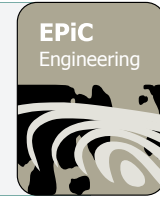




EPiC Series in Engineering

Volume 3, 2018, Pages 1048–1054

HIC 2018. 13th International  
Conference on Hydroinformatics



# Assessing the Effect of Streamflow Estimation at Potential Station Locations in Entropy-Based Hydrometric Network Design

Jongho Keum<sup>1</sup>, Paulin Coulibaly<sup>1</sup>, Alain Pietroniro<sup>2</sup>

<sup>1</sup> McMaster University, Hamilton, Ontario, Canada

<sup>2</sup> National Hydrology Centre, Environment Canada, Saskatoon, Canada

[jkeum@mcmaster.ca](mailto:jkeum@mcmaster.ca), [couliba@mcmaster.ca](mailto:couliba@mcmaster.ca), [Al.Pietroniro@canada.ca](mailto:Al.Pietroniro@canada.ca)

## Abstract

Having an efficient hydrometric network is important not only for successful water resources management but also for dealing with the economic cost of maintaining the network. One of the challenging tasks is to have a reliable dataset at candidate locations of additional monitoring stations. While many have applied regionalization methods, such as spatial interpolation, this study introduced a spatially distributed hydrologic model for generating data at potential locations. The determined optimal networks are compared with those from the use of spatial interpolation. The optimal networks are also evaluated using the outcome of transinformation analysis. The results showed that the optimal results using a spatially distributed model performed better than those using a spatial interpolation method.

## 1 Introduction

An efficient and informative hydrometric network is a fundamental piece of infrastructure for sustainable water resources planning and management. Even though remote sensing techniques can provide viable hydrologic information, the ground-based monitoring network is still essential to collect accurate and reliable in-situ observations and to validate the remotely sensed data. Ideally, hydrometric networks should be able to deliver sufficient and reliable information for hydrologic research and applications. However, the shrinking of hydrometric networks due to financial budget cuts is a global trend [1]. Therefore, the determination of an optimal network by selecting the adequate number and locations of monitoring stations is an important task.

Since the inception of Shannon entropy [2], it has been broadly and successfully adapted to water monitoring network design. However, the calculation of entropy requires estimated time series data

which is still a challenging task that is required to solve network redesign or expansion problems. Many studies have used the spatial interpolation, such as Kriging approach [3] and inverse-distance-weighting (IDW) method [4], to generate flow data at ungauged candidate locations.

In this study, a physically-based hydrologic model namely Modélisation Environnementale Communautaire – Surface and Hydrology (MESH) model is used and compared to the spatial interpolation method. These results are also compared with the transinformation analysis results. The main objectives of this research were to:

- Evaluate the existing network using entropy-based transinformation analysis;
- Determine and compare optimal hydrometric networks obtained using MESH outputs vs. regionalized flow at potential stations.

## 2 Study Area and Data

The study area is limited to the 321,000 km<sup>2</sup> on the Canadian side of the Great Lakes Basin, which will be called CGLB hereafter. There are 347 HYDAT stations that are currently active and maintain at least the last 10 years of streamflow records. Streamflow data from existing stations in the CGLB were retrieved from the HYDAT Database of Environment and Climate Change Canada. Additionally, the outlets of the distinct catchments were chosen as the locations of the potential stations. The stream reaches were determined from the 50,000:1 digital elevation model retrieved from the Natural Resources Canada's GeoGratis database. The threshold value to be defined as stream reaches is 1% of the maximum flow accumulation, which is a simple rule of thumb for stream determination and the default value in ArcHydro toolbox of ArcGIS software. The catchment was delineated for each stream reach, and the outlet which will be used as the potential station was determined. The total number of potential stations determined was 372.

Previous research [5] recommended 10 years of data period to design hydrometric networks using the combined entropy multi-objective optimization technique when daily time series are used. However, the data period of this study is limited to 8 years (2006 to 2013) because (a) the HYDAT Database for the existing stations has not been fully updated for 2015 at the time of analysis; (b) the Canada Daily 10 km Gridded data for Min-Max Temperature and Precipitation dataset is available from 1950 to 2013; (c) the MESH flow series available for the CGLB are limited to 2006 to 2015. Therefore, the common data window was the 8 years period from 2006 to 2013, and the detailed description of the use of these datasets is given in the Methodology section.

## 3 Methodology

### 3.1 Transinformation Analysis

Transinformation analysis is a preliminary station ranking scheme to evaluate the existing hydrometric network using entropy theory. It does not require any potential station data nor complex optimization techniques but employs a simple entropy calculation. The term transinformation expresses the mutual information from two variables, and is calculated by:

$$T(X, \hat{X}) = H(X) - H(X|\hat{X}) = H(\hat{X}) - H(\hat{X}|X) \quad (1)$$

where  $T(X, \hat{X})$  is the transinformation between variables  $X$  and  $\hat{X}$ , and  $H(X|\hat{X})$  is the conditional entropy of variable  $X$  given a variable  $\hat{X}$ . In the transinformation analysis, the marginal entropy of each existing station is calculated by both the observed time series,  $H(X)$ , and the estimated values

using multiple regressions from other stations,  $H(\hat{X})$ . A station with high transinformation indicates that information from that station can be replaced by the information from the neighboring stations. On the other hand, stations with lower transinformation have more independent information from other stations; hence, they are given higher rankings.

### 3.2 Dual Entropy and Multiobjective Optimization

The basic design objectives using the DEMO for the optimal hydrometric network providing a maximum amount of information with a minimum amount of redundant information become:

$$\begin{aligned} & \max[H(S_{N,M}) = H(E_1, E_2, \dots, E_N, X_1, X_2, \dots, X_M)] \\ & \min[C(S_{N,M}) = C(E_1, E_2, \dots, E_N, X_1, X_2, \dots, X_M)] \end{aligned} \tag{2}$$

subject to: N and M are integers;  
 $M \in \{1, 2, \dots, M_{max}\}$

where  $S_{N,M}$  is a network consisting of  $N$  existing stations ( $E$ ) and  $M$  additional stations ( $X$ ).

### 3.3 Streamflow Estimation at Candidate Locations

To estimate the information from potential stations, regionalization approaches have been applied. One of the popular techniques is the inverse distance weighting method combined with the drainage area ratio (IDW-DAR) [6]. The IDW-DAR is one of the simplest spatial regression techniques and predicts streamflow at ungauged locations as follows:

$$Q_p = \sum_{i=1}^n \omega_{i,p} \left(\frac{A_p}{A_i}\right)^\alpha Q_i, \quad \omega_{i,p} = \frac{h_{i,p}^{-2}}{\sum_{i=1}^n (h_{i,p}^{-2})} \tag{3}$$

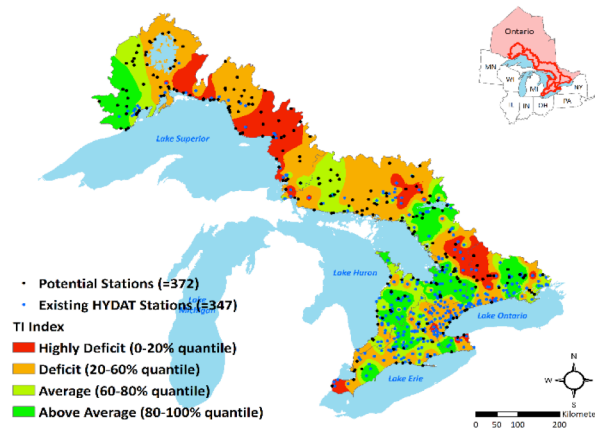
where  $Q_p$  and  $Q_i$  are the interpolated streamflow at the potential station  $p$  and the observed streamflow at the existing station  $i$ ,  $\omega_{i,p}$  is the weighting factor,  $A_p$  and  $A_i$  are the drainage area of the potential station  $p$  and the existing station  $i$ ,  $\alpha$  is the exponent coefficient, and  $h_{i,p}$  is the distance between the existing station  $i$  and the potential station  $p$ , respectively. On the other hand, we also applied the physically-based distributed hydrologic model (namely MESH “Modélisation Environnementale Communautaire – Surface and Hydrology”, [7]), which is an environmental modelling system developed by Environment Canada in order to facilitate the coupling of various physical models specialized in different areas. Here the MESH model outputs specifically the daily flow series were used as estimated streamflow at the potential stations and compared to the transinformation analysis and the optimal networks using IDW-DAR.

## 4 Results

### 4.1 Transinformation Index Map

Figure 2 shows the transinformation index map of the CGLB using the streamflow observations of the existing HYDAT stations. The estimated index for each station was spatially interpolated using the inverse distance weighting method of the ArcGIS software. The gridded transformation index was

then categorized into four levels shown in Figure 2. The northeast of the Lake Superior and the northern region of the Lake Ontario were determined as “Highly deficit” regions meaning that existing networks (if any) were inadequate. “Deficit” regions are also located near the “High Deficit” regions, specifically in the north of the Lakes Superior and Huron. As a preliminary analysis prior to the DEMO applications, the transinformation analysis evaluated the existing hydrometric network and spatially categorized the regions where the monitoring efforts are more or less required.



**Figure 1:** Transinformation index map

## 4.2 DEMO Application Results

After applying DEMO for two cases of streamflow estimation methods (i.e., IDW-DAR and MESH), DEMO has yielded two optimal solutions (or networks) in each case. Figures 2 and 3 show the spatial distributions of the selected stations from each optimal network. The numbers of the selected stations are 9 and 5 using IDW-DAR and 29 and 33 using MESH. In the IDW-DAR results, the selected stations are mostly located inland while they are mainly shown near the Lakeshore from the MESH results, where they may be able to provide more comprehensive information given that most rivers’ outlets are on Lakeshore. In addition, there were two locations in the optimal network using IDW-DAR where the selected stations were very closely located that may cause inefficiency of the network. In comparison, the selected stations of the optimal networks using the MESH model were in better agreement with deficit regions from the transinformation analysis (Figure 4). Conversely, the selected stations from the IDW-DAR optimal networks were mostly located in the southern half of the CGLB and did not contain any stations in the highly deficit areas around the north-eastern shores of Lake Superior (Figure 5).

By comparing the optimal networks obtained using MESH (Figure 4) and IDW-DAR (Figure 5), it appears that the use of MESH allows a better spatial distribution of stations selected by DEMO. Thus highlighting the importance of the method used to obtain the estimated flow at the potential station locations.

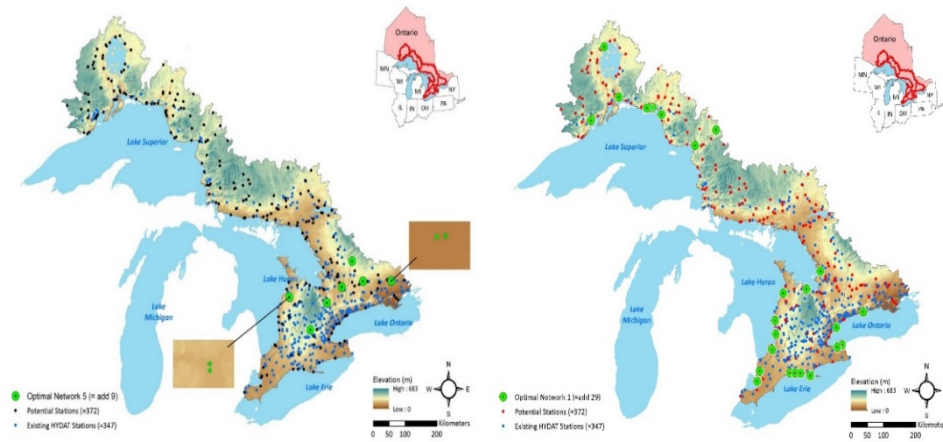


Figure 2 Optimal networks using IDW-DAR

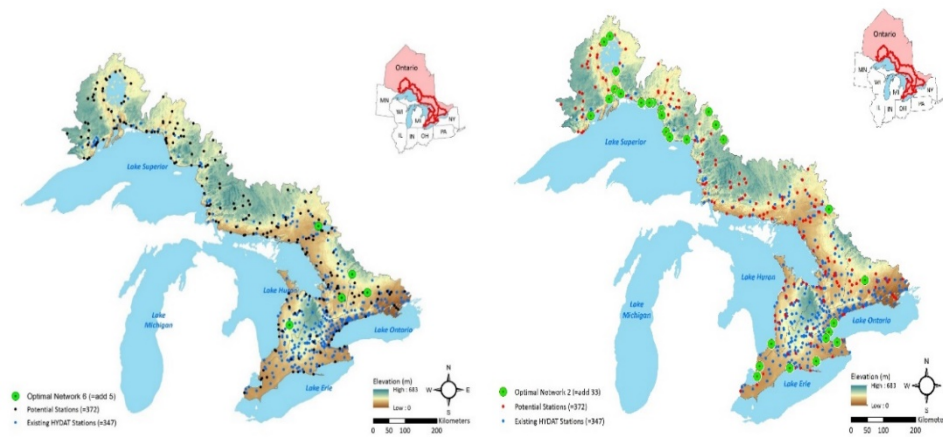


Figure 3 Optimal networks using MESH model

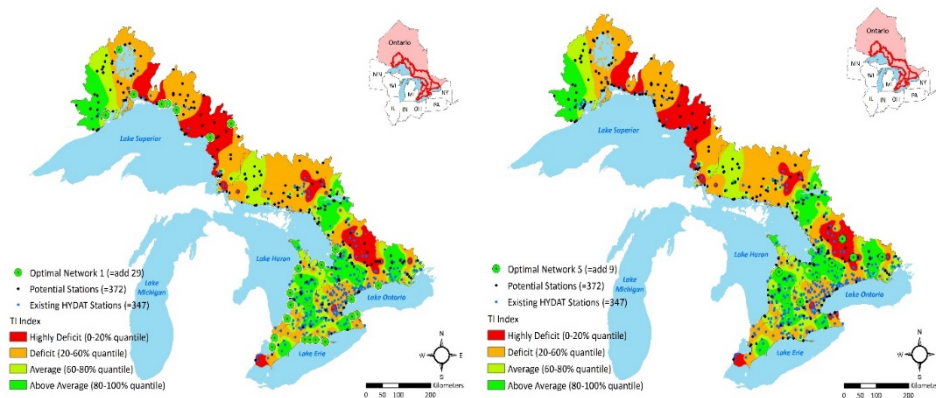


Figure 4 Optimal networks using MESH model on TI index map

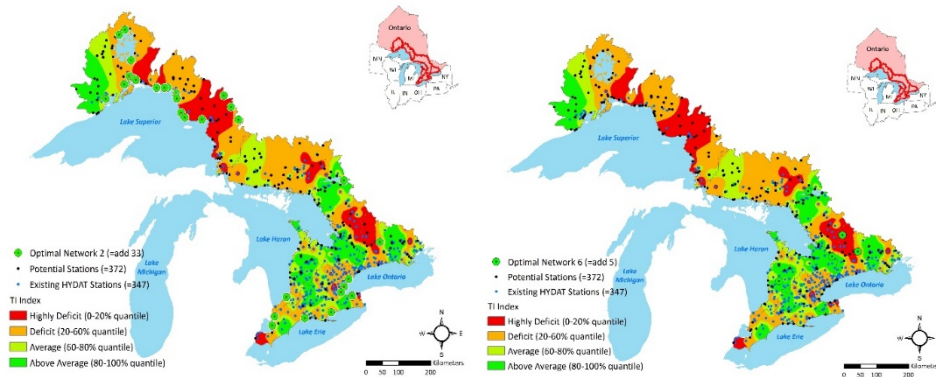


Figure 4 Optimal networks using IDW-DAR model on TI index map

## 5 Conclusions

This study evaluated the existing streamflow network in CGLB and designed optimal networks using simulated streamflow from IDW-DAR and the MESH model at potential station locations. Transinormation analysis results highlighted the critical regions in the northern and the southern parts of CGLB where monitoring efforts are the most needed. The IDW-DAR has been widely used for hydrometric network design problems. However, we found that an advanced spatially distributed hydrologic model, such as MESH, can be more informative after comparing the optimal networks from DEMO and the transinormation analysis map. While the regionalization methods may suffer from spatial correlations due to the interpolation methods, the MESH model was better able to provide streamflow estimations for a large basin, such as CGLB. Therefore, a spatially distributed hydrologic model is recommended to replace the typical regionalization method often used with DEMO.

## Acknowledgment

This work was jointly supported by the Natural Science and Engineering Research Council (NSERC) through the NSERC Canadian FloodNet grant and by Environment and Climate Change Canada.

## References

- [1] Mishra, A. K.; Coulibaly, P. Developments in Hydrometric Network Design: A Review. *Rev. Geophys.* 47 (2009), RG2001, doi:10.1029/2007RG000243.
- [2] Shannon, C. E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* 27 (1948) 379–423, doi:10.1002/j.1538-7305.1948.tb01338.x.
- [3] Yeh, H. C.; Chen, Y. C.; Wei, C.; Chen, R. H. Entropy and kriging approach to rainfall network design. *Paddy Water Environ.* 9 (2011) 343–355, doi:10.1007/s10333-010-0247-x.
- [4] Keum, J.; Coulibaly, P. Information theory-based decision support system for integrated design of multi-variable hydrometric networks. *Water Resour. Res.* 53 (2017) 6239–6259, doi:10.1002/2016WR019981.
- [5] Keum, J.; Coulibaly, P. Sensitivity of Entropy Method to Time Series Length in Hydrometric Network Design. *J. Hydrol. Eng.* 22 (2017), 4017009, doi:10.1061/(ASCE)HE.1943-5584.0001508.
- [6] Samuel, J.; Coulibaly, P.; Kollat, J. B. CRDEMO: Combined Regionalization and Dual Entropy-Multiobjective Optimization for Hydrometric Network Design. *Water Resour. Res.* 49 (2013), 8070–8089, doi:10.1002/2013WR014058.
- [7] Pietroniro, A.; Fortin, V.; Kouwen, N.; Neal, C.; Turcotte, R.; Davison, B.; Versegny, D. L.; Soulis, E. D.; Caldwell, R.; Evora, N.; Pellerin, P. Development of the MESH modelling system for hydrological ensemble forecasting of the Laurentian Great Lakes at the regional scale. *Hydrol. Earth Syst. Sci.* 3 (2007) 1279–1294, doi:10.5194/hess-11-1279-2007.