Predicting the Discharge of Global Rivers

BART NIJSSEN, GREG M. O’DONNELL, DENNIS P. LETTENMAIER
University of Washington, Seattle, Washington

DAG LOHMANN* AND ERIC F. WOOD
Princeton University, Princeton, New Jersey

(Manuscript received 25 April 2000, in final form 3 November 2000)

ABSTRACT

The ability to simulate coupled energy and water fluxes over large continental river basins, in particular streamflow, was largely nonexistent a decade ago. Since then, macroscale hydrological models (MHMs) have been developed, which predict such fluxes at continental and subcontinental scales. Because the runoff formulation in MHMs must be parameterized because of the large spatial scale at which they are implemented, some calibration of model parameters is inevitably necessary. However, calibration is a time-consuming process and quickly becomes infeasible when the modeled area or the number of basins increases. A methodology for model parameter transfer is described that limits the number of basins requiring direct calibration. Parameters initially were estimated for nine large river basins. As a first attempt to transfer parameters, the global land area was grouped by climate zone, and model parameters were transferred within zones. The transferred parameters were then used to simulate the water balance in 17 other continental river basins. Although the parameter transfer approach did not reduce the bias and root-mean-square error (rmse) for each individual basin, in aggregate the transferred parameters reduced the relative (monthly) rmse from 121% to 96% and the mean bias from 41% to 36%. Subsequent direct calibration of all basins further reduced the relative rmse to an average of 70% and the bias to 12%. After transferring the parameters globally, the mean annual global runoff increased 9.4% and evapotranspiration decreased by 5.0% in comparison with an earlier global simulation using uncalibrated parameters. On a continental basis, the changes in runoff and evapotranspiration were much larger. A diagnosis of simulation errors for four basins with particularly poor results showed that most of the error was attributable to bias in the Global Precipitation Climatology Project precipitation products used to drive the MHM.

1. Introduction

Macroscale hydrological models (MHMs), which model the land surface hydrologic dynamics of continental-scale river basins, have rapidly developed over the past decade (e.g., Russell and Miller 1990; Kuhl and Miller 1992; Miller et al. 1994; Dumenil and Todini 1992; Sausen et al. 1994; Liston et al. 1994; Abdulla et al. 1996; Nijssen et al. 1997; Wood et al. 1997; Kite 1998). MHMs are closely related to the land surface parameterization schemes (LSPs) in general circulation models, but focus more on modeling of runoff and streamflow and their interaction with other terms in the land surface water budget. They can act as a link between global atmospheric models and water resource systems on large spatial scales and long (seasonal to interannual) timescales (e.g., Hamlet and Lettenmaier 1999).

One practical problem in application of MHMs (and LSPs) is the determination of model parameters. Most MHMs are a hybrid of physically based and conceptual components. The energy exchange at the atmosphere–land surface interface is usually based on physical principles, while the runoff generation mechanisms tend to be more conceptual in nature. For example, the Variable Infiltration Capacity (VIC) model (Liang et al. 1994; Nijssen et al. 1997) uses physically based formulations for the calculation of the sensible and latent heat fluxes, but uses the conceptual ARNO baseflow model (Todini 1996) to simulate runoff generation from the deepest soil layer, and a conceptual scheme to represent the spatial variability in infiltration capacity (Zhao et al. 1980) and hence production of surface runoff. Determination of parameters can therefore be separated into two parts. The parameters for the physically based part of the model can, in principle, be determined through direct observation. In practice, given the large areas over which MHMs are applied, this is a challenge. Most mod-

* Current affiliation: National Centers for Environmental Predictions, Camp Springs, Maryland.

Corresponding author address: Dennis P. Lettenmaier, Department of Civil and Environmental Engineering, Box 352700, University of Washington, Seattle, WA 98195-2700.
E-mail: dennisl@u.washington.edu

© 2001 American Meteorological Society
models use some form of lookup-table approach in which land surface attributes are kept constant within each land surface class, and vary only between different classes. This simplifies the application and greatly reduces the number of parameters that need to be specified.

It is less obvious whether or how the parameters for the conceptual parts of the model can be specified a priori. Traditionally, conceptual hydrological models are tailored to a specific application through a process of calibration (e.g., Linsley et al. 1986, chapter 12). In this process model parameters are adjusted in successive model simulations until the model output matches observations to within a previously determined error criterion. Although this is a relatively straightforward process for a small river basin, the approach tends to be piecemeal and labor intensive for large river basins. For example, application of the National Weather Service River Forecast System to the Columbia River in the Pacific Northwest required calibration of the model to over 100 subbasins (Riverside Technologies, Inc., Harza Engineering 1994).

The Project for Intercomparison of Land surface Parameterization Schemes (PILPS) Phase 2(c) (Wood et al. 1998) was designed to provide some insight into the effects of calibration on MHMs and LSPs applied over a large river basin. PILPS 2(c) evaluated the ability of 16 models run in offline mode to reproduce observed water and energy fluxes over the Arkansas–Red River basin in the Southern Great Plains region of the United States (Liang et al. 1998; Lohmann et al. 1998c). The runoff produced by each of the 16 models was then routed to the basin outlets using the routing model of Lohmann et al. (1998a) as a postprocessor. As part of PILPS 2(c), the effect of calibration on simulated streamflow was studied for 11 of the 16 participating LSPs (Wood et al. 1998). Selected parameters of each model were calibrated (by the model developer) for a number of small basins, and subsequently the parameters were transferred to the entire Arkansas–Red River basin, which consisted of 61 $1^\circ \times 1^\circ$ grid cells. A set of “verification basins,” of similar size to those used for calibration, was used in a blind experiment to determine the effects of calibration on model performance. It was left to the participants to choose the method for transferring the model parameters. A key result of PILPS 2(c) was that model performance improved significantly for those models that were able to use streamflow information effectively for model calibration. Model performance for the validation catchments was generally much better for those models that calibrated than for those that did not. Consequently, Wood et al. (1998) suggested that there is value in using catchment data to calibrate the parameters of land surface schemes, even for so-called physically based models, whose parameter values, in theory, can be obtained through direct observation. Early work on the transfer of calibrated model parameters from small calibration catchments to continental-scale river basins by Abdulla and Lettenmaier (1997a,b) has demonstrated the potential for such approaches in continental hydrology. The results from these two studies provide the motivation for the current work.

2. Objectives and modeling strategy

The objectives of this paper are twofold. The first objective is to evaluate the change in model performance achieved when a simple parameter transfer method is used, based on model calibration for a limited number of large river basins. The second objective is to provide a dataset of global runoff estimates and accompanying components of the land surface hydrological cycle, such as evapotranspiration, soil moisture storage, and snow water equivalent. The resulting dataset can be used in large-scale water management studies, or as a benchmark for comparison with runoff estimates from large-scale climate simulations in which the land surface hydrology is forced with model meteorology (e.g., Maurer et al. 2001).

The model setup, model parameters, and meteorological forcing dataset from Nijssen et al. (2001) were taken as the starting point for this study. Nijssen et al. (2001) simulated the global, land surface hydrological cycle based on observed meteorological forcings at a $2^\circ \times 2^\circ$ resolution for the period 1979–93, using the VIC model. No calibration was performed in that study.

The selection of continental-scale river basins used in this study was based on the dataset from Graham et al. (1999) as modified by Comanor et al. (2000). To assess the effects of calibration and parameter transfer, the selected basins were partitioned into two groups. The basins in the first set (primary basins) were calibrated using an approach described in section 5. The primary basins represent a range of climate and vegetation types. Calibration focused on matching observed and simulated streamflow. Streamflow is arguably the most easily measured and best documented component of the regional surface water balance. It acts as an integrator of the runoff response from all parts of a river basin. As such, streamflow offers the opportunity to verify the surface water balance simulated by MHMs. Other model-predicted water storage and flux components (e.g., soil water storage, snow cover, evapotranspiration) are rarely observed at spatial and temporal scales suitable for direct comparison with the output from MHMs. Some observational soil moisture datasets exist, which have been used for comparison with GCMs and macroscale hydrological simulations (e.g., Robock et al. 1998, 2000), but both their temporal and spatial extent are limited. On the other hand, streamflow is available for many major rivers, often for extended periods of time. We do, however, show later in this paper that even streamflow observations are often not readily available for many global rivers for the period since 1984.

Basins in the second set (secondary basins) were ini-
tially modeled using parameters transferred from the calibrated primary basins using the parameter transfer method described in section 5. Observed and simulated streamflow for these basins were compared to determine the effectiveness of the parameter transfer method.

The secondary basins were then calibrated in a second stage of the study. This second stage served two purposes. First, calibration of the secondary basins served to further evaluate the effectiveness of the parameter transfer process. Ideally, this calibration should result in minimal improvement of model results, which would indicate that the parameter transfer process was highly successful. Second, the second stage calibration ensured that the global estimates of water balance components were the best possible, given the model and meteorological forcings. In a final step, the calibrated parameters from all basins were transferred to the remaining global land surface grid cells to allow estimation of the continental and global water balance.

3. Model implementation and data sources

The VIC model (e.g., Liang et al. 1994, 1996) is an MHM that has been used in a number of modeling studies of large river basins (e.g., Abdulla et al. 1996; Bowling et al. 2000; Lohmann et al. 1998b; Maurer et al. 2000; Nijssen et al. 2001, 1997; Wood et al. 1997). Distinguishing characteristics of the VIC model include the representation of the following:

- subgrid variability in land surface vegetation classes;
- subgrid variability in the soil moisture storage capacity, which is represented as a spatial probability distribution;
- drainage from the lower soil moisture zone (baseflow) as a nonlinear recession;
- spatial subgrid variability in precipitation.

Results from previous work by Nijssen et al. (2001) were used as the starting point for the current study. In Nijssen et al. (2001) a gridded dataset of daily meteorological model forcings for the period 1979–93 was developed for global land areas (excluding Greenland and Antarctica) at a spatial resolution of $2^\circ \times 2^\circ$. This dataset was then used to drive the VIC model to calculate a set of derived variables (evapotranspiration, runoff, snow water equivalent, and soil moisture) and to study the water balance of each of the continents. One specific purpose of the study was to develop global soil moisture fields constrained by observed meteorological forcings. Although the reader is referred to Nijssen et al. (2001) for details regarding the development of the dataset, a very brief overview is given here.

In Nijssen et al. (2001), daily precipitation and daily minimum and maximum temperatures were derived from station observations, and extended using stochastic interpolation methods for those areas where not enough daily meteorological stations were available. The resulting daily sequences were scaled to match the means of existing global, monthly time series (Hulme 1995; Huffman et al. 1997; Jones 1994). Daily surface wind speeds were obtained from the National Centers for Environmental Prediction–National Center for Atmospheric Research reanalysis project (Kalnay et al. 1996). The remaining meteorological forcings (vapor pressure, incoming shortwave radiation, and net longwave radiation) were calculated based on daily temperature and precipitation using algorithms by Kimball et al. (1997), Thornton and Running (1999), and Bras (1990).

For each $2^\circ \times 2^\circ$ model grid cell land surface characteristics such as elevation, soil, and vegetation were specified. Elevation data were calculated based on the 5-min TerrainBase Digital Elevation Model (Row et al. 1995), using the land surface mask from Graham et al. (1999). Vegetation types were provided by the Advanced Very High Resolution Radiometer–based, 1-km, global land classification from Hansen et al. (2000), which has 12 unique vegetation classes. Vegetation parameters such as height and minimum stomatal resistance were assigned to each individual vegetation class. Soil textural information and soil bulk densities were derived from the 5-min Food and Agricultural Organization (FAO)–United Nations Educational, Scientific, and Cultural Organization digital soil map of the world (FAO 1995), combined with the World Inventory of Soil Emission Potentials pedon database (Batjes 1995). The remaining soil characteristics, such as porosity, saturated hydraulic conductivity, and the exponent for the unsaturated hydraulic conductivity equation were based on Cosby et al. (1984).

In Nijssen et al. (2001), the soil depth was set at 1 m for all grid cells, with an upper horizon of 0.3 m and a lower horizon of 0.7 m. In addition a deeper layer without roots was specified, from which baseflow was generated using the ARNO baseflow formulation (Todini 1996). This layer had a water storage capacity of 100 mm, corresponding to a depth of about 0.25 m, depending on the porosity.

Comparison with previous estimates of the global and continental water balance showed that the components of the water balance (precipitation, evapotranspiration, and runoff), were similar to those earlier estimates, but that runoff was somewhat lower for some of the continents, especially South America.

4. Runoff production and flow routing

a. Data sources

River basins were delineated based on Graham et al. (1999) as corrected by Comanor et al. (2000). This dataset provides river basin outlines at 5-min resolution for 55 major river basins. The resulting 5-min masks and flow accumulation files were aggregated to $1^\circ \times 1^\circ$ routing networks using the algorithm of O’Donnell et al. (1999), which results in routing networks similar to those produced by Oki and Sud (1998). However, the
O’Donnell et al. algorithm allows for automatic extraction of coarser-resolution networks from high-resolution flow accumulation files, so that in principle the model output could be routed at any grid resolution supported by global digital elevation data.

River discharge records were obtained from the Global River Discharge Center (GRDC) in Koblenz, Germany, and from the RivDis 1.1 database (Vörösmarty et al. 1998), both of which provide historical monthly river discharge values globally. Daily discharge data from GRDC were used for the derivation of the ARNO baseflow parameters as discussed in the appendix.

b. Flow routing

The VIC model calculates the daily baseflow and “fast response” flow that are generated within each grid cell. A stand-alone routing model is then used to route these flows downstream. The routing model is described in detail by Lohmann et al. (1996, 1998a). Flow can exit each grid cell in eight directions and all flow must exit in the same direction. The flow from each grid cell is weighted by the fraction of the grid cell that lies within the basin.

Once the water is in the routing network it is effectively removed from the surface hydrological cycle. The routing model does not account for flood plain–channel interaction, evaporation from the channel, or loss of river water through infiltration in the river bed. As pointed out by Kite (1998), this is problematic in those regions where these phenomena significantly affect the river discharge. For example, observed flows on the Niger River in West Africa decrease from about 1540 m$^3$ s$^{-1}$ to 1140 m$^3$ s$^{-1}$ between Koulikoro, Mali, and Gaya, Niger, even though the upstream area is $1.2 \times 10^5$ km$^2$ at Koulikoro and almost 10 times larger ($1.0 \times 10^6$ km$^2$) at Gaya. The processes responsible for these channel losses are not represented by the routing model.

Similarly, the routing model does not account for extractions, diversions, or reservoir operations. Consequently, the simulated flows are natural flows and cannot be compared directly to measured river discharges on heavily regulated rivers. However, the generated stream discharges can be used as input to reservoir and water resources model (e.g., Hamlet and Lettenmaier 1999; Leung et al. 1999).

For this application the flows were routed on a $1^\circ \times 1^\circ$ network. The four $1^\circ \times 1^\circ$ flow cells within each 2$^\circ \times 2^\circ$ VIC model cell were assigned the same daily runoff values. The higher-resolution flow networks allowed a somewhat better approximation of the modeled flow network to the main stem river.

c. Selected rivers

Because observed streamflows were used for calibration, and because the routing model does not account for water resources developments on the river, river basins were selected using the following criteria:

- observed monthly flows had to be available for at least part of the period 1980–93;
- rivers had to be only minimally affected by extractions, diversions, and dams;
- the selected river basins had to cover a variety of climatic regions and continents.

Following these criteria, nine rivers were excluded because of significant flow regulation. The reservoir storage on each of these rivers was greater than the mean annual runoff volume. A further 20 rivers were excluded because of insufficient flow records coincident with the analysis period or similar problems. Although the GRDC and/or RivDis databases contain flow records for all except 4 of the original 55 river basins, the number of available records decreases rapidly after 1984 (Fig. 1).

Table 1 lists the 26 selected rivers (see also Fig. 2), with the categories as described in section 5. Not surprisingly, the final group of rivers overrepresents the colder climates. Generally, more complete streamflow records are available in these areas and reservoir storage is less important than in warmer, drier climates (Vörösmarty et al. 1997).

5. Calibration and parameter transfer

Only a limited number of model parameters were targeted for calibration. The variable infiltration parameter ($b_i$) controls the amount of water that can infiltrate into the soil as a function of soil moisture. Although Dümenil and Todini (1992) assert that the variable infiltration parameter is a function of the topographical properties of the terrain within a model grid cell, Wood et al. (1992)
<table>
<thead>
<tr>
<th>River basin</th>
<th>Gauge location</th>
<th>Predominant climatic zones</th>
<th>Area (km²) upstream of gauge*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon</td>
<td>Obidos, Brazil</td>
<td>Tropical</td>
<td>4 618 746</td>
</tr>
<tr>
<td>Congo</td>
<td>Brazzaville, Congo</td>
<td>Tropical</td>
<td>3 475 000</td>
</tr>
<tr>
<td>Danube</td>
<td>Drobeta-Turnu Severin, Romania</td>
<td>Midlatitude rainy</td>
<td>576 232</td>
</tr>
<tr>
<td>MacKenzie</td>
<td>Normal Wells, Canada</td>
<td>Arctic</td>
<td>1 570 000</td>
</tr>
<tr>
<td>Mekong</td>
<td>Pakse, Laos</td>
<td>Tropical</td>
<td>545 000</td>
</tr>
<tr>
<td>Mississippi</td>
<td>Vicksburg, United States</td>
<td>Midlatitude rainy</td>
<td>2 964 254</td>
</tr>
<tr>
<td>Senegal</td>
<td>Bakel, Senegal</td>
<td>Arid hot</td>
<td>218 000</td>
</tr>
<tr>
<td>Yellow</td>
<td>Yuayuankou, China</td>
<td>Arid cold</td>
<td>730 036</td>
</tr>
<tr>
<td>Yenisei</td>
<td>Igarka, Russia</td>
<td>Arctic</td>
<td>2 440 000</td>
</tr>
</tbody>
</table>

Category 2**

<table>
<thead>
<tr>
<th>River basin</th>
<th>Gauge location</th>
<th>Predominant climatic zones</th>
<th>Area (km²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amur</td>
<td>Komsomolsk, Russia</td>
<td>Arctic</td>
<td>1 730 000</td>
</tr>
<tr>
<td>Brahmaputra</td>
<td>Bahadurabad, Bangladesh</td>
<td>Midlatitude rainy</td>
<td>636 130</td>
</tr>
<tr>
<td>Changjiang</td>
<td>Datong, China</td>
<td>Midlatitude rainy</td>
<td>1 705 383</td>
</tr>
<tr>
<td>Columbia</td>
<td>The Dalles, United States</td>
<td>Midlatitude rainy</td>
<td>690 000</td>
</tr>
<tr>
<td>Indigirka</td>
<td>Vorontsovo, Russia</td>
<td>Arctic</td>
<td>305 000</td>
</tr>
<tr>
<td>Irrawaddy</td>
<td>Sagaing, Myanmar</td>
<td>Midlatitude rainy</td>
<td>117 900</td>
</tr>
<tr>
<td>Kolyma</td>
<td>Srednekolymsk, Russia</td>
<td>Arctic</td>
<td>361 000</td>
</tr>
<tr>
<td>Lena</td>
<td>Kusur, Russia</td>
<td>Arctic</td>
<td>2 430 000</td>
</tr>
<tr>
<td>Ob</td>
<td>Salekhard, Russia</td>
<td>Arctic</td>
<td>2 430 000</td>
</tr>
<tr>
<td>Olenek</td>
<td>Downstream of Pur River, Russia</td>
<td>Midlatitude rainy</td>
<td>198 000</td>
</tr>
<tr>
<td>Paraná</td>
<td>Corrientes, Argentina</td>
<td>Midlatitude rainy</td>
<td>2 300 000</td>
</tr>
<tr>
<td>Pechora</td>
<td>Ust-Tsilm, Russia</td>
<td>Arctic</td>
<td>248 000</td>
</tr>
<tr>
<td>Severnaya Dvina</td>
<td>Ust-Pinega, Russia</td>
<td>Arctic</td>
<td>348 000</td>
</tr>
<tr>
<td>Volga</td>
<td>Volgograd, Russia</td>
<td>Midlatitude rainy</td>
<td>1 360 000</td>
</tr>
<tr>
<td>Xi</td>
<td>Wuzhou, China</td>
<td>Midlatitude rainy</td>
<td>329 705</td>
</tr>
<tr>
<td>Yana</td>
<td>Dzhangky, Russia</td>
<td>Arctic</td>
<td>216 000</td>
</tr>
<tr>
<td>Yukon</td>
<td>Ruby, United States</td>
<td>Arctic</td>
<td>670 810</td>
</tr>
</tbody>
</table>

* Areas are taken from the GRDC and RivDis databases.

** As explained in section 2, category-1 rivers are used in the first round of calibration, and category-2 rivers are used in the second round of calibration.

argue that this parameter must be determined through calibration. The remaining calibration parameters, namely, the depth of the second soil layer ($D_2$), the saturated hydraulic conductivities of the first and second layers ($k_s$ and $k_{s2}$), and the exponents for the unsaturated hydraulic conductivity in the first and second layers ($n_1$ and $n_2$) all describe soil hydrological properties.

Information about the saturated hydraulic conductivity ($k_s$) and the exponent of the unsaturated hydraulic conductivity curve ($n$) can be derived from soil textural information (e.g., Cosby et al. 1984; Rawls et al. 1993), as was done in Nijssen et al. (2001). We identified these parameters as calibration parameters for several reasons. First, it is difficult to determine average soil properties over an area as large as a $0.5 \times 0.5$ grid cell (about 49.5 $\times 10^3$ km² at the equator). For example, Meeson et al. (1995) reported in the documentation that accompanies the CD-ROMs released as part of the International Satellite Land Surface Climatology Project, that they needed to shift all soil hydraulic properties by one soil textural class in order to obtain realistic drainage patterns with their model. Determination of areal average soil hydraulic properties is especially difficult, because most functions linking soil hydraulic properties to soil textural properties are highly nonlinear. Second, soil depths are generally not well known over large areas. Although Dunne and Wilmott (1996) developed a global dataset of plant-available moisture at a 0.5°, which included an estimate of rooting depth, these root zone estimates are strongly linked to the specific vegetation database used in that study.

Furthermore, and most importantly, the soil moisture transport mechanism in the VIC model is quite simple, consisting of gravity drainage with the hydraulic conductivity in each layer a function of the soil moisture storage within that layer. We implemented the VIC model in such a way that the plant roots could draw water only from the top two soil layers and the baseflow was generated only from the third layer. Consequently, the flux of water from the second layer into the third layer determines how much baseflow will occur. The ARNO model routing parameters (see the appendix) only determine how quickly the water stored in the third layer is evacuated. The transport of water through the second
layer is largely determined by the parameters selected for calibration, because the VIC model calculates the hydraulic conductivity as

$$k = k\left(\frac{\theta}{\phi}\right)^n = k\left(\frac{W_n}{W_{n_{max}}}\right)^n,$$

with $k$ the hydraulic conductivity, $\theta$ the moisture content, $\phi$ the porosity, $W_n$ the moisture storage, and $W_{n_{max}}$ the maximum moisture storage equal to $\phi D$, with $D$ the depth of the layer. The moisture storage is dynamically determined by the model. A change in the thickness of the second layer affects not only the hydraulic conductivity, but also the maximum storage available in the second layer and consequently the water available for transpiration.

Climatic characteristics were selected as the basis for the transfer of calibration parameters under the premise that hydrological processes and the parameters used to describe them are more similar within than between different climate zones. To this end, each of the model grid cells was classified into one of five climatic zones (Table 2). These zones were constructed by regrouping the zones of the Köppen classification (e.g., Hubert et al. 1998).

An initial set of nine basins was selected for calibration on the basis of the following:

- the quality of the observed discharge record;
- the geographical location, to represent each climatic zone and continent, and
- the number of climate zones in each basin.

We tried to represent each climate zone and to have at least one basin on each continent. Australia was not represented, due to the high amount of regulation on the Australian rivers for which discharge data were available. To facilitate calibration and parameter transfer, the primary basins were selected to be located as much as possible within a single climate zone.

Calibration was performed manually and focused on matching the total annual flow volume and the shape of the mean monthly hydrograph. In the first round of calibration, all primary basins were grouped by climate zone and all basins within the same climate zone were calibrated using the same parameters. Although this approach was successful in the arctic and temperate climate zones, it did not work as well in the tropical climate zone. In particular, the calibrated parameters for the Amazon did not improve the model results for the Congo and vice versa. The tropical climate zone was therefore subdivided into three geographical regions: an American region from 180° to 30°W, an African region from 30°W to 60°E, and an Asian–Australian region from 60°E to 180°. It should be noted that the tropical climate zone encompasses all climates that are not defined as dry and where the mean daily maximum temperature is greater than 18°C in all months. However, this climate
when 70% of the annual precipitation falls in the coldest season, and the superscripts min and max indicate the coldest and warmest months.

** Table 2. Grouping of Köppen climate zones into parameter transfer zones. **

<table>
<thead>
<tr>
<th>Parameter transfer zone</th>
<th>Köppen* climate zones</th>
<th>Climate characteristics**</th>
</tr>
</thead>
<tbody>
<tr>
<td>1—Tropical</td>
<td>A</td>
<td>Tropical climate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not a B climate and (T_a \geq 18^\circ C)</td>
</tr>
<tr>
<td>2—Arid hot</td>
<td>Bh</td>
<td>Dry, warm climate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B climate with (T_a &gt; 18^\circ C)</td>
</tr>
<tr>
<td>3—Arid cold</td>
<td>Bk</td>
<td>Dry, cold climate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B climate with (T_a \leq 18^\circ C)</td>
</tr>
<tr>
<td>4—Midlatitude rainy</td>
<td>C</td>
<td>Rainy, midlatitude climate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not a B climate and (-3^\circ C &lt; T_{a\min} &lt; 18^\circ C)</td>
</tr>
<tr>
<td></td>
<td>Da</td>
<td>Continental climate with hot summer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not a B climate and (T_{a\min} &lt; -3^\circ C) and (T_{a\max} &gt; 22^\circ C)</td>
</tr>
<tr>
<td></td>
<td>Db</td>
<td>Continental climate with cool summer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not a B climate and (T_{a\min} &lt; -3^\circ C) and (T_{a\max} &gt; 10^\circ C) for at least 4 months</td>
</tr>
<tr>
<td>5—Arctic</td>
<td>Dc</td>
<td>Continental climate with short cool summer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not a B climate and (T_{a\min} &lt; -3^\circ C) and (T_{a\max} &gt; 10^\circ C) for less than 4 months</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>Polar climate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Not a B climate and (T_{a\min} &lt; 10^\circ C)</td>
</tr>
</tbody>
</table>

* Here B climates are defined as those climates for which \(P_a > 20T_a + 140\) for those climates with equidistributed precipitation, \(P_a > 20T_a\) when 70% of the annual precipitation falls in the coldest season, and \(P_a > 20T_a + 280\) when 70% of the annual precipitation falls in the warmest season.

** Here \(T\) is the temperature (°C), \(P\) is the precipitation (mm), the subscripts \(m\) and \(a\) are used to indicate the monthly or annual period, and the superscripts min and max indicate the coldest and warmest months.

zone includes a number of different precipitation regimes. For instance, the Amazon basin is much wetter than the Congo basin, while the precipitation in Southeast Asia is monsoon dominated.

During the calibration process, the infiltration parameter \(b_i\) and the thickness of the second soil layer \(D_2\), which were treated as the primary calibration parameters, were changed to a uniform value in a given climate zone. For example, all grid cells in the American tropical zone were given the same \(b_i\) and \(D_2\). The outer calibration parameters, which were used to fine tune the calibration, were changed from their original, spatially varying values using a regionally uniform multiplier. Consequently, after calibration the texture-based soil hydraulic properties \((k_i\) and \(n\)) varied spatially, while \(b_i\) and \(D_2\) were constant within each region.

Generally, the thickness of the second soil layer was increased to allow for more storage. Only in the arctic climate zone was the thickness of the second soil layer reduced to 0.5 m. The unsaturated hydraulic conductivity and its exponent were changed to reduce the soil moisture flux in the arid climates, and to increase the soil moisture flux in the other climate zones. The calibrated infiltration parameter \(b_i\) tended to be smallest in the arid climates, in an effort to reduce runoff production.

Parameters were transferred from the primary to the secondary basins based on climate zone. For example, arctic cells in a secondary basin were modeled using parameters from the arctic zone, while temperate cells were modeled using parameters from the temperate zone. This meant that \(b_i\) and \(D_2\) were set to the value for the climate zone, and that the soil hydraulic properties were modified by the multiplier for that zone.

To evaluate the effectiveness of the parameter transfer process and to provide the best possible global water balance estimates, the secondary basins were then further calibrated before transferring the parameters from all of the calibrated basins to the remaining global land surface grid cells. This second calibration was also performed by climate zone, although the primary basins were not modified. After this second calibration parameters were transferred to the rest of the global land surface area using the nearest calibrated grid cell within the same climate zone.

6. Results

a. Primary basins

Figure 3 shows the mean monthly hydrographs for the nine primary basins. The uncalibrated or base case simulation results are based on Nijssen et al. (2001). By construct, calibration improved the results in all instances, although in some cases the final calibration was still unsatisfactory, especially for arid basins such as the Senegal (Table 3). Some of these problems are related to the adopted routing scheme (section 4), since water is “removed” from the hydrologic cycle once it enters the stream channel. In addition, the Senegal flows through a region with strong precipitation gradients, which are not well resolved by the \(2^0 \times 2^0\) forcing dataset. Further, the VIC model does not represent capillary rise in the soil zone. Although this process is usually of minor importance to the water budget in hu-
mid areas, it may be significant in arid and semiarid regions. Excluding the Senegal, calibration reduced the mean bias from 28.9% to 9.9% and the relative root-mean-square error (rrmse) from 62.0% to 37.2%.

Figure 4 shows the annual runoff volumes as well as the annual precipitation for each of the basins. Although the changes in annual runoff volume between the base case and calibrated run are sometimes small (e.g., Mississippi and Danube), the calibrated model runs generally reflect the seasonal runoff cycle considerably better (Fig. 3). Except for the Congo River, the calibrated hydrographs peak within a month of the observed peak.

b. Secondary basins

Figures 5 and 6 are similar to Figs. 3 and 4, respectively, and show the parameter transfer and calibration results for the 17 secondary basins. Four of the 17 basins were excluded from the dataset after further examination revealed serious problems with the precipitation forcings. These four basins, the Yukon, Columbia, Brahmaputra, and Irrawaddy, are further discussed in section 7. Unfortunately, most of the remaining basins are in the northern mid- to high latitudes, which makes it difficult to evaluate the performance of the parameter transfer scheme for climate zones 1 and 2 (Table 2).

For the remaining 13 basins, the parameter transfer process improved the simulated flow volume in six cases (Amur, Indigirka, Kolyma, Lena, Olenek, and Yana), resulted in little or no change in three cases (Xi, Changjiang, and Volga) and resulted in a worse simulation in four cases (Paraná, Ob, Pechora, and Severnaya Dvina). Three of these last four basins are located in the western part of the Russian Arctic (from west to east of the Severnaya Dvina, Pechora, and Ob) and they differ from the basins immediately to the east in that their annual average precipitation is considerably higher. Whereas these three basins have a mean annual precipitation of about 600–800 mm, the basins to the east (such as the Yenisei and Lena) are much drier, with mean annual precipitation of about 400–500 mm. In the second round of calibration, these three basins were grouped and calibrated using a common set of parameters. For these three basins, was set to 0.25, the depth of the second soil layer to 2.5 m, and the hydraulic conductivities and exponents were reset to their precalibration values. This resulted in much improved flow volumes in all three cases. The depth of the second layer is unrealistic, but
at present the VIC model does not allow for surface depression storage, which plays an important role immediately after snowmelt when the ground is partially frozen in these cold climate basins. The deep soil layer acts partially as a surrogate for this surface storage.

Overall, parameter transfer reduced the rmse from 120.8% to 96.3% and the mean bias from 40.6% to 35.9%. Further calibration of these basins reduced the rmse to 69.8% and the bias to 11.5%. Four of the arctic basins (Indigirka, Kolyma, Lena, Yana) as well as the Amur and the Xi were not further calibrated in the second round, since the model results based on the parameter transfer were deemed sufficient.

c. Continental water balance

The calibrated parameters from the 22 calibrated basins were transferred to the rest of the globe, based on a nearest neighbor search within the same climate zone. The resulting continental water balances, summarized in Table 4, generally show an increase in runoff and, consequently, a decrease in evaporation compared with the base case simulation.

Globally, annual runoff increased 23 mm (9.4%) and evapotranspiration decreased 24 mm (5.0%). Because long-term storage changes are close to zero, the absolute change in runoff and evapotranspiration are about equal, but of opposite sign. The greatest decrease in evapotranspiration is in South America, where the modeled, annual evapotranspiration decreased by 119 mm (11.9%) and runoff increased 26.2% (119 mm). The runoff estimates are close to observation-based estimates from Lvovitch (1973) and Baumgartner and Reichen (1975) (Table 4), which is not surprising, because the calibration process ensured that the simulated runoff volumes are close to the observations. Table 4 clearly shows the large spread in continental precipitation and evapotranspiration estimates. The differences in evapotranspiration estimates largely stem from differences in the precipitation estimates. Although we have used recent global precipitation estimates in the construction of our model forcing dataset (Nijssen et al. 2001), the next section discusses some remaining problems.

7. Precipitation forcings

The precipitation dataset (Nijssen et al. 2001) was constructed on a 2° × 2° grid combining station obser-
vations with the monthly precipitation values from the 2.5° × 2.5° Global Precipitation Climatology Project (GPCP); (Huffman et al. 1997) and the 3.75° × 2.5° dataset from Hulme (1995). By construct, the monthly precipitation totals are the same as the monthly GPCP precipitation time series for the period covered by the GPCP dataset (1987 onward).

As mentioned earlier, it became apparent that the precipitation forcings at 2° × 2° resolution were suspect for four of the basins. For example, the mean annual precipitation over the Brahmaputra basin is 1144 mm, while the observed mean annual runoff is about 1133 mm for the same period (Fig. 6). If we assume that the long-term changes in terrestrial water storage are close to zero, this would leave only 10 mm yr⁻¹ for evapotranspiration, which is clearly unrealistic. The Brahmaputra flows through the foothills of the Himalaya, where orographic effects result in high precipitation during the monsoon season. Clearly, the precipitation is underestimated. Consequently, our simulations result in modeled streamflow that is much lower than the observed streamflow. The same pattern, in even more extreme form, is observed for the Irrawaddy in Southeast Asia (mean annual precipitation of 900 mm and observed annual streamflow of 2203 mm for the area upstream of the gauge at Sagaing, Myanmar) and in somewhat less extreme form for the Columbia River in North America.

For some rivers the opposite effect was observed. For the Yukon River in North America the observed annual flow is about 159 mm, while the mean annual precipitation is about 627 mm. This would mean that the annual evapotranspiration in that region would be about 468 mm, which seems high given the short growing season and the long, cold winters. Consequently, our model results overestimated the mean annual streamflow in the Yukon. The high precipitation was largely due to model grid cells along the Gulf of Alaska, where a few stations, on which the GPCP climatology is based, show very high precipitation amounts (in excess of 2000 mm yr⁻¹).

To further diagnose problems with the precipitation forcings, the Columbia River simulation was evaluated in more detail. The Columbia River has been extensively studied with the VIC model at high spatial resolutions (0.25°, Matheussen et al. 2000 and 0.125°, Hamlet and Lettenmaier 1999) and good estimates of precipitation are available for this area. Figure 7 shows the mean annual precipitation over the Columbia River basin for the 2° × 2° resolution global dataset (based on the GPCP dataset), and a regional climatology produced by Daly et al. (1994) aggregated to 0.125° and 2°. The Parameter-elevation Regressions on Independent Slopes Model (PRISM) developed by Daly et al. (1994) produces a precipitation climatology at a resolution of 2.5 min, based on more than 8000 precipitation stations over the
continental United States. Its distinguishing feature is that it accounts for orographic effects by calculating local regression relationships between precipitation and elevation, and using these relationships to estimate mean areal precipitation. Whereas the GPCP-based precipitation generally decreases from west to east over the Columbia River basin, the PRISM-based precipitation shows a much more complex spatial pattern with the highest precipitation along the northwestern (Cascade Mountains), and the northern and eastern (Rocky Mountains) perimeters of the basin.

To further diagnose whether the undersimulation of streamflow in the Columbia River was due to the precipitation or to a poor combination of parameters, we
forced the VIC model with the PRISM-based precipitation aggregated to $2^\circ \times 2^\circ$. Additionally, we used a station-based temperature dataset (Hamlet and Lettenmaier 1999) to ensure that biases in the temperature were not the cause for the poor simulation results. Figure 8 shows the observed Columbia River hydrograph, and the results from four separate model runs. Note that the “observed” flows have been adjusted to account for the effects of reservoir regulation, that is, they are the flows that would have occurred in the absence of reservoirs or other water management effects (A. G. Crook Company 1993). The first model run uses the forcing data used in this study. The second model run uses the same precipitation, with the station-based temperature data from Hamlet and Lettenmaier (1999). The third model run uses the same temperature data as the first run, but uses the PRISM-based precipitation, aggregated to $2^\circ$ spatial resolution. The fourth model run uses the station-based temperature data from Hamlet and Lettenmaier (1999) and the PRISM precipitation. All other model parameters are identical for each of the four runs. The figure demonstrates clearly that the modeled streamflow is highly sensitive to the quality of the precipitation estimates. If the high-resolution PRISM dataset, which includes the effects of orography on precipitation, is taken as the “true” precipitation, the model results are
compared to those for the other basins and the parameter transfer could have been regarded as moderately successful. However, using the GPCP-based precipitation, the streamflow is severely overpredicted.

Unlike the Columbia River basin, the two precipitation datasets agreed reasonably well with respect to the mean annual precipitation over the Mississippi River basin and the mean annual simulated streamflow was within 15% of the observed flow. In the Mississippi basin, the effects of the orography are largely limited to grid cells along the northwestern edge of the basin (Rocky Mountains).

Generally it appears that the greatest weakness of the precipitation dataset is in areas with complex topography, such as coastal mountain ranges, where very strong gradients in precipitation exist. In these areas, the small number of stations per grid cell used in the construction of the GPCP dataset do not adequately reflect the areal average precipitation. In addition, the relatively coarse modeling grid can lead to erroneous precipitation forcings when, for example, a river drains the lee side of a mountain range. If the mountain range is narrow, the coarse-resolution precipitation data includes precipitation from the windward side of the mountain range, resulting in an overprediction of the total precipitation input. Because of the problems with the precipitation inputs, the four rivers mentioned earlier in this section were not further calibrated and their parameters were excluded from the nearest neighbor search in the parameter transfer process.

8. Conclusions

The ability to represent runoff and streamflow from large river basins, and from continents in total, is critical to an understanding of the global hydrologic cycle. Streamflow is important not only for water resource studies, but it is also the most commonly available component of the surface water balance, and therefore can be used in a variety of diagnostic studies (e.g., Maurer et al. 2001). The runoff formulation in most MHMs is quasi-conceptual, which necessitates some calibration of model parameters. We found that calibration of a restricted number of VIC model parameters reduced the relative root-mean-square error (rrmse) of the monthly flows from 62% to 37% and the mean bias in annual flows from 29% to 10%. Problems in model predictions remain in hot, arid areas, where the VIC model tends to overestimate the mean annual runoff.

However, calibration is a time-consuming process and quickly becomes infeasible when the model area or the number of basins increases. Therefore, a methodology of model parameter transfer was developed to limit the number of basins for which calibration was needed. As a first attempt, model cells were grouped by climate zone, and model parameters were transferred within these groups. Although, the results were mixed, the transferred parameters reduced the rrmse from 121% to 96% and the mean bias from 41% to 36%. Further calibration of these basins reduced the rrmse to 70% and the bias to 12%.

After the parameter transfer from all calibrated basins to the rest of the globe, the mean annual global runoff increased 9.4% and the evapotranspiration decreased 5.0% relative to the values reported by Nijssen et al. (2001) for a model run without calibration. However, on a continental basis the changes were much larger. For South America, the simulated evapotranspiration was 119 mm (12%) smaller than in the base case sce-
nario. This corresponds to a decrease in the mean latent heat flux of 9.4 W m⁻². Because evapotranspiration shows a strong diurnal cycle, the maximum instantaneous change in the latent heat flux is much larger.

Our study also pointed out some shortcomings in the global datasets used. Lack of available streamflow observations as well as regulation on many of the world’s largest rivers halved the number of rivers in our original dataset. There is a great need for recent (post 1980) streamflow data for these rivers. Difficulties with the precipitation forcing eliminated an additional four river basins. Some of the precipitation problems were due to
the coarse resolution of the GPCP-based dataset ($2° \times 2°$), which failed to resolve strong spatial gradients in the precipitation. However, a more persistent problem was the underestimation of the precipitation in areas with complex terrain. Because most precipitation stations in those areas tend to be located in areas with low elevation, the effects of orographic enhancement are not adequately represented. Comparison of the mean annual precipitation field over the Columbia River basin with a PRISM-based precipitation field showed that the annual biases could be as large as 1000 mm.

The parameter transfer method resulted in only a modest improvement in the streamflow simulations. It is clear from these early results that many questions remain. What are the best criteria for grouping basins or grid cells? How can parameters be transferred within and between these groups? Which parameters can safely and successfully be transferred? However, despite these difficulties, we think that the parameter transfer process is one of the few realistic options for performing large-scale or global hydrological simulations.

Both the meteorological forcing dataset and the model-simulated water balance components are available to the scientific community through the World Wide Web (information is available online at http://www.hydro.washington.edu). The resulting 347 baseflow parameter sets were included as one of the optimized variables. The remaining baseflow parameters ($d_2$, $d_1$, and $d_3$) were obtained by fitting (A1) to all baseflow sequences through optimization using the shuffled complex evolution (SCE) algorithm (Duan et al. 1992, 1993, 1994). As suggested by Abdulla et al. (1999), the optimum parameter values were determined by ordinary least squares applied to transformed (Box–Cox), prewhitened residuals. The Box–Cox transformation parameter $\lambda$ was included as one of the optimized variables.

The resulting 347 baseflow parameter sets were in-
interpolated to the individual 2° × 2° grid cells using a moving window with a radius of 250 km. If more than one baseflow parameter set was available within 250 km from the center of the grid cell, the baseflow parameter values were averaged, otherwise the nearest baseflow parameter set was used.

REFERENCES


——, ———, and ———, 1998b: Regional scale hydrology: II.
Application of the VIC-2L model to the Weser River, Germany. 


