The influence of spatial discretization on HEC-HMS modelling: a case study

Abstract

The spatial discretization of a watershed can directly influence the derived outputs of hydrological models. Conclusive guidelines regarding the proper level of spatial discretization have not yet been established due to the great diversity of models, calibrated parameters and topographic and geospatial characteristics of study areas. The objective of this study was to investigate the influence of different levels of watershed spatial discretization (2, 4, 8, 16 and 32 sub basins) on the HEC-HMS model's performance, considering the Environmental Protection Area (APA) of the Uberaba River Basin region. The SCS-CN method was applied to the different levels of spatial discretization for two distinctive periods (calibration and validation) and its performance was evaluated according to seven goodness-of-fit measures: MAE, RMSE, RSR, NSE, PBIAS, R² and KGE. The results obtained from a qualitative and quantitative perspective, based on the visual analysis of the simulated hydrographs and the obtained goodness-of-fit values; indicate the HEC-HMS model represents the hydrological processes of the basin under investigation efficiently, regardless of the level of spatial discretization. The results suggest the subdivision of a watershed does not result in improvement of the HEC-HMS model's performance without significant differences in physiographic characteristics (slope, land use and types of soil). Considering its good performance for the study area, it is suggested the application of the HEC-HMS model for future works that aim to investigate the drainage capacity of the APA of the Uberaba River.

Keywords: catchment size, GIS, soil conservation service, hydrological modelling, spatial discretization.

Abbreviations: ANA, brazilian national water agency; BDMEP, meteorological database for education and research; CN, curve number; DEM, digital elevation model; EPA, environmental protection area; GIS, geographic information system; HEC, hydrologic engineering center; HMS, hydrologic modeling system; KGE, kling-gupta efficiency; MAE, mean absolute error; NSE, nash-sutcliffe efficiency; PBIAS, percent bias; R², coefficient of determination; RMSE, root mean squared error; RSR, RMSE-observations standard deviation ratio; SCS, soil conservation service; UH, unit hydrograph; USACE, united states army corps of engineers.

Introduction

A watershed is a natural catchment area, composed of hill slopes and a drainage network, where the collected rainfall is converted into a single riverbed, whose outlet is known as river mouth. It can be regarded as a system, where the input is the precipitation and the output is the volume of water that leaves the river basin through its mouth, considering as losses the evaporation, transpiration and infiltration. A watershed is the hydrological unit most suitable for studies, since it provides an ideal modeling environment where you have a known input, precipitation data, and a known output, stream gage data. Hydrological modelling aims to understand and represent the hydrological processes that occur in a watershed, based on the simulation of different scenarios. It can be described as a tool that enables the prediction and estimation of extreme events, as well as to assist in water resource planning and management processes.

Hydrological simulation depends on how well the watershed processes are represented by the model and on how representative the input parameters values are. Consequently, it is common practice to discretize the watershed into subbasins, where each subbasin is presumed as a homogeneous spatial unit, in order to better represent its heterogeneities, such as the channel network, the topography, the types of soils, the land use, and the climate inputs. The spatial discretization of a water catchment directly affects the model outputs, such as peak discharge, outflow hydrograph, time to reach peak discharge, and stream routing. This is due to the changes in the input parameters, seeing as: the basin’s boundaries predetermines aggregation patterns of spatial information; the topographic parameter values are influenced by the basin’s size and extent; and the characteristics of the drainage network can cause misrepresentation of the rainfall-runoff process. Guidelines for the most appropriate discretization scheme are often poorly defined and left to the modelers’ discretion. This can be related to the diversity of existing models, calibrated parameters and topographic and geospatial characteristics of different study areas. This paper aimed to investigate the influence of different levels of spatial discretization (2, 4, 8, 16 and 32 subbasins) on the HEC-HMS model’s performance, considering the Environmental Protection Area (APA - Área de Proteção Ambiental) of the Uberaba River region as the study area. This region is under intense pressure from urban development companies, which demonstrates the relevance of hydrological studies in the area, since these can help prevent problems like those Uberaba city faces today, such as floods caused by extreme rainfall events.
Material and methods

Study Area

In order to run the HEC-HMS model in the drainage area concerning the Uberaba gaging station (code 61794000) belonging to the Brazilian National Water Agency (Agência Nacional das Águas – ANA), located in the longitude 47°58’42.96”W and latitude 19°43’41.16”S, the study was conducted more precisely in an area that exceeds the legally established territorial boundaries for the Environmental Protection Area (Área de Proteção Ambiental – APA) of the Uberaba River Basin. The APA of the Uberaba River Basin, acknowledged as a Sustainable Land Use Conservation Unit by Brazilian Federal Law n° 9.985/2000, is located on Cerrado Biome, in the city of Uberaba, Triângulo Mineiro Region, State of Minas Gerais, Brazil. This area, located upstream of the city’s water catchment point for public supply, was created with the purpose of guaranteeing the conservation of the Uberaba River Basin’s water resources, seeking the maintenance of its quality and rational use. Uberaba is inserted in the plateau of the Paraná Basin – Bauru Sub-basin, characterized by the incidence of tablelands and scarps. Geology is dominated by the presence of the Uberaba and Marília formations, belonging to the Bauru Group, overlaying the basaltic rocks belonging to the Serra Geral Formation. Involving the Marília Formation are the Cenozoic’s sediments attributed to the Nova Ponte Formation, which compose the municipality’s lands of higher altitudes and lower slope.

According to the classification of Köppen, three climatic classes characterize the Triângulo Mineiro e Alto Paranaíba mesoregion: Aw (Tropical Wet and Dry or Savanna climate), Cwa (Subtropical-Dry Winter) and Cwb (Dry Winter Subtropical Highland climate). The Aw, which dominates 78% of this mesoregion, is characterized by a dry winter (May – September), presenting less than 40 millimeters of precipitation a month, and a rainy summer (October-March), with 60% of the annual rainfall concentrated in the months of November, December, January and February. According to Brazilian system of soil classification, the most predominant soil units found in the catchment area of the Uberaba River are: the Latossolos, in the suborders Vermelho e Vermelho Amarelo; the Cambissolos, in the suborders Háplico and Húmico; the Gleissolos, in the suborder Melânico; the Neossolos, in the suborders Regolítico and Flúvico; the Argissolos, in the suborder Amarelo, and the Plintossolo, in the suborder Háplico.

Data preprocessing

The preprocessing steps adopted in this study were carried out by using the ArcGIS software, based on a 30 meters resolution Digital Elevation Model (DEM) obtained from the Shuttle Radar Topography Mission (SRTM), developed by the National Aeronautics and Space Administration (NASA). The methodology is summarized in Figure 1 and explained subsequently. Initially the Fill tool, available in the Spatial Analyst toolbox, was used to fill the DEM’s surface depressions and to eliminate its imperfections, thereby generating the Hydro DEM of the study area. Afterwards the Flow Direction tool, run with the D8 method (Figure 2), was used to generate a raster which assigns a flow from each cell to one of its eight neighboring pixels, either adjacent or diagonal, in the direction of the steepest downward slope. Based on the flow direction grid, the Flow Accumulation tool was used to generate a new grid (Figure 3) which gives the flow as the sum of the amount of water flowing into each downslope cell. Subsequently, the Set Null tool was run in order to create a raster with the drainage network, by using the condition “Value <1000”, and thus considering 1000 cells as the threshold of contributing pixels. To convert the raster network into a vector the Stream to Feature tool was used. Based on the hydrological, topological and pedological features of the study area, the subbasins’ outlets were created for each level of spatial discretization under study and its delineations (Figure 4-11) were obtained running the Watershed tool.

The influence of spatial discretization on HEC-HMS modelling: a case study

Copyright: ©2019 Barbosa et al.


Hydrological modeling (HEC-HMS)

Developed by the Hydrologic Engineering Center (HEC) of the United States Army Corps of Engineers (USACE), the Hydrologic Modeling System (HEC-HMS) is a software that contains an array of mathematical models designed to simulate the hydrologic processes of watershed systems, such as rainfall, snowmelt, evapotranspiration, infiltration, excess precipitation transformation, baseflow and stream routing. The HEC-HMS is one the most widely used and documented hydrological model, run several different studies.

Input data

Daily stream flow data from 1995 through 2017 was acquired from the Uberaba gage station, available at ANA's portal HIDROWEB. Precipitation data for the same period was obtained at a daily time step from the station 83577 - UBERABA - MG (19°43'48.0"S; 47°57'00.0" W), available at the Meteorological Database for Education and Research (BDMEP - Banco de Dados Meteorológico para Ensino e Pesquisa) belonging to the Brazilian National Institute of Meteorology (INMET – Instituto Nacional de Meteorologia). The physiographic parameters required by the adopted models, such as area, channel length, slope and curve number, were derived from each subbasin in the ArcGIS environment, for the five levels of discretization under study.

Loss model

Precipitation losses, which occurs through the runoff/infiltration processes, was determined using the SCS Curve Number method. The method calculates the effective precipitation of a hydrological event based on an empirically derived relationship between location, soil-type, land use, antecedent moisture conditions and runoff (Eq. 1).

\[
P_e = \frac{(P - 0.2S)^2}{P + 0.8S}
\]

Where: \(P_e\), accumulated precipitation excess at time \(t\); \(P\), accumulated rainfall depth at time \(t\); \(S\), potential maximum retention.

The initial abstraction \((I_a)\) was estimated as a constant 20% of the maximum potential retention \(S\), as recommended by the Technical Reference Manual. The curve number (commonly abbreviated CN) is an intermediate parameter, which relates the maximum retention, and the watershed characteristics (Eq. 2). The CN values, ranging from 0 to 100, were estimated based on the tables published by the SCS and knowledge of the soil type and land use of the study area.

\[
S = \frac{25400 - 254CN}{CN}
\]

Considering that Brazilian soils do not fit the SCS’s hydrological groups descriptions, it was adopted a new classification proposal. This classification takes into account not only the soil’s texture, but also its characteristics and profile, reducing the modeler’s doubts. Although this classification was created based on soils situated in the State of São Paulo, it can be applied throughout Brazil where similar soils occur. The CN Map (Figure 5) was generated based on the data of the Land Use and Occupation Map (Figure 6), created by Uniube’s Geoprocessing Laboratory, and the Soil Map (Figure 7), adapted from the Soil Map of the State of Minas Gerais.

![Figure 3 Flow accumulation coding.](image1)

![Figure 4 2 Subbasin configuration.](image2)

![Figure 5 4 Subbasin configuration.](image3)
The influence of spatial discretization on HEC-HMS modelling: a case study

**Figure 6** Subbasin configuration.

**Figure 7** 16 Subbasin configuration.

**Figure 8** 32 Subbasin configuration.

**Figure 9** Curve Number Map.

**Figure 10** Land Use and Occupation Map.

**Figure 11** Soil Classification Map.

The influence of spatial discretization on HEC-HMS modelling: a case study

The performance evaluation of 22 Tc formulas indicates the Bransby-Williams is the most suitable method for large watershed.\(^5\) As such, the Tc was determined using the Bransby-Williams formula (Eq. 5).

\[
Tc = 58.5 . L . A^{-0.1} - S^{-0.2}
\]

Where: L = mainstream length; A = catchment area; S = equal area slope.

Routing method

To calculate the downstream outflow hydrograph, given an upstream hydrograph as a boundary condition, it was selected the Muskingum-Cunge routing method. This method uses a finite difference approximation of the partial derivatives of the continuity equation and the diffusion form of the momentum equation to generate the Eq. 6.\(^14\)

\[
O_i = C_f I_i + C_i O_{i-1} + C_{2i} (q_L \Delta x)
\]

The coefficients are:

\[
C_f = \frac{\Delta t}{k} + \frac{2 X}{k} + 2 (1 - X)
\]

\[
C_i = \frac{2 (1 - X)}{k} - \frac{\Delta t}{k} + 2 (1 - X)
\]

\[
C_{2i} = \frac{2 (1 - X)}{k} + 2 (1 - X)
\]

\[
C_{2i} = \frac{2 (1 - X)}{k} \frac{\Delta t}{k} + 2 (1 - X)
\]

Where: K= travel time of the flood wave through routing reach; X= attenuation of the flood wave; \(\Delta t\) = time step.

In which the parameters K and X are: \(^14\)

\[
K = \frac{\Delta x}{c}
\]

\[
X = \frac{1}{2} \left( \frac{1 - Q}{B S c \Delta x} \right)
\]

Where: Q = reference flow; B = top width of the water surface; So= friction slope or bed slope; c= wave celerity (speed); \(\Delta x\)= the length of reach.

Model calibration

Based on several simulations, the event of greater adherence was selected for the calibration period, comprised between 1st of November 2003 and the 31st of December 2003. In this step,
the objective function Sum of squared residuals (Eq.13) and the Univariate-Gradient search algorithm were adopted in order to obtain the best fit between the observed and calculated flow values.

\[ Z = \sum_{i=1}^{NQ} \left[ q_s(i) - q_o(i) \right]^2 \]  

(Eq.13)

Where: \( Z \) = objective function; \( NQ \) = number of computed hydrograph ordinates; \( q_s \) = observed flows; \( q_o \) = calculated flows.

Model validation

The validation period, comprised between comprised between 1st of November 2004 and the 31st of December 2004, was selected to verify the reliability of the calibration.

Model performance evaluation

To evaluate the performance of the hydrological model it was used the hydroGOF package, within the R software environment, to calculate seven goodness-of-fit measures: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Nash-Sutcliffe Efficiency (NSE), Percent Bias (PBIAS), Coefficient of Determination (R2) and Kling-Gupta Efficiency (KGE). The MAE (Eq. 14) records in real units the level of agreement between the observed and modelled datasets, evaluating all the deviations from the observed values, without considering the magnitude of the event. It is a non-negative metric that indicates a perfect simulation when the result is zero.\(^{24}\)

\[ MAE = \frac{1}{n} \sum_{i=1}^{n} \left| Q_{obs} - Q_{sim} \right| \]  

(Eq.14)

Where: \( n \) = number of samples; \( Q_{obs} \) = observed flow; \( Q_{sim} \) = simulated flow.

The RMSE (Eq. 15) computes on squared differences the mean magnitude of the error between the observed and modelled values, in which the largest deviations contribute the most. Zero indicates a perfect fit between the simulated and observed data.\(^{24}\)

\[ RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left( Q_{obs} - Q_{sim} \right)^2} \]  

(Eq.15)

The RSR (Eq. 16) is calculated as the ratio of the RMSE and the standard deviation of measured data. It varies from 0 to \( \infty \), in which the lower the RSR value the better the hydrological model performance.\(^{25}\)

\[ RSR = \frac{\sum_{i=1}^{n} \left( Q_{obs} - Q_{sim} \right)^2}{\sqrt{\sum_{i=1}^{n} \left( Q_{obs} - Q_{obs} \right)^2}} \]  

(Eq.16)

Where \( Q_{obs} \) represents the mean of the observed data;

The NSE (Eq. 17) has been widely used to evaluate the performance of hydrological models.\(^{26}\) It ranges from \(-\infty \) and 1, with the NSE=1 indicating a perfect fit between simulated and observed data. Values between 0 and 1 are generally regarded as acceptable levels of performance, while values \( \leq 0 \) suggest unacceptable performance.\(^{24}\)

\[ NSE = 1 - \frac{\sum_{i=1}^{n} \left( Q_{obs} - Q_{sim} \right)^2}{\sum_{i=1}^{n} \left( Q_{obs} - Q_{obs} \right)^2} \]  

(Eq.17)

The PBIAS (Equation 18) evaluates the average tendency of the simulated values to be higher or lower than those observed. The ideal value of PBIAS is 0; positive values indicate a model bias toward underestimation; and negative values indicate a bias toward overestimation.\(^{27}\)

\[ PBIAS = 100 \times \frac{\sum_{i=1}^{n} \left( Q_{obs} - Q_{sim} \right)}{\sum_{i=1}^{n} \left( Q_{obs} \right)} \]  

(Eq.18)

General model evaluation guidelines for systematic quantification of accuracy in watershed simulations were created based on RSR, NSE and PBIAS performance ratings.\(^{24}\) These values are presented in Table 1.

<table>
<thead>
<tr>
<th>Performance Rating</th>
<th>RSR</th>
<th>NSE</th>
<th>PBIAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very good</td>
<td>0.00≤RSR≤0.50</td>
<td>0.75&lt;NSE≤1.00</td>
<td>PBIAS±10</td>
</tr>
<tr>
<td>Good</td>
<td>0.50&lt;RSR≤0.60</td>
<td>0.65&lt;NSE≤0.75</td>
<td>±10&lt;PBIAS&lt;±15</td>
</tr>
<tr>
<td>Satisfactory</td>
<td>0.60&lt;RSR≤0.70</td>
<td>0.50&lt;NSE≤0.65</td>
<td>±15&lt;PBIAS&lt;±25</td>
</tr>
<tr>
<td>Unsatisfactory</td>
<td>RSR&gt;0.70</td>
<td>NSE≤0.50</td>
<td>PBIAS±25</td>
</tr>
</tbody>
</table>

The R2 (Eq. 19) records as a ratio the proportion of the total statistical variance in the observed dataset that can be explained by the model. It varies from 0 (poor model) to 1 (perfect model).\(^{27}\)

\[ R^2 = \left[ \frac{\sum_{i=1}^{n} \left( Q_{obs} - Q_{sim} \right)^2}{\sum_{i=1}^{n} \left( Q_{obs} - Q_{obs} \right)^2} \right]^2 \]  

(Eq.19)

In which \( Q_{sim} \) represents the mean of the simulated data.

The KGE (Eq. 20) is based on the decomposition of the Nash-Sutcliffe coefficient and the Mean Squared Error (MSE) into three different terms that represent the bias, the correlation and a measure of relative variability in the simulated and observed values. Analogously to the NSE, the KGE ranges from \(-\infty \) to 1, with an ideal value of 1.\(^{24}\)

\[ KGE = 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \]  

(Eq.20)

Where \( r \) represents the Pearson’s correlation coefficient, \( \alpha \) is the ratio between the mean simulated and mean observed flows and \( \beta \) represents the bias.

Results and discussion

It was adopted five distinctive basin model configurations in the HEC-HMS desktop (Figure 12), in order to represent the watershed...
under study. In which: the subbasins are represented by a reservoir with an inlet and an outlet; lines represent the channels, with arrows indicating the flow direction; and a reservoir with two inlets and an outlet represents the junction of different elements. Based on the qualitative analysis of the simulated (Qsim) and observed (Qobs) hydrographs generated for the calibration and validation period (Figure 13), it can be noted that although the calculated hydrographs underestimate the peak flow magnitudes of the observed hydrographs, the overall fit of the model is good, indicating the model represents the hydrological processes in the watershed efficiently. It can also be noted the modeled hydrographs overlap each other, showing no significant change, thus demonstrating that different levels of spatial discretization did not interfere in the watershed’s outlet flow.

### Table 2 Evaluation parameters for calibration and validation periods

<table>
<thead>
<tr>
<th>Configuration</th>
<th>MAE</th>
<th>RMSE</th>
<th>RSR</th>
<th>NSE</th>
<th>PBIAS %</th>
<th>R²</th>
<th>KGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 subbasins</td>
<td>C</td>
<td>4.74</td>
<td>7.1</td>
<td>0.61</td>
<td>0.63</td>
<td>-11.8</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>V</td>
<td>8.95</td>
<td>14.42</td>
<td>0.53</td>
<td>0.72</td>
<td>5.6</td>
<td>0.72</td>
</tr>
<tr>
<td>4 subbasins</td>
<td>C</td>
<td>4.58</td>
<td>7.01</td>
<td>0.6</td>
<td>0.64</td>
<td>-11.2</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>V</td>
<td>9.03</td>
<td>14.49</td>
<td>0.53</td>
<td>0.71</td>
<td>7.5</td>
<td>0.72</td>
</tr>
<tr>
<td>8 subbasins</td>
<td>C</td>
<td>4.59</td>
<td>7.1</td>
<td>0.61</td>
<td>0.63</td>
<td>-11.6</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>V</td>
<td>8.81</td>
<td>14.34</td>
<td>0.52</td>
<td>0.72</td>
<td>5.6</td>
<td>0.72</td>
</tr>
<tr>
<td>16 subbasins</td>
<td>C</td>
<td>4.57</td>
<td>7.05</td>
<td>0.6</td>
<td>0.63</td>
<td>-11.3</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>V</td>
<td>8.85</td>
<td>14.4</td>
<td>0.53</td>
<td>0.72</td>
<td>5.8</td>
<td>0.72</td>
</tr>
<tr>
<td>32 subbasins</td>
<td>C</td>
<td>4.62</td>
<td>7.19</td>
<td>0.61</td>
<td>0.62</td>
<td>-11.8</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>V</td>
<td>8.79</td>
<td>14.22</td>
<td>0.52</td>
<td>0.72</td>
<td>5.1</td>
<td>0.73</td>
</tr>
</tbody>
</table>

Similar studies support the qualitative and quantitative analysis results, indicating the level of spatial discretization does not significantly affect the model’s performance. This can be explained by the watershed similar physiographic characteristics (such as slope, land use and occupation and soil types), that result in model’s standardization. As such, the process of watershed spatial discretization should always consider the presence of a reservoir in a tributary channel, the significant difference in the level of urbanization, or the significant difference in physiographic characteristics. Moreover, the homogeneity of outcomes no matter the subbasins configuration, may have been influenced by the intrinsic deterministic nature of the HMS models and the use of only one discharge station to evaluation the simulation results.

### Acknowledgments

None.

### Conflicts of interest

The authors declare that there is no conflict of interest.

### Funding

None.

### References

2. Braud JA. Impact of watershed delineation detail on hydrologic process modeling in low slope areas Recommended Citation. *Louisiana State University and Agricultural and Mechanical College*. 2009.

Citation:


