Improvement of SWAT Auto-Calibration for Accuracies in High and Low Flow Regime using K-means Clustering Algorithm
Improvement of SWAT Auto-Calibration for Accuracies in High and Low Flow Regime Using K-means Clustering Algorithm

2010 International SWAT Conference
August 4, 2010

Hyunwoo Kang, Youn Shik Park, Jonggun Kim, Won Seok Jang, Ji Chul Ryu, Nam Won Kim, Dong Su Gong, Kyoung Jae Lim

Kangwon National University
**Introduction**

Nash and Sutcliff coefficient = NSE

NSE = 1 → Perfect match
Introduction

\[ NSE = 1 - \frac{\sum_{i=1}^{n} \left( y^M_{t_i} - y_{t_i} \right)^2}{\sum_{i=1}^{n} \left( y^M_{t_i} - \bar{y} \right)^2} \]
The coefficient of river regime in South Korea is very high because of intensive precipitation events during summer. This indicates that there are higher chances of greater NSE value although simulated data do not match measured data reasonably well for all flow regimes.

<table>
<thead>
<tr>
<th>Name of river</th>
<th>The coefficient of river regime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Almost 3,500 CMS</td>
<td>390</td>
</tr>
<tr>
<td>Thames River</td>
<td>8</td>
</tr>
<tr>
<td>Seine River</td>
<td>34</td>
</tr>
<tr>
<td>Rhein River</td>
<td>16</td>
</tr>
<tr>
<td>Missouri River</td>
<td>75</td>
</tr>
</tbody>
</table>


Almost 3,500 CMS
Introduction

- SWAT auto-calibration
- Parameter solution (Parasol)
- Best parameter
Parameter solution (Parasol)

Goal function = NSE

It has a many chances to make a problem!!
Introduction

K-means clustering algorithm

Simplest algorithm that solve the well known clustering problem
Objectives of this study

- To modify SWAT Auto-calibration using K-means clustering algorithm to improve accuracy in flow estimation for all flow regime

- To evaluate enhanced SWAT Auto-calibration module developed with K-means clustering algorithm by applying it to study watershed
Two SWAT auto-calibration modules

- SWAT auto-calibration using Parameter Solution (Parasol) in current SWAT 2005 engine → Original auto-calibration

- Enhanced SWAT Auto-calibration module using K-means algorithm → K-means auto-calibration
Study area

Area: 2,703 km²

Forest: 89.6%

Agricultural: 5.3%
Development of K-means auto-calibration module and simple GIS interface
Weakness of Nash-Sutcliffe coefficient in evaluating flow simulation

Park et al., (2007)

- NSE value for flow group I = 0.058
- NSE value for flow group II = -0.482

Unacceptable Results!!!!!!!
Modification of auto-calibration using K-means algorithm

Calculate NSE (Low flow)

Calculate NSE (High flow)

Validation of High and Low flow separately
Objective function in Parasol

Parameter Solution (Parasol)
Objective function => Sum of the squares of the residuals (SSQ)

\[
SSQ = \sum_{t_i=1}^{n} (y^M_{t_i} - y_{t_i})^2
\]

\[
NSE = 1 - \frac{SSQ}{\sum_{t_i=1}^{n} (y^M_{t_i} - \overline{y})^2}
\]

- \(y^M_{t_i} = \) Measured Data
- \(y_{t_i} = \) Simulated Data
- \(\overline{y} = \) Average of measured Data

\[
\sum_{t_i=1}^{n} (y^M_{t_i} - \overline{y})^2 = \overline{Y}
\]
Enhancement of Parasol algorithm using K-means clustering

- SWAT simulated data
- K-means Algorithm
- Clustering data into two groups using K-means algorithm
- Calculating NSE_High flow
- Calculating NSE_Low flow
- Estimation of new objective function considering NSE_H and NSE_L

Parasol

Observed data
Substitution of objective function considering NSE_H and NSE_L

Original clustering

\[ SSQ \]

\[ \bar{Y} \]

K-means clustering

\[ SSQ_H \]

\[ SSQ_L \]

\[ \bar{Y}_H \]

\[ \bar{Y}_L \]
Substitution of objective function considering **NSE_H** and **NSE_L**

\[
\frac{SSQ_{-H}}{Y_H} < 4.0 \quad \text{&} \quad \frac{SSQ_{-L}}{Y_L} < 4.0
\]

\[
\text{HSSQ} \quad \text{&} \quad \text{LSSQ}
\]

Accept if:

\[
\text{NSE}_H = 1 - \frac{SSQ_{-H}}{Y_H} > 0.6
\]

\[
\text{NSE}_L = 1 - \frac{SSQ_{-L}}{Y_L} > 0.6
\]

---

**Donigan and Love (2003)**

<table>
<thead>
<tr>
<th></th>
<th>Poor</th>
<th>Fair</th>
<th>Good</th>
<th>Very Good</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSE for Daily Simulation</td>
<td>&lt; 0.60</td>
<td>0.60 ~ 0.70</td>
<td>0.70 ~ 0.80</td>
<td>&gt; 0.80</td>
</tr>
</tbody>
</table>
Substitution of objective function considering \textbf{NSE\textsubscript{H}} and \textbf{NSE\textsubscript{L}}

New objective function

\[
\frac{SSQ\_H}{\bar{Y}_H} + \frac{SSQ\_L}{\bar{Y}_L}
\]

\textbf{NSE\textsubscript{H}} and \textbf{NSE\textsubscript{L}} together
Determination of the number of main outlet in watershed automatically
Application of K-means auto-calibration
## The parameters used in auto-calibration

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN2</td>
<td>SCS runoff curve number for Moisture condition II</td>
<td>35 ~ 98</td>
</tr>
<tr>
<td>AHPAH_BF</td>
<td>Baseflow alpha factor</td>
<td>0 ~ 1</td>
</tr>
<tr>
<td>SURLAG</td>
<td>Surface runoff lag time</td>
<td>1 ~ 24</td>
</tr>
<tr>
<td>CH_N</td>
<td>Manning’s “n” value for the main channel</td>
<td>-0.01 ~ 0.3</td>
</tr>
<tr>
<td>CH_K2</td>
<td>Effective hydraulic conductivity in main channel alluvium</td>
<td>-0.01 ~ 150</td>
</tr>
<tr>
<td>GW_DELAY</td>
<td>Groundwater delay</td>
<td>0 ~ 500</td>
</tr>
<tr>
<td>GW_REVAP</td>
<td>Groundwater “revap” coefficient</td>
<td>0.02 ~ 0.2</td>
</tr>
<tr>
<td>GWQMN</td>
<td>Threshold depth of water in the shallow aquifer required for return flow to occur</td>
<td>0 ~ 5000</td>
</tr>
<tr>
<td>SOL_AWC</td>
<td>Available water capacity of the soil layer</td>
<td>1 ~ 24</td>
</tr>
</tbody>
</table>
Comparison of SWAT auto-calibration

The NSE values of Total flow are similar.

K-means Auto-calibration

\[ \text{NSE} = 0.889 \]
### Comparison of SWAT auto-calibration

<table>
<thead>
<tr>
<th>Type of Auto Calibration</th>
<th>NSE - Total Flow</th>
<th>NSE - Low Flow</th>
<th>NSE - High Flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donigan and Love (2003)</td>
<td>0.699</td>
<td>0.460</td>
<td></td>
</tr>
<tr>
<td>NSE for Daily Simulation</td>
<td>0.889</td>
<td>0.608</td>
<td>0.604</td>
</tr>
<tr>
<td></td>
<td>&lt; 0.60</td>
<td>0.60 ~ 0.70</td>
<td>0.70 ~ 0.80</td>
</tr>
</tbody>
</table>
Conclusions

- The NSE values from K-means auto-calibration are 0.608 and 0.604 for low and high flow regimes.

- However the NSE value for high flow from original auto-calibration is 0.460 (Poor calibration result).

  Donigan and Love (2003)
Conclusions

- SWAT simulated flow using original auto-calibration may not match measured flow data because objective function in original auto-calibration is affected by big number.

- With K-means auto-calibration, developed in this study, would provide better estimation for all flow regimes.
Thank you for your Attention!!

Flow duration curve

Huge difference!!

NSE = 0.683