

CITATION:

Liu, H., B. A. Tolson, J. R. Craig, and M. Shafii (2016), A priori discretization error metrics for distributed hydrologic modeling applications, *Journal of Hydrology*, 543, Part B, 873-891.

1        ***A priori* discretization error metrics for distributed hydrologic modeling applications**

2                                Hongli Liu<sup>a</sup>, Bryan A. Tolson<sup>a</sup>, James R. Craig<sup>a</sup> and Mahyar Shafii<sup>b</sup>

3                                <sup>a</sup> *Dept. of Civil and Environmental Engineering, University of Waterloo, Waterloo, ON, Canada*

4                                <sup>b</sup> *Dept. of Earth and Environmental Sciences, University of Waterloo, Waterloo, ON, Canada*

5

6

7    Corresponding author: Hongli Liu.

8    Tel.: +1 519 888 4567 x37876.

9    Postal address: Department of Civil and Environmental Engineering, University of Waterloo,

10   200 University Avenue West, Waterloo, ON, Canada. N2L 3G1

11

12   E-mail addresses:

13   [hongli.liu@uwaterloo.ca](mailto:hongli.liu@uwaterloo.ca) (H. Liu)

14   [btolson@uwaterloo.ca](mailto:btolson@uwaterloo.ca) (B. A. Tolson)

15   [jrcraig@uwaterloo.ca](mailto:jrcraig@uwaterloo.ca) (J. R. Craig)

16   [mshafiih@uwaterloo.ca](mailto:mshafiih@uwaterloo.ca) (M. Shafii).

*The final publication is available at Elsevier via <http://dx.doi.org/10.1016/j.jhydrol.2016.11.008> © 2016. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/> is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>*

17

**Abstract:**

18 Watershed spatial discretization is an important step in developing a distributed hydrologic  
19 model. A key difficulty in the spatial discretization process is maintaining a balance between the  
20 aggregation-induced information loss and the increase in computational burden caused by the  
21 inclusion of additional computational units. Objective identification of an appropriate  
22 discretization scheme still remains a challenge, in part because of the lack of quantitative  
23 measures for assessing discretization quality, particularly prior to simulation. This study  
24 proposes *a priori* discretization error metrics to quantify the information loss of any candidate  
25 discretization scheme without having to run and calibrate a hydrologic model. These error  
26 metrics are applicable to multi-variable and multi-site discretization evaluation and provide  
27 directly interpretable information to the hydrologic modeler about discretization quality. The first  
28 metric, a subbasin error metric, quantifies the routing information loss from discretization, and  
29 the second, a hydrological response unit (HRU) error metric, improves upon existing *a priori*  
30 metrics by quantifying the information loss due to changes in land cover or soil type property  
31 aggregation. The metrics are straightforward to understand and easy to recode. Informed by the  
32 error metrics, a two-step discretization decision-making approach is proposed with the advantage  
33 of reducing extreme errors and meeting the user-specified discretization error targets. The  
34 metrics and decision-making approach are applied to the discretization of the Grand River  
35 watershed in Ontario, Canada. Results show that information loss increases as discretization gets  
36 coarser. Moreover, results help to explain the modeling difficulties associated with smaller  
37 upstream subbasins since the worst discretization errors and highest error variability appear in  
38 smaller upstream areas instead of larger downstream drainage areas. Hydrologic modeling  
39 experiments under candidate discretization schemes validate the strong correlation between the

40 proposed discretization error metrics and hydrologic simulation responses. Discretization  
41 decision-making results show that the common and convenient approach of making uniform  
42 discretization decisions across the watershed performs worse than the proposed non-uniform  
43 discretization approach in terms of preserving spatial heterogeneity under the same  
44 computational cost.

45

46 KEY WORDS: *a priori* spatial discretization error metrics; distributed hydrologic modeling;  
47 spatial heterogeneity; information loss; routing errors; discretization decision-making

48

## 49 **1. Introduction**

50 In distributed hydrologic modeling, a watershed is treated as a number of small homogeneous  
51 units to address the spatial heterogeneity which results from variability of physical processes and  
52 physical character across a watershed (Singh & Frevert, 2005). This spatial heterogeneity is often  
53 attributed to the uneven distribution of a hydrological properties across a watershed (Anselin,  
54 2010). The spatial discretization process, whereby we separate a watershed into homogeneous  
55 computational units for depiction in a hydrological model, is really the effort of determining how  
56 to characterize the inherent spatial heterogeneity found in a watershed. In general, spatial  
57 discretization should be detailed enough to capture the dominant processes and natural variability,  
58 while it also needs to be as concise as possible to save computation time and respect data  
59 availability (Booij, 2005). Excessively detailed spatial discretization increases model complexity  
60 (i.e., number of computational units) and thus increases model computation time. However, an  
61 overly coarse aggregation can lead to substantial information losses and give rise to increased  
62 model structural uncertainty, whose impacts on hydrological predictions are far more adverse

63 than those of parameter and data uncertainty (Liu & Gupta, 2007; Ludwig et al., 2009).  
64 Therefore, defining an appropriate level of discretization is a critical task in distributed  
65 hydrologic modeling.

66

67 In order to investigate spatial discretization, it is necessary to first clarify the components of  
68 watershed discretization. For this paper, we will be examining the common subbasin-HRU  
69 discretization approach. In this approach, a watershed is discretized into a set of one or more  
70 subbasins, which can be further discretized into a number of contiguous or non-contiguous  
71 hydrological response units (HRUs), defined as areas with hydrologically unique response to  
72 meteorologic events. Subbasins are referred to by different names in the literature, including grid  
73 cell, subcatchment, and subwatershed (Tuppad, 2006). Here we recursively define a subbasin as  
74 the drainage area of a location on a stream network minus the drainage areas of one or more  
75 upstream subbasins which flow directly into the subbasin. Headwater subbasins are those which  
76 do not have any subbasins upstream, i.e., those whose drainage areas are equal to their subbasin  
77 area. An HRU is the basic computational unit of hydrological simulation and typically defined as  
78 a unique combination of hydrological response determinants such as soil, land cover, terrain type,  
79 and management policy (Flügel, 1995), often generated from readily available mapping products.  
80 The HRU is conceptually similar to other computational units such as the Representative  
81 Elementary Area (REA), Representative Elementary Watershed (REW), Grouped Response Unit  
82 (GRU), hydro-landscape unit, and field (Dehotin & Braud, 2008; Fenicia, Kavetski, Savenije, &  
83 Pfister, 2016; N Kouwen, Soulis, Pietroniro, Donald, & Harrington, 1993; Reggiani, Sivapalan,  
84 & Hassanizadeh, 1998; Wood, Sivapalan, Beven, & Band, 1988), and therefore the approach  
85 developed here will port over to models which are discretized using these alternative definitions

86 of the smallest computational unit. In recent decades, the traditional approach for watershed  
87 discretization has been to use Geographic Information Systems (GISs) such as ESRI's ArcGIS  
88 software or ArcGIS-based toolkits such as Arc HYDRO, ArcSWAT, and HEC-GeoHMS (Doan,  
89 2000; ESRI, 2014; Maidment, 2002; Winchell, Srinivasan, Di Luzio, & Arnold, 2007). While  
90 such automatic techniques make watershed discretization easy to practically implement, they do  
91 not have an explicit mechanism to account for, or assess, spatial input data information losses  
92 due to discretization choices. Here, information loss refers to the content change between  
93 candidate discretization schemes and the original, fully detailed, input data layers. Instead,  
94 modelers can only explicitly assess the model complexity under candidate discretization schemes  
95 based on the number of modelled homogeneous areas (subbasin or HRU computational units).

96

97 Haghnegahdar et al. (2015) claim that most modelers make discretization decisions in an ad hoc  
98 fashion. This approach is often based on the past experience of the modeler, rules of thumb or  
99 default discretization settings in specialized ArcGIS-based toolkits for creating a distributed  
100 hydrologic model (e.g., ArcSWAT (Winchell, et al., 2007)). The shortcoming with all ad hoc  
101 approaches is that there is no quantitative or formal justification of the selected discretization  
102 over other potential discretization choices. More sophisticated discretization approaches found in  
103 the literature use a cumbersome trial-and-error approach of building and then possibly calibrating  
104 multiple candidate models with different discretization levels in order to identify the most  
105 appropriate choice. For example, Arnold et al. (2010) compared the calibration and validation  
106 period flow simulation results of an enhanced SWAT model with four landscape delineations,  
107 and Petrucci and Bonhomme (2014) tested the calibration and validation period water quantity  
108 and water quality simulation results of six different discretization scenarios of the Stormwater

109 Management Model. Haghnegahdar et al. (2015) followed a similarly intensive but improved  
110 process except that they took into account the computational time spent for calibrating  
111 (calibration budget) and focused on the model performance in ungauged basins under four  
112 discretization schemes for a land-surface hydrologic modelling application. All of these  
113 approaches require model calibration in order to assess the quality of a given discretization  
114 scheme.

115

116 Given the above limitations, other studies have instead focused on designing *a priori*  
117 discretization error metrics to quantify the information loss incurred from spatial discretization.  
118 Such metrics are advantageous in that they do not require model runs. Haverkamp et al. (2002)  
119 provided an entropy based statistical tool, the Subwatershed Spatial Analysis Tool (SUSAT), to  
120 estimate the information loss for subwatershed and HRU discretization, respectively. Booij  
121 (2003) utilized the bias of the variance of aerielly averaged variables under different correlation  
122 lengths to decide the appropriate modeling scale. Dehotin and Braud (2008) used Manhattan  
123 distance to measure the composition descriptor (e.g., histogram, mean, standard deviation, or  
124 matrix of co-occurrence) similarity between each mapping cell and the reference zones. There  
125 are three main shortcomings of the existing *a priori* discretization error metrics. First, the metrics  
126 do not directly correlate to the information required by hydrologic modeling applications, in  
127 particular for semi-distributed modeling. For example, entropy represents spatial disorder from  
128 the systematic perspective, but spatial heterogeneity essentially describes spatial pattern  
129 variability (Journel & Deutsch, 1993). Changes in system disorder cannot fully reflect the (more  
130 hydrologically important) changes in spatial heterogeneity and hence entropy is not a directly  
131 interpretable indicator for hydrologic modeling. Second, their property change identification

132 process fails to refer to the original spatial input data in a complete way (i.e., cell-by-cell  
133 comparison). Instead, they use the overall heterogeneity statistics difference between a candidate  
134 discretization scheme and the original spatial input data as the information loss, which may lead  
135 to the equifinality problem. Finally, the existing *a priori* approaches are all aggregated (e.g., over  
136 the entire study watershed) and do not provide spatially distributed evaluations of candidate  
137 discretizations. The importance of evaluating distributed model behaviors rather than an  
138 integrated value (e.g., runoff at the watershed outlet) for distributed models has been highlighted  
139 by numerous researchers (Beven & Binley, 1992; Grayson, Blöschl, Moore, & Singh, 1995;  
140 Refsgaard, 1997; Shrestha & Rode, 2008). Just like multi-site calibration provides an efficient  
141 framework for spatially distributed evaluations (Madsen, 2003), multi-site discretization quality  
142 assessment is intrinsically valuable to reduce the prevalence of aggregation or compensation  
143 effects in distributed hydrologic modeling. With such shortcomings in mind, this study is  
144 focused on developing *a priori* discretization error metrics that are directly interpretable,  
145 spatially distributed, and hydrologically relevant, providing a direct measurement of information  
146 loss relative to the original spatial input data, where the original spatial data is presumed to have  
147 the highest information content.

148

149 In addition to the information loss induced by the extensively studied HRU discretization,  
150 another type of information loss occurs due to subbasin discretization which affects the routing  
151 processes of semi-distributed and distributed models, hereinafter called routing information loss.  
152 In a finely discretized fully distributed model, channel structure, channel roughness, and  
153 therefore network travel times can be well-respected. As the watershed is discretized into  
154 subbasins, stream network branches are implicitly merged, replaced, and shortened. As far as we

155 know, in the published literature, the routing information loss has never been quantified though  
156 its significance has been highlighted by many studies. For example, Haverkamp (2002) indicates  
157 that the influences of the routing structure through subbasins to the watershed outlet should be  
158 considered in discretization evaluations when the effect of the routing on model results is not  
159 negligible. Dehotin and Braud (2008) emphasize the prospect of inclusion of linear  
160 discontinuities, including river reaches, hedges, ditches, and dikes, in order to properly describe  
161 networks in spatial discretization. Here, we address this need through the introduction of  
162 additional error metrics to estimate the routing information loss due to subbasin discretization.

163

164 The specific goals of this study are to (1) introduce *a priori* discretization error metrics to  
165 quantify the information loss due to subbasin and HRU discretization, respectively; (2) propose a  
166 two-step decision-making approach to identify an appropriate discretization scheme; (3) apply  
167 the error metrics and decision-making approach to the discretization of the Grand River  
168 watershed in Ontario, Canada. The simplicity of the error metrics allows for easy recoding and  
169 adoption into the preprocessing of a wide range of distributed models, including all semi-  
170 distributed models, such as HBV (Bergström, 1976, 1992), TOPMODEL (Beven & Kirkby,  
171 1979), WATFLOOD (Nicholas Kouwen, 1988), the Soil and Water Assessment Tool (SWAT) (J.  
172 G. Arnold, Srinivasan, Muttiah, & Williams, 1998), and Modélisation Environnementale–Surface  
173 et Hydrologie (MESH) (Pietroniro et al., 2007). The error metrics may also be useful for fully  
174 distributed models, e.g., System Hydrologique Europeen (SHE) (Abbott, Bathurst, Cunge,  
175 O'Connell, & Rasmussen, 1986), TOPKAPI (Ciarapica & Todini, 2002), and Soil Moisture  
176 Distributed and Routing (SMDR) (Srinivasan, Gérard-Marchant, Veith, Gburek, & Steenhuis,  
177 2005) when the model cell scales are greater than the resolution of original spatial input data.



178

179 The reminder of the paper is organized as follows. Section 2 describes in detail the *a priori*  
180 discretization error metrics and the two-step discretization decision-making approach. Section 3  
181 explains the error metric applications to the Grand River watershed discretization. Section 4  
182 provides an effective discussion of the proposed methods. Section 5 summarizes conclusions.

183

## 184 **2. Methodology**

### 185 **2.1. Discretization Error Metrics**

186 Our *a priori* discretization error metrics provide a novel and simple quantitative measurement of  
187 the information loss in the process of spatial discretization. They are introduced for the purpose  
188 of assessing candidate discretization schemes and finding an appropriate discretization level in  
189 data preprocessing without having to rely on computationally intensive hydrologic model  
190 building exercises. For each candidate discretization scheme, the metrics are designed to  
191 compare the user-defined key model input variable properties with that of a reference  
192 discretization scheme. The reference scheme is defined as a scheme that fully retains the  
193 information of the original spatial input data or, in special cases, the finest plausible  
194 discretization. Both a subbasin discretization error metric and a HRU discretization error metric  
195 are proposed.

#### 196 **2.1.1. Subbasin Discretization Error Metric**

197 In general, the routing process has two components: in-catchment routing and in-channel routing.  
198 In-catchment routing occurs within a subbasin, and refers to the means of handling the delayed  
199 release of water from runoff, interflow, and baseflow to a subbasin outlet. This time delay is

200 typically described by a unit hydrograph. In contrast, in-channel routing is the means by which  
 201 water is exchanged downstream between subbasins and within the main channel of each  
 202 subbasin. These definitions are applied by other models like ArcSWAT and HEC-GeoHMS  
 203 (Doan, 2000; Winchell, et al., 2007). Our subbasin discretization assessment focuses on the  
 204 influences of discretization only on in-channel routing. The approach assumes that in-channel  
 205 routing is unidirectional (i.e., water moves downstream only through a branching stream  
 206 network), each subbasin has one outlet and one main channel, headwater subbasins have no main  
 207 channel for routing, and non-headwater subbasins have upstream subbasin flows added to the  
 208 beginning of their respective main channels. Should any of these assumptions not hold in other  
 209 modelling case studies, the error metric procedures detailed below would need to be adjusted  
 210 accordingly.

211  
 212 Calculation of subbasin discretization errors requires a high resolution reference subbasin  
 213 discretization scheme. For the drainage area upstream of a subbasin outlet, the in-channel routing  
 214 length error ( $\Delta L_s$ ) equals to the in-channel routing length difference between the reference  
 215 scheme (scheme 0) and the evaluated discretization (scheme  $s$ ) as shown in Equation 1.

$$216 \quad \Delta L_s = L_0 - L_s = \frac{\sum_{i=1}^n A_{i0} L_{i0}}{\sum_{i=1}^n A_{i0}} - \frac{\sum_{j=1}^m A_{js} L_{js}}{\sum_{j=1}^m A_{js}} \quad (1)$$

217 where  $L_0$  and  $L_s$  are respectively the area-weighted in-channel routing length of scheme 0 and  
 218 scheme  $s$ . For scheme 0, there are  $n$  subbasins within the evaluated drainage area and  $i =$   
 219  $1, 2, \dots, n$  represents subbasin indices.  $A_{i0}$  is the area of subbasin  $i$  in scheme 0, and  $\frac{\sum_{i=1}^n A_{i0} L_{i0}}{\sum_{i=1}^n A_{i0}}$  is  
 220 the area-weighted sum of the in-channel routing length of subbasin  $i$  from the subbasin  $i$  outlet to  
 221 the drainage area outlet of interest. For scheme  $s$ , there are  $m$  subbasins within the evaluated

222 drainage area and  $j = 1, 2, \dots, m$  represents subbasin indices.  $A_{js}$  is the area of subbasin  $j$  in  
223 scheme  $s$ , and  $\frac{\sum_{j=1}^m A_{js} L_{js}}{\sum_{j=1}^m A_{js}}$  is the area-weighted sum of the in-channel routing length of subbasin  $j$   
224 from the subbasin  $j$  outlet to the drainage area outlet of interest. The total area of the drainage  
225 area is  $A = \sum_{i=1}^n A_{i0} = \sum_{j=1}^m A_{js}$ .

226

227 The calculation of the in-channel routing length difference between schemes is best described in  
228 Figure 1 below with a visual example. The example in Figure 1 demonstrates the in-channel  
229 routing length difference ( $\Delta L_s$ ) between scheme 0 and scheme  $s$  as the difference in the thick  
230 routing arrows between the two discretization options. For example, in scheme 0, flows from  
231 headwater subbasins 1, 2 and 3 are all routed in the main channel of subbasin 7 for 2 km. In  
232 comparison, with the coarser discretization scheme  $s$ , the flows from this region of the watershed  
233 (subbasins 1, 2 and 3 in scheme 0) are no longer routed in-channel for this distance and thus  
234 treated as a discretization error. A similar error occurs for the subarea including subbasins 4, 5  
235 and 6. In our metric, in-channel routing length errors are computed for subbasin outlets of  
236 interest and in this example, the ‘outlet’ is the site of interest in Figure 1. If all flows reaching the  
237 outlet had a 2 km shorter in-channel routing length in scheme  $s$  versus scheme 0, then  $\Delta L_s$  would  
238 be 2 km at the outlet. This is not typically the case and so the representative change in routing  
239 length,  $\Delta L_s$ , must account for this using area-weighting.

240

241

Figure 1. Here.

242

243

244

### 245 2.1.2. HRU Discretization Error Metric

246 As explained before, the information loss from spatial discretization is due to the diminished  
247 representation of spatial data content between a candidate discretization scheme and the original,  
248 fully detailed, input data layer. To quantify the relevant (case study specific) information loss  
249 derived from HRU discretization, the dominant hydrologic processes should first be identified by  
250 considering the modeling purpose, physiographic characteristics and management measures  
251 within the watershed. These dominant processes can be linked to dominant hydrologic model  
252 input variables derived from map inputs which will be used to evaluate information losses. For  
253 example, in rainfall-runoff modeling, if infiltration is identified as a critical process then the most  
254 relevant variables to compute information losses for can be hydraulic conductivity and/or  
255 available water content.

256  
257 For a drainage area above an outlet, assume there are  $n$  HRUs in the reference scheme (scheme  
258 0), and  $m$  HRUs in the evaluated discretization (scheme  $s$ ), and thus  $n \geq m$ . In order to  
259 effectively consider the spatial pattern changes between the two schemes, the evaluated scheme  
260 layer needs to be overlaid with the reference scheme layer using vector overlay tools (e.g., union)  
261 for vector maps or raster overlay tools (i.e., weighted overlay) for raster maps in ArcGIS (ESRI,  
262 2014). After overlay, each polygon or cell of the output possesses both the evaluated and  
263 reference scheme HRU properties. Assume there are  $v$  polygons (cells) ( $u = 1, \dots, v$ ) of the  
264 output. HRU discretization error metrics are designed to go through each polygon (cell) and  
265 measure the relative error of variable change between scheme  $s$  and scheme 0. Two different *a*  
266 *priori* discretization error metrics corresponding to nominal (categorized) and quantitative  
267 (continuous) data are developed.

268

269 Figure 2 shows an example of the overlay comparison process required for computing HRU  
270 discretization errors for an example subbasin, corresponding to subbasin 1 of scheme 0 in Figure  
271 1, and uses a nominal variable (land cover) as an example. In scheme 0, there are four different  
272 land covers scattered over the entire subbasin (Figure 2a), but only two land covers remain in the  
273 coarser scheme  $s$  (Figure 2b). After overlay and property comparison, four cells show a property  
274 change as highlighted in Figure 2c, in which one cell of coniferous forest turns into deciduous  
275 forest, and one cell of coniferous forest and two cells of pasture turn into crop. The information  
276 loss due to recategorization is considered as a discretization error, expressed in terms of  
277 recategorized area (i.e., 4 km<sup>2</sup> in this example). For quantitative variables, the only difference is  
278 the absolute values of the property changes are utilized as shown in the following equations.

279

280

Figure 2. Here.

281

282 For nominal input variables (e.g., soil and land cover), the relative error equals to the sum of  
283 areas with property change from scheme 0 to scheme  $s$  divided by the total drainage area.

284 
$$\delta_{HRUs} = \frac{\sum_{u=1}^v \Delta_u A_u}{\sum_{u=1}^v A_u} \quad (2)$$

285 
$$\Delta_u = \begin{cases} 0, & \text{if variable property is unchanged relative to the reference scheme} \\ 1, & \text{if variable property is changed relative to the reference scheme} \end{cases} \quad (3)$$

286 where  $\delta_{HRUs}$  is relative error (0-1) of the evaluated scheme  $s$  describing the proportion of the  
287 drainage area where the variable property is changed and thus incorrect relative to the original  
288 spatial data.  $A_u$  is the area of the  $u^{th}$  polygon (cell) of the overlay output.

289

290 For quantitative input variables (e.g., hydraulic conductivity and available water content), the  
 291 relative error equals to the area-weighted sum of the absolute values of input variable differences  
 292 of all polygons (cells) between scheme  $s$  and scheme 0 divided by the area-weighted mean input  
 293 variable value of scheme 0 within the drainage area. It is expressed as:

$$294 \quad \delta_{HRUs} = \frac{\sum_{u=1}^v w_u |x_{us} - x_{u0}|}{\sum_{i=1}^n w_i x_{i0}} \quad (4)$$

295 where  $\delta_{HRUs}$  is relative error (0-1) of the evaluated scheme  $s$  indicating the level of absolute  
 296 input variable value change relative to the mean value of scheme 0.  $x_{us}$  and  $x_{u0}$  are the input  
 297 variable values of the  $u^{th}$  polygon (cell) in scheme  $s$  and scheme 0, respectively.  $w_u$  is the area  
 298 weight of the  $u^{th}$  polygon (cell) of the total drainage area. It is calculated by:

$$299 \quad w_u = \frac{A_u}{\sum_{u=1}^v A_u} \quad (5)$$

300 where  $\sum_{u=1}^v A_u$  is the total area of the evaluated drainage area.

301  
 302 The absolute value operation utilized in Equation (4) is to properly track all spatial heterogeneity  
 303 changes once the input variable property differs from the original spatial input data. In other  
 304 words, compensation effects (two errors cancelling each other) are not allowed.

305

## 306 **2.2. Sensitivity of Hydrologic Model Simulation Results to Discretization Error Metrics**

307 To validate the impact of the *a priori* error metrics on hydrologic model simulation results,  
 308 multiple hydrologic models were built (one for each candidate discretization scheme). The only  
 309 difference between these models exists in discretization. We chose to build our simulation  
 310 models for different discretization levels in the Raven hydrological modeling framework (Craig  
 311 et al., 2016). All the models are semi-distributed with two buckets and simulate water transfer

312 between soil (upper and lower layers) and atmosphere through a series of hydrologic processes.  
313 The models simulate on an hourly time step and in-channel routing is based on a non-linear level  
314 pool routing approach using Manning's equation. Specific details of the hydrologic model are  
315 provided in Appendix A.1 of this paper.

316

317 Similar to discretization error metrics, hydrologic simulation results are assessed relative to a  
318 reference simulation result. The reference simulation result corresponds to the model using the  
319 reference discretization scheme (scheme 0). All other model simulation results are compared  
320 relative to the reference result using error indices such as the peak flow rate error, the peak flow  
321 timing error, and the cumulative flow volume error. The peak flow rate error is computed as the  
322 absolute peak flow rate difference between scheme  $s$  and scheme 0 divided by the peak flow rate  
323 of scheme 0. The peak flow timing error is the time of peak flow occurrence with scheme 0  
324 minus the time of peak flow occurrence with scheme  $s$ . The cumulative flow volume error is the  
325 absolute cumulative flow volume difference between scheme  $s$  and scheme 0 divided by the  
326 cumulative flow volume of scheme 0. Non-zero values for these indices are the direct result of  
327 different discretization choices.

328

329 The relationship between discretization errors and model errors is estimated by the Spearman's  
330 rank correlation coefficient ( $r_s$ ) which ranges from -1 to +1. The objective of this analysis is to  
331 validate that changes in our proposed error metrics indeed impact hydrologic model simulation  
332 results. Note that our analysis necessarily avoids the issue of model calibration and validation  
333 decisions confounding the analysis. A future larger scale, multi-basin study would be required to

334 properly validate the role discretization errors have in terms of their net impact on model  
335 predictive accuracy.

336

### 337 **2.3. Discretization Decision-Making Approach**

338 We demonstrate one of many ways modelers can utilize the proposed *a priori* error metrics by  
339 using them within a structured two-step approach to watershed discretization decision-making.  
340 The two-step approach is applicable to both subbasin and HRU discretization decisions and is  
341 described in the following two sections.

#### 342 **2.3.1. Subbasin Discretization Decision-Making Approach**

343 Step 1: Select a subbasin scheme from candidate discretization schemes (Candidacy step).  
344 Candidate subbasin schemes would typically first be generated by placing subbasin outlets at the  
345 sites of interest within the watershed (e.g., gauge stations and/or reservoirs) and at stream  
346 junctions, with subbasin boundaries determined using standard terrain analysis algorithms. The  
347 subbasin boundaries will vary depending on stream network resolution. Here, we generate the  
348 stream network and junctions based on a flow accumulation threshold as done in ArcSWAT  
349 (Winchell, et al., 2007). Other approaches to junction generation could be used, for example,  
350 truncating the stream network based upon Strahler stream order. The relationships between the  
351 flow accumulation threshold and the coarseness of the stream network are monotonic – as the  
352 accumulation threshold increases, stream network becomes less detailed and fewer subbasins are  
353 included. In this step, typically users should vary the spatially consistent flow accumulation  
354 threshold (uniformly applied for the entire watershed) and assess the resulting routing length  
355 errors.



356

357 A routing length error threshold (referred to as the preliminary error threshold) is then specified  
358 to select a subbasin scheme from candidates. The selected scheme meets the criteria that all sites  
359 of interest satisfy the preliminary error threshold at the minimum discretization complexity cost  
360 (i.e., the total number of subbasins) among all candidate schemes. Setting the preliminary error  
361 threshold to a very large value would function to select the most coarsely defined candidate  
362 scheme among the candidates.

363

364 Step 2: Refine subbasin discretization for the areas with extreme discretization errors (Polishing  
365 step).

366 This step is used to refine the candidate subbasin discretization selected in Step 1 for the areas  
367 with the most extreme discretization errors. It can also be used to focus on minimizing  
368 discretization errors at modeler-specified critical sites of interest where smaller discretization  
369 errors are desired for some reason. Functionally speaking, this step is optional. If utilized, this  
370 step involves specifying a second, stricter routing length error threshold (referred to as extreme  
371 error threshold) and requires the stream junction locations of other finer resolution candidate  
372 schemes. Given a subbasin scheme from *Step 1*, the complete process of *Step 2* is demonstrated  
373 by Figure 3.

374

375 Figure 3. Here.

376

377 Step 2a: Identify the sites of interest with extreme discretization errors.

378 Sites of interest with discretization errors not satisfying the extreme error threshold are identified.

379 These sites are referred to as extreme sites.

380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
  
394  
395  
396  
397  
398  
399  
400  
401

Step 2b: Replace junctions in the upstream refinement areas of extreme sites with those of the nearest, more detailed satisfactory discretization scheme.

There are three cases in identifying the upstream refinement area for each extreme site:

- Case 1. If the extreme site has no upstream sites of interest, increased resolution of the stream network is applied to the whole drainage area above the extreme site.
- Case 2. If the extreme site has satisfactory (non-extreme) upstream site(s) of interest, increased network resolution is only applied to the intermediate area between upstream site(s) of interest and the extreme site.
- Case 3. If the extreme site has an upstream extreme site(s), the network is not refined. What will happen in this case is that the upstream extreme site(s) will first be refined (under Case 1) and then in a future discretization refinement iteration, the intermediate area(s) will only be refined if the new discretization error(s) for the site in question remains extreme (the extreme site will be re-categorized into Case 2).

Once the upstream refinement area is determined, replace the junctions within it with those of the nearest more detailed satisfactory scheme. More detailed alternate candidate schemes would typically be available from the candidacy selection step (Step 1) but if not, the modeler would be required to generate one or more detailed schemes (e.g., by decreasing the flow accumulation threshold). It is worth explaining the reason why there is no need to replace junctions for the extreme sites of Case 3. In Case 3, the influence of the upstream refinement on the downstream error metric result is unclear unless the new errors are recalculated. If the extreme site located downstream can take the advantage of upstream refinement and obtain a satisfactory error result

402 without junction replacements, this will be the most cost-effective solution in terms of model  
403 complexity.

404 Step2c: Re-discretize subbasins and re-calculate errors for the sites of interest.

405 In order to get the systematic upstream-downstream flow path relation among subbasins, re-  
406 discretize the watershed with the updated junctions and re-calculate the error metric results. The  
407 detailed re-discretization processes are provided in the Appendix A.2 of this paper.

- 408 • If the new error metric results in all the previously extreme sites are satisfactory (less  
409 than the extreme error threshold), adopt these junctions. *Step 2* ends.
- 410 • If some extreme sites do not satisfy the extreme error threshold, return to *Step 2b*.  
411 Iterate *Step 2b* and *Step 2c* until all the extreme sites are satisfactory.

412

413 Because the polishing step introduces non-uniformity to the discretization scheme (i.e., the  
414 refined areas have finer subbasin discretization than the non-refined areas), we refer to this  
415 discretization scheme as a non-uniform scheme.

416

### 417 **2.3.2. HRU Discretization Decision-Making Approach**

418 Similar to subbasin discretization decision-making, modelers can also choose an appropriate  
419 HRU discretization following the two-step decision-making approach outlined in Section 2.3.1.  
420 *Step 1* is selecting a uniform HRU scheme from candidates based on some predefined uniform  
421 HRU discretization preliminary error threshold(s). As with subbasin discretization, the candidate  
422 HRU discretization schemes should each be based on some uniform level of detail across the  
423 watershed. As described in Section 2.3.1, we identified candidate HRU schemes by varying an

424 HRU size threshold, below which the small HRUs in that subbasin are merged and replaced with  
425 more dominant HRU types. Again, the relationship between this size threshold and the model  
426 complexity is monotonic. Unlike the subbasin discretization step, there may be multiple  
427 hydrologic model input variables for which a modeler wishes to compute HRU discretization  
428 errors. In this case, the metric results of multiple input variables can be treated equally or  
429 assigned different weights based on their importance in decision-making.

430

431 *Step 2* is polishing HRU discretization. The only difference from subbasin discretization  
432 refinement is that, in *Step 2b*, HRUs can be directly replaced without junction replacement. *Step*  
433 *2c* simply involves merging all resultant HRUs into an output layer and re-calculating errors for  
434 the sites of interest.

435

### 436 **3. Results of Discretization Error Metrics Application**

437 This study is conducted in the Grand River watershed in southwestern Ontario, Canada. With  
438 drainage area of 6704 km<sup>2</sup>, the Grand River flows south to Lake Erie and is mainly covered by  
439 agricultural land. The applications are presented in two sections. Section 3.1 shows the  
440 application of the subbasin discretization error metric, and Section 3.2 shows the application of  
441 the HRU discretization error metric.

442

### 443 3.1. Subbasin Discretization Error Metric Application

#### 444 3.1.1. Candidate Subbasin Discretization Schemes

445 In this study, subbasins were represented in subwatershed format and derived from  $10m \times 10m$   
446 digital elevation model (DEM) data. Subbasins were discretized based on the ArcSWAT  
447 (Winchell, et al., 2007) flow accumulation threshold approach as described in Section 2.3.1.  
448 Research shows that, reducing the flow accumulation threshold below 0.5% of the maximum  
449 flow accumulation doesn't improve model performance but complicates remaining preprocessing,  
450 whereas increasing it significantly above 1% might lead to performance ramifications (Djokic,  
451 2008). According to these findings, we took the percentage of the maximum flow accumulation  
452 across the entire watershed as the subbasin discretization threshold and treated 0.5% as the  
453 minimum flow accumulation threshold value. Therefore, twelve candidate subbasin schemes  
454 were generated corresponding to twelve successively increasing flow accumulation thresholds.  
455 The detailed subbasin discretization results are listed in Table 1.

456

457 Scheme 0 was defined as the reference scheme because subbasin discretization with threshold  
458 0.5% is the finest scheme of all the candidates and we assume the channel information loss  
459 between the real full channel scheme (i.e., one channel for each DEM cell) and scheme 0 is  
460 irremediable. Scheme Max only used the 32 sites of interest as subbasin outlets. The 32 sites  
461 include 24 gauge stations, 7 dams, and the watershed outlet, and their detailed information has  
462 been listed in Table 2.

463

464

**Table 1.** Candidate subbasin discretization schemes

Scheme	Flow accumulation threshold (%)	Number of subbasins
0	0.5	130
1	0.6	110
2	0.7	100
3	0.8	94
4	0.9	92
5	1.0	90
6	2.0	60
7	3.0	46
8	5.0	44
9	6.0	40
10	10.0	38
Max	Only sites of interest	32

465

466

**Table 2.** Details of 32 sites of interest and their drainage areas

Site of interest	Site name	Drainage area (km <sup>2</sup> )	Site of interest	Site name	Drainage area (km <sup>2</sup> )
1	02GA041	66	17	02GA015	565
2	Luther Dam	45	18	02GA038	313
3	02GA014	654	19	Laurel Creek Dam	31
4	02GA039	272	20	02GA024	59
5	Shand Dam	775	21	02GA047	757
6	02GA016	776	22	02GA048	2477
7	Conestogo Dam	559	23	Shades Mill Dam	96
8	02GA028	564	24	02GA018	536
9	02GA040	178	25	02GA003	3490
10	Woolwich Dam	60	26	02GA010	1028
11	Guelph Dam	241	27	02GB006	157
12	02GA034	1148	28	02GB007	384
13	02GA031	40	29	02GB001	4784
14	02GA023	113	30	02GB008	378
15	02GA029	226	31	02GB010	170
16	02GA006	769	32	Watershed outlet	6704

467

468 In this study, we consider all 32 sites of interest as locations where the discretization error

469 metrics will be assessed. Each drainage area is the combined total upstream area draining to the

470 site as illustrated in Figure 4. For instance, drainage area 3 is defined to include subbasins 1, 2

471 and 3.

472

473

Figure 4. Here.

474

### 475 **3.1.2. Subbasin Discretization Error Metric Results**

476 In distributed hydrologic modeling applications, most of the time modelers will only pay  
477 attention to the information loss at the sites of interest. Moreover, it is unnecessary to analyze  
478 error metric results for the sites above which candidate subbasin discretizations are always as  
479 fine as the reference one because in this situation the error metric result is always zero. As a  
480 result, we limited the subbasin error metric results analysis to the 32 sites as introduced in  
481 Section 3.1 and then excluded the 13 sites whose upstream subbasins do not change from scheme  
482 0 to scheme Max. The remaining 19 sites for analysis are sites 3, 5, 6, 7, 8, 9, 11, 12, 16, 17, 18,  
483 21, 22, 24, 25, 26, 28, 29, 32, and their error metric results for the twelve discretization schemes  
484 were computed. For brevity, only the results from nine representative subbasin schemes are  
485 shown in Figure 5.

486

487

Figure 5. Here.

488

489 In each subplot of Figure 5, the routing length errors of the 19 sites are plotted versus their  
490 drainage areas. Figure 5(a) shows that when discretization is detailed at the reference scheme  
491 level, no error exists. Then in Figure 5(b-i), as subbasin discretization gets coarser, the number of  
492 subbasins within a drainage area decreases, and the routing length error increases. This is  
493 reflected by the ranges of error values of Figure 5(b-i). Moreover, in each subplot, the  
494 downstream sites with the largest drainage areas typically have intermediate error values rather  
495 than the maximum value of all the errors at the 19 sites of interest. For example, moving

496 downstream in the Grand River watershed, site 22 (drainage area 2477 km<sup>2</sup>), site 25 (drainage  
497 area 3490 km<sup>2</sup>), and site 29 (drainage area 4784 km<sup>2</sup>) all get intermediate error values for all the  
498 subbasin schemes. This trend can be explained by the fact that, for in-channel routing, the  
499 downstream error integrates all its upstream errors in an area-weighted fashion (see Equation 1),  
500 so the drainage area outlet is not necessarily the point that has the largest information loss. This  
501 implies that if modelers are concerned about the multi-site discretization quality or the multi-site  
502 hydrologic model performance, multiple sites rather than just the watershed outlet are worth  
503 considering in subbasin discretization evaluation.

504

### 505 **3.1.3. Sensitivity of Hydrologic Model Simulation Results to Subbasin Discretization**

#### 506 **Error Metric**

507 To access the sensitivity of model simulation results to the proposed subbasin error metric, we  
508 built twelve hydrologic models corresponding to all the subbasin schemes of Table 1, in which  
509 their only difference is subbasin discretization and the connectivity between subbasins. We  
510 focused the analysis on a short period (Jan 4 – Jan 20, 2008) of peak or near peak measured  
511 flows over the last ~15 year period across the Grand River watershed. The reference simulation  
512 result corresponds to the model using the reference discretization scheme (scheme 0 of Table 1)  
513 and all simulation model results were compared relative to the reference result using the peak  
514 flow rate error and peak flow timing error.

515

516 Taking the watershed outlet as an example, Figure 6 summarizes the relationship between the *a*  
517 *priori* routing length error metric and the hydrologic model error indices where each data  
518 corresponds to one of the eleven candidate subbasin discretization schemes. As the routing



519 length error increases, both model error indices increase (almost monotonically) to practically  
520 significant levels. Correlation ( $r_s$ ) between the routing length error and the peak flow rate error  
521 is 0.99, and correlation ( $r_s$ ) between the routing length error and the peak flow timing error is  
522 also 0.99. This strong correlation is observed for the majority of sites of interest (e.g.,  
523 considering the correlation between the routing length error and the peak flow rate error, 15 sites  
524 show  $r_s$  values of 0.8 or more).

525

526 Figure 6. Here.

527

#### 528 **3.1.4. Subbasin Discretization Decision-making**

529 Based on the error metric results of all candidate subbasin discretization schemes, we applied the  
530 two-step decision-making approach to get an appropriate subbasin discretization scheme. It was  
531 assumed that all of the 19 sites of interest are equally important, and 21 km is selected as the  
532 preliminary routing length error threshold. The subjective value of 21 km was selected for  
533 demonstration purposes and based on balancing travel time error implications (assuming a  
534 reference velocity of 1 m/s) and computational complexity (limiting number of subbasins).

535 Step 1: Select a subbasin scheme from candidate discretization schemes

536 Scheme 6 (number of subbasins=60) was chosen as the uniform threshold subbasin scheme  
537 because the error metric values of all the 19 sites of scheme 6 are satisfactory (less than 21  
538 km) and the number of subbasins is the minimum among all the satisfactory schemes  
539 (schemes 1-6).

540 Step 2: Refine subbasin discretization for the areas with extreme discretization errors

- 541 • Step 2a: Identify the sites of interest with extreme discretization errors (extreme sites).  
542 The extreme error threshold was 10.7 km, defined as the 90<sup>th</sup> percentile of the error  
543 distribution of scheme 6, and the resultant extreme sites that have the highest 10% errors  
544 were sites 26 and 28, which are highlighted in Figure 7a and Figure 7b.
- 545 • Step 2b: Replace junctions in the upstream refinement areas of extreme sites with those of  
546 the nearest, more detailed satisfactory discretization scheme. Specifically, different sites have  
547 different upstream refinement areas:
- 548 Case 1: Site 28 has no upstream sites of interest, thus junction replacement is applicable  
549 to the whole drainage area above site 28. Since the error of site 28 in scheme 5 is 4.2 km  
550 (less than 10.7 km), scheme 5 is the nearest satisfactory scheme compared with scheme 6.
- 551 Case 2: Site 26 has a satisfactory upstream site of interest, site 24, so junction  
552 replacement only takes place in the intermediate area between the site 24 and site 26.  
553 Since the error of site 26 in scheme 5 is 5.0 km (less than 10.7 km), scheme 5 is also the  
554 nearest satisfactory scheme relative to scheme 6.
- 555 • Step2c: Re-discretize subbasins and re-calculate errors for the sites of interest.  
556 After re-discretization, the subbasin compositions within the upstream refinement areas were  
557 changed to the new more detailed subbasins as shown in Figure 7c. Meanwhile, the total  
558 number of subbasins for the entire Grand River watershed increased from 60 to 66. The  
559 routing length errors of sites 26 and 28 became satisfactory (less than 10.7 km as shown in  
560 Table 3).

561

562

Figure 7. Here.

563

564 Table 3 shows the routing length errors of scheme 6, refined scheme 6, and scheme\*. Scheme\*  
565 has the same number of subbasins as refined scheme 6 but was generated with a uniform flow  
566 accumulation threshold of 1.55%. In addition to the purposeful reduction of routing errors at the  
567 two extreme sites, Table 3 also shows the substantial error decrease of all the associated  
568 downstream sites (e.g., sites 29 and 32) in refined scheme 6. Moreover, comparing refined  
569 scheme 6 with scheme\*, the error mean and standard deviation of refined scheme 6 are lower  
570 than those of scheme\*. This indicates that the refined subbasin discretization better represents the  
571 in-channel routing structure than the uniform discretization under the same number of  
572 computational (subbasin) units.

573

574 **Table 3.** Subbasin discretization error metric results for three subbasin discretization schemes.  
575 Scheme 6 is based on a flow accumulation threshold of 2.0%, while Scheme\* is based on a  
576 threshold of 1.55%. Sites of interest that are discretized the same way under all three schemes  
577 are not included. Highlighted errors for refined scheme 6 are lower than corresponding errors in  
578 one or both of Scheme 6 and Scheme\*. Note that site 32 corresponds to the watershed outlet.

Site of interest	Scheme 6		Refined scheme 6		Scheme*	
	Number of subbasins	Error (km)	Number of subbasins	Error (km)	Number of subbasins	Error (km)
7	4	4.0	4	4.0	6	0.9
8	5	4.0	5	4.0	7	0.9
16	6	6.4	6	6.4	2	3.9
17	7	1.7	7	1.7	8	4.1
18	3	0.0	3	0.0	7	1.7
22	25	5.9	25	5.9	27	5.2
25	37	4.7	37	4.7	39	4.2
26 <sup>e</sup>	5	20.7	9	5.0	5	20.7
28 <sup>e</sup>	1	20.4	3	4.2	3	4.2
29	47	8.0	51	4.6	49	7.6
32	60	8.5	66	5.2	66	7.0
<b>Error mean</b>		7.7		4.2		5.5
<b>Error St. deviation</b>		6.8		1.8		5.5
<b>Error 90<sup>th</sup> percentile<sup>f</sup></b>		10.7 <sup>f</sup>				

579 <sup>e</sup> denotes an extreme site under scheme 6 based on exceeding the 90<sup>th</sup> percentiles of the error  
580 metric. The subbasin discretization within this site's drainage area is refined based on *Step2*.

581 <sup>f</sup> The 90<sup>th</sup> percentile computed based on errors across all 19 sites considered (see Section 3.1.2).

## 582 3.2. HRU Discretization Error Metric Application

### 583 3.2.1. Candidate HRU Discretization Schemes

584 In this study, HRU is discretized after subbasin, and an HRU is defined as the unique  
585 combination of subbasin and soil and land cover categories. Subbasin input was one of the  
586 candidate subbasin schemes generated in Section 3.1. Soil spatial input data was from the  
587 Canadian Soil Information Service (CANSIS) available from Agriculture and Agri-Food Canada  
588 (2013) and subdivided into fourteen classes in terms of soil profile. Each soil profile except  
589 water is built up by a unique soil horizon combination from three mineral horizons A, B, C, and  
590 an organic horizon O. Soil profile A-B-C covers more than 70% of the Grand River watershed  
591 (Table 4a). Land cover spatial input data was from Canada's National Land Cover Database  
592 available from Natural Resources Canada (2014) and subdivided in seven classes, in which  
593 cropland is dominant across the watershed (Table 4b). Soil and land cover inputs used here are  
594 vector coverages derived from 1:20,000 to 1:60,000 scale county-level soil maps attained from  
595 CANSIS and 1:50,000 scale land cover maps from Canada's National Land Cover Database.

596

597 The map obtained by the overlay (union) of the above subbasin, soil, and land cover layers  
598 defines the reference HRU scheme (scheme 0). Since the map algebra union operation usually  
599 leads to a very fragmented set of sliver HRUs, these sliver HRUs can be suppressed for  
600 aggregation based on certain HRU size threshold. Here, the HRU size threshold was defined as  
601 the HRU area percentage of its affiliated subbasin. The HRU whose area percentage is less than  
602 the size threshold was merged with its neighboring HRU sharing the longest border within the  
603 same subbasin. In order to investigate the influence of the subbasin discretization input on HRU  
604 discretization, we chose two representative subbasin schemes (scheme 5 and scheme Max) as

605 subbasin inputs to discretize HRUs, respectively. The generated candidate HRU schemes are  
 606 listed in Table 5. For the HRU candidates under 90 subbasins, HRU scheme 0 (number of  
 607 HRUs=2706) is the reference scheme; while for the HRU candidates under 32 subbasins, HRU  
 608 scheme 0 (number of HRUs=1232) is the reference scheme. Each reference scheme retains 100%  
 609 of land cover and soil data as the reference scheme does not eliminate/aggregate sliver HRUs. In  
 610 HRU scheme Max, each subbasin is represented by the dominant HRU. Table 5 shows that  
 611 subbasin discretization choice significantly affects HRU discretization complexity (i.e., the  
 612 number of HRUs) because under the same HRU size threshold, the number of HRUs with 90  
 613 subbasins input is always two to three times more than that with 32 subbasins input.

614

615 **Table 4.** Grand River watershed (a) Soil classes (b) Land cover classes and their percent  
 616 coverage of the watershed.

<b>Soil class</b>	<b>Area percentage (%)</b>	<b>Land cover class</b>	<b>Area percentage (%)</b>
A B C	72.29	Annual Cropland	40.70
Water	8.17	Perennial Cropland and Pasture	33.91
A B BC C	7.76	Deciduous Forest	14.74
O B	3.40	Urban	5.43
A B	3.32	Mixed Forest	2.98
A AB B C	2.51	Wetland	1.24
A B AB B C	1.13	Water	1.00
AB	0.64		
C	0.27		
A C	0.25		
O C	0.12		
A	0.09		
C A C	0.04		
A AB C	0.03		

617

618

619 **Table 5.** Candidate HRU discretization schemes with two subbasin discretization schemes (90  
 620 and 32 subbasins)

HRU Scheme	HRU size threshold (% of subbasin area)	Number of HRUs	
		Number of subbasins=90	Number of subbasins=32
0	0	2706	1232
1	1	852	333
2	2	625	234
3	3	502	190
4	4	433	156
5	5	385	135
6	6	346	121
7	7	318	109
8	8	290	99
9	9	252	90
10	10	234	84
Max	One HRU per subbasin	90	32

621

622 **3.2.2. HRU Discretization Error Metric Results**

623 In this study, infiltration and evapotranspiration were identified as the two dominant  
 624 hydrological processes, thus vertical hydraulic conductivity (Kz), available water content (AWC),  
 625 and land cover were defined as the key hydrologic model input variables of interest. For each  
 626 soil class of Table 4a, Kz and AWC are the weighted harmonic mean values of the Kz and AWC  
 627 of its soil horizon components. The detailed soil horizon information is available from  
 628 Agriculture and Agri-Food Canada (2013). The area-weighted mean values of Kz and AWC of  
 629 the entire watershed are 0.9 cm/h and 12.6% (except the soil class water), respectively. Figure 8  
 630 demonstrates the discretization error metric results of Kz, AWC, and land cover at the watershed  
 631 outlet versus HRU size thresholds. As the HRU size threshold increases, discretization gets  
 632 coarser, meanwhile the relative errors of all the three variables increase. However, the same  
 633 HRU size threshold imposes different impacts on the information losses of different variables.  
 634 For example, under the same HRU schemes (before HRU scheme Max), the relative errors of Kz

635 and land cover are always similar in magnitude (Figure 8a, Figure 8c), but the relative errors of  
636 AWC are comparatively smaller (less than 0.05 in Figure 8b). In HRU scheme Max, land cover  
637 error jumps to 0.55, while Kz and AWC errors are 0.15 and 0.05, respectively. Land cover errors  
638 jump to much higher values compared to Kz and AWC because some merged HRUs only  
639 experience land cover changes but no change in soil properties. The results show that,  
640 unsurprisingly, relative discretization errors are positively correlated with HRU size threshold.

641  
642 The subbasin discretization decision between 90 or 32 subbasins has a substantial influence on  
643 HRU discretization complexity (100%-200% increase in number of HRUs seen in Table 5).  
644 However, this decision does not make a big difference for information loss as Figure 8 indicates  
645 that two error metric results (AWC and land cover) of the three variables are almost identical and  
646 only one variable (Kz) obtains slightly different error metric results under different subbasin  
647 inputs.

648  
649 Figure 8. Here.

650  
651 Figure 8 supports how a modeler might make decisions based on a single watershed outlet.  
652 However, in distributed or semi-distributed modelling applications where distributed watershed  
653 responses are of interest, discretization errors should be assessed at multiple sites beyond just the  
654 outlet. Figure 9 is a more robust comparative approach than Figure 8 as it compares  
655 discretization errors at all the 32 sites of interest across the Grand River watershed under  
656 subbasin scheme 5 (number of subbasins = 90). The interesting pattern in Figure 9 is that for all  
657 the three variables of interest (Kz, AWC, and Land cover), the largest discretization errors (and

658 the highest variance) appear in the relatively small drainage areas, and as drainage area increases,  
659 errors approach some constant level. Therefore, while errors for the watershed outlet might be  
660 sufficiently small, they can be unacceptably large in some small upstream subbasins. Although  
661 results are not shown, this pattern persists across all HRU discretization levels.

662

663

Figure 9. Here.

664

### 665 **3.2.3. Sensitivity of Hydrologic Model Simulation Results to HRU Discretization Error**

#### 666 **Metrics**

667 Similar to the sensitivity analysis in Section 3.1.3, we checked the sensitivity of hydrologic  
668 model simulation results to the proposed HRU error metrics based on twelve hydrologic models.  
669 These models correspond to all the HRU schemes under 90 subbasins of Table 5, and the only  
670 difference between these models is the property of HRUs. The model output with scheme 0  
671 (Number of HRUs =2706) is the reference simulation result in model errors calculation. The  
672 peak flow rate error and cumulative flow volume error were computed.

673

674 Figure 10 presents the relationship between the *a priori* HRU discretization error metrics and the  
675 model error indices where each data corresponds to one of the eleven candidate HRU  
676 discretization schemes at the watershed outlet (subbasin 32 outlet). The two model errors are  
677 plotted versus the HRU discretization errors of Kz, AWC, and land cover. Clearly, both model  
678 errors indices monotonically increase with the HRU discretization errors of the three variables.  
679 Correlations ( $r_s$ ) between the three HRU discretization errors (Kz, AWC, and land cover) and the



680 peak flow rate error are all 0.99. Similarly, correlations ( $r_s$ ) between the three HRU discretization  
681 errors and the cumulative flow volume error are also 0.99. This strong correlation also shows up  
682 in most sites of interest (e.g., considering the correlation between the land cover error metric and  
683 the peak flow rate error, 23 sites show  $r_s$  values of 0.8 or more).

684

685

Figure 10. Here

686

687 Figure 11 provides a more complete description of hydrologic simulation responses by plotting  
688 all sites of interest model errors against their drainage areas under the same three representative  
689 HRU schemes of Figure 9. The upstream sites with relatively small drainage areas obtain a high  
690 variance of model errors, in which some of them have three or more times errors than their  
691 downstream sites. This observation appears in both the peak flow rate error and the cumulative  
692 flow volume error and is consistent with results from Figure 9 (indicating the largest HRU  
693 discretization errors are also associated with small drainage areas).

694

695

Figure 11. Here

696

#### 697 **3.2.4. HRU Discretization Decision-making**

698 An alternative to the commonly applied uniform discretization framework demonstrated above is  
699 to make discretization decisions differently in different parts of the watershed, in response to  
700 excessively high error metric values. This relies on the two-step HRU discretization decision-  
701 making approach (see Section 2.3.2) where different subareas can use different HRU delineation  
702 thresholds. To demonstrate, assume subbasin scheme 5 (number of subbasins=90) is the subbasin

703 input for HRU discretization; all the 32 sites of interest and all the three hydrologic model input  
704 variables of interest are equivalently important in HRU scheme decision-making; and 0.40 is the  
705 preliminary error threshold for all the three variables. The subjective value of 0.40 was selected  
706 for demonstration purposes only and selected with the goal of generating a modest number of  
707 HRUs relative to the range of candidate HRU discretizations.

708

709 Step 1: Select an HRU scheme from candidate discretization schemes.

710 HRU scheme 10 (number of HRUs=234) was chosen as the uniform HRU scheme because  
711 the relative errors of all the sites of interest in scheme 10 are satisfactory (less than 0.40) and  
712 the number of HRUs is minimum among all the satisfactory schemes (schemes 1-10).

713 Step 2: Refine HRU discretization for the areas with extreme discretization errors.

714 • Step 2a: Identify the sites of interest with extreme discretization metrics (extreme sites).

715 The extreme error thresholds were defined as the 90<sup>th</sup> percentiles of the error distributions of the  
716 three variables (0.13, 0.06, and 0.31 for Kz, AWC, and land cover, respectively). As a result, the  
717 sites having the highest 10% Kz, AWC, or land cover errors were identified as the extreme sites  
718 of interest to have their discretization refined (i.e., sites 10, 13, 15, 19, 20, 23, 28, and 31). The  
719 drainage areas above sites 19 and 20 are highlighted for discretization refinement demonstration  
720 in Figure 12.

721

722 • Step 2b: Replace HRUs in the upstream refinement areas of extreme sites with those of the  
723 nearest, more detailed satisfactory discretization scheme, in which different extreme sites  
724 have different upstream refinement areas.

725 Case 1: Extreme sites 10, 13, 19, 23, 28, and 31 have no upstream sites of interest, so the  
 726 HRU refinement areas cover the whole drainage areas above these sites.

727 Case 3: Extreme sites 15 and 20 have extreme upstream sites of interest (site 13 and 19,  
 728 respectively), and it is unnecessary to replace HRUs across their entire drainage area in  
 729 the first refinement iteration.

730 Then, only Case 1 sites had the HRUs within them replaced with those of the nearest more  
 731 detailed satisfactory HRU scheme relative to HRU scheme 10. This replacement step was  
 732 applied independently for each of the refined extreme sites. The detailed HRU replacement  
 733 results are summarized in Table 6.

734

735 **Table 6.** HRU replacement results for the extreme sites in the first HRU discretization  
 736 refinement iteration

HRU replacement area	Original number of HRUs	Nearest satisfactory HRU Scheme	New number of HRUs
Drainage area above site 10	2	HRU scheme 1	10
Drainage area above site 13	4	HRU scheme 9	5
Drainage area above site 19	4	HRU scheme 5	8
Drainage area above site 23	2	HRU scheme 7	5
Drainage area above site 28	10	HRU scheme 5	16
Drainage area above site 31	3	HRU scheme 3	7

737  
 738

739 • Step 2c: Merge all resultant HRUs into an output layer and re-calculate discretization errors  
 740 for the sites of interest.

741 After the first refinement iteration, it was found that all Case 1 sites became satisfactory, but  
 742 some errors of sites 15 and 20 were still extreme. Therefore, *Step 2b* needed to be repeated to  
 743 replace HRUs for sites 15 and 20.

744 • Step 2b: Replace some or all HRUs in the upstream refinement areas of extreme sites with  
 745 those of the nearest, more detailed satisfactory discretization scheme.

746 Case 2: Extreme sites 15 and 20 became Case 2 after the first refinement iteration, and  
 747 thus the intermediate area between sites 13 and 15 and the intermediate area between  
 748 sites 19 and 20 were identified as the HRU replacement areas in the second refinement  
 749 iteration. The HRUs within these intermediate areas were replaced with those of the  
 750 nearest satisfactory schemes relative to HRU scheme 10. The detailed HRU replacement  
 751 results are summarized in Table 7.

752

753 **Table 7.** HRU replacement results for the extreme sites in the second HRU discretization  
 754 refinement iteration

HRU replacement area	Original number of HRUs	Nearest satisfactory HRU Scheme	New number of HRUs
Intermediate area 13-15 <sup>1</sup>	10	HRU scheme 8	16
Intermediate area 19-20	1	HRU scheme 4	3

755 <sup>1</sup> Intermediate area 13-15 means the intermediate area between site 13 and site 15. This also applies to  
 756 intermediate area 19-20.  
 757

758 • Step 2c: Merge all resultant HRUs into an output layer and re-calculate errors across the  
 759 watershed.

760 After the second iteration, the errors of all the originally identified extreme sites became  
 761 satisfactory, thus the refinement process ended and this scheme was the refined HRU scheme.

762 Figure 12 provides a visual comparison for the HRUs before and after refinement of the drainage  
 763 areas of sites 19 and 20 under subbasin scheme 5 (90 subbasins).

764

765 Figure 12. Here.

766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781

Table 8 shows the error metric results of HRU scheme 10, refined scheme 10, and scheme\*. Scheme\* has the same number of HRUs as refined scheme 10 (number of HRUs=271), but was generated with a uniform HRU size threshold of 8.4%. Through discretization refinement, the extreme errors identified in Step 2b above are reduced, and the discretization quality for site 32 representing the entire watershed is also improved for all the three variables. Moreover, the discretization error means and standard deviations of the three variables across all the sites of interest of the refined scheme also decrease in contrast with scheme\*. Therefore, the non-uniform discretization functions to retain more input data information than the uniform discretization under the same discretization complexity. In addition, to get a sense of how non-uniform the HRU discretization is in refined scheme 10, the average HRU sizes of the uniformly discretized areas and the non-uniformly discretized areas within the watershed were respectively calculated as 28.8 km<sup>2</sup> and 13.6 km<sup>2</sup>. The latter is more than 50% smaller than the former, which means the HRUs within the refined areas are obviously finer than those of the uniformly discretized areas.

782 **Table 8.** Discretization error metric results for three HRU discretization schemes (using 90 subbasins).  
783 Scheme 10 is based on an HRU size threshold of 10%, while Scheme\* is based on a threshold of 8.4%.  
784 Note that site 32 corresponds to the watershed outlet and sites of interest 1, 9, 11, 18, 24, 27 and 30 are  
785 not included because they are discretized the same way under all three schemes. Highlighted errors for  
786 Refined scheme 10 are lower than corresponding errors in one or both of Scheme 10 and Scheme\*.

Site of interest	Scheme 10				Refined scheme 10				Scheme*			
	Number of HRUs	Kz	AWC	Land cover	Number of HRUs	Kz	AWC	Land cover	Number of HRUs	Kz	AWC	Land cover
2	3	0.06	0.02	0.25	3	0.06	0.02	0.25	4	0.02	0.01	0.22
3	17	0.06	0.03	0.20	17	0.06	0.03	0.20	22	0.06	0.02	0.14
4	7	0.02	0.01	0.12	7	0.02	0.01	0.12	8	0.02	0.01	0.07
5	20	0.06	0.03	0.22	20	0.06	0.03	0.22	25	0.05	0.02	0.17
6	22	0.06	0.03	0.22	22	0.06	0.03	0.22	28	0.05	0.02	0.17
7	15	0.02	0.01	0.10	15	0.02	0.01	0.10	16	0.02	0.01	0.07
8	18	0.02	0.01	0.10	18	0.02	0.01	0.10	19	0.02	0.01	0.07
10 <sup>c</sup>	2	0.11	0.06	0.11	10	0.06	0.02	0.02	2	0.11	0.06	0.11
12	29	0.06	0.03	0.19	29	0.06	0.03	0.19	35	0.06	0.02	0.15
13 <sup>c</sup>	4	0.02	0.00	0.31	5	0.01	0.01	0.22	6	0.02	0.00	0.14
14	4	0.12	0.05	0.14	12	0.10	0.03	0.08	4	0.12	0.05	0.14
15 <sup>c</sup>	14	0.03	0.01	0.36	21	0.03	0.01	0.26	20	0.03	0.01	0.30
16	24	0.07	0.01	0.09	24	0.07	0.01	0.09	26	0.06	0.01	0.08
17	33	0.05	0.02	0.27	40	0.05	0.02	0.23	40	0.05	0.02	0.23
19 <sup>c</sup>	4	0.23	0.03	0.35	8	0.13	0.01	0.17	4	0.23	0.03	0.35
20 <sup>c</sup>	5	0.23	0.03	0.34	11	0.13	0.02	0.17	5	0.23	0.03	0.34
21	36	0.07	0.03	0.26	43	0.07	0.03	0.23	43	0.07	0.03	0.23
22	89	0.08	0.02	0.15	103	0.08	0.02	0.15	102	0.08	0.02	0.12
23 <sup>c</sup>	2	0.13	0.07	0.39	8	0.03	0.01	0.07	5	0.05	0.02	0.18
25	141	0.08	0.03	0.19	168	0.08	0.02	0.17	166	0.08	0.02	0.16
26	19	0.12	0.03	0.16	19	0.12	0.03	0.16	20	0.12	0.02	0.16
28 <sup>c</sup>	10	0.21	0.06	0.13	16	0.12	0.03	0.08	11	0.20	0.05	0.13
29	176	0.09	0.03	0.19	203	0.09	0.02	0.17	207	0.09	0.02	0.16
31 <sup>c</sup>	3	0.24	0.13	0.04	7	0.06	0.04	0.05	3	0.24	0.13	0.04
32	234	0.12	0.04	0.16	271	0.10	0.03	0.15	271	0.11	0.04	0.14
<b>Error mean</b>		0.09	0.03	0.20		0.07	0.02	0.15		0.09	0.03	0.16
<b>Error Std. deviation</b>		0.07	0.03	0.09		0.04	0.01	0.07		0.07	0.03	0.08
<b>Error 90<sup>th</sup> percentile<sup>f</sup></b>		0.13	0.06	0.31								

787 <sup>c</sup> denotes an extreme site under scheme 10 based on exceeding the 90<sup>th</sup> percentiles of the error metrics.

788 The HRU discretization within this site's drainage area is refined based on *Step2*.

789 <sup>f</sup> The 90<sup>th</sup> percentile computed based on errors across all 32 sites of interest.

790 **4. Discussion**

791 **4.1. Reference Discretization Scheme Determination**

792 The reference scheme is defined as a scheme that fully retains the information of the original  
793 spatial input data or, in special cases, the finest plausible discretization. The implication is that  
794 the modeler is interested in quantifying how much information is lost relative to the reference  
795 scheme.

796

797 Our subbasin reference scheme was defined based on a subjective flow accumulation threshold  
798 to determine reference main channel lengths. Alternatively, the real full flow path information  
799 (i.e., flow path of each cell in the DEM) can be obtained, for example, by the flow length tool of  
800 ArcGIS, and the corresponding discretization could be used as the reference scheme. For HRU  
801 discretization, our reference scheme retained all raw input spatial data and thus avoided any  
802 subjective decisions. Alternatively, the reference HRU scheme could be subjectively defined as  
803 a discretization that addresses some numerical and topological problems if this discretization is  
804 the one that modelers will practically apply and want relative errors computed against (Sanzana  
805 et al., 2013). While absolute discretization error metric values will be impacted by what can be  
806 a subjective reference scheme choice, the relative error values among candidate discretization  
807 choices should not change significantly.

808

809 **4.2. Discretization Error Metrics**

810 The subbasin discretization error metric estimates the in-channel routing length difference  
811 relative to the reference scheme. An improved approach would be to instead consider travel time  
812 error by using a reference flow velocity. Assuming the reference flow velocity as a constant is an  
813 easy and common method for practical purposes (De Lavenne, Boudhraâ, & Cudennec, 2015;  
814 Rigon, Bancheri, Formetta, & de Lavenne, 2016; Sanzana, et al., 2013). However, this  
815 assumption is questionable as typical watersheds have faster velocity upstream reaches compared  
816 to lower velocity downstream reaches. As such, a travel time error could instead be based on a  
817 spatially variable reference flow velocity. In addition, other available roughness, geometry,  
818 channel slope information can be incorporated into the routing information loss estimation by  
819 being linked to flow velocity (e.g., with Manning's Equation).

820

821 The *a priori* metrics (both nominal and quantitative) are able to provide directly meaningful  
822 descriptions on information loss because they explicitly characterize how much area or value of  
823 the hydrologic model input variable is changed after discretization. Moreover, they are unique as  
824 compared to the existing *a priori* metrics in the way they identify the property change.  
825 Haverkamp et al.(2002) , Booij (2003), and Dehotin and Braud (2008) all define discretization  
826 information loss as the overall statistics difference between the candidate scheme and the  
827 reference scheme, failing to conduct the cell-by-cell comparison with the original spatial input  
828 data. In contrast, the metrics proposed here correspond one-to-one with information loss during  
829 discretization. The overlay comparison process is a straightforward technique and is feasible for



830 both raster and vector spatial input data. This enables other hydrologic variables of interest such  
831 as land surface slope and aspect to be analyzed with similar *a priori* discretization error metrics.

832

833 Figure 5 and Figure 9 show that discretization errors are highest in smaller upstream subbasins  
834 while Figure 11 shows that hydrologic model error indices (for peak flow and cumulative  
835 volume) are also highest in smaller upstream subbasins. This observation explains to some extent  
836 the modelling difficulties associated with small upstream subbasins in semi-distributed  
837 modelling. Although past studies such as Andersen (2001) and Tuppad (2006) have attributed the  
838 poor relative performance of calibrated upstream gauges to calibrated downstream locations to  
839 factors like more uncertain rainfall, our observation reveals that relatively poor performance in  
840 the smaller upstream subbasins of our case study can be expected since the discretization errors  
841 (and hydrologic model error indices) of these subbasins have a high variance and can be three  
842 times larger than the corresponding errors of the downstream larger drainage areas in the uniform  
843 threshold discretization framework. This demonstrates the utility of multi-site discretization  
844 evaluation in distributed modelling applications, and also suggests that a non-homogenous  
845 approach to watershed discretization decision-making would be beneficial.

846

847 Our HRU error metric approach for nominal input data did not disaggregate the individual area  
848 changes of the different categories of nominal data. For example, the area changes of the crop  
849 land or the deciduous forest land. However, modelers may only care about the area change of a  
850 certain category in their watershed (e.g., the change from forest to suburban may be of  
851 consequence but the change from wetland to swamp may be immaterial). Although not

852 demonstrated in this work, the error metric for nominal variables (Equation 2) can be readily  
853 modified to assess the relative error of a specific category of nominal input data.

854

### 855 **4.3. Variations to Discretization Approach**

856 Our approach first generated all candidate subbasin schemes by ArcSWAT, and then generated  
857 candidate HRU schemes by sliver area aggregation. Any other candidate schemes generated by  
858 different discretization methods can also be evaluated by our proposed *a priori* discretization  
859 error metrics. Additional checks could be added to the subbasin discretization step, for example,  
860 checking the reference scheme against additional data such as orthophotos or hydrographic  
861 survey maps. Another variation is related to handling the small but potentially important sliver  
862 HRUs in HRU discretization simplification. For instance, in periurban areas where the land  
863 cover is very heterogeneous some small HRUs can be meaningful in terms of hydrology, thus  
864 such HRUs should be protected from merging in discretization simplification. One approach to  
865 preserve these key HRUs is to introduce an importance factor that would artificially increase the  
866 areas of key HRUs so that they would exceed the HRU discretization threshold used to aggregate  
867 small HRUs.

868

## 869 **5. Conclusions**

870 This study proposed *a priori* discretization error metrics that can estimate the information loss  
871 for any candidate discretization scheme. These metrics do not require model simulation, are  
872 independent of any specific modelling software, provide modelers with directly interpretable  
873 information on discretization quality, and allow for multi-site and multi-variable discretization

874 evaluations prior to model development. In particular, the subbasin error metric provides the first  
875 attempt at quantifying the routing information loss from discretization; the HRU error metrics  
876 improves upon the existing *a priori* metrics in variable property change identification by the  
877 overlay comparison process. The proposed error metrics are straightforward to understand and  
878 easy to recode into the preprocessing of any semi-distributed hydrologic models and the fully  
879 distributed models using spatial input data aggregation. As a potential application of the  
880 proposed *a priori* discretization error metrics, a two-step decision-making approach was  
881 formulated to help modelers to get the appropriate subbasin and HRU discretization schemes,  
882 respectively. The approach does not only allow choosing a traditional spatially uniform-threshold  
883 discretization scheme based on the modeler-defined error threshold(s), but also enables  
884 compressing extreme errors to satisfy the modeler-specified discretization error targets.

885  
886 These *a priori* discretization error metrics were applied to the discretization of the Grand River  
887 watershed. Results indicated that the discretization-induced information loss as measured by our  
888 discretization error metrics monotonically increases as discretization gets coarser. Hydrologic  
889 modeling under candidate discretization schemes validates the strong correlation between our  
890 discretization error metrics and model predictions (peak flow rate, cumulative flow and peak  
891 flow timing). Discretization evaluation results show that model accuracy moving from larger  
892 downstream locations to smaller upstream locations would be expected to increase since the  
893 largest discretization errors and highest error variability occur in smaller upstream locations.  
894 This pattern is also evident when changes in hydrologic model outputs were used in place of  
895 HRU discretization error metrics. Finally, results show that the common and convenient  
896 approach of applying uniform discretization across the watershed domain performs worse

897 compared with the metrics-informed non-uniform discretization approach as the latter is able to  
898 preserve more input data information using the same number of computational units. However,  
899 the influence of non-uniform discretization on hydrologic model outputs should be further  
900 studied using a number of hydrologic models and case studies.

901

902 In applying the proposed *a priori* discretization error metrics to discretization decision-making,  
903 accounting for input forcing data (e.g., precipitation and temperature) resolution is also an  
904 important future consideration. This will require comparing the spatial and temporal distributions  
905 of the forcing input data under candidate schemes and those under the reference scheme. Beyond  
906 the application in discretization decision-making, future studies can utilize the discretization  
907 error metrics in other ways. For instance, the discretization error metrics may be useful in trying  
908 to account for the uncertainty induced by watershed discretization decisions which is commonly  
909 ignored. Furthermore, the discretization error metrics should prove useful even when they are not  
910 calculated *a priori* in that they could serve an important role in diagnosing the causes of model  
911 prediction errors in distributed modeling applications.

912

### 913 **Acknowledgement**

914 This research was funded by NSERC Canadian FloodNet grant (Tolson as co-investigator) used  
915 to support PhD student Liu. We are grateful to the four anonymous reviewers for their thoughtful  
916 comments that helped us improve our manuscript significantly.

917

918 **References**

- 919 Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E., & Rasmussen, J. (1986). An  
920 introduction to the European Hydrological System—Systeme Hydrologique  
921 Europeen,“SHE”, 1: History and philosophy of a physically-based, distributed modelling  
922 system. *Journal of Hydrology*, 87(1), 45-59.
- 923 Agriculture and Agri-Food Canada. (2013). Detailed Soil Survey (DSS) Compilations.  
924 Retrieved September 15, 2014, from <http://sis.agr.gc.ca/cansis/nsdb/dss/v3/index.html>
- 925 Anselin, L. (2010). Thirty years of spatial econometrics. *Papers in regional science*, 89(1), 3-25.
- 926 Arnold, J., Allen, P., Volk, M., Williams, J., & Bosch, D. (2010). Assessment of different  
927 representations of spatial variability on SWAT model performance. *Trans. ASABE*, 53(5),  
928 1433-1443.
- 929 Arnold, J. G., Srinivasan, R., Mutiah, R. S., & Williams, J. R. (1998). Large area hydrologic  
930 modeling and assessment part I: Model development1: Wiley Online Library.
- 931 Bergström, S. (1976). Development and application of a conceptual runoff model for  
932 Scandinavian catchments. Norrköping, Sweden: Swedish Meteorological and  
933 Hydrological Institute (SMHI).
- 934 Bergström, S. (1992). The HBV model: Its structure and applications. Norrköping, Sweden:  
935 Swedish Meteorological and Hydrological Institute (SMHI).
- 936 Beven, K., & Binley, A. (1992). The future role of distributed models: model calibration and  
937 predictive uncertainty. *Hydrological Processes*, 6, 279-298.
- 938 Beven, K., & Kirkby, M. (1979). A physically based, variable contributing area model of basin  
939 hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin  
940 versant. *Hydrological Sciences Journal*, 24(1), 43-69.

941 Booij, M. (2003). Determination and integration of appropriate spatial scales for river basin  
942 modelling. *Hydrological Processes*, 17(13), 2581-2598.

943 Booij, M. (2005). Impact of climate change on river flooding assessed with different spatial  
944 model resolutions. *Journal of Hydrology*, 303(1), 176-198.

945 Ciarapica, L., & Todini, E. (2002). TOPKAPI: A model for the representation of the rainfall -  
946 runoff process at different scales. *Hydrological Processes*, 16(2), 207-229.

947 Craig, J. R., Jenkinson, W., Jost, G., Serrer, M., Snowdon, A. P., Sgro, N., . . . Tolson, B. A.  
948 (2016). Flexible watershed simulation with the Raven hydrological modeling framework.  
949 Submitted to *Environmental Modeling and Software*.

950 De Lavenne, A., Boudhraâ, H., & Cudennec, C. (2015). Streamflow prediction in ungauged  
951 basins through geomorphology-based hydrograph transposition. *Hydrology Research*,  
952 46(2), 291-302.

953 Dehotin, J., & Braud, I. (2008). Which spatial discretization for distributed hydrological models?  
954 Proposition of a methodology and illustration for medium to large-scale catchments.  
955 *Hydrology and Earth System Sciences*, 12(3), 769-796.

956 Djokic, D. (2008). Comprehensive terrain preprocessing using arc hydro tools. *US: ESRI*.

957 Doan, J. (2000). Geospatial hydrologic modeling extension HEC-GeoHMS-user's manual-  
958 version 1.0. *Davis, California*.

959 ESRI. (2014). ArcGIS 10.2.2 for Desktop. *Redlands, CA: Environmental Systems Research*  
960 *Institute* Retrieved September 15, 2014, from  
961 <http://www.esri.com/software/arcgis/arcgis-for-desktop>

962 Fenicia, F., Kavetski, D., Savenije, H. H., & Pfister, L. (2016). From spatially variable  
963 streamflow to distributed hydrological models: Analysis of key modeling decisions.  
964 *Water Resources Research*, 52, 954-989.

965 Flügel, W. A. (1995). Delineating hydrological response units by geographical information  
966 system analyses for regional hydrological modelling using PRMS/MMS in the drainage  
967 basin of the River Bröl, Germany. *Hydrological Processes*, 9(3-4), 423-436.

968 Grayson, R., Blöschl, G., Moore, I., & Singh, V. (1995). Distributed parameter hydrologic  
969 modelling using vector elevation data: THALES and TAPES-C. *Computer models of*  
970 *watershed hydrology*, 669-696.

971 Haghnegahdar, A., Tolson, B. A., Craig, J. R., & Paya, K. T. (2015). Assessing the performance  
972 of a semi - distributed hydrological model under various watershed discretization  
973 schemes. *Hydrological Processes*, 29, 4018-4031.

974 Haverkamp, S., Srinivasan, R., Frede, H. G., & Santhi, C. (2002). Subwatershed spatial analysis  
975 tool: Discretization of a distributed hydrologic model by statistical criteria. *Journal of the*  
976 *American Water Resources Association*, 38(6), 1723-1733.

977 Journel, A. G., & Deutsch, C. V. (1993). Entropy and spatial disorder. *Mathematical Geology*,  
978 25(3), 329-355.

979 Kouwen, N. (1988). WATFLOOD: a micro-computer based flood forecasting system based on  
980 real-time weather radar. *Canadian Water Resources Journal*, 13(1), 62-77.

981 Kouwen, N., Soulis, E., Pietroniro, A., Donald, J., & Harrington, R. (1993). Grouped response  
982 units for distributed hydrologic modeling. *Journal of Water Resources Planning and*  
983 *Management*, 119(3), 289-305.

984 Liu, Y., & Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data  
985 assimilation framework. *Water Resources Research*, 43(7).

986 Ludwig, R., May, I., Turcotte, R., Vescovi, L., Braun, M., Cyr, J.-F., . . . Chartier, I. (2009). The  
987 role of hydrological model complexity and uncertainty in climate change impact  
988 assessment. *Advances in Consciences*, 21, 63-71.

989 Madsen, H. (2003). Parameter estimation in distributed hydrological catchment modelling using  
990 automatic calibration with multiple objectives. *Advances in Water Resources*, 26(2), 205-  
991 216.

992 Maidment, D. R. (2002). *Arc Hydro: GIS for water resources* (Vol. 1): ESRI, Inc.

993 Natural Resources Canada. (2014). Land Cover, circa 2000-Vector. Retrieved September 15,  
994 2014, from <http://www.geobase.ca/>

995 Petrucci, G., & Bonhomme, C. (2014). The dilemma of spatial representation for urban  
996 hydrology semi-distributed modelling: Trade-offs among complexity, calibration and  
997 geographical data. *Journal of Hydrology*, 517, 997-1007.

998 Pietroniro, A., Fortin, V., Kouwen, N., Neal, C., Turcotte, R., Davison, B., . . . Evora, N. (2007).  
999 Development of the MESH modelling system for hydrological ensemble forecasting of  
1000 the Laurentian Great Lakes at the regional scale. *Hydrology and Earth System Sciences*,  
1001 11(4), 1279-1294.

1002 Refsgaard, J. C. (1997). Parameterisation, calibration and validation of distributed hydrological  
1003 models. *Journal of Hydrology*, 198(1), 69-97.

1004 Reggiani, P., Sivapalan, M., & Hassanizadeh, S. M. (1998). A unifying framework for watershed  
1005 thermodynamics: balance equations for mass, momentum, energy and entropy, and the  
1006 second law of thermodynamics. *Advances in Water Resources*, 22(4), 367-398.



1007 Rigon, R., Bancheri, M., Formetta, G., & de Lavenne, A. (2016). The geomorphological unit  
1008 hydrograph from a historical - critical perspective. *Earth Surface Processes and*  
1009 *Landforms*, 41(1), 27-37.

1010 Sanzana, P., Jankowfsky, S., Branger, F., Braud, I., Vargas, X., Hirschfeld, N., & Gironás, J.  
1011 (2013). Computer-assisted mesh generation based on hydrological response units for  
1012 distributed hydrological modeling. *Computers & Geosciences*, 57, 32-43.

1013 Shrestha, R. R., & Rode, M. (2008). Multi-objective calibration and fuzzy preference selection of  
1014 a distributed hydrological model. *Environmental Modelling & Software*, 23(12), 1384-  
1015 1395.

1016 Singh, V. P., & Frevert, D. K. (2005). *Watershed models*: CRC Press.

1017 Srinivasan, M., Gérard-Marchant, P., Veith, T. L., Gburek, W. J., & Steenhuis, T. S. (2005).  
1018 Watershed scale modeling of critical source areas of runoff generation and phosphorus  
1019 transport1. *Journal of the American Water Resources Association*, 41(2), 361-377.

1020 Tuppad, P. (2006). *Hydrologic modeling response to NEXRAD and raingage spatial variability*  
1021 *and strategic watershed management*. Ph.D. thesis, Kansas State University, Manhattan,  
1022 Kansas, USA.

1023 Winchell, M., Srinivasan, R., Di Luzio, M., & Arnold, J. (2007). ArcSWAT interface for  
1024 SWAT2005 User's guide. *Texas Agricultural Experiment Station and United States*  
1025 *Department of Agriculture, Temple, TX*.

1026 Wood, E. F., Sivapalan, M., Beven, K., & Band, L. (1988). Effects of spatial variability and scale  
1027 with implications to hydrologic modeling. *Journal of Hydrology*, 102(1), 29-47.