans, H., Kashaigili, J. J., & Conway, D. (2018). Going local: Evaluating and regionalizing a global hydrological model's simulation of river flows in a medium-sized East African basin. *Journal of Hydrology: Regional Studies*, *19*(October), 349–364. https://doi.org/10.1016/j.ejrh.2018.10.007

Waring, A. (1996). *Practical systems thinking*. Thomson Learning. Retrieved from books.google.lk/books?redir_esc=y&id=ltwxkiGX2YwC&q=input+process+output#v=snippet&q=input process output&f=false

Weiler, M., & Beven, K. (2015). Do we need a Community Hydrological Model? *Water Resources Research COMMENTARY*, *51*, 7777–7784. https://doi.org/10.1002/2014WR016731.Received

Xia, J., Falconer, R. A., Lin, B., & Tan, G. (2011). Modelling flash flood risk in urban areas. *Water Managment*, *164*(WM6), 267–282. https://doi.org/10.1680/wama.2011.164.6.267

Yu, D., Yin, J., & Liu, M. (2016). Validating city-scale surface water flood modelling using crowd- sourced data. *Environmental Research Letters*, *11*, 1748–9326. https://doi.org/10.1088/1748-9326/11/12/124011