HRU aggregation and its effects on model outputs

Michael Strauch, Robert Otto, Martin Volk

2015 International SWAT Conference
Sardinia, June 24–26, 2015
HRU definition in ArcSWAT

Which threshold values should be used?

“For most applications, the default settings for land use threshold (20%) and soil threshold (10%) and slope threshold (20%) are adequate.” (Winchell et al., 2007, p. 126)
HRU definition in ArcSWAT

Theoretical example

a) No thresholds applied
HRU definition in ArcSWAT

Theoretical example

b) Absolute area thresholds applied (land use: 5 ha, soil: 0 ha, slope: 1.5 ha)
HRU definition in ArcSWAT

Theoretical example

c) Percentage thresholds applied
(land use: 30 %, soil: 30 %, slope: 30 %)

Area in ha

Subbasin
Land use
Soil
Slope

Categories
A
B
Discarded
**Error measure**

Average Absolute Error of Aggregation:

\[
\frac{(10 \text{ ha (Lu)} + 6 \text{ ha (Soil)} + 14 \text{ ha (Slope))}}{3} = 10 \text{ ha}
\]

\[
\frac{(4 \text{ ha (Lu)} + 2 \text{ ha (Soil)} + 4 \text{ ha (Slope))}}{3} = 3.34 \text{ ha}
\]

Relative Error:

\[
\frac{10 \text{ ha}}{50 \text{ ha}} = 0.2
\]

\[
\frac{3.34 \text{ ha}}{50 \text{ ha}} = 0.067
\]
Working hypotheses

1. **HRU number reduction can significantly distort input data.**

2. **The larger the input data error, the larger is the model error** (assuming the model equations are right).

3. **Different threshold combinations can lead to similar totals of HRUs with different input data errors** and thus different model errors.
Study catchment 1

Saale River Basin

Area: 23,666 km²
Elevation: 14 – 1144 m
Study catchment 2

Upper Amazon Basin

Area: 1,021,859 km²
Elevation: 25 – 6599 m
Saale River Basin
full HRU Model

HRUs (n=5,256)
Upper Amazon Basin
full HRU Model

HRUs (n=15,989)
Model calibration (full HRU model)

Saale

NSE: 0.84
PBIAS: 0.2

NSE: 0.82
PBIAS: 3.4
Model calibration (full HRU model)

Upper Amazon

NSE: 0.75
PBIAS: -6.5

NSE: 0.77
PBIAS: -2.6
Method:

Define HRUs with thousands of different threshold combinations using R

⇒ **Input for R-Script:** *hrus* table (project.mdb) without thresholds

⇒ the **R-script** then...

- defines HRUs using *ha* or % threshold values for all threshold combinations (land use / soil / slope) within a pre-defined range
- calculates the average Relative Error of Aggregation (*aREA*) compared to the full HRU distribution
HRU-Definition using R

- percentage method
- land use 0-39%, soil 0-39%, slope 0-39%
- increment 1% (64,000 solutions)
HRU-Definition using R

- absolute area method
- range 0-23,010 ha (Saale) / 0-74,100 ha (Amazon)
- increment 590 ha (Saale) / 1,900 ha (Amazon) = 64,000 solutions
HRU-Definition using R

- non-dominated solutions as best solutions
- absolute area method is preferable over percentage method
Study design to test effects on model outputs

- test 18 solutions out of 128,000 (all percentage & absolute area solutions)
  ⇒ S = standard values solution (20%/10%/20%), S’ = respective best solution
  ⇒ R1-R8 = stratified random samples, R1’-R8’ = respective best solutions
<table>
<thead>
<tr>
<th>Solution</th>
<th>Saale (2.37 × 10^6 ha)</th>
<th>Upper Amazon (1.02 × 10^8 ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Land use</td>
<td>Soil</td>
</tr>
<tr>
<td>F</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>S</td>
<td>20%</td>
<td>10%</td>
</tr>
<tr>
<td>R1</td>
<td>0%</td>
<td>3%</td>
</tr>
<tr>
<td>R2</td>
<td>0%</td>
<td>5%</td>
</tr>
<tr>
<td>R3</td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>R4</td>
<td>17%</td>
<td>1%</td>
</tr>
<tr>
<td>R5</td>
<td>1180 ha</td>
<td>590 ha</td>
</tr>
<tr>
<td>R6</td>
<td>9%</td>
<td>7%</td>
</tr>
<tr>
<td>R7</td>
<td>20060 ha</td>
<td>2360 ha</td>
</tr>
<tr>
<td>R8</td>
<td>31%</td>
<td>38%</td>
</tr>
<tr>
<td>S'</td>
<td>2360 ha</td>
<td>2360 ha</td>
</tr>
<tr>
<td>R1'</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>R2'</td>
<td>0 ha</td>
<td>0 ha</td>
</tr>
<tr>
<td>R3'</td>
<td>590 ha</td>
<td>590 ha</td>
</tr>
<tr>
<td>R4'</td>
<td>0 ha</td>
<td>590 ha</td>
</tr>
<tr>
<td>R5'</td>
<td>590 ha</td>
<td>1180 ha</td>
</tr>
<tr>
<td>R6'</td>
<td>1180 ha</td>
<td>2360 ha</td>
</tr>
<tr>
<td>R7'</td>
<td>3540 ha</td>
<td>7080 ha</td>
</tr>
<tr>
<td>R8'</td>
<td>4720 ha</td>
<td>8850 ha</td>
</tr>
</tbody>
</table>
Model outputs – example surface runoff

Saale

S - 20%, 10%, 20%

Average absolute deviation
S: 43.2 %
S': 5.4 %

F - full HRU model
Surface runoff (mm)
- <= 10
- > 10 - 20
- > 20 - 30
- > 30 - 40
- > 40 - 60
- > 60

S' - 2.360ha, 2.360ha, 2.360ha

Deviation from F (%)
- <= 2.5
- > 2.5 - 5
- > 5 - 10
- > 10 - 15
- > 15 - 20
- > 20 - 30
- > 30 - 40
- > 40 - 50
- > 50 - 75
- > 75 (hatched if positive)
Model outputs – example surface runoff

Upper Amazon

Average absolute deviation

S: 14.5%
S': 6.8%

F - full HRU model
Surface runoff (mm)
- <= 25
- > 25 - 50
- > 50 - 100
- > 100 - 200
- > 200 - 300
- > 300 - 400
- > 400 - 500
- > 500 - 1000
- > 1000 - 1500
- > 1500

S' - 7,600ha, 15,200ha, 15,200ha
Deviation from F (%)
- <= 2.5
- > 2.5 - 5
- > 5 - 10
- > 10 - 15
- > 15 - 20
- > 20 - 30
- > 30 - 40
- > 40 - 50
- > 50 - 75
- > 75
(hatched if positive)
Output ~ Input error:
Deviation from F averaged for basin ~ average Relative Error of Aggregation (aREA)

Surface runoff (%)

Saale

Upper Amazon

Lateral flow (%)

Groundwater flow (%)

$R^2 = 0.57$

$R^2 = 0.95$

$R^2 = 0.95$

$R^2 = 0.96$

$R^2 = 0.93$

$R^2 = 0.97$
Back to the hypotheses: conclusions

1. **HRU number reduction can significantly distort input data.**
   - aREA as a direct measure for input data distortion, e.g. 19 % average error for the Saale Basin when using standard thresholds (5256 => 359 HRUs)
   - Input data error increases exponentially with HRU number reduction

2. **The larger the input data error, the larger is the model error (assuming the model equations are right).**
   - Linear correlation between input data distortion and output error
   - Remarkably high errors for single flow components

3. **Different threshold combinations can lead to similar totals of HRUs with different input data errors and thus different model errors.**
   - Average input error can vary strongly (e.g. 5-20%) for similar totals of HRUs
   - Choose thresholds wisely! (=> use R-script!)
Questions?

Number of HRUs reduced from 6245
HRU aggregation

- a necessary evil for efficient simulations
- two ways for reducing complexity

1. Generalize (reclassify) single inputs

2. Discard HRUs below area threshold

computation speed vs. modeling error
How to measure input data distortion?

**average Relative Error of Aggregation (\(aREA\))**

\[
REA_{lu} = \frac{0.5 \times \sum_{i,k=1}^{I,K} |y_{agg,i,k} - y_{ref,i,k}|}{A}
\]

\(y\) = HRU area (\(agg\): aggregated, \(ref\): reference)

\(i\) = land use category

\(k\) = subbasin number

\(REA_{soil}\) and \(REA_{soil}\) are calculated analogously with \(i\) as soil and slope category, respectively.

\(aREA\) is an overall indicator for the level of aggregation and directly interpretable in any watershed.

\(aREA\) can range between 0 and 1. Multiplied with 100 it is the average percentage of the modified area for land use, soil, and slope on the total watershed area.
Model outputs – example streamflow

Saale

Streamflow dissimilarity to F
(reduction of Nash-Sutcliff-Efficiency)

- <= 0.01
- > 0.01 - 0.02
- > 0.02 - 0.03
- > 0.02 - 0.03
- > 0.03
Model outputs – example streamflow

Upper Amazon

Streamflow dissimilarity to F
(reduction of Nash-Sutcliffe-Efficiency)

- <= 0.01
- > 0.01 - 0.05
- > 0.05 - 0.1
- > 0.1 - 0.2
- > 0.2

S - 20%, 10%, 20%
S' - 7,600ha, 15,200ha, 15,200ha