Application of WEHY-HCM for Modeling Interactive Atmospheric-Hydrologic Processes at Watershed Scale to a Sparsely Gauged Watershed

Suhyung Jang 1,*, Shuichi Kure 2, Noriaki Ohara 3, M. Levent Kavvas 4, Z. Q. Chen 5, Kara J. Carr 4 and Michael L. Anderson 6

1 Water Resources Research Center, K-Water Institute, Daejeon 34045, Korea
2 Department of Environmental Engineering, Toyama Prefectural University, Toyama 939-0398, Japan; kure@pu-toyama.ac.jp
3 Department of Civil and Architectural Engineering, University of Wyoming, Laramie, WY 82071, USA; nohara1@uwyo.edu
4 Department of Civil and Environmental Engineering, University of California, Davis, CA 95616, USA; mlkavvas@ucdavis.edu (M.L.K.); kjcarr@ucdavis.edu (K.J.C.)
5 California Department of Water Resources, Sacramento, CA 95814, USA; zchen@water.ca.gov
6 California Department of Water Resources, Sacramento, CA 95821, USA; michael.l.anderson@water.ca.gov

* Correspondence: kwaterjang@kwater.or.kr; Tel.: +82–42-870–7413

Received: 29 June 2017; Accepted: 30 August 2017; Published: 1 September 2017

Abstract: A lack of observations within watersheds can make the production of streamflow data via hydrologic models a big challenge. This study evaluates the model performance of the Watershed Environmental Hydrology Hydro-Climate Model (WEHY-HCM), reproducing streamflow in a sparsely gauged watershed. The fifth generation mesoscale model (MM5) is utilized within WEHY-HCM as an atmospheric module coupling with its process-based hydrologic module, WEHY. The WEHY-HCM is set up over a sparsely gauged watershed and the spatially downscaled reconstructed atmospheric data to a 3-km horizontal grid resolution with an hourly time increment, is obtained by the fifth generation mesoscale model (MM5) from NCAR/NCEP global reanalysis data (reanalysis I). Hydrologic simulations by WEHY-HCM were applied to the Upper Putah Creek watershed based on the reconstructed atmospheric data and the estimated WEHY model parameters. The simulation results of WEHY-HCM were evaluated by means of statistical tests for both calibration and validation periods. The results of statistical tests performed using observed and simulated values indicated that the model performance can be considered as exhibiting an acceptable accuracy during both calibration and validation periods. The spatial maps of the evapotranspiration rate and runoff volume showed that the WEHY-HCM can represent a sparsely gauged watershed with unique topography well. This study found that the WEHY-HCM can be a useful tool to simulate the hydrologic processes in a sparsely gauged watershed.

Keywords: sparsely gauged watershed; runoff; atmospheric-hydrologic process; WEHY-HCM

1. Introduction

In most watersheds, observations are not sufficient for watershed assessments due to sparsely located observation stations and observation records with a relatively coarse time scale (normally daily and monthly). Such limited data are not suitable for hydrologic applications which require long-term continuous simulations in order to be utilized for the design of flood defenses and general infrastructures as well as the strategic planning of water resources, agriculture, ecosystems, and water-related disasters [1,2].
One can develop the planning and management of floods, droughts, water quality, and water resources for those watersheds (ungauged or sparsely gauged) by means of using a physically-based distributed hydrologic model (e.g., [3–5]). Physically-based distributed hydrologic models can incorporate hydrological processes within a natural watershed due to the complicated effects from heterogeneity in land surface properties. However, the physically-based distributed models totally rely on atmospheric data (e.g., temperature, precipitation, relative humidity, wind fields, radiation, etc.). As such, it is still difficult and challenging to model in ungauged or sparsely gauged watersheds [6–8].

Recently, satellite, radar, and multi-sensor-based precipitation data with fine spatial and time resolutions have been made publically available across the globe. For example, JAXA (Japan Aerospace Exploration Agency) provides satellite-driven precipitation data such as AMSR-2 (advanced microwave scanning radiometer 2 available at [9]) and GSMaP (global satellite mapping of precipitation available at [10]). NASA (National Aeronautics and Space Administration) provides integrated multi-satellite precipitation data such as TMPA (TRMM multi-satellite precipitation analysis available at [11]) and IMERG (integrated multi-satellite retrievals for GPM available at [12]). The NCEP (National Centers for Environmental Prediction) provides multi-sensor (radar and rain gauge) precipitation products available at [13]. Many studies have obtained hydrological information from those satellite and multi-sensor products in order to apply them in ungauged or sparsely gauged watersheds [14–18]. However, these satellite, radar, and multi-sensor-based precipitation data were mostly made available after the 2000s, and can only cover the historical period. Additionally, spatially distributed atmospheric data other than precipitation, such as radiation, wind fields, and relative humidity [19], are very limited or unavailable, while they provide important information on the application of land surface hydrological modeling [18].

The coupling of a physically-based distributed model and a regional climate model (RCM) is appropriate for sparsely gauged or ungauged watersheds because RCMs can account for the physical interactions of the atmosphere and land surface processes in order to create various atmospheric variables (e.g., temperature, precipitation, relative humidity, wind fields, radiation) at fine spatio-temporal resolutions considering the heterogeneity in topography, soil, vegetation, and atmospheric variables over a watershed [18,20,21]. From this perspective, the WEHY-HCM (Watershed Environmental Hydrology Hydro-Climate Model, [18,21]), which is a coupled model of atmospheric and hydrologic processes at the watershed scale, based on the WEHY [3,18,20–24] model coupled with the fifth generation mesoscale model (MM5, [25]), can be usefully applied to sparsely gauged or ungauged watersheds.

This study evaluates the utility of WEHY-HCM through the estimation of runoff from a sparsely gauged watershed. The application is unique in that the watershed is a small medium-size watershed, with unique geomorphological and vegetation characteristics, while mercury from abandoned mines over the study area is of great concern. Furthermore, the surface waters flow swiftly downstream during flood events and transport a variety of contaminants. Therefore, it is important to deliberate the resolution spatial and temporal spatio-temporal variability of land surface properties in order to estimate runoff and the associated transport of nonpoint source contamination from this watershed.

2. Study Area

The Upper Putah Creek (drainage area of 663 km$^2$) watershed is the major water resource for Monticello Dam, operated by Solano County Water Agency/Solano Irrigation District (see Figure 1). The management of water quality and quantity from this watershed is important because the Lake Berryessa reservoir, located downstream of the watershed and upstream of Monticello Dam, irrigates approximately 388 km$^2$ of land. Management is particularly important during the late summer, when the water supply is often short. Although Lake Berryessa and the downstream reaches of Putah Creek offer fishing opportunities, mercury from abandoned mines is a major concern in the Upper Putah Creek watershed where there are over 100 abandoned mine sites, most of which have not been adequately remediated. Hence, a rigorous assessment by an appropriate model is required to support the watershed managements and plans for the target watershed.
The Upper Putah Creek watershed has a unique topography and a variety of vegetation (see Figures 1–3). The central area, surrounded by mountain ridges, comprises most of the low-lying terrain. This area is mainly covered by local meadows and croplands with regions of open field and rather flat hillslope. The overland flow and subsurface flow from mountainous areas are stored in a groundwater aquifer in the central area and serve as an important resource for this watershed and Lake Berryessa. The region stretching from the northwestern to southwestern boundaries of the watershed is highly elevated and heavily vegetated, comprised of mostly hardwood and conifer forests. The tree density is more than 60% and its individual hillslopes are very steep. These areas account for the majority of precipitation over the watershed, and a large amount of water from these areas swiftly flows toward the groundwater aquifer in the central area. In turn, the subsurface flow plays an important role in variable-source-area runoff processes in high elevation and heavy vegetation areas [26]. Meanwhile, the region stretching from the northeastern to southeastern boundaries of the watershed is highly elevated, but the vegetation is sparse (mostly shrub). The tree density is less than 10% and its individual hillslopes are steep. In this area, both variable-source-area runoff processes and hortonian runoff processes are present [27,28].

Figure 1. Location, observation sites, and digital elevation model (DEM) of the Upper Putah Creek watershed.

Figure 2. Surface slope at individual hillslopes.
3. Reconstruction of Historical Atmospheric Conditions

The processes of model components (an atmospheric model component, a land surface model component, a hillslope process model component, and a coupled groundwater flow-river channel flow routing model component) in WEHY-HCM are mutually interacting [21]. Spatially distributed atmospheric data (e.g., temperature, precipitation, relative humidity, wind fields, radiation) with a fine time interval are required for the application of the WEHY model; however, there are no weather gauges with hourly time intervals and a long duration within the study watershed. For this reason, the historical atmospheric data were reconstructed based on the land surface and atmospheric model components of the WEHY-HCM during both calibration and validation periods at hourly intervals and at a 3-km horizontal grid resolution over the watershed.

3.1. Methodology for Generating a Reconstructed Atmospheric Data

The MM5 as an atmospheric component in WEHY-HCM was set up over the Upper Putah Creek watershed domain, downscaling the National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR) reanalysis data (NCEP/NCAR reanalysis data in this study) from a 210 km by 210 km to a 3 km by 3 km horizontal grid spacing, in order to obtain the fine scale of reconstructed atmospheric data (e.g., precipitation, temperature, radiation, relative humidity) as an input for a hydrologic module, WEHY in WEHY-HCM. The NCEP/NCAR reanalysis data are appropriate for testing the model performance by means of comparing them against observations because NCEP/NCAR reanalysis data are the results of assimilating observations from in situ sources such as satellites, aircraft, pibal, rawinsonde, ships, and land surface [29,30].

3.2. Bias Correction and Reconstruction of Atmospheric Data

Although the NCAR/NCEP reanalysis data are assimilated by observed values, and the MM5 downscales these assimilated data to regional scale data, the reconstructed atmospheric data may have biases. Bias-correction should be conducted to improve the quality of the reconstructed atmospheric data, especially for precipitation.

Bias correction of precipitation between ground observations from the California Data Exchange Center (CDEC, [31]) and reconstructed values downscaled by the MM5 simulations at five rain-gauge stations (Whispering Pines, WSP; Guenoc, GNC; Napa valley basin, NVB; Homestake mining, HSM; and Clear lake highlands, CLH) was performed by spatially interpolating them over the entire simulation grids with the inverse distance squared method. Bias correction was conducted at daily time intervals to coincide with the available long-term daily precipitation data from the five rain gauges (see Figure 1). Because the WEHY-HCM module simulates the hydrologic processes in hourly intervals, the simulation results were post-processed to create daily data.
Figure 4 shows the observed and bias corrected reconstructed values of daily precipitation at five rain gauges (CLH, GNC, HSM, NVB, WSP), from October 2001 to September 2006. As shown in Figure 4, the simulated and observed daily precipitation values are similar with respect to the timing and magnitude of peak precipitation. Figure 5 shows the observed and reconstructed daily mean air temperature at KNO CDEC weather gauge from October 2001 to September 2006. This was the only gauge used for air temperature validation, and the reconstructed data were not bias corrected. As shown in Figures 4 and 5, the reconstructed atmospheric data by the MM5 represent the observed values well at all comparison sites.

As an example of the distributed atmospheric components created by the model, Figure 6 shows reconstructed values of precipitation, air temperature, wind speed, and solar radiation over the study watershed for February 2005. These reconstructed atmospheric components incorporate the influences of the heterogeneous characteristics such as land cover and topography over the watershed (see Figures 1–3). For example, the amount of precipitation is diverse over the watershed according to the elevation due to the orographic effect. The highly vegetated areas with hardwood and conifer (areas from the northwestern to southwestern parts) show relatively high precipitation fields compared to the rarely vegetated areas (areas from the northeastern to southeastern parts). The wind speed is fast at the mountain ridges, but the wind speed at the central area of the watershed is slow, as expected due to the flat terrain surrounded by high elevation mountains.

The reconstructed atmospheric data in Figure 6 show a good performance of spatial variability over the study watershed. Such spatial variability of atmospheric variables may not be captured by the observation stations due to the lack of observation stations in most watersheds. In a sparsely gauged watershed like Upper Putah Creek, observation stations may not capture such spatial variability of atmospheric variables due to the lack of observation stations. Furthermore, ground observation stations are usually located at low elevation areas where they can be easily accessed from the road, managed, and operated.

Figure 4. Cont.
Figure 4. Observed and simulated daily precipitation for the five sites in the Upper Putah Creek watershed for (a) Clear lake highlands; (b) Guenoc valley; (c) Homestake mining; (d) Napa valley basin; and (e) Whispering pines.
Figure 5. Observed and simulated daily mean air temperature at Knoxville creek gauge in the Upper Putah Creek watershed.

Figure 6. Sample of distributed atmospheric components (February 2005) from the database, (a) precipitation (mm); (b) air temperature (°C); (c) wind speed (m/s); and (d) solar radiation (W/m²).
4. Implementation of WEHY Model over Upper Putah Creek Watershed

4.1. Delineation of Stream Reaches and Model Computational Units

The upscaled hydrologic conservation equations utilized in the WEHY model enable us to describe the heterogeneity within the natural watersheds [32]. The WEHY model subdivides a watershed first into model computational units (MCUs, [23]). These MCUs are either individual hillslopes or first-order sub-watersheds within a watershed in order to account for the surface and subsurface hillslope hydrologic processes [23,32]. Thus, the surface and subsurface flows of each individual hillslope as its MCU in the WEHY model are directly drained into an adjacent stream reach, and routed further to the watershed outlet [23]. The details of delineation of MCUs and stream reaches are described in [22,23]. As shown in Figure 7, the stream reaches and MCUs (56 and 112, respectively) were delineated for the study watershed using the USGS (U.S. Geological Survey) NED (National Elevation Dataset, [33]) in 1-arc-second (approximately 30 m horizontal grid spacing).

![Figure 7. Delineation of (a) stream reaches and (b) MCUs for the Upper Putah Creek watershed.](image)

4.2. Estimation of Geomorphologic, Land Surface, and Soil Hydraulic Parameters

The WEHY model requires detailed topographic information for each MCU and stream reach. These geomorphologic parameters are estimated based on the prepared GIS database, and define the flow domains and the configurations of rill and interrill areas for MCUs of the WEHY model over the study watershed.

The WEHY model also requires land surface parameters (e.g., roughness height, albedo, emissivity, leaf area index, and vegetation root depth) and soil hydraulic parameters (e.g., bubbling pressure, soil porosity, soil depth, and saturated hydraulic conductivity) with time and spatial variability describing the heterogeneity within the watershed. Land surface parameters were estimated using the multi-source land cover data from California Wildlife Habitat Relationships (CWHR, available at [34]) and reference information from [35–37]. The monthly mean Leaf Area Index (LAI) was obtained from MODIS satellite driven data available at [38]. Soil parameters were estimated using the Soil Survey Geographic (SSURGO) Database (Natural Resources Conservation Service, NRCS available at [39]) and reference information from [37,40–42].

These parameters were acquired and stored in the Upper Putah Creek watershed GIS database. Figure 8 shows samples of the land surface and soil hydraulic parameters from the GIS database.
5. Discussion of WEHY-HCM Application Results

For the hydrologic modeling over the study watershed, the hydrologic model components in WEHY-HCM were implemented using the estimated WEHY model parameters and the reconstructed atmospheric data. The USGS Guenoc (PCG) gauge (see Figure 1) is the only gauge for stream flow data which was used for model calibration and validation.

The performance of the WEHY-HCM was evaluated based on the recommendations from [43], such as graphical comparison, deviation of runoff volume \( (D_v) \) [44], and the Nash-Sutcliffe coefficient \( (E_{NS}) \) [45]. Additional appropriate evaluation methods, such as the Chi-square goodness-of-fit test, root mean square error \( (RMSE) \), standard deviation \( (STDEV) \), and correlation coefficient \( (R) \) were applied.

The model acceptance limits and performance were applied, as suggested by [46], in which the \( D_v \) of the simulated value from the observed value is used to decide the under prediction or over prediction limit for the model simulation. The acceptable accuracy limit of \( D_v \) was equal to or less than 20% and under prediction or over prediction was considered as low (slight), moderate, and severe when the limit was \( \leq 10\% \), 10–20%, and 20–30% of the measured values, respectively [47]. \( E_{NS} \) and \( R \) values can vary from 0 to 1, with 1 indicating a perfect fit. An \( RMSE \) value of zero, meaning that the model predicts observations with perfect accuracy, is the ideal, but it is practically never possible. The Chi-square goodness-of-fit test indicates how well the simulations fit to a set of observations.

Figures 9 and 10 show the model calibration results (October 2005–September 2006) and the model validation results (October 2001–September 2004). These figures compare graphically observed and simulated hourly discharge values at the USGS Guenoc (PCG) gauge. As shown in Figures 9 and 10, the model-simulated discharge matched the corresponding observed discharge at the Guenoc USGS
The simulated and observed hourly flow discharge values are similar in the calibration and validation periods with respect to the rising and receding limbs of both hydrographs, as are the timing and magnitude of peak discharge.

Table 1 shows the statistical test results of the observed and simulated daily discharge for both the calibration and validation periods. The mean and STDEV values show that the model produced slightly more variability in the simulated values. The R values of 0.95 and 0.92, ENS values of 0.89 and 0.84, small values of RMSE (8.66 and 7.53), for the calibration and validation, respectively, indicate a close correspondence between the simulated and observed discharge values. The overall De (16.68% and 2.77%) is within the acceptable limits of accuracy. Furthermore, the Chi-square goodness-of-fit tests were passed for both calibration and validation periods.

Figure 9. Calibration results of hourly flow discharge at USGS Guenoc (PCG) station during October 2005–September 2006.

Figure 10. Validation results of hourly flow discharge at USGS Guenoc (PCG) station during October 2001–September 2004.

Evapotranspiration (ET) is the amount of water that is evaporated and transpired from the soil surface and vegetation. Therefore, ET depends on not only vegetation and soil moisture availability (e.g., LAI, root depth, vegetation height, etc.), but also atmospheric conditions (e.g., radiation, air temperature, humidity, wind speed, precipitation, etc.). Vegetation, soil characteristics, and atmospheric data are plugged into WEHY-HCM and the estimation of ET values is from the concept of
atmospheric conditions and moisture availability in WEHY-HCM, as described in detail in previous works [48–50]. Although most precipitation occurs during the wet period (October–March), the main parameter values for the ET estimation are quite small during the wet period. Most ET occurs during the dry period (April–September) due to the high values of parameters such as air temperature, solar radiation, wind speed, and LAI (leaf area index).

Table 1. Statistical test values of observed and simulated daily discharge for model calibration and validation.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (observed/simulated)</td>
<td>14.00/11.67</td>
<td>6.63/6.45</td>
</tr>
<tr>
<td>STDEV (observed/simulated)</td>
<td>26.68/27.00</td>
<td>18.94/19.26</td>
</tr>
<tr>
<td>RMSE</td>
<td>8.66</td>
<td>7.93</td>
</tr>
<tr>
<td>Correlation coefficient, R</td>
<td>0.95</td>
<td>0.92</td>
</tr>
<tr>
<td>Nash-Sutcliffe efficiency, ENS</td>
<td>0.89</td>
<td>0.84</td>
</tr>
<tr>
<td>Volume deviation, Dv (%)</td>
<td>16.68</td>
<td>2.77</td>
</tr>
<tr>
<td>Chi-square test, α = 0.05 (critical/calculated)</td>
<td>18.31/16.08</td>
<td>18.31/12.75</td>
</tr>
</tbody>
</table>

Figure 11 shows the annual ET values simulated by WEHY-HCM during October 2005–September 2006 at each MCU. Higher ET values are found in the central region which has a deep soil zone, is flat, and has a high air temperature and high solar radiation compared to other areas within the watershed. During the wet period, water from the mountainous areas may therefore be stored in the soil in the central area of watershed, allowing large quantities of water to be evapotranspired from the soil surface and vegetation during the dry period due to water availability, a high air temperature, and solar radiation in this area.

Figure 11. Evapotranspiration simulated by WEHY-HCM during October 2005–September 2006 at each model computational unit (MCU).

Figure 12 shows the annual runoff volume simulated by WEHY-HCM during October 2005–September 2006 at each MCU. Higher runoff volumes are found in the mountainous areas in the region spanning from the northwest to southwest during water year 2006, due to the high elevation, heavy vegetative cover, high vegetation density (mostly hardwood and conifer), and high precipitation. The region spanning from the northeastern to southeastern watershed boundaries is also highly elevated and mountainous; however, the vegetation is sparse and the magnitude of precipitation is relatively small. For this reason, the runoff volume is relatively small compared to the region spanning from the northwest to southwest.

Figure 12. Runoff volume simulated by WEHY-HCM during October 2005–September 2006 at each model computational unit (MCU).
Physically-based distributed hydrologic models are commonly used for hydrological applications as tools for water resources and environmental assessments. However, distributed hydrologic models without an atmospheric component are limited by their inability to represent sparsely gauged or ungauged watersheds. It is difficult to account for modeling interactive hydrologic processes on the land surface without a representation of the atmospheric processes. Hydrologic models generally have various conceptual components and point-scale governing equations which have a limited ability to interpret the spatial and temporal variability needed when considering the interactions in both atmospheric and hydrologic processes over a heterogeneous land surface. Meanwhile, the WEHY-HCM can account for the land surface heterogeneities by means of the upscaled conservation equations and parameters estimated directly from land and soil properties. In this respect, the WEHY-HCM is able to create the historical atmospheric and hydrologic data at fine spatio-temporal resolutions and can be a useful tool to simulate hydrological processes in areas with heterogeneous topography, heterogeneous land use and land cover, and even for sparsely gauged watershed or ungauged watersheds.

6. Conclusions

In order to explore the utility of applying WEHY-HCM to a sparsely gauged watershed with unique geomorphological and vegetation characteristics, the WEHY-HCM was implemented for the Upper Putah Creek watershed. The historical atmospheric data were reconstructed by the atmospheric model components in WEHY-HCM from NCAR/NCEP global reanalysis data at a 3-km horizontal grid resolution and hourly time increment. Then, the WEHY hydrologic module in WEHY-HCM was implemented using the reconstructed atmospheric data and the estimated WEHY model parameters for applying it to the target watershed in northern California. The graphical comparisons and statistical tests were conducted to evaluate the performance of WEHY-HCM using model simulation results and corresponding historical observations during both calibration and validation periods. The model performed well with respect to daily discharge, and the model simulation results were considered as exhibiting an acceptable accuracy based on the statistical tests during both calibration and validation periods. Spatial maps of evapotranspiration and runoff volume from the WEHY-HCM can account for the effect of heterogeneity within a watershed like the Upper Putah Creek watershed.

The above statistical test results for the calibration and validation periods indicated that the model simulations are within an acceptable accuracy. The spatial maps of evapotranspiration and runoff volume during water year 2006 show that the WEHY-HCM can represent a heterogeneous watershed such as the Upper Putah Creek watershed well.

Figure 12. Runoff volume simulated by WEHY-HCM during October 2005–September 2006 at each model computational unit (MCU).
Acknowledgments: This work is partially supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by Ministry of Land, Infrastructure and Transport (Grant No. 17AWMP-B0830660-04).

Author Contributions: Z. Q. Chen set up the regional climate model; Suhyung Jang and Shuichi Kure reconstructed and simulated the atmospheric and hydrological data; Suhyung Jang wrote the paper with support from Noriaki Ohara, Kara J. Carr, Michael L. Anderson, and M. Levent Kavvas.

Conflicts of Interest: The authors declare no conflict of interest.

References


47. Mishra, A.; Kar, S.; Singh, V.P. Determination of runoff and sediment yield from a small watershed in sub-humid subtropics using the HSPF model. *Hydrol. Process.* 2007, 21, 3035–3045. [CrossRef]


© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).