≋USGS
science for a changing world

Hydropower Assessment of Bolivia—A Multisource Satellite Data and Hydrologic Modeling Approach

U.S. Department of the Interior
U.S. Geological Survey

Hydropower Assessment of Bolivia—A Multisource Satellite Data and Hydrologic Modeling Approach

By Naga Manohar Velpuri, Md Shahriar Pervez, and W. Matthew Cushing

Open-File Report 2016–1156

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
SALLY JEWELL, Secretary

U.S. Geological Survey
Suzette M. Kimball, Director

U.S. Geological Survey, Reston, Virginia: 2016

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit http://www.usgs.gov or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit http://www.usgs.gov/store.usgs.gov/.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Acknowledgments

Contents

Figures

Appendix Figures

Tables

Appendix Tables

Conversion Factors

International System of Units to U.S. customary units

Multiply	By	To obtain
Length		
millimeter (mm)	0.03937	inch (in.)
meter (m)	3.281	foot (ft)
meter (m)	1.094	yard (yd)
kilometer (km)	0.6214	mile (mi)
kilometer (km)	0.5400	mile, nautical (nmi)
Area		
square kilometer (km^2)	247.1	acre
square kilometer (km^2)	0.3861	square mile (mi^2)
Flow rate		
cubic meter per second (m^3/s)	70.07	acre-foot per day (acre-ft/d)
cubic meter per second (m^3/s)	35.31	cubic foot per second (ft^3/s)
cubic meter per second (m^3/s)	22.83	million gallons per day (Mgal/d)
millimeter per month (mm/month)	0.03937	inch per hour (in/month)
millimeter per year (mm/year)	0.03937	inch per year (in/year)
Density		
kilogram per cubic meter (kg/m^3)	0.06242	pound per cubic foot (lb/ft^3)

Supplemental Information

The acceleration of gravity is given in meters per square second (m/s^2).

The resolution of the Shuttle Radar Topography Mission digital elevation model is given in arc seconds.

The total theoretical hydropower potential is given in megawatts (MW).

Abbreviations

ANOVA	analysis of variance
ANA	Brazilian Agência Nacional de Águas
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
CHIRP	Climate Hazards Group InfraRed Precipitation
CHIRPS	CHIRP with Stations
CHPClim	Climate Hazards Group Precipitation Climatology
CMAP	Climate Prediction Center Merged Analysis of Precipitation
CPC	Climate Prediction Center
DEM	digital elevation model
GDEM	Global DEM
GIS	geographic information system
GPCC	Global Precipitation Climatology Center
GPCP	Global Precipitation Climatology Project
GPI	Geostationary Operational Environmental Satellite Precipitation Index
IR	infrared
ITCZ	intertropical convergence zone
MAE	mean absolute error
MSU	microwave sounding unit
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
NSCE	Nash-Sutcliffe Coefficient of Efficiency
OLR	Outgoing Longwave Radiation
OPI	Outgoing Longwave Radiation Precipitation Index
PSENAMHI	Peruvian Servicio Nacional de Meteorología e Hidrología
R^2	coefficient of determination
RMSE	root-mean-square error
SENAMHI	Bolivian Servicio Nacional de Meteorología e Hidrología
SIR–C	Shuttle Imaging Radar with payload C
SRTM	Shuttle Radar Topography Mission
SSE	sum of squared error
SSM/I	Special Sensor Microwave Imager
TRMM	Tropical Rainfall Measuring Mission
TRMM–D	disaggregated TRMM
TTHP	total theoretical hydropower potential
TukeyHSD	Tukey's honest significant difference
USGS	U.S. Geological Survey
X–SAR	Synthetic Aperture Radar with payload X

Hydropower Assessment of Bolivia—A Multisource Satellite Data and Hydrologic Modeling Approach

By Naga Manohar Velpuri, Shahriar Pervez, and W. Matthew Cushing

Abstract

This study produced a geospatial database for use in a decision support system by the Bolivian authorities to investigate further development and investment potentials in sustainable hydropower in Bolivia. The study assessed theoretical hydropower of all 1-kilometer (km) stream segments in the country using multisource satellite data and a hydrologic modeling approach. With the assessment covering the 2 million square kilometer (km²) region influencing Bolivia's drainage network, the potential hydropower figures are based on theoretical yield assuming that the systems generating the power are 100 percent efficient. There are several factors to consider when determining the real-world or technical power potential of a hydropower system, and these factors can vary depending on local conditions. Since this assessment covers a large area, it was necessary to reduce these variables to the two that can be modeled consistently throughout the region, streamflow or discharge, and elevation drop or head. First, the Shuttle Radar Topography Mission high-resolution 30-meter (m) digital elevation model was used to identify stream segments with greater than 10 km² of upstream drainage. We applied several preconditioning processes to the 30-m digital elevation model to reduce errors and improve the accuracy of stream delineation and head height estimation. A total of 316,500 1-km stream segments were identified and used in this study to assess the total theoretical hydropower potential of Bolivia. Precipitation observations from a total of 463 stations obtained from the Bolivian Servicio Nacional de Meteorología e Hidrología (Bolivian National Meteorology and Hydrology Service) and the Brazilian Agência Nacional de Águas (Brazilian National Water Agency) were used to validate six different gridded precipitation estimates for Bolivia obtained from various sources. Validation results indicated that gridded precipitation estimates from the Tropical Rainfall Measuring Mission (TRMM) reanalysis product (3B43) had the highest accuracies. The coarse-resolution (25-km) TRMM data were disaggregated to 5-km pixels using climatology information obtained from the Climate Hazards Group Infrared Precipitation with Stations dataset. About a 17-percent bias was observed in the disaggregated TRMM estimates, which was corrected using the station observations. The bias-corrected, disaggregated TRMM precipitation estimate was used to compute stream discharge using a regionalization approach. In regionalization approach, required homogeneous regions for Bolivia were derived from precipitation patterns and topographic characteristics using a k-means clustering approach. Using the discharge and head height estimates for each 1-km stream segment, we computed hydropower potential for 316,490 stream segments within Bolivia and that share borders with Bolivia. The total theoretical hydropower potential (TTHP) of these stream segments was found to be 212 gigawatts (GW). Out of this total, 77.4 GW was within protected areas where hydropower projects cannot be developed; hence, the remaining total theoretical hydropower in Bolivia (outside the protected areas) was estimated as 135 GW. Nearly 1,000 1-km stream segments, however, were within the boundaries of existing hydropower projects. The TTHP of these stream segments was nearly 1.4 GW, so the residual TTHP of the streams in Bolivia was estimated as 133 GW. Care should be exercised to understand and interpret the TTHP identified in this study because all the stream segments identified and assessed in this study cannot be harnessed to their full capacity; furthermore, factors such as required environmental flows, efficiency, economics, and feasibility need to be considered to better identify a more real-world hydropower potential. If environmental flow requirements of 20–40 percent are considered, the total theoretical power available reduces by 60–80 percent. In addition, a 0.72 efficiency factor further reduces the estimation by another 28 percent. This study provides the base theoretical hydropower potential for Bolivia, the next step is to identify optimal hydropower plant locations and factor in the principles to appraise a real-world power potential in Bolivia.

Introduction

In this study, we describe the work that the U.S. Geological Survey (USGS) completed to produce a geospatial database of theoretical hydropower potential at every 1-kilometer (km) stream segment for Bolivia. This report describes the methodology applied to produce this geodatabase. The geodatabase contains all the stream segments that are within the territory of Bolivia or drain into Bolivian territory. The geodatabase also contains, for each 1-km stream segment, the drops in elevation and estimates of the potentially available
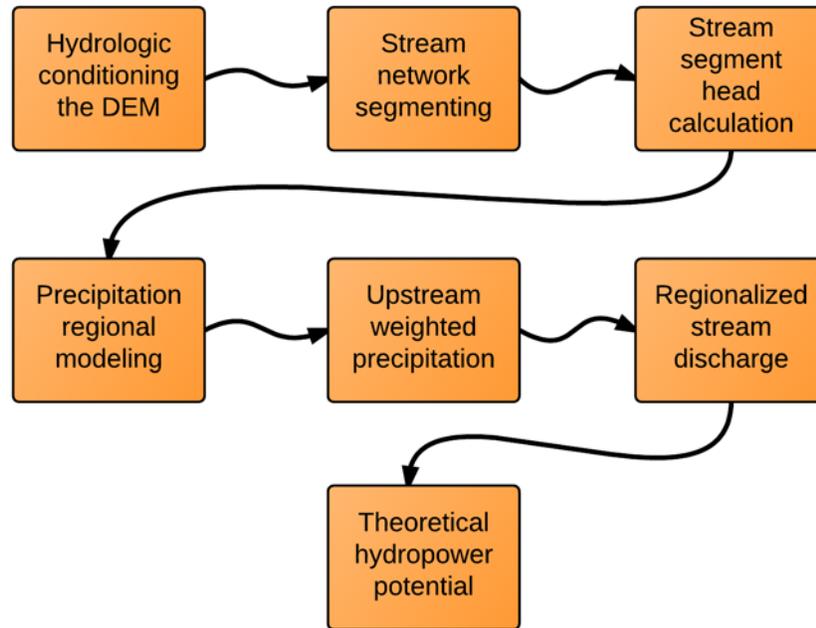
Figure 1.   Flowchart of the total theoretical hydropower potential methodology [DEM, digital elevation model].

stream discharge associated with exceedance probability, which are the two primary parameters for estimating theoretical hydropower potential.

A flowchart of the methodology followed to estimate the total theoretical hydropower potential for Bolivia is shown in figure 1. The first step is the hydrologic conditioning of the digital elevation model (DEM) data. This conditioning process corrects anomalies in the DEM by filling data voids, smoothing out artificial surfaces, and filling erroneous sinks. During the second step, the drainage or stream network from the DEM is modeled, and each stream is split into 1-km segments. After segmenting, the head, or elevation drop, for each segment is calculated in step three. During the fourth step, a regionalized precipitation climatology model for the area is generated. This step is completed by first validating and comparing several global gridded climatology precipitation datasets and combining the best-available climate datasets of the region to generate bias-corrected mean annual precipitation climatology models. Based on these corrected rainfall models, step five is completed to generate the upstream weighted annual precipitation dataset. For step six, these results are used in combination with regionalized drainage regions to determine the stream discharge volumes for each 1-km segment. The final step (step seven) is completed to determine the theoretical hydropower potential for each of these segments using the elevation drop (computed in step three) and stream discharge volumes (computed in step six).

In the subsequent sections of this report, we introduce the data and methods used to produce a hydropower potential assessment for all the stream segments within Bolivian territory. The objective of the work was to build a geospatial database that will serve as a hydropower decision support system. The database and results from this study will provide initial hydropower assessments for the region that can be used subsequently for more detailed feasibility studies (Phase II) on the development of hydropower plants. These Phase II studies will be completed by a third party at a later date.

Assessment Area

Hydropower assessment was completed for all streams within the territory of Bolivia and streams flowing into Bolivia. The extent of the assessment area in relation to Bolivia is shown in figure 2. Bolivia is a landlocked country in western South America, bordered by Brazil, Peru, Chile, Argentina, and Paraguay. Its combination of topography and climate gives the country varied hydrologic regions. The climate varies from tropical in the northeast, semiarid in the southeast, mountainous tropical forests along the eastern slopes of the Andes, and arid highlands to the west. In the northeast lies the Oriente region, covering the southwestern reaches of the Amazon River Basin. It includes several land formations: low alluvial plains, swamps, flooded bottomlands, open savannas, and tropical forest. Southern Bolivia, where the Andes begin to widen, is the Puna region. The area consists of high sharp-descending escarpments on the western slopes and gentler descending escarpments on the eastern slopes. In between these escarpments are fertile valleys called Valles. This formation wanes southeastward, transitioning to the western extent of the Gran Chaco. The Gran Chaco is a mix of subtropical forest and grasslands with variable seasonality. During the rainy season from October to March, low-lying areas become swamps; during the dry season the region is
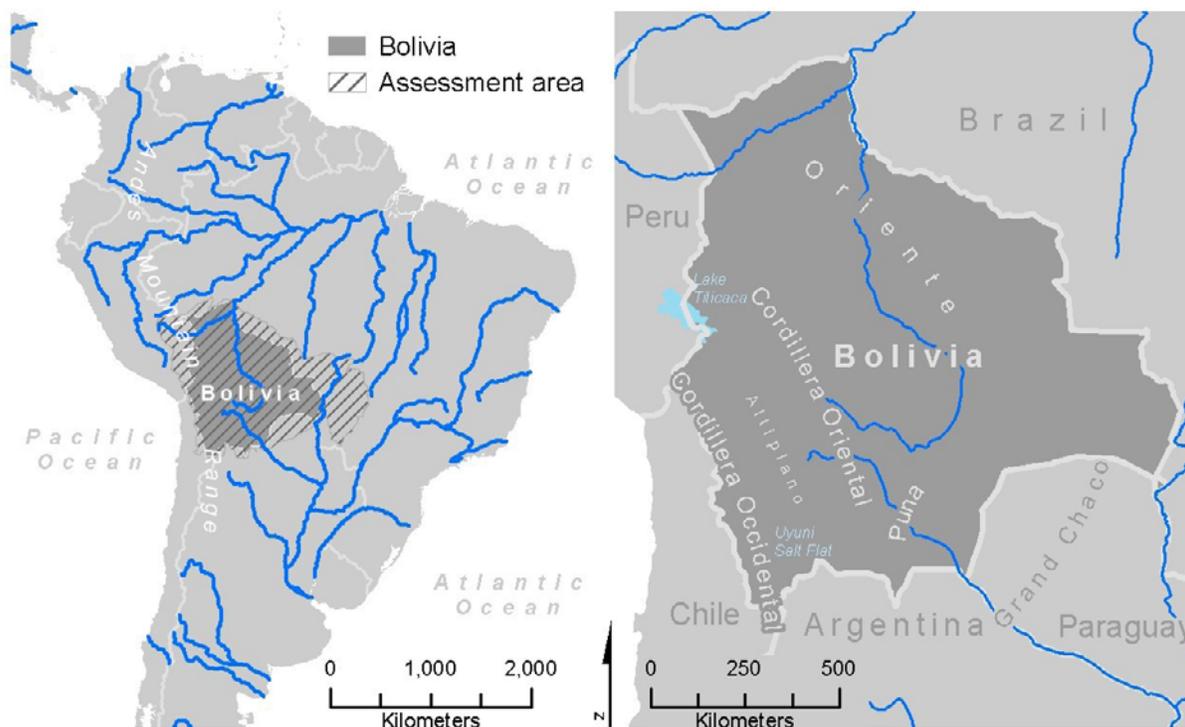
Figure 2.   Area where the hydropower assessment was completed, highlighting Bolivia, the assessment area, and major rivers.

a hot semidesert. Western Bolivia is occupied by two parallel mountain ranges. Bordering with Chile is the Cordillera Occidental range, home to the Uyuni Salt Flat and the site of many active volcanoes. To the east runs the Cordillera Oriental with its impressive snow-covered peaks surpassing 6,000 meters (m) in elevation. Between these two massive ranges is the Altiplano or High Plateau, a flat depression at 3,800 m, stretching from Peru to the north and Argentina to the south.

The climate in the region is determined by seasonal oscillation in the intertropical convergence zone (ITCZ). During austral winter, the ITCZ is north of Bolivia, and tropical anticyclones produce a somewhat cold, dry season. During austral summer, from December to March, the ITCZ proceeds to its most southerly position, producing a wet and warm season, coinciding with the eastern intertropical flux that brings water vapor from the Atlantic Ocean (Francou and others, 2003).

Rainfall is high variable depending on the elevation. On average, summer precipitation (October–March) contributes the most (greater than [>] 50 percent) to annual rainfall; however, small amounts of rain are possible year-round that contribute to the total annual rainfall. Peak summertime precipitation (December–February) is the only source of water for the high-altitude Altiplano regions (>2,500 m) in Bolivia. The glaciers in Bolivia contribute to the runoff in the rivers; however, runoff contribution varies seasonally, with low discharges in the dry season (May–August) and high values in the humid season (October–March).

Data

The following section describes the types of data used in this analysis and the sources of those data, which include both digital elevation and rainfall data.

Digital Elevation Data

Potential head drops and stream connectivity were calculated from the Shuttle Radar Topography Mission (SRTM) DEM, which has a 1-arc-second resolution (about 30 m at the equator). This 11-day mission was launched on February 11, 2000, onboard the space shuttle Endeavour. From the payload, two radars swept most of the Earth's surface and recorded its topography. The data are available from the National Geospatial-Intelligence Agency archived at the USGS Earth Resources Observation and Science Center.

The SRTM used a technique called interferometry to collect the topographic data (National Aeronautics and Space Administration, 2005a). The images collected from interferometry are not typical photographs of the Earth's surface. For SRTM, these images were produced by two synthetic aperture radar systems: C-band (Shuttle Imaging Radar with payload C [SIR–C]) and X-band (Synthetic Aperture Radar with payload X [X–SAR]) (Farr and others, 2007). For this discussion, we do not cover the detailed mechanics of data collection other than to state that the SRTM data referenced in this document

were collected by the C-band. To acquire these images during the mission, a main antenna was installed in the shuttle's cargo bay, and a second outboard antenna was deployed on a mast that extended 60 m from the cargo bay once in orbit (fig. 3) (National Aeronautics and Space Administration, 2005a). The main antenna transmitted a beam of radar waves toward the Earth. When these waves hit the surface, they scattered. The scattered waves that bounced back toward either antenna (received wave) were collected, and the distances traveled were measured. Using this information and knowing the constant distance between the two antennas (baseline), an accurate elevation of the Earth's surface could be calculated (figs. 3 and 4) (National Aeronautics and Space Administration, 2005b).
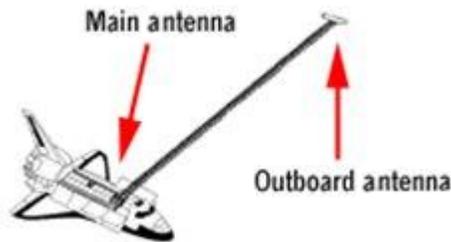
Figure 3.   Schematic diagram of radar antennas (National Aeronautics and Space Administration, 2005b).
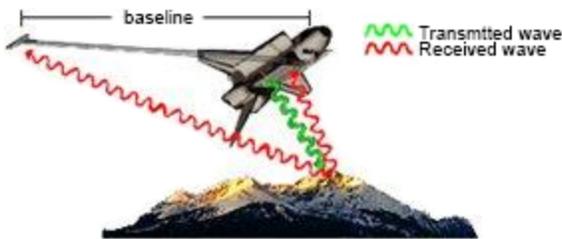
Figure 4.   Schematic diagram of transmission and receiving the beam of radar waves (National Aeronautics and Space Administration, 2005b).

The SRTM was designed with specific mapping accuracy thresholds to help ensure a consistent and accurate global topographic dataset. For South America, the SRTM dataset has an average horizontal error of 9.0 m and an average absolute vertical (height) error of 6.2 m (Rodriguez and others, 2006). South America's mapping accuracy has improved tenfold (in both spatial resolution and vertical accuracy) with the release of the SRTM DEM compared to the last continental DEM, the USGS GTOPO30 dataset (Gesch and others, 1999).

The SRTM DEM data have some data voids in areas that are in the shadow of the Andes Mountains. Most SRTM data voids are less than five pixels in size (Hall and others, 2005). Larger voids are either over water or in mountainous areas. Water surfaces produce radar signal scattering, which makes it impossible for the interferometer to detect meaningful reflections. In mountainous terrain on the eastern side of the Andes,

surface inclinations above 20 degrees (°) cause data voids because this steep terrain shadows the radar signal. The DEM data voids were filled with data from the USGS GTOPO30 and the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global DEM (GDEM) database. Using the Delta Surface Fill method (fig. 5) developed by the National Geospatial-Intelligence Agency (Grohman and others, 2006), the ASTER or GTOPO30 datasets were used to fill these voids. The GTOPO30 global topographic dataset (Gesch and others, 1999) has a resolution of 30 arc-seconds of latitude and longitude (about 1 km), and the GDEM version 1 (METI and others, 2009) has a resolution of 1 arc-second (30 m at the equator). The SRTM 30-m dataset was used in this study.
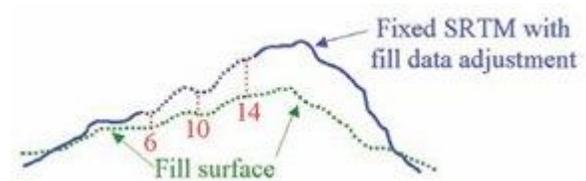
Figure 5.   Delta Surface Fill adjustment to fill the voids in the Shuttle Radar Topography Mission digital elevation model (Grohman and others, 2006).

The SRTM 30-m DEM for Bolivia and surrounding regions can become a large volume of data to process. To avoid the problem of handling large volumes of data, we divided the study region into five processing basins (fig. 6). This division enables easy processing of the DEM, and handling SRTM derivatives becomes feasible.

Rainfall Data

Two different types of rainfall datasets were used in this study—rainfall observations and rainfall estimates. The rain gage data used were obtained from Bolivian Servicio Nacional de Meteorología e Hidrología (SENAMHI), the Brazilian Agência Nacional de Águas (ANA), and the Peruvian Servicio Nacional de Meteorología e Hidrología (PSENAMHI). The gridded global monthly and annual rainfall climatology estimates were obtained from six different rainfall datasets.

Station Rainfall Data

Initially, data for 619 stations that cover Bolivia and neighboring basins and that have records for more than 1 year during 1981–2013 were obtained. Of these stations, 313 were obtained from SENAMHI, 259 from ANA, and 47 from PSENAMHI. All stations that met the criteria for data completeness, redundancy, and consistency were used for computing annual climatology. Out of the 619 stations, 41 were dropped because they had missing observations (for example, a station had data for only 4 years, but all the data for
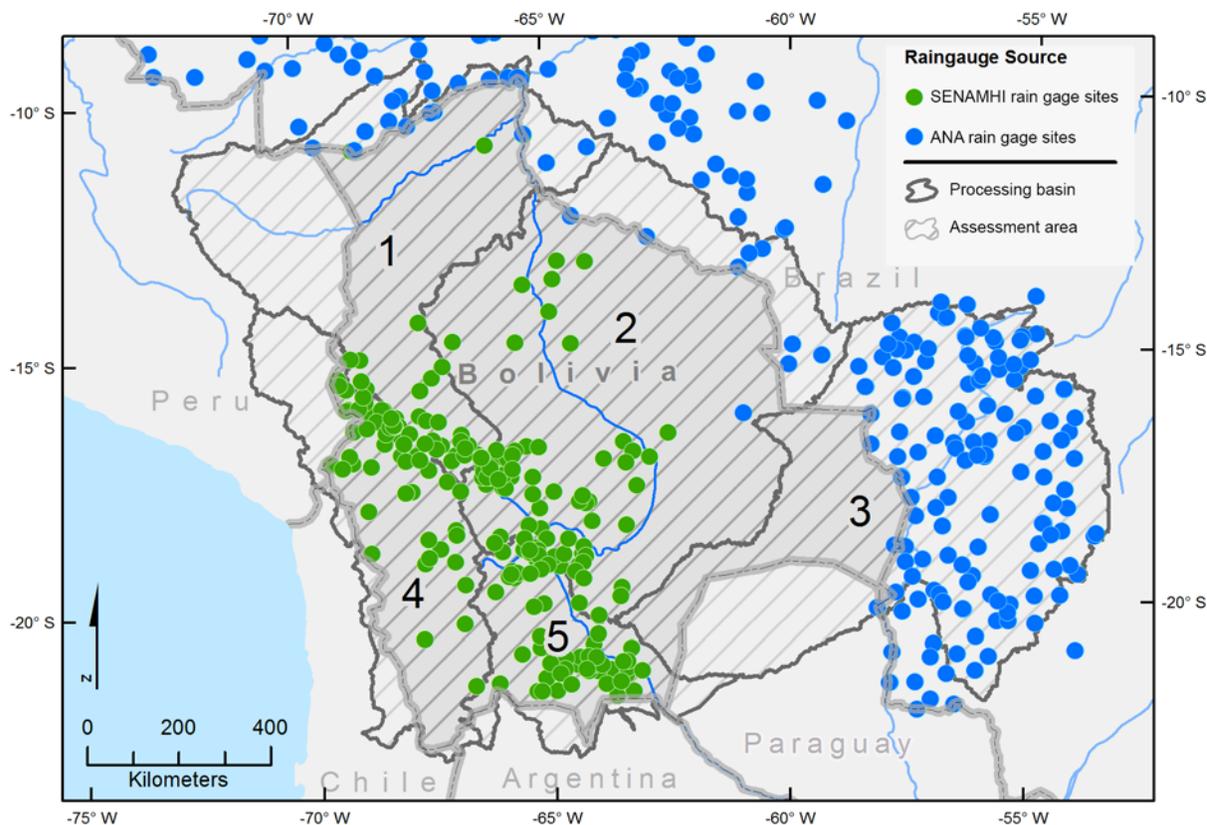
Figure 6. Distribution of rain gage data from Bolivia and Brazil across the processing basins identified for the hydropower analysis [ANA, Agência Nacional de Águas; SENAMHI, Servicio Nacional de Meteorología e Hidrología].

a station were obtained in the wet season). Such data were not used to compute annual climatology because of its bias toward high rainfall. All 47 PSENAMHI stations were dropped because only their mean annual values were available. Another 36 stations were omitted because of data redundancy (data for the same station obtained from two different sources), and 7 stations were dropped because station coordinates were missing. In sum, 154 stations out of 619 stations were omitted.

Finally, 233 stations from Bolivia and 235 stations from Brazil were chosen for computing climatology (fig. 6). The breakdown and distribution of stations for each processing basin (fig. 6) is provided in table 1. The climate rainfall information for each of these stations was used to validate global gridded climatology precipitation datasets.

Satellite-Based Rainfall Data

Six gridded rainfall climatology datasets were used to identify the best rainfall climatology dataset for Bolivia. The characteristics of all the rainfall datasets used in this study are provided in table 2.

The first dataset used in this study, the Global Precipitation Climatology Project (GPCP) Version 2.2 Combined Precipitation Dataset, is produced by combining observations and satellite precipitation data into 2.5x2.5° global grids

(Alder and others, 2003). This dataset was developed and computed by the National Aeronautics and Space Administration (NASA)/Goddard Space Flight Center's Laboratory for Atmospheres as a contribution to the Global Energy and Water Cycle Experiment GPCP. The GPCP precipitation data were obtained from http://www.esrl.noaa.gov/psd/.

The second dataset is the Climate Prediction Center Merged Analysis of Precipitation (CMAP) dataset. This dataset provides monthly global gridded precipitation climatology. The precipitation estimates are obtained from five kinds of satellite estimates (Geostationary Operational Environmental Satellite Precipitation Index [GPI], Outgoing

Table 1. Breakdown of the number of stations by processing basin of data used in the validation and bias correction of gridded precipitation datasets.

Basin (fig. 6)	Number of stations
Processing basin 1	63
Processing basin 2	85
Processing basin 3	128
Processing basin 4	45
Processing basin 5	63
Outside processing basins	81

Table 2.   Characteristics of gridded rainfall climatology datasets used in this study.

[GPCP, Global Precipitation Climatology Project; V, version; CPDS, combined monthly precipitation dataset; ~, about; km, kilometer; NOAA, National Oceanic and Atmospheric Administration; OAR, Oceanic and Atmospheric Research; ESRL, Earth Systems Research Laboratory; CMAP, Climate Prediction Center Merged Analysis of Precipitation; CPC, Climate Prediction Center; GPCC, Global Precipitation Climatology Center; TRMM, Tropical Rainfall Measuring Mission; TMPA, TRMM Multi-Satellite Precipitation Analysis ; NASA, National Aeronautics and Space Administration; CHIRPS, Climate Hazards Group InfraRed Precipitation with Stations; USGS, U.S. Geological Survey; FEWSNET TRMM Multi-Satellite Precipitation Analysis]

Rainfall dataset	Climatology period	Resolution	Frequency used	Source
GPCP V2.2 CPDS	1981–2010	~250 km	monthly, annual	NOAA/OAR/ESRL
CMAP (enhanced)	1981–2010	~250 km	monthly, annual	NOAA CPC
GPCC V6	1981–2010	~50 km	monthly, annual	NOAA GPCC
TRMM TMPA (3B43 V7)	1998–2013	~25 km	monthly, annual	NASA
CHIRPS	1981–2010	~5 km	monthly, annual	USGS/FEWSNET

Longwave Radiation [OLR] Precipitation Index [OPI], Special Sensor Microwave Imager [SSM/I] scattering, SSM/I emission, and Microwave Sounding Unit [MSU]) (Xie and Arkin, 1997; Saha and others, 2010). The enhanced file also includes blended National Center for Environmental Prediction/National Center for Atmospheric Research Reanalysis Precipitation values. The CMAP precipitation data were obtained from http://www.esrl.noaa.gov/psd/data/gridded/data.cmap.html.

The Global Precipitation Climatology Center (GPCC) produces monthly climatology precipitation datasets. The version 6 dataset (1901–present [2016]) is produced by combining precipitation observations from quality-controlled data from nearly 67,000 stations globally that feature record durations of 10 years or longer (Schneider and others, 2011).

The Tropical Rainfall Measuring Mission (TRMM) Multi-Satellite Precipitation Analysis (computed at monthly intervals as 3B43) combines 3-hourly integrated high-quality data and infrared (IR) estimates (3B42) with the monthly accumulated Climate Assessment Monitoring System or GPCC rain gage analysis, produced by the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center (CPC), the global rain gage product produced by the GPCC (Huffman and others, 2007), or both. Compared to version 6 precipitation data, several improvements were implemented in version 7: additional satellite data, a uniform data reprocessing and calibration scheme, and single use of GPCC monthly rain gage analysis. Further details on data processing and improvements are in the technical document available at ftp://precip.gsfc.nasa.gov/pub/trmmdocs/3B42_3B43_doc.pdf.

The Climate Hazards Group InfraRed Precipitation (CHIRP) with Stations (CHIRPS) dataset (Peterson and others, 2013) is a quasiglobal (50° S–50° N, 180° E–180° W), 0.05° resolution dataset, available from January 1981 to the present (2016). Initially, the CHIRP dataset (without station data correction) is produced by combining a long-term/climate (more than 30 years) rainfall accumulation dataset Climate Hazards Group Precipitation Climatology [CHPClim]) with three geostationary thermal IR satellite observations. These IR observations are the NOAA CPC IR (0.5-hour temporal resolution, 4-km spatial resolution, 2000–present), the National Climate

Data Center B1 IR (3-hourly, 8-km spatial resolution, 1981–2008) (Janowiak and others, 2001), and the NASA TRMM 3B42 product (Huffman and others, 2007). Rainfall estimates from CHIRPS are then corrected using in situ precipitation observations obtained from the World Meteorological Organization Global Telecommunication Systems station rainfall network to produce the corrected dataset known as CHIRPS. Rainfall data are compiled as 5-day rainfall accumulations.

All six precipitation datasets are available in monthly increments. Monthly precipitation is summed to compute annual climatology. Finally, gridded monthly and annual precipitation climatology datasets are used for further analysis (validation and bias correction).

Hydrological Conditioning

Despite being one of the most accurate global scale elevation models, the SRTM still has its challenges when modeling hydrologic features (Guth, 2006). First, radar shadows cause data voids (discussed in the "Digital Elevation Data" section). Another challenge with SRTM is that it is a surface model, not bare earth, which presents a vegetation bias in the elevation in areas of dense forest cover (Kellndorfer and others, 2004). Although the SRTM 30-m resolution dataset is considered "high-resolution" on a global scale, 30-m resolution still presents challenges for extracting hydrologic features. To compensate for these challenges and improve stream channel modeling and better represent ground conditions, the SRTM DEM was hydrologically conditioned to reinforce drainage pathways.

Drainage Network Development

The SRTM DEM worked well for modeling drainage networks in areas with moderate to high relief and well-developed, broad stream channels. There were some cases in high relief areas where stream channels formed narrow gorges at SRTM resolution. These features were too narrow to be recognized, resulting in inaccurately filling the narrow
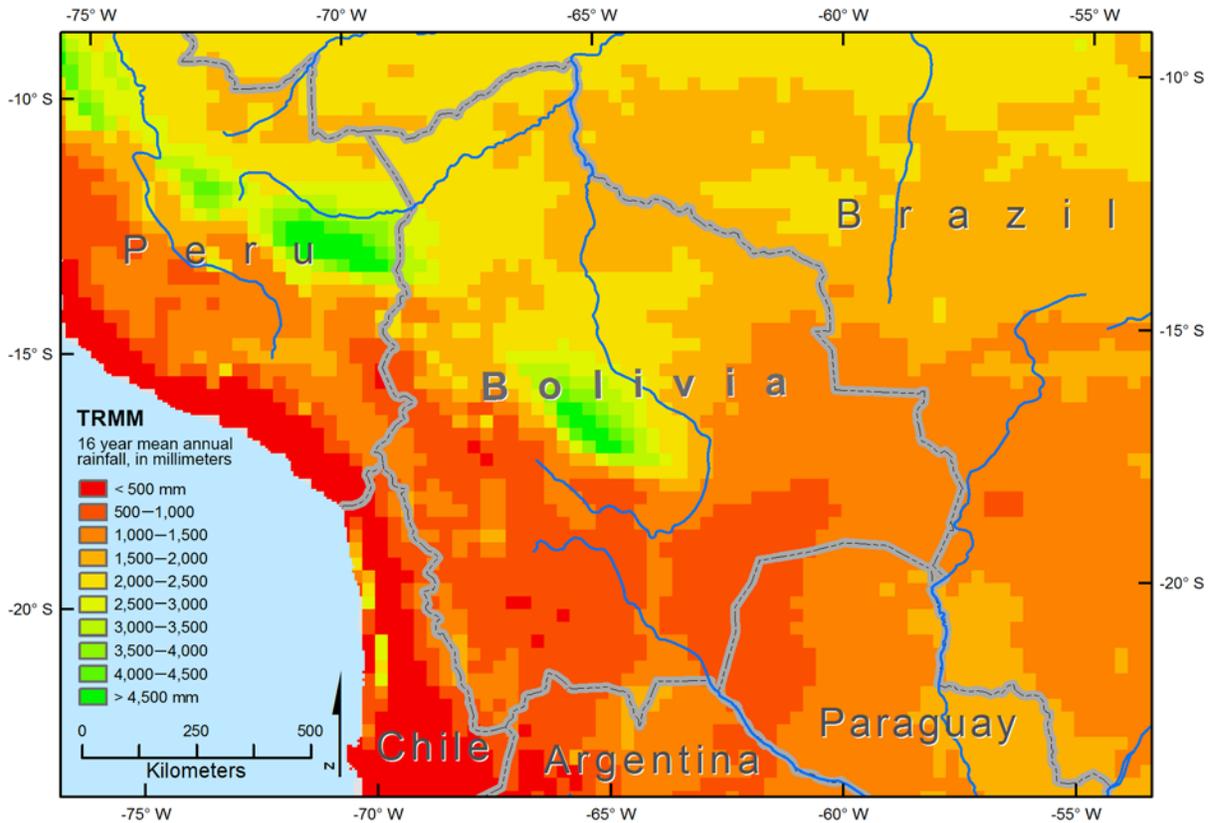
Figure 7.   Mean annual rainfall for the study region derived from Tropical Rainfall Measuring Mission (TRMM) 3B43 version 7 rainfall (1998–2013) at a horizontal resolution of 25 kilometers.

channel. Through the modeling process before hydrologically conditioning, these false elevation records generated artificial barriers resulting in stream channel displacement. The hydrologic conditioning steps are illustrated in figure 8. Panel A represents the original SRTM elevation with the channel's elevation profile at the bottom. The red line in panel B is the segment, interpreted from Landsat 8 imagery, used in the hydrologic conditioning process to reinforce the proper stream channel. Panel C shows the results of the process with the corrected channel and elevation profile at the bottom.

The inverse was true in the Oriente region. With its alluvial plains, swamps, flood plains, and open savanna, there is little topographic relief to influence drainage patterns. This is reflected in the 2015 Landsat 8 imagery, with swamps peppering the region and the Mamoré River meandering through the middle (fig. 9). In these cases, there is not enough relief expressed in the DEM in either the vertical or horizontal resolution to show these minute elevation changes. As a result, the hydrologic filling algorithm (appendix 1) tends to generate smooth flat surfaces. These smooth surfaces create a fishbone stream pattern and often misrepresent the streamflow (fig. 10, panel A). To mitigate the fishbone patterns and correct streamflow errors, the same conditioning process is used that corrected the filled gorges in high relief areas; however, unlike those areas where only small line segments were used, the flat plains require a more extensive channel network to reinforce the drainage. The source for defining these channels was the

SENAMHI 1:24,000 topographic hydrography data and Landsat 8 visually interpreted stream channels. For more details on this step of the conditioning process, see appendix 1. The inputs used in the conditioning process are illustrated in figure 10: panel A shows the unconditioned drainage model, panel B shows the SENAMHI topographic hydrography data layer, panel C shows the selected conditioning channels used for the final conditioning step, and panel D shows the results of the conditioning process.

In both situations, the topographic stream data provided by SENAMHI and stream channels interpolated from Landsat 8 imagery were applied to the DEM to reinforce the drainage network and better reflect ground conditions. This conditioning was a semiautomated process and was repeated until the modeled streams simulated what was observed in the imagery and completed for the study area. Overall, the hydrologic modeling yielded a drainage network reflecting a good spatial agreement with the hydrography data layer provided by SENAMHI (fig. 10, panel B).

Stream Segmentation

Once the conditioning process was completed, each contiguous stream in the drainage network was split into 1-km segments. The segmentation process moves upstream, starting at the confluence of two or more streams. A typical contiguous
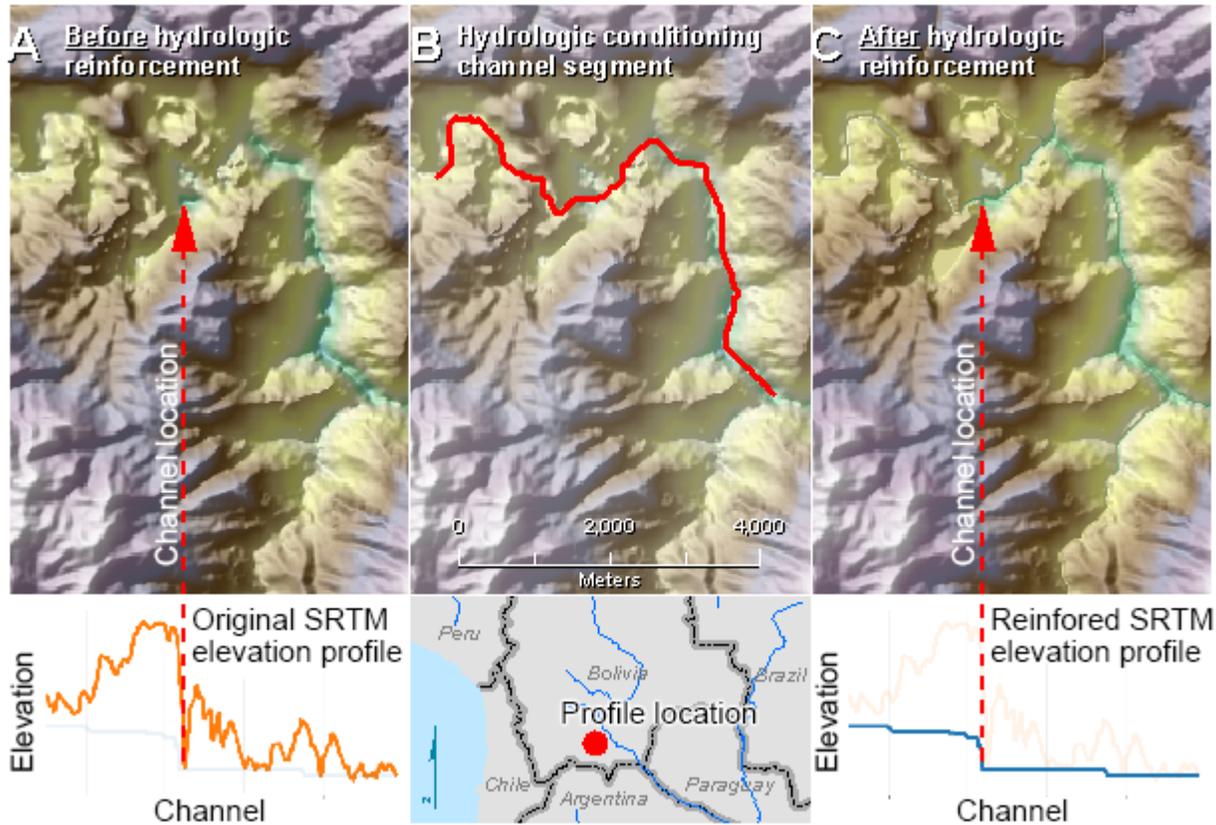
Figure 8.   Before and after steps of the hydrologic conditioning process. *A*, before hydrologic reinforcement. *B*, hydrologic conditioning channel segment. *C*, after hydrologic conditioning.
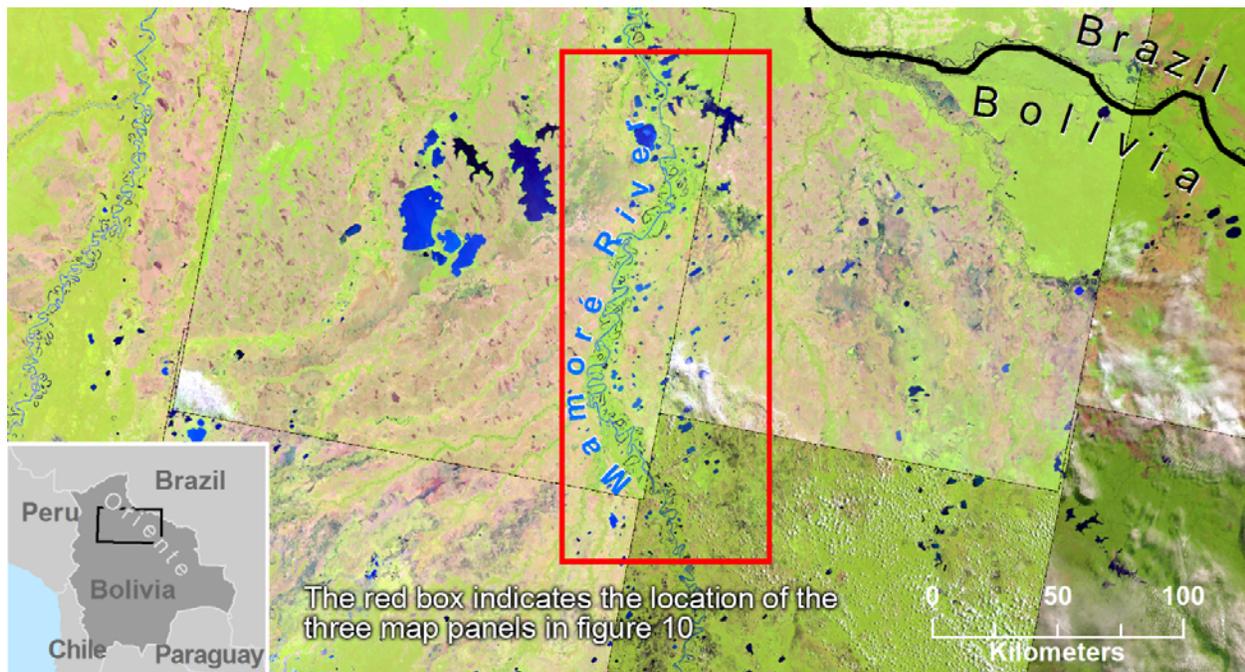
Figure 9.   Circa 2015 Landsat 8 imagery shows the flat alluvial plains, swamps, and savanna of the Oriente region.
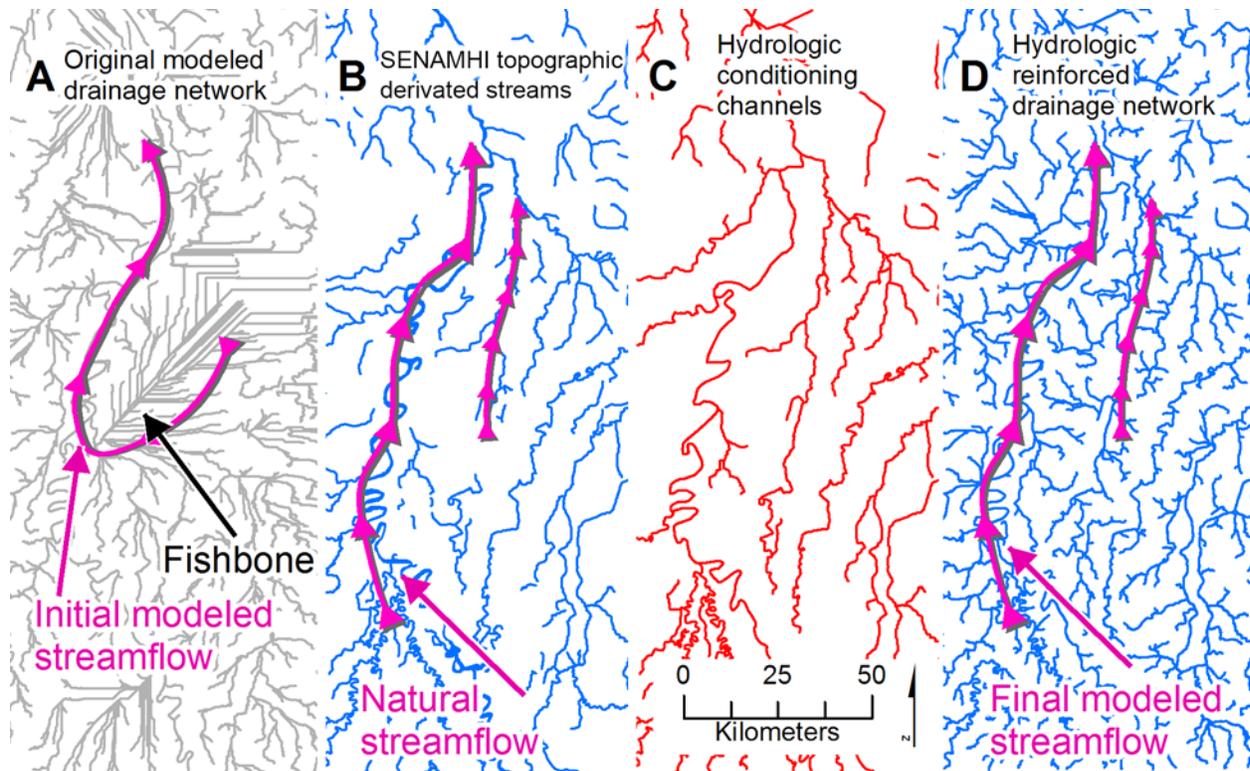
Figure 10. Hydrologic conditioning drainage network modeling inputs and results. *A*, the initial uncorrected drainage network with fishbone pattern and misrepresented streamflow. *B*, Servicio Nacional de Meteorología e Hidrología (SENAMHI) topographic hydrography used to aid in correcting errors from *A*. *C*, the conditioning channels derived from the hydrography and imagery interpretation. *D*, the results of the conditioning process after channels were applied in the model.

stream is split into 1-km segments with the final upstream segment being less than 1 km. This stream truncation is a result of the stream bifurcation shown in figure 11. Inversely, a small fraction of segments are greater than 1 km. These segments fall within areas affected by SRTM data voids (see the "Digital Elevation Data" section); in these cases, the segments were lengthened to cover the extent of the data void. For more details on the segmentation process, see appendix 1.

Generation of Best Rainfall Dataset for Bolivia

One of the key hydrological variables required for assessment of hydropower is the discharge/streamflow estimate for each 1-km river segment; however, accuracy and reliability of any hydropower assessment study depends heavily on the availability of good-quality precipitation estimates that are used, in turn, to estimate flow discharges. In theory, availability of long-term station (rain gage) records and good spatial distribution of the station network are required to produce a reliable rainfall dataset for the hydropower assessment model. Often, absence of station records or lack of good spatial

coverage leads to inaccuracies in the precipitation estimates, especially in data-scarce regions.

In Bolivia, we do not have a good spatial distribution of station records, especially in the high rainfall regions of processing basin 1 and in topographically complex southern parts of Bolivia (fig. 6). Most of the regions in the processing basins that lack station records, moreover, show high variability in terms of topography and hydroclimatology; hence, in this study using station observations, we first validate six gridded precipitation climatology datasets to identify the best precipitation model and further improve (bias correct) it using station observations.

Validation and Bias Correction of Rainfall Estimates

All satellite-based rainfall estimates are produced from the synthesis of observations from microwave- and IR-based sensors, and are often merged with station observations available globally. Although satellite-based precipitation datasets provide consistent and seamless estimates of rainfall on daily to annual time scales, they have errors (both random and consistent [bias] errors) when compared to the in situ
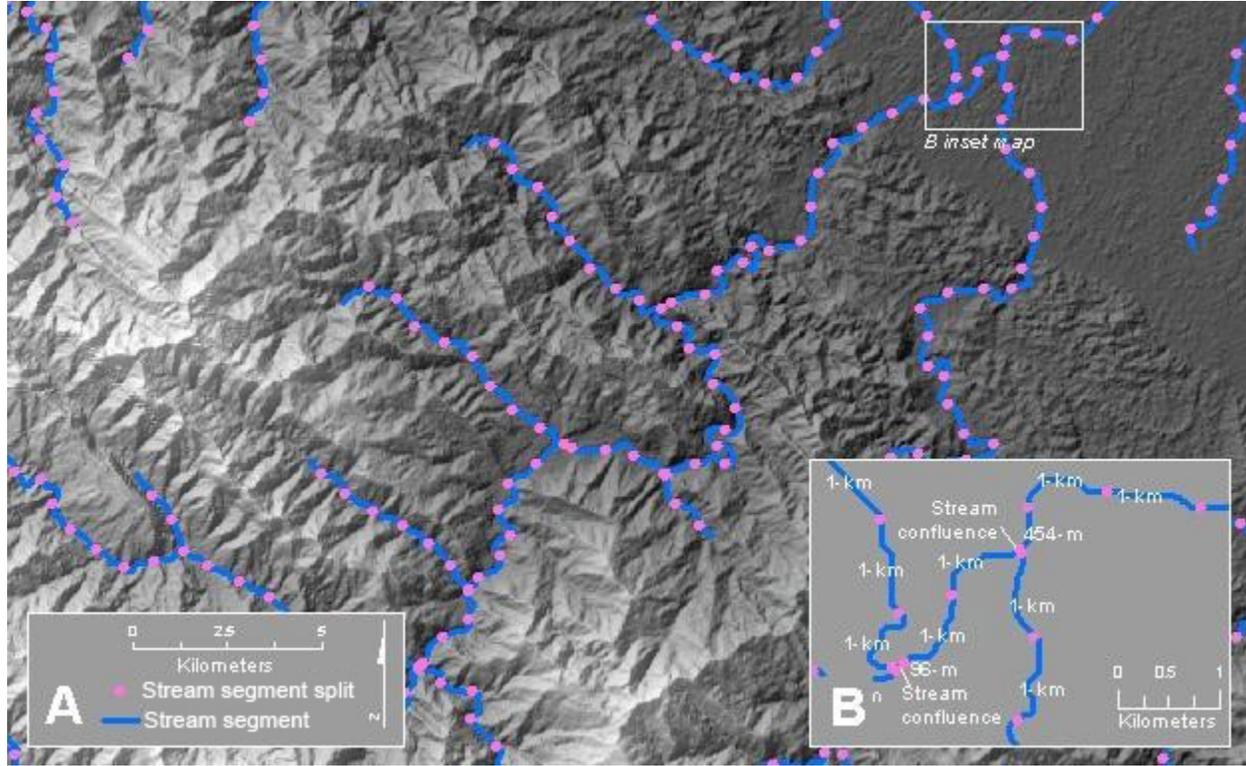
Figure 11. The larger panel *A* identifies the locations along the drainage network where the streams are split into 1-kilometer (km) segments. Inset panel *B* shows two cases where segments are less than 1 km immediately downstream from confluences [m, meter].

observations; hence, the errors in the rainfall dataset must be quantified.

In this study, we use data from 465 stations to validate and quantify the error in six gridded precipitation climatology datasets. Validation results are presented by month, year, processing basin (basin number 1, 2, 3, 4, and 5), elevation zone, and monthly and annual rainfall range.

The results of the validation are summarized using error indices. First, coefficient of determination (R^2, unitless) was computed to observe the degree of relation between the gridded precipitation climatology and observed station data.

To understand the goodness of fit, root-mean-square error (RMSE, in millimeters) was computed using equation 1. In this study, it is important to understand the RMSE in each gridded precipitation dataset because this index provides an overall estimate of the error including the influence of large errors (caused by the disagreement in precipitation estimates when compared to observed station data). Research indicates that satellite-based rainfall datasets poorly estimate precipitation in topographically complex regions. This is mainly because complex topography introduces high spatiotemporal variability in rainfall. This is particularly true in the mountainous and topographically complex regions of Bolivia. For this hydropower assessment study, we want to identify precipitation datasets that accurately modeled rainfall in these topographically complex regions characterized by high variability.

The next error index computed was the mean absolute error (MAE, in millimeters) using equation 2. This index is often not influenced by large errors but provides overall estimates of error in the dataset.

Another widely used measure in hydrology, the Nash-Sutcliffe Coefficient of Efficiency (NSCE), was used to compute the model efficiency (equation 3). The advantage of the NSCE is that it accounts for the model errors in estimating the mean of the observed datasets. The NSCE is an indicator of the model's ability to predict the 1:1 line (Nash and Sutcliffe, 1970). A value of 1 represents a perfect match, and a value of 0 or less is no more accurate than predicting the mean value.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Pg_i - Ps_i)^2}{n}} \qquad (1)$$

$$MAE = \frac{\sum_{i=1}^{n}|Pg_i - Ps_i|}{n} \qquad (2)$$

$$NSCE = 1 - \frac{\sum_{i=1}^{n}(Pg_i - Ps_i)^2}{\sum_{i=1}^{n}(Ps_i - \overline{Ps})^2} \qquad (3)$$

where

Pg is the gridded rainfall estimate,
Ps is the station-based rainfall estimate,

n is the total number of observations,

i represents the time step, and

$\overline{Ps}$ is the mean value of the station-based
           precipitation for a processing basin.

To enable comparison of errors, we computed percent error (*RMSE* [in percent] or *MAE* [in percent]) by dividing the error by the mean of the observation; hence, error statistics are reported as percent estimates.

Validation of Gridded Climatology Rainfall Estimates—Results

Monthly Validation of Gridded Rainfall Estimates

Monthly rainfall estimates from the gridded rainfall climatology datasets were extracted over the station location. The point estimates corresponding to all stations were pooled and compared between station and individual gridded precipitation datasets. The results (error indices) for each month are summarized in figure 12. Results indicate that accuracies for each dataset vary from month to month with lower errors in the wet season and higher errors in the dry season months. Precipitation datasets, such as GPCP and CMAP, had the lowest overall accuracies. On the other hand, GPCC and TRMM had higher accuracies, whereas accuracies of the CHIRPS dataset lie in between. The TRMM had higher accuracies and lower errors during the rainy season. Monthly validation results for each dataset for each month are presented in appendix 2.

Validation of Gridded Annual Rainfall Climatology Estimates (1981–2013)

Processing basin-wide validation results are presented in figure 13, which compares individual error indices for different processing basins. Results indicate that GPCP, CMAP, and CHIRPS have low R^2 and high errors, whereas TRMM and GPCC have higher R^2 and lower errors. The NSCE was highest for TRMM.

Areas or locations with the most potential hydropower are often in high elevation zones or in areas of rugged terrain. Areas with complex terrain, however, have complex physiography, and it is difficult to capture the rainfall intensity and

variability accurately; hence, analysis of accuracies and errors in the gridded climatology precipitation dataset by different elevation zones is presented in figure 14.

Results indicate that all datasets had higher accuracies at elevations lower than 1,000 m with R^2 about 0.8; however, accuracies decrease as elevation increases. The GPCP, CHIRPS, and CMAP had lower R^2 in elevation ranges of 1,000–3,000 m than in higher elevations. The GPCP had the lowest accuracy and highest RMSE error (>100 percent) in high elevations (>3,000 m). The TRMM had the highest overall R^2 and lowest errors; however, CHIRPS had the overall lowest relative bias in the precipitation estimates.

Validation of Gridded Rainfall Climatology Datasets by Rainfall Amounts

Since this study estimates theoretical hydropower potential of stream segments, it is important to accurately capture the dynamic range of precipitation. Because precipitation datasets exhibit lower accuracies in high elevation regions, it is important to understand the accuracies and errors for different amounts of annual rainfall. We evaluated all the gridded precipitation datasets based on different rates of rainfall. First, we validated datasets based on monthly rainfall rates (less than [<] 200, 200–500, and >500 millimeters per month [mm/month]) (fig. 15). The evaluation results indicate that all precipitation datasets had high accuracies (R^2>0.7) at low monthly rainfall rates (<200 mm/month); however, at higher rates (>200 mm/month) of rainfall, R^2 decreased. In terms of errors, GPCP had the highest errors when the rainfall rate was <200 mm/month; however, CMAP had the lowest accuracies at other rainfall rates (200–500 and >500 mm/month). The CHIRPS had the fewest errors at <200 mm/month (when compared to others). On the other hand, TRMM and GPCC had consistently higher accuracies and lower RMSE compared to CHIRPS. A comparison of individual scatterplots is provided in appendix 2.

Similarly, we evaluated accuracy at different rainfall rates (<3,000 millimeters per year [mm/year] and >3,000 mm/year) (fig. 16) obtained from gridded annual precipitation datasets. Results indicated that TRMM and GPCC had higher accuracies at higher rainfall rates.
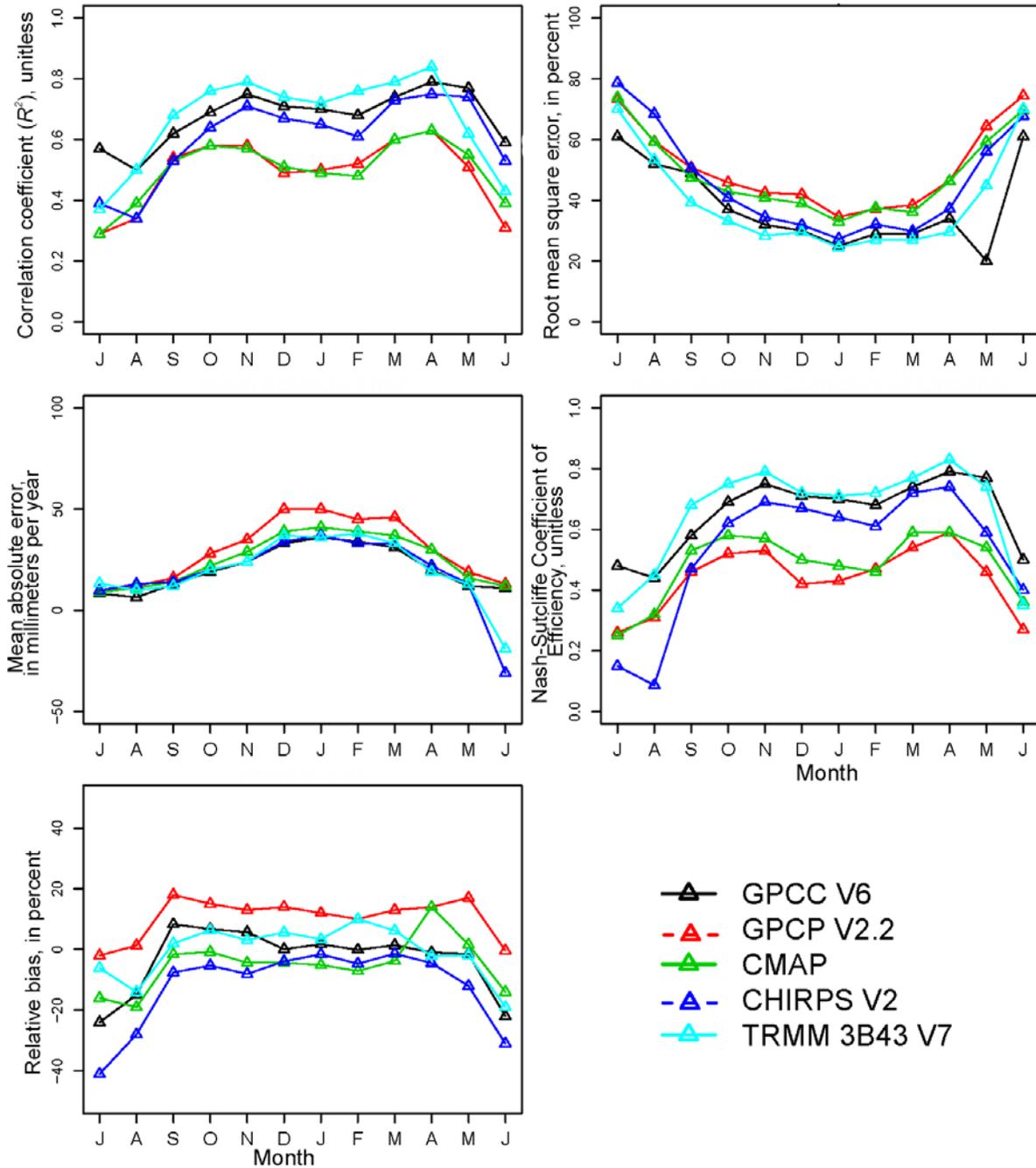
Figure 12.   Validation of monthly climatology rainfall estimates (1981–2013) using station observations: Monthly distribution of error indices [GPCC, Global Precipitation Climatology Center; V, version; GPCP, Global Precipitation Climatology Project; CMAP, Climate Prediction Center Merged Analysis of Precipitation; CHIRPS, Climate Hazards Group InfraRed Precipitation with Stations; TRMM, Tropical Rainfall Measuring Mission].
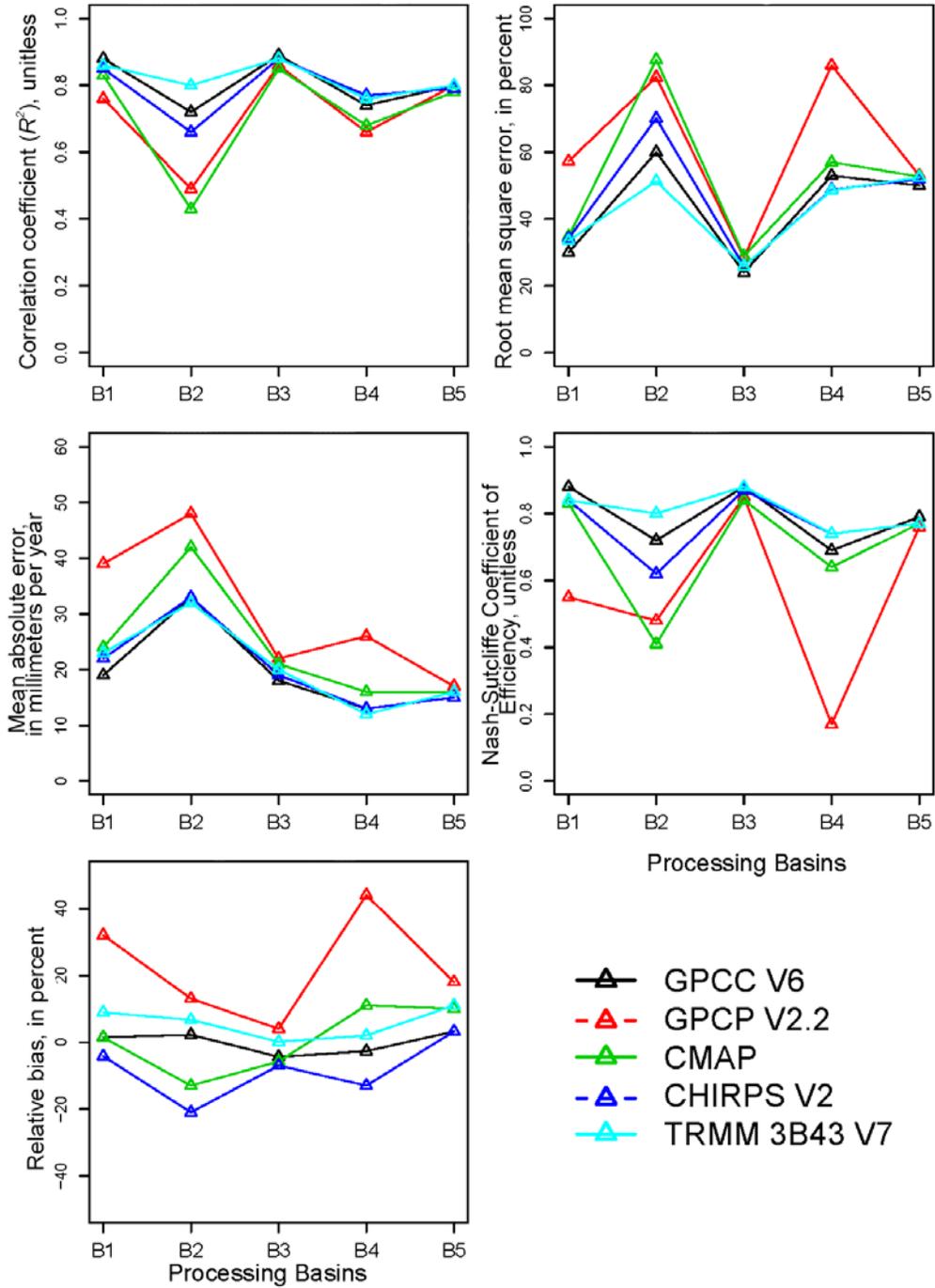
Figure 13.   Validation of gridded annual rainfall climatology estimates (1981–2013) against station observations aggregated by processing basin [GPCC, Global Precipitation Climatology Center; V, version; GPCP, Global Precipitation Climatology Project; CMAP, Climate Prediction Center Merged Analysis of Precipitation; CHIRPS, Climate Hazards Group InfraRed Precipitation with Stations; TRMM, Tropical Rainfall Measuring Mission].
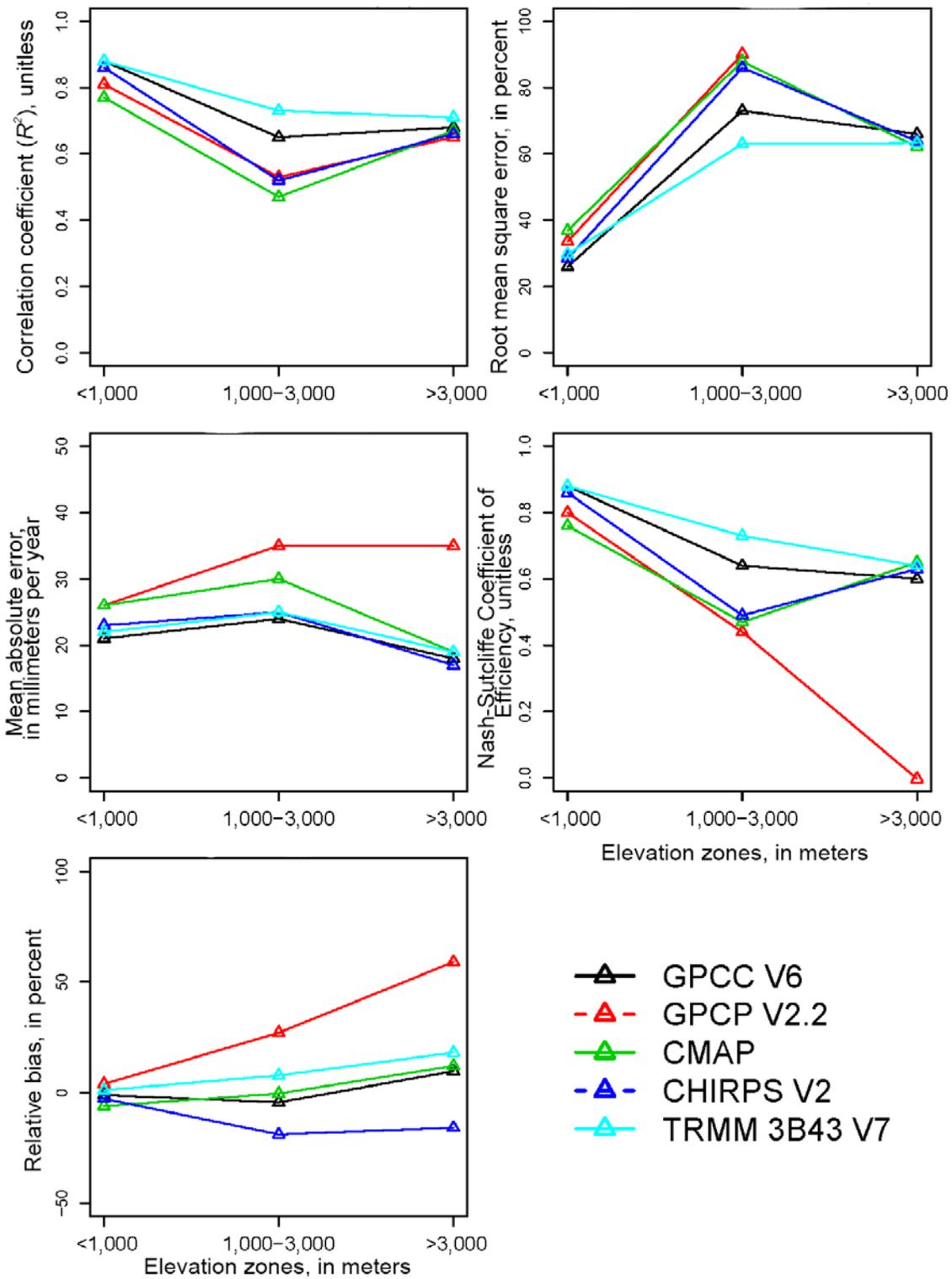
Figure 14.   Validation of gridded annual rainfall climatology estimates (1981–2013) against station observations aggregated by two elevation ranges [GPCC, Global Precipitation Climatology Center; V, version; GPCP, Global Precipitation Climatology Project; CMAP, Climate Prediction Center Merged Analysis of Precipitation; CHIRPS, Climate Hazards Group InfraRed Precipitation with Stations; TRMM, Tropical Rainfall Measuring Mission].
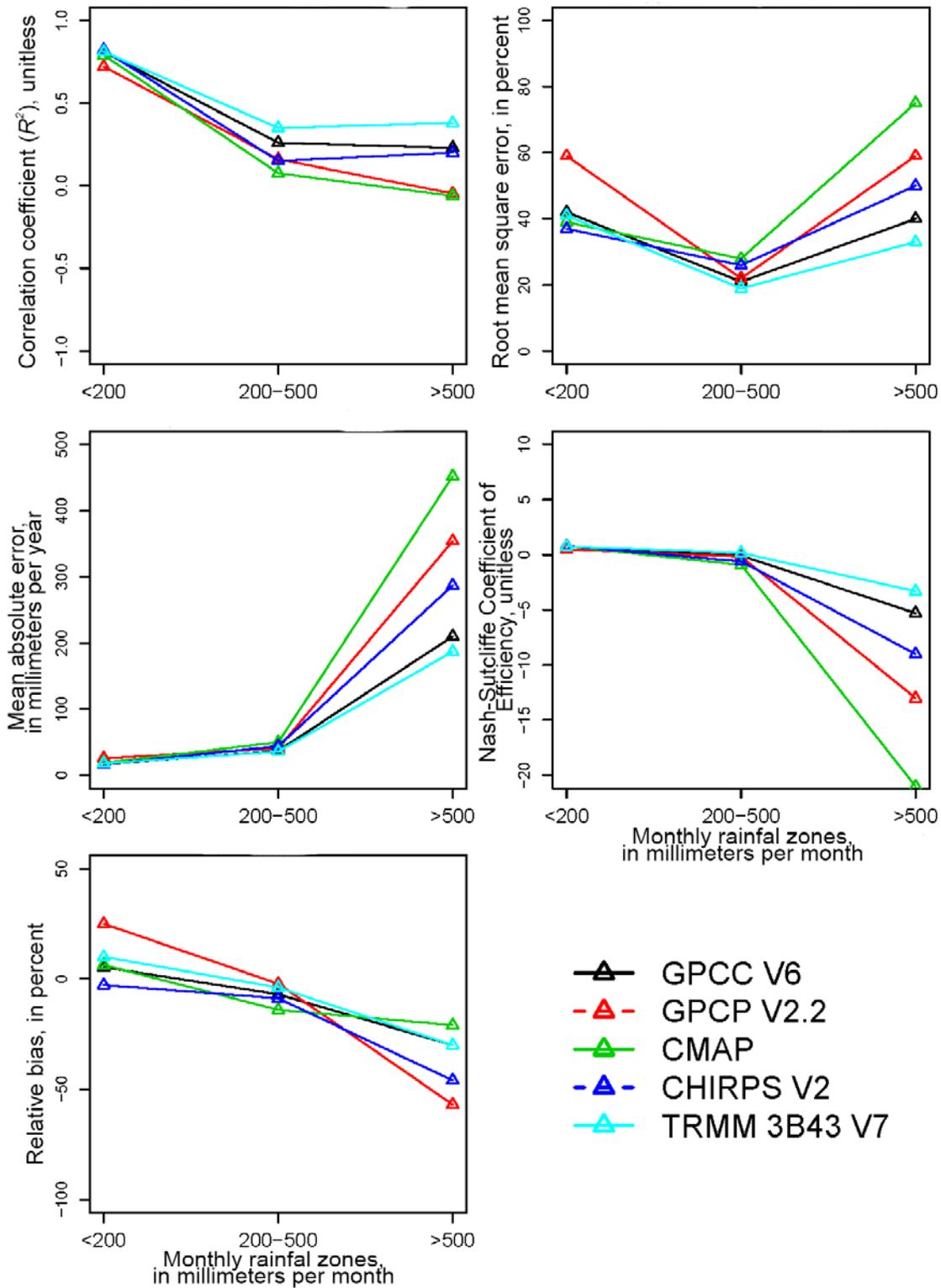
Figure 15.   Validation of gridded monthly rainfall climatology estimates (1981–2013) against station observations aggregated by different monthly rainfall totals [GPCC, Global Precipitation Climatology Center; V, version; GPCP, Global Precipitation Climatology Project; CMAP, Climate Prediction Center Merged Analysis of Precipitation; CHIRPS, Climate Hazards Group InfraRed Precipitation with Stations; TRMM, Tropical Rainfall Measuring Mission].
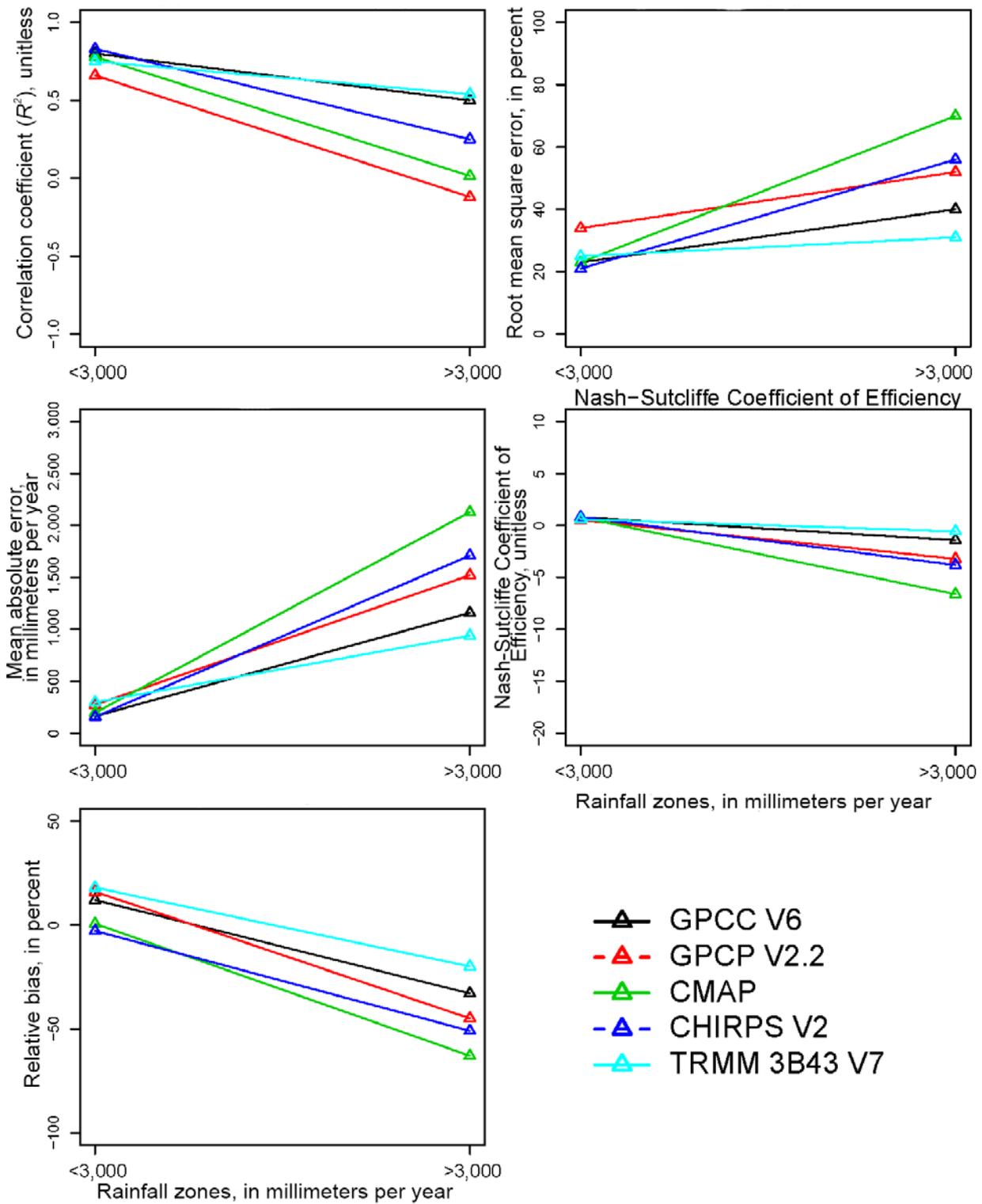
Figure 16.   Validation of gridded annual rainfall climatology estimates (1981–2013) against station observations aggregated by different annual rainfall totals [GPCC, Global Precipitation Climatology Center; V, version; GPCP, Global Precipitation Climatology Project; CMAP, Climate Prediction Center Merged Analysis of Precipitation; CHIRPS, Climate Hazards Group InfraRed Precipitation with Stations; TRMM, Tropical Rainfall Measuring Mission].

Summary of Validation Results

Summaries of all validation results are presented in table 3. This comparison summary is based on the validation results obtained from gridded climatology precipitation datasets.

Table 3. Validation results (error indices) obtained from annual gridded precipitation datasets summarized for different datasets.

[R^2, coefficient of determination; RMSE, root mean squared error; MAE, mean absolute error; NSCE, Nash-Sutcliffe Coefficient of Efficiency; GPCC, Global Precipitation Climatology Center; GPCP, Global Precipitation Climatology Project; CMAP, Climate Prediction Center Merged Analysis of Precipitation; CHIRPS, Climate Hazards group InfraRed Precipitation with Stations; TRMM, Tropical Rainfall Measuring Mission]

Dataset	R^2	RMSE (percent)	MAE (millimeter)	NSCE	Relative bias percent
GPCC	0.77	26.5	186	0.77	0.50
GPCP	0.66	38.6	302	0.50	13.00
CMAP	0.57	36.0	240	0.57	-3.30
CHIRPS	0.71	30.0	191	0.70	-5.80
TRMM	0.81	24.1	190	0.81	17.00

The R^2 (the goodness of fit) was highest for TRMM. The GPCC had a comparable but slightly lower R^2 than TRMM. Other datasets, such as CHIRPS and GPCP, had an intermediate range of R^2 values, whereas CMAP had the lowest R^2.

The RMSE error in precipitation ranged from 24.1 to 38.6 percent, and MAE ranged from 186 to 302 mm/year. The NSCE demonstrates the ability of the dataset to simulate the 1:1 line. The dataset that is closer to 1 is considered the best. Results indicate that TRMM was the closest to the observed rainfall.

Hence, based on the results of R^2, RMSE, and NSCE, the TRMM precipitation dataset had higher accuracy than other rainfall datasets; however, the TRMM demonstrated a positive but consistent bias (17.0 percent), highlighting the need for bias correction.

Disaggregating Tropical Rainfall Measuring Mission Using Climate Hazards Group InfraRed Precipitation with Stations Rainfall Data

Although TRMM data had the highest R^2, highest NSCE, and the lowest RMSE, its climatology is computed from 15 years of data (1998–2013); moreover, the resolution of the dataset is 25 km. At this resolution, it is difficult to capture the variability in rainfall, especially in the high elevation regions of Bolivia where the complex topography can introduce high variability in rainfall over short distances. To overcome this problem, we used CHIRPS rainfall data because of its higher resolution (5 km) and because its climatology is computed using data from a 33-year record (1981–2013).

We used a simple disaggregation method that uses the mean from the TRMM precipitation dataset and variance from the CHIRPS dataset; hence, the end result is high-resolution TRMM with similar accuracies as the original TRMM. The disaggregation method can be explained using equation 4.

$$P(Td)_{x',y'} = (P(T)_{x,y} + [P(C)_{x',y'} - \overline{P(C)_{x,y}}]) \qquad (4)$$

where
<div style="padding-left:2em">

$P(Td)$ is the disaggregated TRMM (5 km),

$P(T)$ is the original TRMM (25 km) resampled to 5 km,

$P(C)$ is the CHIRPS precipitation (5 km),

x,y is the longitude and latitude of a pixel from the original 25-km TRMM data resampled to 5 km,

x',y' is the longitude and latitude of a pixel from the 5-km CHIRPS data, and

$\overline{P(C)_{x,y}}$ is the CHIRPS mean precipitation estimate obtained from a 5x5 km window (which is equal to one resampled TRMM pixel).

</div>

The disaggregated TRMM (5 km) annual precipitation climatology dataset is referred to as disaggregated TRMM (TRMM–D).

Bias Correction of Disaggregated TRMM Precipitation Dataset

Errors in satellite-based precipitation products are often due to systematic bias (instrument error) and random error. Although random error cannot be corrected, most of the bias error can be corrected using station records. Bias correction of the best gridded precipitation model is important for this study because the correct high precipitation events and high flow estimates are needed to produce reliable estimates of hydropower. Bias correction improves the overall accuracy and reliability of satellite-based precipitation (Vernimmen and others, 2012). In this study, we use data from 465 stations to correct the bias in the TRMM–D dataset.

The annual TRMM–D climatology precipitation dataset was analyzed for the presence of overall bias. The comparison of station observations and annual TRMM–D estimates is shown in figure 17. The two red lines in figure 17 indicate the presence of differences in bias estimates for different rates of annual rainfall; hence, a bias correction was applied for annual rates of <2,500 mm/year and >2,500 mm/year separately. We estimated bias in the processing basin scale precipitation estimates by combining data for all stations within a processing basin using equation 5.

$$Br = \frac{\sum_{i=1}^{n}(Pg_i - Ps_i)}{\sum_{i=1}^{n} Ps_i} \times 100 \qquad (5)$$

where
<div style="padding-left:2em">

Br is the bias estimate

n is the number of observations,

i is the time period,

Pg is the gridded precipitation dataset (TRMM–D), and

</div>

Ps   is the precipitation estimate obtained from stations.

Using equation 5, bias was estimated for rainfall rates as shown in table 4.

Table 4.   Relative bias in CHIRPS identified for all the processing basins.

[mm/yr, millimeter per year]

Rate	Relative bias percent	Description of bias
$\leq 2{,}500$ mm/year	18.0	Overestimation.
$>2{,}500$ mm/year	-4.5	Underestimation.
All rainfall	17.0	Overestimation.

$$Pg_{i,corr} = (100 - B_{r,i}) \, x \, Pc_i \qquad (6)$$

where

$Pg_{i,corr}$ is the bias-corrected gridded rainfall,
$B_{r,i}$ is the bias rate for a time period i, and
Pc_i is the precipitation estimate for the time period i.

The bias-corrected and TRMM–D is incorporated into the geodataset.

Comparison of Uncorrected and Bias-Corrected CHIRPS Rainfall

The comparison of uncorrected and bias-corrected TRMM–D rainfall is shown in figure 18. At all rain rates (combined data), results indicate that bias correction improves the bias estimates from an overestimation of 17 to -2.9 percent. The bias in the rainfall for the annual rates of less than or equal to ($\leq$) 2,500 mm/year was corrected (bias reduced from 18 to -3 percent).

This -3 percent underestimation is an overall estimate, and a close look at the scatterplot reveals that at much lower rates of rainfall (<1,000 mm/year), the bias-corrected TRMM–D still slightly overestimates at higher rates of rainfall. Similarly, the bias in the rainfall for the annual rates of >2,500 mm/year was corrected (bias improved from -4.5 to -0.1 percent). The RMSE, MAE, and NSCE for high rainfall regions (>2,500 mm/year) were much improved; however, the overall accuracy remains mostly unchanged except for the improvement in the bias.
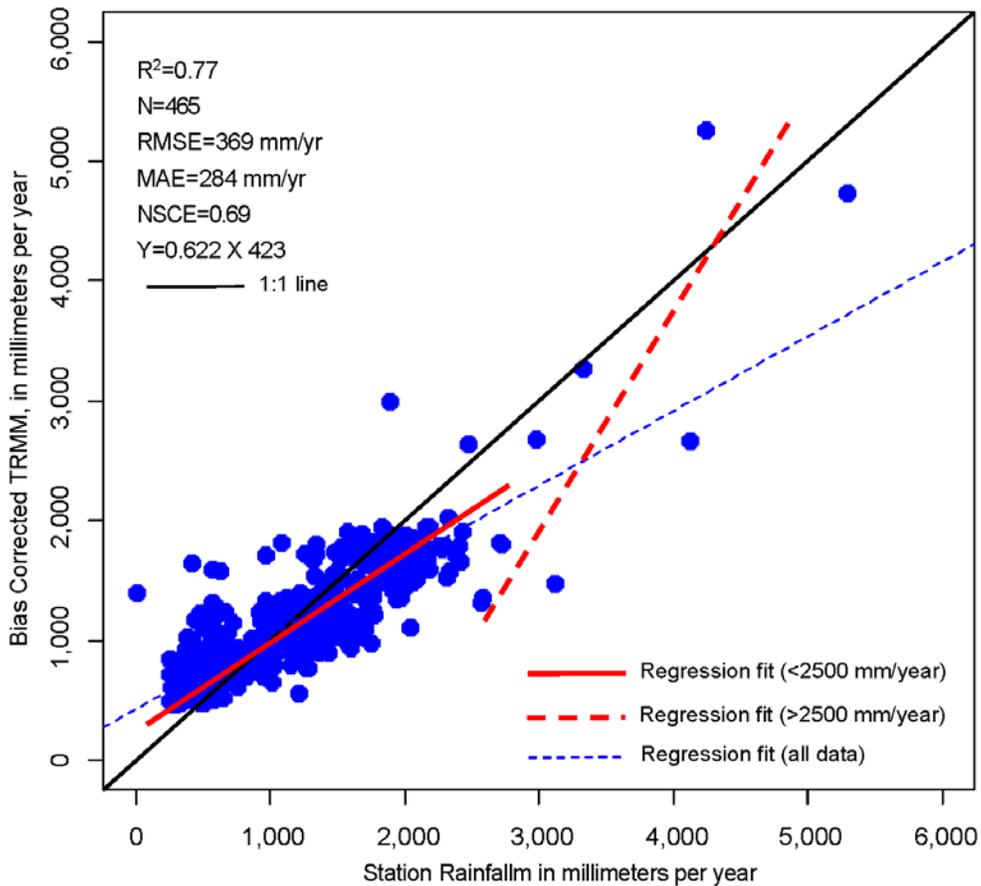
Figure 17.   Scatterplot showing comparison of disaggregated Tropical Rainfall Measuring Mission (TRMM–D) precipitation estimates against station observations [R^2, coefficient of determination; N, number; RMSE, root-mean-square error; MAE, mean absolute error; NSCE, Nash-Sutcliffe Coefficient of Efficiency].
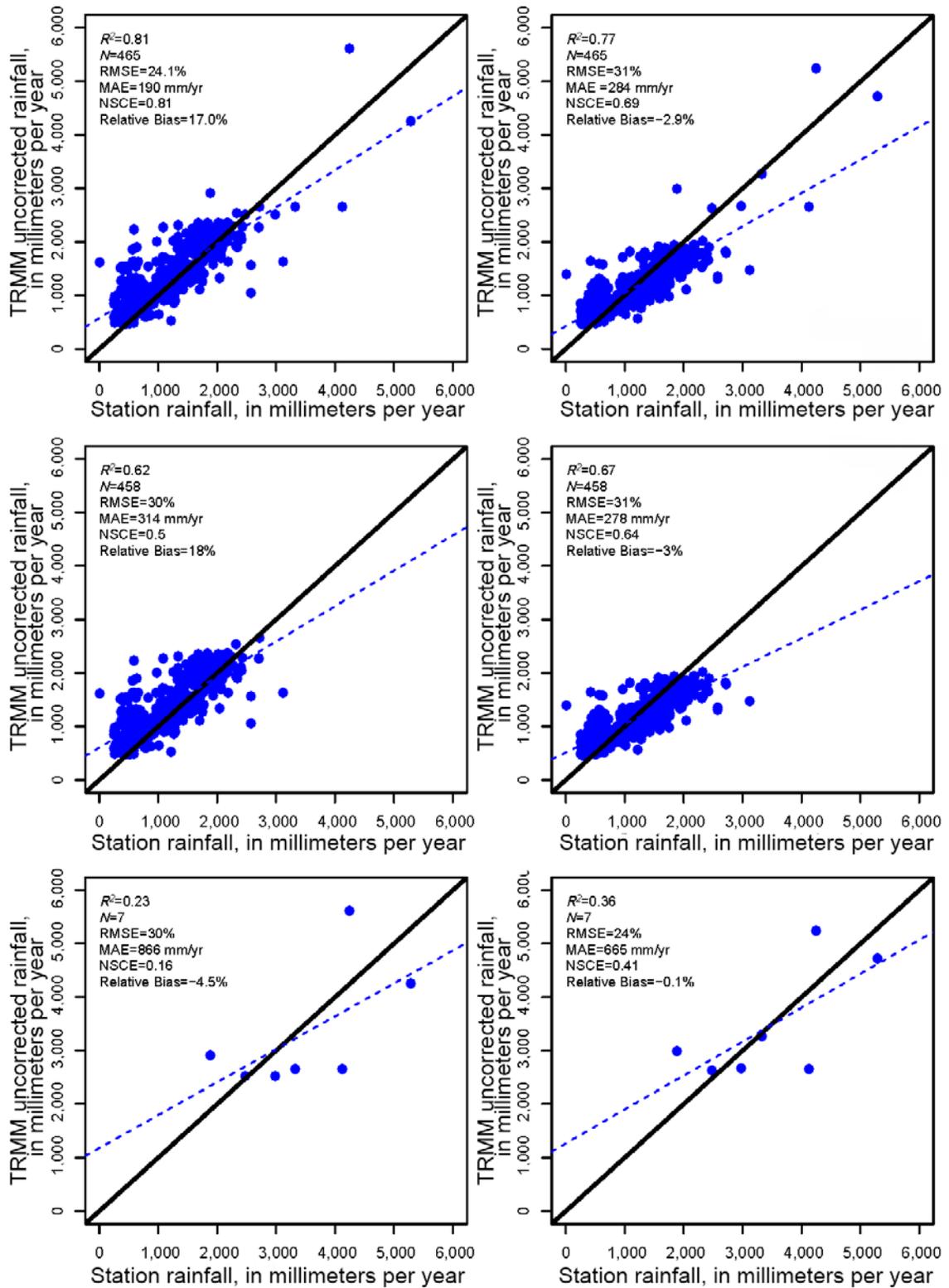
Figure 18. Scatterplots showing the comparison of uncorrected and bias-corrected disaggregated Tropical Rainfall Measuring Mission (TRMM–D) rainfall for (a) all rainfall rates (combined data), (b) rain rates ≤2,500 mm/year, and (c) rain rates >2,500 millimeters per year (mm/year) [R^2, coefficient of determination; N, number; RMSE, root-mean-square error; MAE, mean absolute error; NSCE, Nash-Sutcliffe Coefficient of Efficiency].

Basin Regionalization

Introduction

Reliable estimates of streamflow are crucial for the determination of hydropower potential for any river segment. Streamflow can be estimated using established rating curves and measured river stage at gage stations or through hydrologic modeling calibrated with spatially distributed observed discharge records. For ungaged subbasins or subbasins with very few observation stations, streamflows are best estimated using regression models based on basin characteristics, which involves the transfer of streamflow information from gaged to ungaged sites within a prespecified hydrologic homogeneous region. A homogeneous region is formed by a group of subbasins that can be considered to be similar in terms of basin hydrologic response. Through meaningful homogenization, a strong relation in streamflow properties and basin characteristics can be determined and can be used to estimate streamflow at ungaged subbasins with similar patterns (Chiang, 1996). Similarities in basin climatic, topographic, and physiographic characteristics are often considered most responsive to basin streamflow characteristics (Abida and Ellouze, 2006); therefore, for Bolivia, we have used more than 30 years of rainfall-pattern and topography data to group subbasins in a meaningful way. Streamflow varies widely from stream to stream and is influenced by the spatial distribution of rainfall, evaporation demand, travel times of water from runoff source areas through surface and subsurface reservoirs and channels to the stream gage, and human management (Chiang, 1996); therefore, the selection of variables in grouping the subbasins influences the homogenization results (Fovell and Fovell, 1993; Stooksbury and Michaels, 1991). In general, the greatest similarities are associated with the smallest distances (Tasker, 1982); however, for the hydrologic homogeneous regions, the grouped subbasins may or may not be geographically contiguous (Burn, 1997; Ercan and others, 2008). Many different approaches have been used for hydrologic homogenization, such as K-means, Ward's method, L-moments, self-organizing map, and fuzzy classification. In this study, we have used the most widely used K-means clustering approach to identify homogeneous regions in the study area.

Input Data for Regionalization

In preparation for the homogeneous regionalization, the study area was subdivided into 271 subbasins, delineated by light gray lines in figure 19. The location and size of the basins were based on upstream areas greater than 6,000 square kilometers (km^2). Because this measurement is based on accumulated area and not surface area, a basin's surface area can vary widely based on the occurrence and location of confluences.
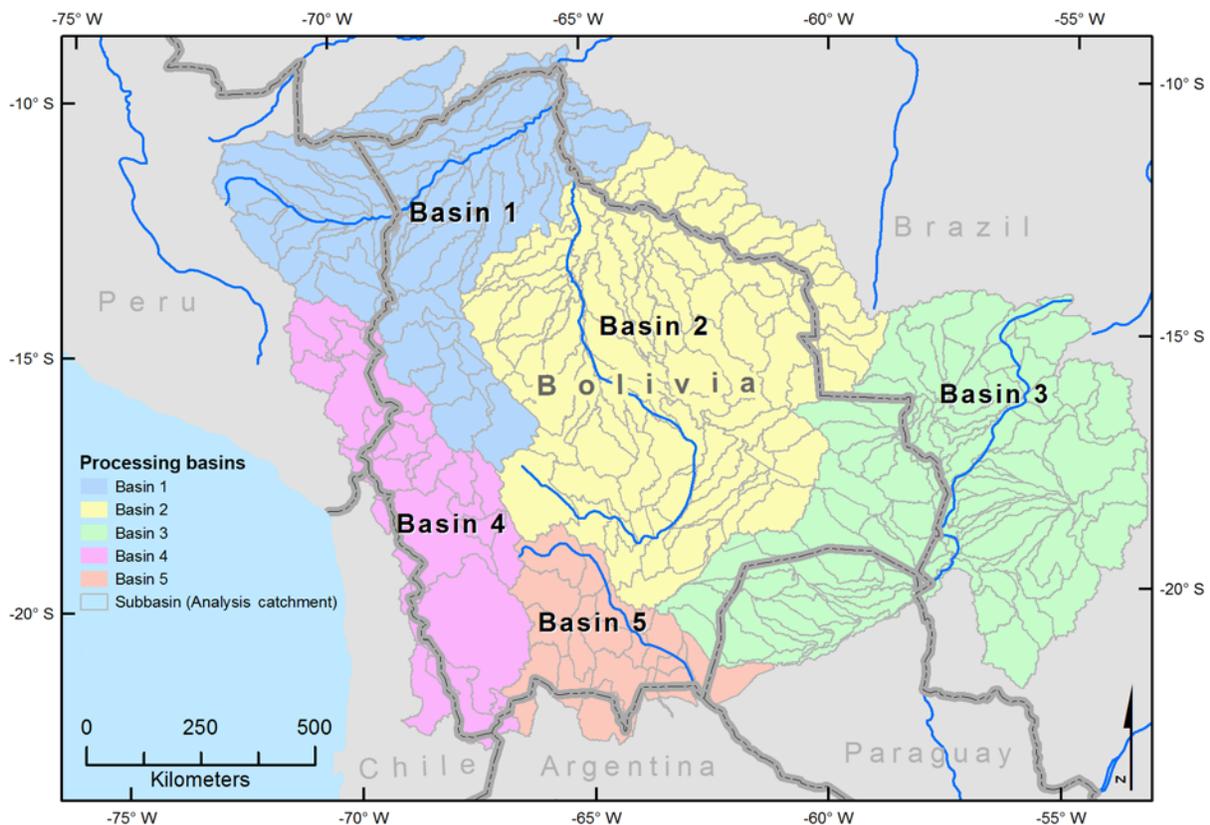
Figure 19. The subbasin map of the study area color-coded by the corresponding large basin and overlaid by country boundaries.

Table 5. Summary statistics of rainfall and topography in the study area.

[mm, millimeter; m, meter]

Subbasins	Annual total rainfall (mm)	Range of annual total rainfall (mm)	Rainfall standard deviation (mm)	Mean elevation (m)	Range of elevation (m)	Elevation standard deviation
Minimum	471	39	9	97	40	6
Maximum	4,729	4,804	1,642	4,609	6,021	1,686
Mean	1,410	544	134	797	1,131	195
Standard deviation	714	778	215	1,201	1,369	294

The basins' surface areas ranged from 1,000 to 22,000 km² with a median surface area of 6,000 km². Three catchments exceeded 22,000 km²: two of them were closed basins in the Uyuni Salt Flats and one was an aggregation of smaller catchments encompassing Lake Titicaca. The target area for these catchments was between 6,000 and 9,000 km². Excluding the three large basins, the final areas of each basin varied between 612 and 21,000 km², with a mean of 7,000 km² and a standard deviation of 5,300 km².

These 271 subbasins were aggregated to form Basins 1, 2, 3, 4, and 5, and each subbasin was assigned a unique identification number (fig. 19). Because homogeneous regions may be found for any set of variables, the possible weighting of the most appropriate set of variables constitutes an important step in the regionalization process. Since streamflow is an integrated variable of atmospheric and land processes (Ercan and others, 2008), we used rainfall and topographic characteristics to identify homogeneous regions among these 271 subbasins. Because of insufficient spatial distribution of observed rainfall stations with consistent periods of observations, we used bias-corrected TRMM–D rainfall along with SRTM DEM (Grohman and others, 2006) at 30-m spatial resolution in homogenization. We computed total rainfall climatology at resampled 5 km resolution from 16 years (1998–2013) of TRMM records merged with rainfall variability taken from 33 years (1981–2013) of CHIRPS records. The mean annual total rainfall was then spatially averaged for each of the subbasins and attributed to the subbasins. Standard deviation of the annual total rainfall and rainfall range within the subbasins was also computed and attributed to the subbasins. Similarly, mean elevation and range of the elevations for each of the subbasins have been computed and attributed to the subbasins. The distribution of rainfall and elevation patterns in the study area, which covers 1.9 million km² (with an average subbasin size of 7,219 km²), is characterized in table 5.

The summary statistics in table 5 suggest that the rainfall in the region is characterized by high spatial and temporal variability. The mean annual total rainfall in the region is 1,410 millimeters (mm) with a high degree of spatial variability between a minimum of 471 mm and a maximum of 4,729 mm. The standard deviation and range of subbasin rainfall also indicates a high degree of spatial variability among the basins. The standard deviation varies between 9 and 1,642 mm, and the range of annual total rainfall varies between 39 and 4,804 mm. The topography of the study area is characterized by flatlands and high altitude mountainous areas. Although the subbasin mean elevation varies between 97 and 4,609 m, the range of elevations within a subbasin can be as high as 6,021 m. The high range of elevations suggests a high topographic contrast and spatial rainfall variability at the subbasin level. The mean rainfall range at the subbasin level is 544 mm. The subbasin-specific parameters of the variables shown in table 5 have been used in cluster analysis using the K-means algorithm to identify distinct regions of subbasins of similar characteristics. Analysis of variance has been applied to ascertain the statistical significance of the identified regions.

K-Means Clustering

We used the K-means clustering for homogenization, which is an unsupervised hard clustering method. The K-means algorithm (Forgy, 1965) uses the centroids of clusters to characterize the data. They are determined by minimizing the sum of squared errors between the objects in the clusters and their respective cluster centroids (c) (Ding and He, 2004). The K-means algorithm is defined in equation 7.

$$J_K = \sum_{k=1}^{k} \sum_{i \epsilon C_k} (X_i - m_k)^2 \qquad (7)$$

where
$(X_1, \ldots, Xn) = X$ is the data matrix,
$m_K = \sum_{i \epsilon C_k} X_i / n_k$ is the centroid of cluster C_k, and
n_k is the number of points in C_k.

The centers of the clusters are initialized by randomly selecting cluster seeds from the dataset. Then the dataset is clustered in the process of assigning each point to the nearest cluster seed to form a set of temporary clusters. When the dataset has been assigned, the average position of the data points within each cluster is calculated, and the cluster seeds are then replaced by the average position of the cluster. The points are reassigned, and this process of assigning and averaging is repeated until no further changes occur in the cluster. The process is then said to have converged (Lin and Chen, 2006). The K-means algorithm is sensitive to the selection of the number of clusters to determine (Goyal and Gupta, 2014). To

obtain well separated and compact clusters, it is important to choose the optimal value of the number of clusters. Although several measures are available (Pham and others, 2005), we determined the optimal value of the number of clusters from the sum of squared error (SSE) diagram. The SSE is defined as the sum of squared distance between each member of a cluster and its cluster centroid; thus, SSE can be considered as a global measure of error and computed as:

$$SSE = \sum_{i=1}^{N}(m_i - m_{ik})^2 \qquad (8)$$

where

$(m_i - m_{ik})$ is the difference between i member of cluster and cluster mean that belongs to i, and

N is the number of clusters.

The K-means algorithm (Hartigan and Wong, 1979) has been implemented in the R statistical package.

Homogenization Results

Selection of Optimal Number of Clusters

To determine the optimal cluster size for the K-mean, the SSE was computed for as many as 15 clusters. The SSE against cluster solutions is shown in figure 20. As the number of clusters increases, the SSE decreases because the number of clusters is getting smaller. An appropriate cluster solution could be defined as the solution at which the reduction in SSE slows dramatically. According to figure 20, the reduction in SSE slows down after five cluster solutions, which suggests that cluster solutions greater than five do not have a substantial effect on the total SSE; therefore, five is chosen as the optimal number of clusters for the K-means algorithm.

Homogeneous Regions

The K-means algorithm yields five distinctly different homogeneous regions utilizing the subbasin-specific mean annual total rainfall range of the annual total rainfall, mean
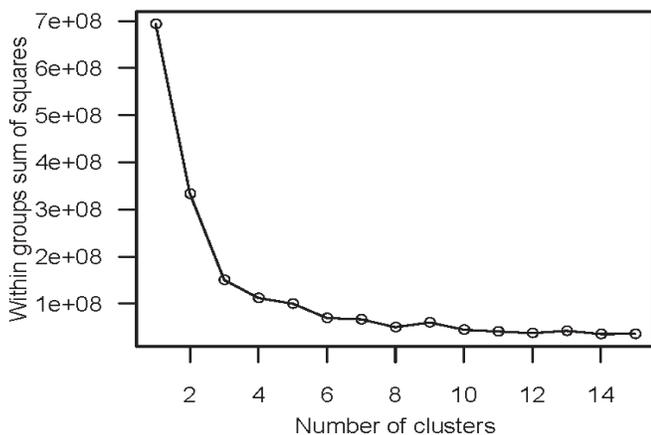
Figure 20.   Within groups sum of squared error (SSE) diagram for 15-cluster solutions.

elevation, and elevation range. The subbasins are considered potential candidates for membership in the same homogeneous region if they are similar in topographic and climatic characteristics. The results of the K-means analysis are presented in figure 21, which illustrates five distinctive regions, each showing hydrologic homogeneous regions across the study area. The homogeneous regions extend mostly from northwest to southeast. Regions 1, 4, and 5 are relatively narrow, and Regions 2 and 3 form a relatively wide stripe split into two separate areas. To analyze the topographic and climatic patterns in these homogeneous regions, the components of the K-means are plotted and the group means (by region) of the parameters computed. The group mean values are presented in table 6, and the components of the five clusters are shown in figure 22. The mean values in table 6 suggest that the homogeneous regions are distinctively located in different identifiable elevation and rainfall zones. The homogeneous regions are divided into mean elevation bands between 223 and 3,887 m and mean annual total rainfall between 756 and 3,272 mm with no specific orders of separation for the in-between regions. The mean annual rainfall for Regions 3 and 5 is nearly the same (1,030 mm and 1,071 mm, respectively); however, similar rainfall occurs at two distinct elevation zones (285 m and 2,082 m, respectively), which possibly makes it possible to identify them separately. Region 3 is the largest region by area followed by Regions 2, 1, 4, and 5. Region 4, located at the moderate elevation level (2,976 m), receives the highest amount of rainfall (3,272 mm), whereas Region 1, located at a higher elevation (3,887 m), receives the lowest amount of rainfall (756 mm) (table 6).

The scatterplots in figure 22 illustrate the climate and topographic characteristics of the subbasins that belong to the same homogeneous region. Homogeneous regions are color-coded separately on the scatterplot. Ideally, if the clusters are grouped together well at the selected level, there should not be substantial overlap in the distributions of the samples on the plots. The plot suggests that the clusters are grouped together well and that the groups are distinctly different from each other. The principal two components explain 86.3 percent of the sample variability. There is a general agreement in the rainfall distribution with respect to the topographic features, particularly the elevation. The characteristics of the regions with respect to the topographic features can be described as follows. Finally, the homogeneous regions' shapefile was imported into the geodatabase.

- *Region 1.*—This dry mountainous region is in the southwestern part of the study area with a mean elevation of over 3,000 m (fig. 22). Rainfall in this region is consistently the lowest among all the regions, averaging 756 mm annually (table 6).

- *Region 2.*—This region covers small hills and plains in the north and has the lowest mean elevation and elevation variability (fig. 22). Low flat areas receive about 1,725 mm of rainfall annually (table 6).
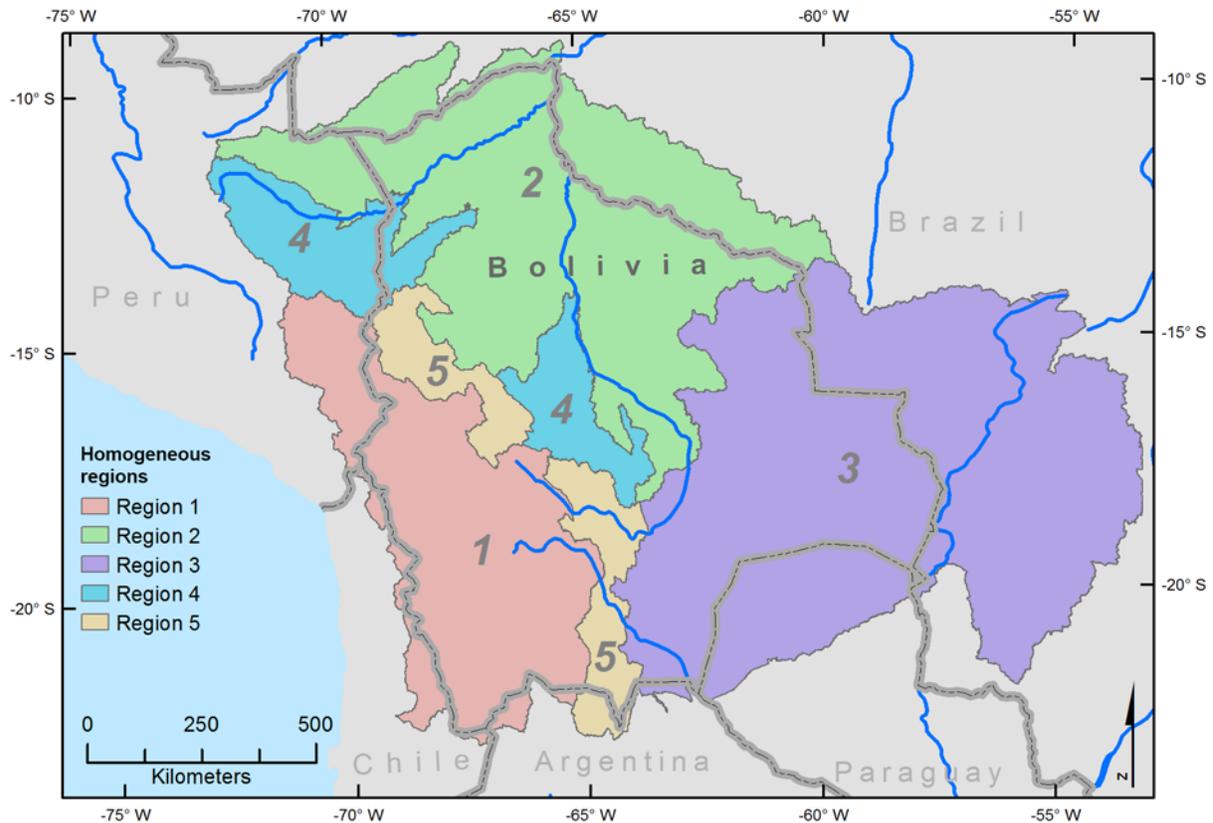
Figure 21.   Map of the homogeneous regions derived using *K*-means clustering from climate and topographic variables.

- *Region 3.*—This region, the largest by area, is also relatively flat and dry, although mean elevation in this region is slightly higher than Region 2 and rainfall is much lower than Region 2 (fig. 22).

- *Region 4.*—This region is at the foothills and receives the highest amount of rainfall annually (3,272 mm) (table 6). The region coincides with two pockets of high rainfall areas. The foothills elevation of the region

is moderate (930 m) with relatively moderate elevation variability (fig. 22).

- *Region 5.*—This region transitions between high mountains and plains and has the highest elevation variability: 4,329 m (table 6). The region has the second highest mean elevation but receives slightly higher rainfall than Region 1 (fig. 22).
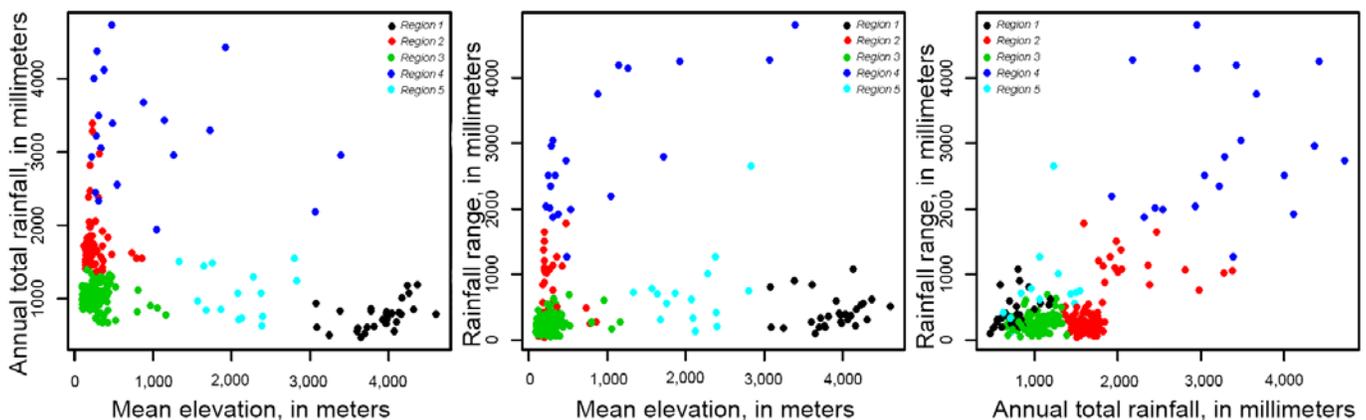
Figure 22.   Climate and topographic characteristics of the subbasins belonging to the same homogeneous region, coded by separate colors.

Evaluation of Homogeneous Regions

Climate and Topographic Characterization by Region

To evaluate the climatic and topographic characteristics of the regions, we plotted the component distribution by region in figure 23. The boxplot shows that, in general, the mean rainfall and elevation are quite different between the regions; however, the boxplot also shows that each region presents a different amount of variability for the components individually. Even though the mean annual total rainfall for Regions 2 and 3 is different, for example, the variations in the mean annual total rainfall are quite similar; furthermore, the distribution of mean elevation between Regions 2 and Region 3 is similar.

One-Way Analysis of Variance Test

In order to check the statistical significance of the identified homogeneous regions, a one-way analysis of variance (ANOVA) was applied between the homogeneous regions with the null hypothesis that the component means for the homogeneous regions are all equal. The F values and corresponding significance (p-value) of the test are provided in table 7. If the p-value for F is less than or equal to the significance level, the null hypothesis is rejected. The null hypothesis can be rejected (table 7), which means, in general, the component means between the homogeneous regions are statistically and significantly different because the p-values are all less than 0.05; however, table 7 does not indicate between which homogeneous regions the means are significantly different. To determine which pairs of homogeneous regions' component means are different, the Tukey's honest significant difference (TukeyHSD) post hoc test was used. The summary statistics for the TukeyHSD test are provided in appendix 3. Based on the mean difference annual rainfall and elevation between the regions and their corresponding p-value, it can be concluded

Table 7. F values and significance (p-values) from the analysis of variance test by component.

[p-value, probability]

Components	F value	p-value
Total annual rainfall	248.8	<0.0001
Rainfall variability	191.6	<0.0001
Mean elevation	845.1	<0.0001

that, except between Regions 3 and 5 for the mean annual rainfall and between Regions 2 and 3 for the mean elevation, all the other regions have significantly different rainfall and elevation.

The TukeyHSD significance test results are also shown graphically in figure 24. As illustrated in figure 24, weighting on the importance of the input components can be determined when regrouping the homogeneous regions or implementing the same regression model on multiple homogeneous regions if necessary.

Comparison with Köppen-Geiger Global Climate Map

Three main climate types—arid, temperate, and tropical—dominate the study area and are oriented diagonally from southwest to northeast similar to the pattern in the rainfall gradient of the homogenized regions (fig. 25; table 6). Homogeneous Region 1 is mostly characterized by an arid cold climate with the lowest expected rainfall. Correspondingly and as expected, the mean annual total rainfall for Region 1 is the lowest among all the regions. Region 5 is characterized by temperate dry winter and warm summer climate types that produce low annual rainfall. Region 4 has the highest mean annual total precipitation in the study area as it coincides with the tropical monsoon climate zone. Region 3 is characterized by a tropical savannah climate zone, which coincides with
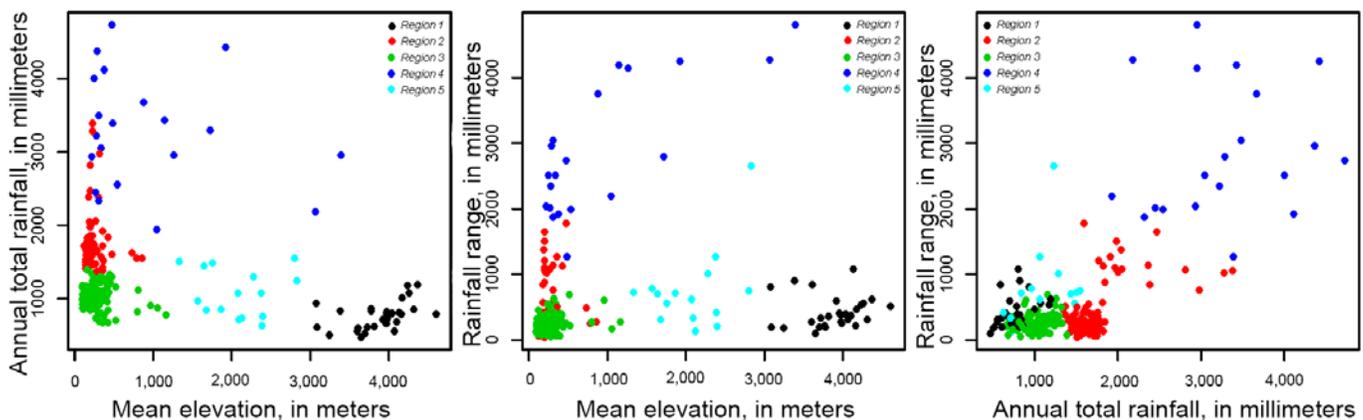
Figure 23. Climatic and topographic characteristics by homogeneous region.
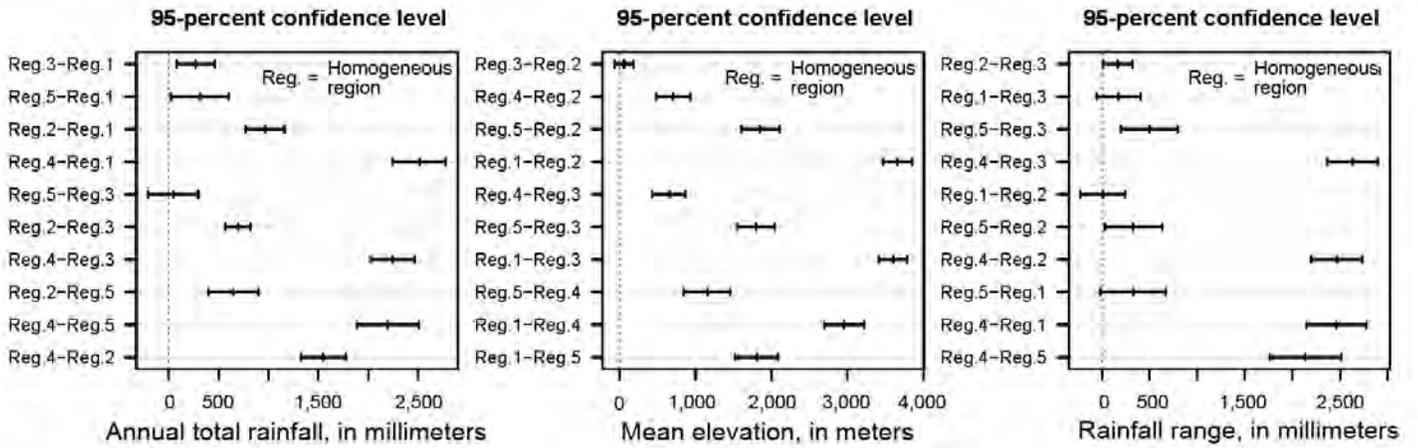
Figure 24. Representation of the Tukey's honest significant difference (TukeyHSD) test results by component.

low rainfall that is less than that found in Region 2, which is characterized by a tropical monsoon climate type (fig. 25; table 8). Overall, the identified homogeneous regions agree relatively well with the climate types of the region given the differences in resolution and deriving methodology between the homogeneous regions and this climate map. Thus, these results bolster confidence on the delineation methodology of the homogeneous regions.

Estimation of Mean Annual Streamflow

Another key variable in estimating theoretical potential hydropower is the mean annual streamflow. As the theoretical potential hydropower will be estimated at each 1-km segment of the streams, it is essential to compute mean annual streamflow at the outlet of each 1-km segment of the streams.
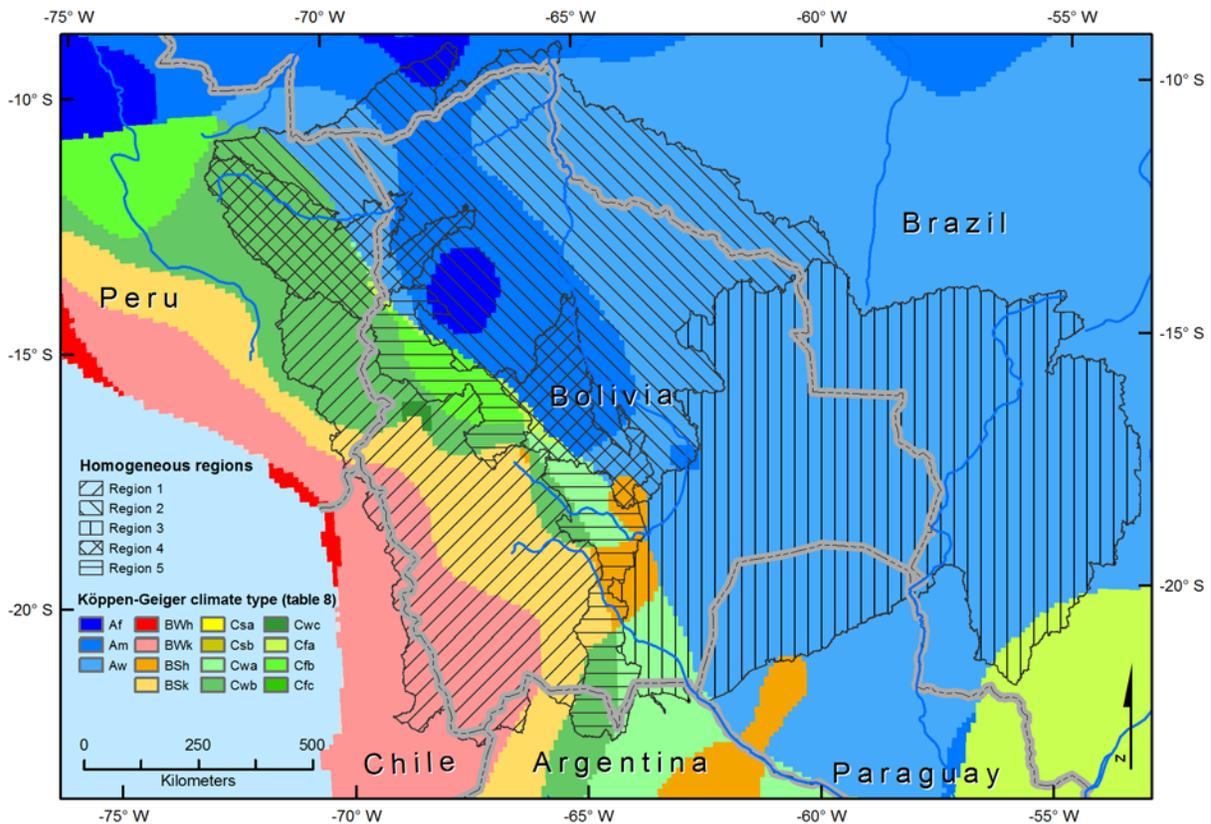
Figure 25. Köppen-Geiger climate types map for the study area. The description of the climate symbols and their defining criteria are provided in table 8.

Table 8.   Description of Köppen climate symbols and defining criteria.

[T_{cold}, temperature of the coldest month; P_{dry}, precipitation of the driest month; MAP, mean annual precipitation; $P_{threshold}$, varies according to the following rules (if 70 percent of MAP occurs in winter then $P_{threshold}$, 2 x MAT, if 70 percent of MAP occurs in summer then $P_{threshold}$ = 2 x MAT + 28, otherwise $P_{threshold}$ = 2 x MAT + 14) MAT, mean annual temperature; T_{hot}, temperature of the hottest month; P_{sdry}, precipitation of the driest month in summer; P_{wdry}, precipitation of the driest month in winter; T_{mon}10, number of months where the temperature is above 10; Summer (winter) is defined as the warmer (cooler) six-month period of ONDJFM and AMJJAS. All temperatures are in egrees Celsius]

Type	Description	Criterion
Af	Tropical rainforest	$T_{cold} \geq 18$ and $P_{dry} \geq 60$
Am	Tropical monsoon	$T_{cold} \geq 18$ and $P_{dry} \geq 100\text{-MAP}/25$
Aw	Tropical savannah	$T_{cold} \geq 18$ and $P_{dry} < 100\text{-MAP}/25$
BWh	Arid desert hot	$\text{MAP} < 5 \times P_{threshold}$ and $\text{MAT} \geq 18$
BWk	Arid desert cold	$\text{MAP} < 5 \times P_{threshold}$ and $\text{MAT} < 18$
BSh	Arid steppe hot	$\text{MAP} \geq 5 \times P_{threshold}$ and $\text{MAT} \geq 18$
BSk	Arid steppe cold	$\text{MAP} \geq 5 \times P_{threshold}$ and $\text{MAT} < 18$
Csa	Temperate dry hot summer	$T_{hot} > 10$ and $0 < T_{cold} < 18$ and $P_{sdry} < 40$ and $P_{sdry} < P_{wwet}/3$ and $T_{hot} \geq 22$
Csb	Temperate dry warm summer	$T_{hot} > 10$ and $0 < T_{cold} < 18$ and $P_{sdry} < 40$ and $P_{sdry} < P_{wwet}/3$ and $T_{mon}10 \geq 4$
Cwa	Temperate dry winter hot summer	$T_{hot} > 10$ and $0 < T_{cold} < 18$ and $P_{sdry} < 40$ and $P_{wdry} < P_{wwet}/3$ and $T_{hot} \geq 22$
Cwb	Temperate dry winter warm summer	$T_{hot} > 10$ and $0 < T_{cold} < 18$ and $P_{sdry} < 40$ and $P_{wdry} < P_{wwet}/3$ and $T_{mon}10 \geq 4$
Cwc	Temperate dry winter cold summer	$T_{hot} > 10$ and $0 < T_{cold} < 18$ and $P_{sdry} < 40$ and $P_{wdry} < P_{wwet}/3$ and $1 \leq T_{mon}10 \leq 4$
Cfa	Temperate without dry season hot summer	$T_{hot} > 10$ and $0 < T_{cold} < 18$ and $P_{sdry} < 40$ and $T_{hot} \geq 22$
Cfb	Temperate without dry season warm summer	$T_{hot} > 10$ and $0 < T_{cold} < 18$ and $P_{sdry} < 40$ and $T_{mon}10 \geq 4$
Cfc	Temperate without dry season cold summer	$T_{hot} > 10$ and $0 < T_{cold} < 18$ and $P_{sdry} < 40$ and $1 \leq T_{mon}10 \leq 4$

This section describes in detail the two-step process used to generate the mean annual streamflow along with the analysis of the exceedance probability of the generated streamflow by the homogenized regions in the study area.

Contributing Upstream Area Rainfall

To compute mean annual streamflow, the mean annual rainfall of the contributing upstream area at the outlet point of each 1-km segment of the streams was first computed. The contributing upstream area rainfall was extracted from the bias-corrected TRMM mean annual rainfall grid. A continuous parameterization technique (Harvey and Eash, 1996) has been implemented to extract contributing upstream mean rainfall. The continuous parameterization technique uses the same rationale of the flow direction and flow accumulation calculation algorithms, but instead of treating every pixel as one unit, each pixel is weighted by the rainfall value calculated from the bias-corrected TRMM rainfall grid. Before initiating the continuous parameterization, the mean annual TRMM gridded rainfall was resampled to 30-m resolution and aligned with the stream grid. The continuous parameterization technique was implemented using geographic information systems (GIS). Pixel-to-pixel flow direction and flow accumulation calculated from the 30-m DEM provided the framework for the calculation from the upstream weighted mean rainfall. The continuous parameterization process yielded contributing upstream area weighted rainfall for each pixel in the study area. An example from the upstream contributing area weighted rainfall in Bolivia is shown in figure 26.

The annual average rainfall values along the stream cells are much lower than the annual average rainfall values from the surrounding areas of the streams (fig. 26) because the contributing upstream area for the stream cells is much larger than the contributing upstream area for its surrounding cells. This change in rainfall values can be seen in the example shown in figure 26, where the Pilaya River flows from west to east, increasing the number of tributaries that contribute to it; therefore, typically contributing upstream weighted rainfall downstream gradually reduces as the upstream drainage area increases.

The computed upstream area weighted bias-corrected TRMM rainfall for each 1-km segment of the streams is added into the geodatabase.

Development of Regional Regression Equations

Stream Discharge Data

A rich set of daily streamflow records was made available for this study. The daily streamflow discharge records at 146 stations across Bolivia were provided by SENAMHI, and daily discharge records at another 16 stations in Brazil along the Brazil-Bolivia border were provided by ANA. The discharge records span from 1944 to current (2015) with missing days and years of records. The stations are mostly in central
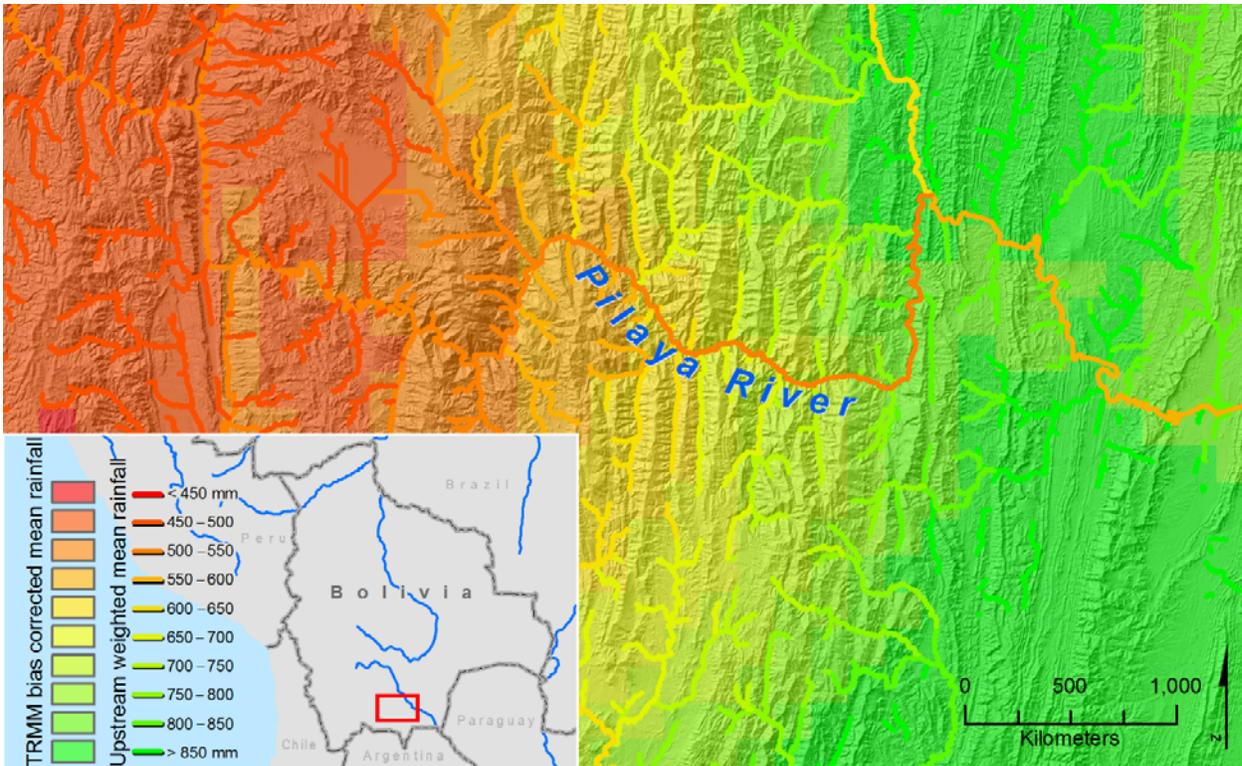
Figure 26.   Contributing upstream area weighted annual average rainfall in a portion of Bolivia. [TRMM, Tropical Rainfall Measuring Mission; mm, millimeter].

Bolivia covering homogeneous Regions 1, 4, and 5. There are several discharge stations around the northern tip of Bolivia in homogeneous Region 2; however, discharge stations in the eastern part of the country corresponding to homogeneous Region 3 were sparse. The spatial distribution of discharge stations in the study area is shown in figure 27, and figure 28 shows the data availability for those stations. Because the prime purpose of using the discharge data collected in these stations was to derive statistical relations between river discharge, corresponding drainage area, and areal average rainfall over the drainage area, it was important that the discharge observations and rainfall estimates be temporally synchronized; however, as the rainfall model developed for this study is based on climatology from TRMM–D enhanced with variability from CHIRPS, we set two criteria to select the discharge station's data to be used in the analysis: (a) the station has daily records between 1970 and 2013, and (b) the station has at least 5 complete years of data. The selection criteria yielded 108 stations across Bolivia and in parts of Brazil. The number of discharge stations by homogeneous region is shown in table 9. Homogeneous Regions 1, 2, and 5 were well represented by the discharge stations; however, homogeneous Regions 3 and 4 had only six and seven discharge stations, respectively. The observed daily discharge records of these stations were quality checked for ambiguous records and for data inconsistencies. Many of these discharge stations, especially in Region 1, were not on the main river channel

and, therefore, represent small upstream areas. Because the discharge records will be used to develop a regression model representative of the entire region, we excluded another 23 stations from Region 1 based on stations that have only 5 years of records and stations that have a small upstream area or inconsistent data. That left 21 stations for homogeneous Region 1 to be used in the regression model.

Streamflow Modeling

We developed regression equations based on statistical relations between the climatic and physical variables to estimate mean annual streamflow (Vogel and others, 1999;

Table 9.   Observed discharge stations with at least 5 years of records between 1970 and 2013 by homogeneous region.

Homogeneous regions	Discharge stations with ≥ 5 years of data
Region 1	44
Region 2	30
Region 3	6
Region 4	7
Region 5	21
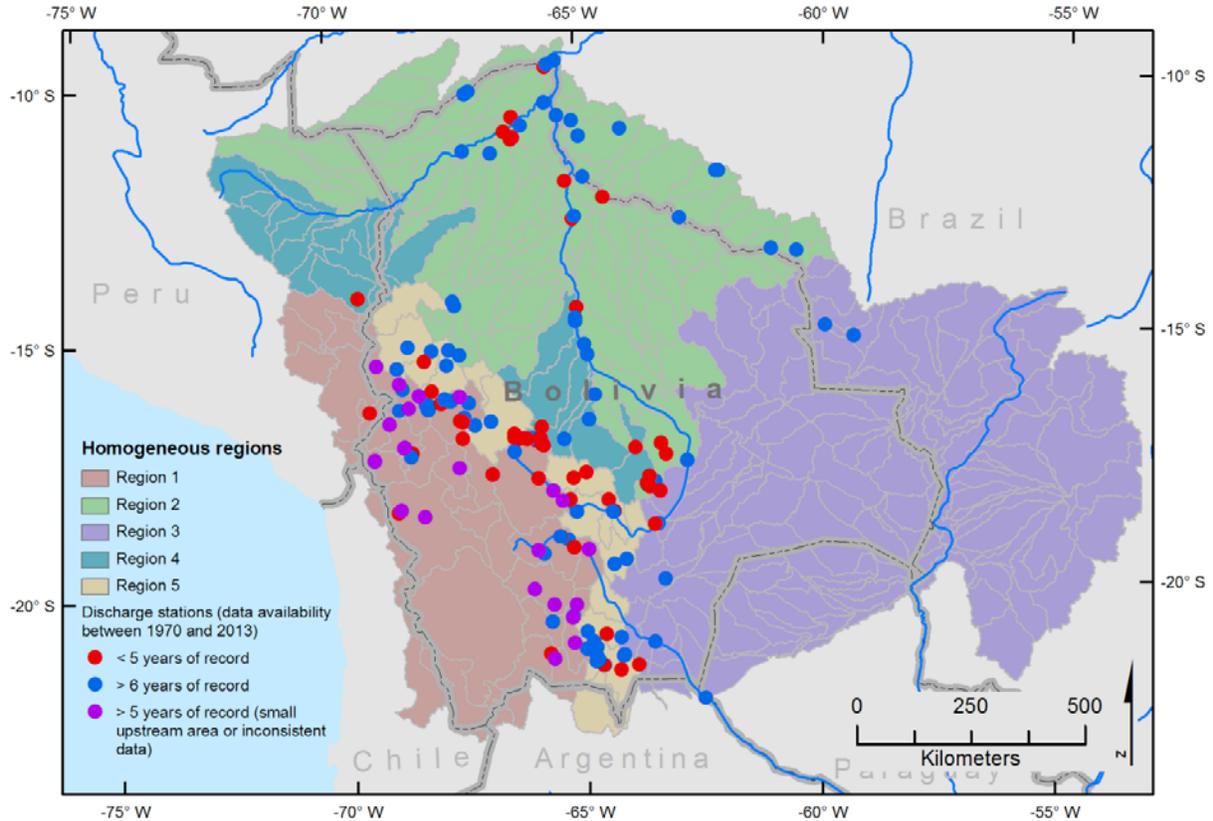
Figure 27.   Observed discharge stations across the study area by homogeneous region.

Parks and Madison 1985). The regression equations are of the general form (Vogel and others, 1999):

$$Q = e^{c0}X_1^{c1}X_2^{c2}\ldots X_n^{cn}e^{\varepsilon} \qquad (9)$$

where

Q is the observed annual streamflow in a gaged basin in m3s-1,

X_i are the various physical and climatic parameters (for example, upstream basin area, rainfall),
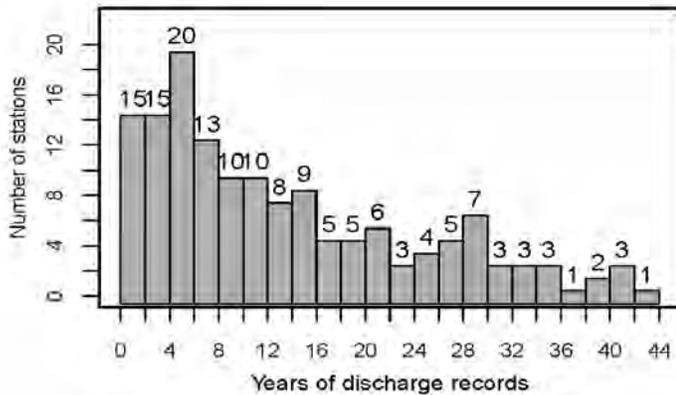
Figure 28.   Observed daily discharge data availability between 1970 and 2013 by station in the study area.

ci are the ordinary least square regression coefficients, and

ε is the residual of the model.

As a lognormal distribution model is suggested for the best fit to the annual streamflow and basin climatic characteristics (Vogel and others, 1999), we developed lognormal distributed regression equations using only upstream drainage area and mean annual rainfall as independent variables. Unique regression equations were developed for each of the five homogeneous regions. The best fit regionalized regression equation to estimate annual streamflow took the form

$$\log(Q) = \log(a) + b * \log(DA) + c * \log(P) \text{ or } Q = 10a * DA^b * P^C \qquad (10)$$

where

DA is the total upstream drainage area, in square kilometers;

P is the mean annual rainfall for the upstream drainage area, in millimeters;

a,b,c are the coefficients of the regression.

The derived regression coefficients for the homogeneous regions along with correlation coefficients are presented in table 10.

The final number of observed discharge stations used in fitting the model and general characteristics of the fitted model by comparing simulated and observed discharge

Table 10. Regression model coefficients by homogeneous regions.

[m³/s, cubic meter per second; *n*, number of stations; R^2, coefficient of determination; *p*-value, probability; RMSE, root-mean-square error; MAE, mean absolute error; *Q*, observed annual streamflow; *DA*, total upstream drainage area; *P*, mean annual rainfall]

Homogeneous region	Mean annual flow (m³/s)	Number of stations (*n*)	Adjusted R^2	*p*-value	RMSE percent	MAE percent
Region 1	$Q = (10^{-14.7714}) * (DA^{1.1537}) * (P^{4.2702})$	21	0.84	<0.001	33.5	27.1
Region 2	$Q = (10^{-6.67766}) * (DA^{1.0504}) * (P^{1.46482})$	30	0.95	<0.001	9.6	7.4
Region 3	$Q = (10^{-0.24433}) * (DA^{0.44992}) * (P^{0.14725})$	6	0.94	<0.001	2.1	1.7
Region 4	$Q = (10^{-6.8282}) * (DA^{0.8085}) * (P^{1.9021})$	7	0.96	<0.001	10.2	7.7
Region 5	$Q = (10^{-9.32122}) * (DA^{0.75632}) * (P^{2.86822})$	21	0.94	<0.001	13.2	10.5

values are identified in table 10. Each of the fitted models was statistically significant, and the agreement (adjusted) between observed and simulated discharge was excellent— RMSE ranged between 2.1 and 33.5 percent and MAE ranged between 1.7 and 27.1 percent of the mean annual streamflow. The observed and model-fitted simulated values of the mean annual streamflow by homogeneous region are plotted in figure 29.

The characteristics of the regression model fit were also analyzed with the help of residual plots shown in figure 30. The residual plots inform how well or how poorly a model simulation fits the observed data. If the model is appropriate for the data, the plot should show an even scatter, and any visible pattern in the plot means that the regression equation does not describe the data correctly. A scatter pattern forms when the residuals are unevenly distributed about the regression line. We used four types of residual plots to assess the model fit. The diagnostic plots we used are the simple residual versus fitted values, the scale-location plot, the normal quantile-quantile (Q–Q) plot, and residual versus leverage or Cook's distance. The simple residual is a plot of residuals with the calculated discharge values. The scale-location plot is similar to the residuals versus fitted values plot, but it uses the square root of the standardized residuals. When there is discernible pattern in this plot, the proposed fitted model fails because of the lack of homoscedasticity (the probability distribution for the response variable does not have the same standard deviation regardless of the predictor value). The last two plots are for standardized residuals. The residuals are in units of the dependent variable and there are no cutoff points for defining what is a "large" residual; using the standardized residuals makes it easy to compare the fits of different models in various regions. The normal quantile-quantile plot of the standardized residuals indicates whether the errors of the model follow a normal distribution; if it follows the normal quantile-quantile plot, the residuals will follow a straight line. The Cook's distance plot is useful to identify points that have more influence than other points and points that are distant from other points in the data, either for the dependent variable or one or more independent variables.

The final results of the regression analysis by homogeneous region are summarized in figure 30. The diagnostic plots

suggest that all the selected regression models fulfilled the requirements of the regression model validity and every model had good predictive skill of the mean annual streamflow. Because of the limitation of long-term streamgage-measured discharges that were used to develop the regionalization models, there is room to improve the models if more streamflow data become available. We used two inputs to estimate mean annual streamflow at the outlet of each 1-km segment of the stream using the regionalization method. The inputs were the upstream drainage area and the mean annual rainfall from the upstream contributing area at the stream outlet; therefore, the regional interbasin flow transfers and consumptive water use that took place were not considered mostly because of lack of recorded information after the streamflow data were recorded but not considered in our analysis.

The mean annual discharge for each 1-km segment of the stream was estimated using the homogeneous region-specific regression model shown in table 10 and added into the geodatabase. The estimated discharge represents the expected mean annual discharge for that specific 1-km segment of the stream depending on the rainfall upstream from that segment. The expected mean annual discharge typically increases for the downstream segments as the upstream basin area increases. For the streams that cross between two or multiple homogeneous regions, however, the estimated mean annual discharge over the specific homogeneous region is also influenced by the hydrometeorological and topographical characteristics of the region; therefore, the mean annual discharge of the tributaries expanding over multiple homogeneous regions may not equal the mean annual discharge at the mouth of that particular stream. This can be considered a limitation for the regionalized regression method applied in this assessment. We expect this will have minimal impact on the geodatabase assessing theoretical hydropower as the hydropower potential will be computed for each 1-km segment of the stream.

Streamflow Validation

We used a cross-validation technique to validate the estimated streamflow for each individual homogeneous region separately. Cross-validation provides measures of predictive
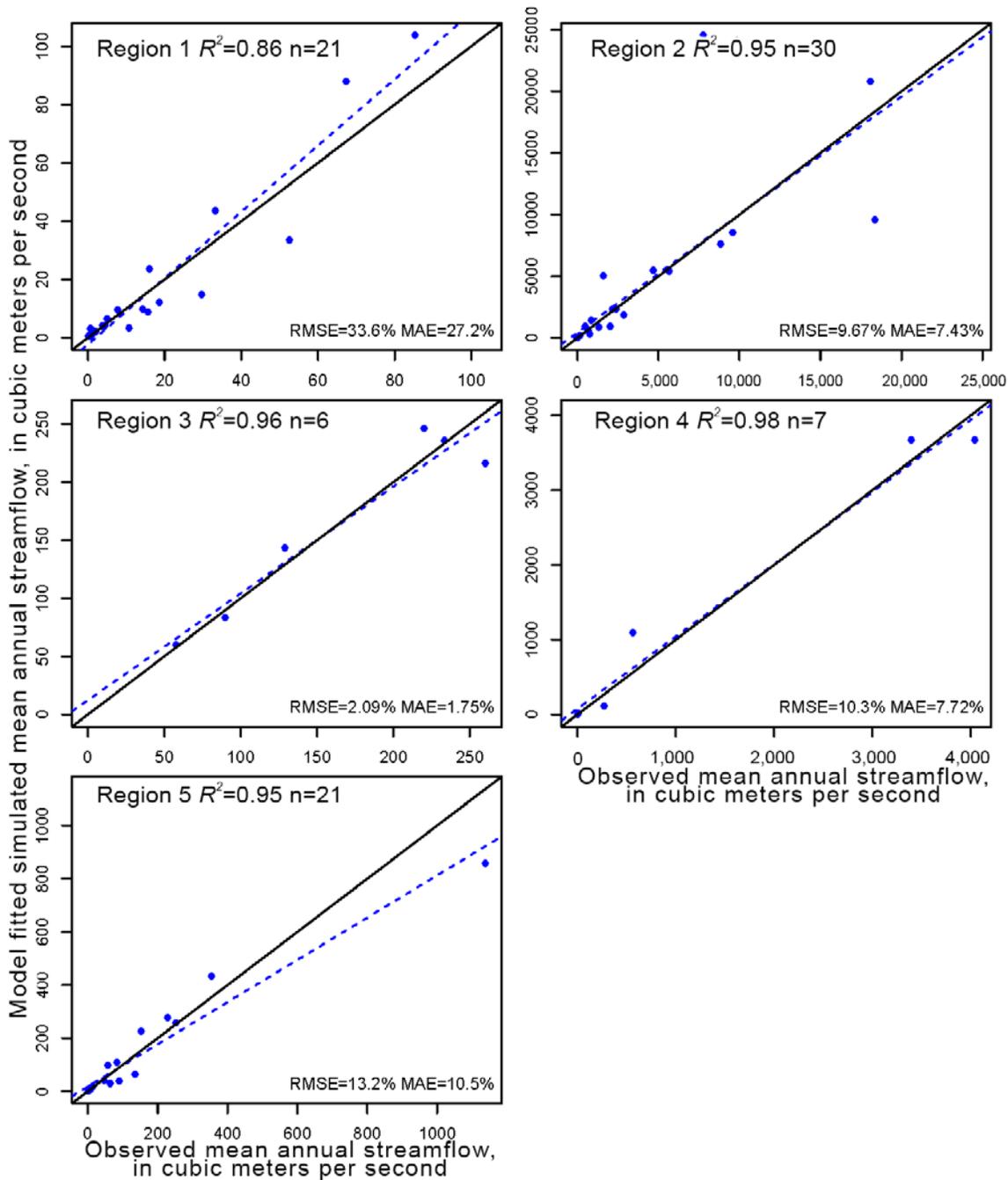
Figure 29.   Observed and model-fitted simulated mean annual streamflow by homogeneous region [R^2, coefficient of determination; n, number; RMSE, root-mean-square error; %, percent; MAE, mean absolute error].

accuracy for multiple linear regressions. This technique focuses on not using the entire dataset when building a model. The data are randomly assigned to a number of "folds." Each fold is removed, in turn, while the remaining data are used to refit/build the regression model and to predict the deleted observations. The data points, which were removed, are then used to test the performance of the model. In this study, because of the small number of observed discharge stations by homogeneous regions, we used 3-fold application for cross-validation. The number of stations for each fold by homogeneous region is shown in table 11.

In the end, the predicted values for each fold by homogeneous region were grouped together and compared with the corresponding observed values. Successively, the RMSE and MAE were computed in evaluating the validation of the mean annual streamflow predicted by the regression model. The RMSE, MAE and correlation coefficients are also listed in table 11 by homogeneous region. The correlation between
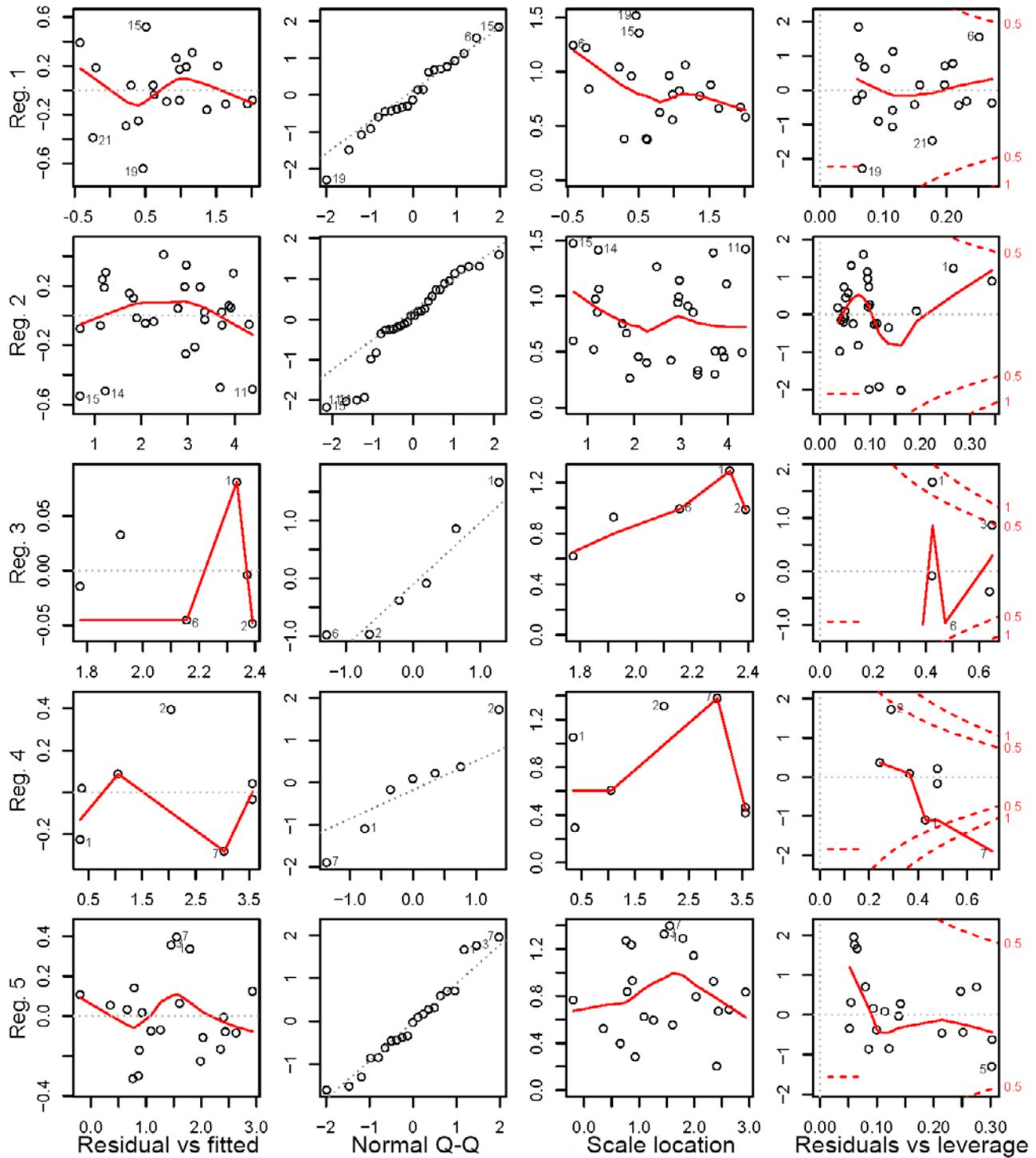
Figure 30. Region-specific regression model diagnostic plots [Reg., homogeneous region].

observed and predicted mean annual streamflow values estimated during cross-validation is shown in figure 31.

The streamflow validation results presented in figure 31 suggest that the regression models for homogeneous Regions 2, 3, and 4 are in part simulating discharge from upstream drainage area and average rainfall in the drainage area; however, the streamflow in Region 1 was slightly overestimated, and in Region 5 streamflow was slightly underestimated. The effect of these overestimations and underestimations is expected to be minimal in computation of theoretical hydropower potential because the average observed discharge at stations in these two regions is among the lowest of all the homogeneous regions because it constitutes dry mountainous regions.

Table 11.   Cross-validation results of the simulated streamflow.

[R^2, coefficient of determination; p-value, probability; RMSE, root-mean-square error; MAE, mean absolute error]

Homogeneous region	Number of stations by fold	Cross-validation R^2	p-value	RMSE (percent)	MAE (percent)
Region 1	7, 7, 7	0.82	<0.001	37.8	30.5
Region 2	10,10,10	0.94	<0.001	10.8	8.3
Region 3	2, 2, 2	0.94	<0.001	10.8	8.3
Region 4	2, 2, 3	0.88	<0.001	24.6	19.7
Region 5	7, 7, 7	0.94	<0.001	14.0	11.4

Streamflow Characterization

In addition to estimating mean annual streamflow at each 1-km segment of the streams, we also provide analysis of the long-term pattern in the mean annual streamflow. The long-term pattern of the mean annual streamflow can be useful when defining available discharge for hydropower, determining the proper size and type of turbine, and establishing minimum flows in the stream for environmental reasons. Statistical characterization of the mean annual streamflow provides the basis for long-term pattern analysis, and flow-duration data are commonly used to statistically characterize streamflow. Flow-duration data are mean daily flow values measured over a specified period that have exceeded various percentages of the specified time interval. For example, a 5-percent exceedance probability represents a high flow that has been exceeded only 5 percent of all years of the flow record; conversely, a 95-percent exceedance probability would characterize low-flow conditions in a stream because 95 percent of all mean annual flows in the record are greater than that amount.

We summarized the daily mean streamflow into mean annual streamflow and then computed the exceedance probability using the following equation, which is also referred to as the flow-duration percentile. Flow-duration percentile was computed for mean annual streamflow because the final hydropower potential will be estimated based on the annual streamflow at each 1-km segment of the river:

$$P = 100 * (m/[n+1]) \qquad (11)$$

where

> P is the exceedance probability,
> m is the ranking from highest to lowest of all annual mean flows for the specified period (between 1970 and 2013),
> n is the number of mean annual flows.

Flow-duration statistics are points along a flow-duration curve and reflect only the period for which they are calculated; however, the statistics can be an indicator of probable future conditions if the period-of-record used to compute the statistics is sufficiently long—typically at least 10 years (Searcy,

1959). We generated the flow-duration exceedance curve for each of the stations by homogeneous region. To generate a flow-duration exceedance curve representative of the corresponding homogeneous region, we averaged the flow-duration exceedance curves estimated from the station-observed daily data within that region. To determine the average flow for a specific flow-duration percentile for a specific homogeneous region, the discharges associated with percentiles on either side of the specific percentile was interpolated between each station before averaging them to compute the regional flow-duration exceedance curve. The flow-duration exceedance curve for each of the stations by homogeneous region is presented in appendix 4. The last plot in each graph represents the mean of all station-specific plots and, therefore, considered as the flow exceedance curve for the corresponding homogeneous region.

The duration statistics used in this study were the 40, 60, 70, and 90 percent exceedance quantiles on the flow-duration curve estimated with a log-Pearson Type III distribution. The log-Pearson type III distribution has been the most used frequency distribution for hydrologic analyses since the recommendation of the Water Resources Council (1967) and the Interagency Advisory Committee on Water Data Hydrology Subcommittee (1982). In addition, we provide homogeneous region-specific fraction of mean flow exceedance probability statistics at 5 percent increments (table 12) computed from the regional flow exceedance curve. These statistics can be used to compute duration statistics at any 5-percent increment exceedance quantiles between 5 and 100. Flow exceedance curves for individual stations and average exceedance curves for the regions are provided in appendix 4.

The statistics were determined using annual mean flow at each station for the period-of-record from 1970 to 2013. For hydropower assessment, what is of most interest is the shape of the flow-duration curve in the low-flow regions. The shapes of the flow-duration curves in the high-flow regions indicate the type of flow regime likely to be seen in the stream where the data come from during years of high flow. By using annual average flows instead of daily values, the resulting flow-duration curve is flatter because of the averaging process. A curve with a steep slope indicates a highly variable stream, whereas a curve with a flat slope denotes a stream that does not vary
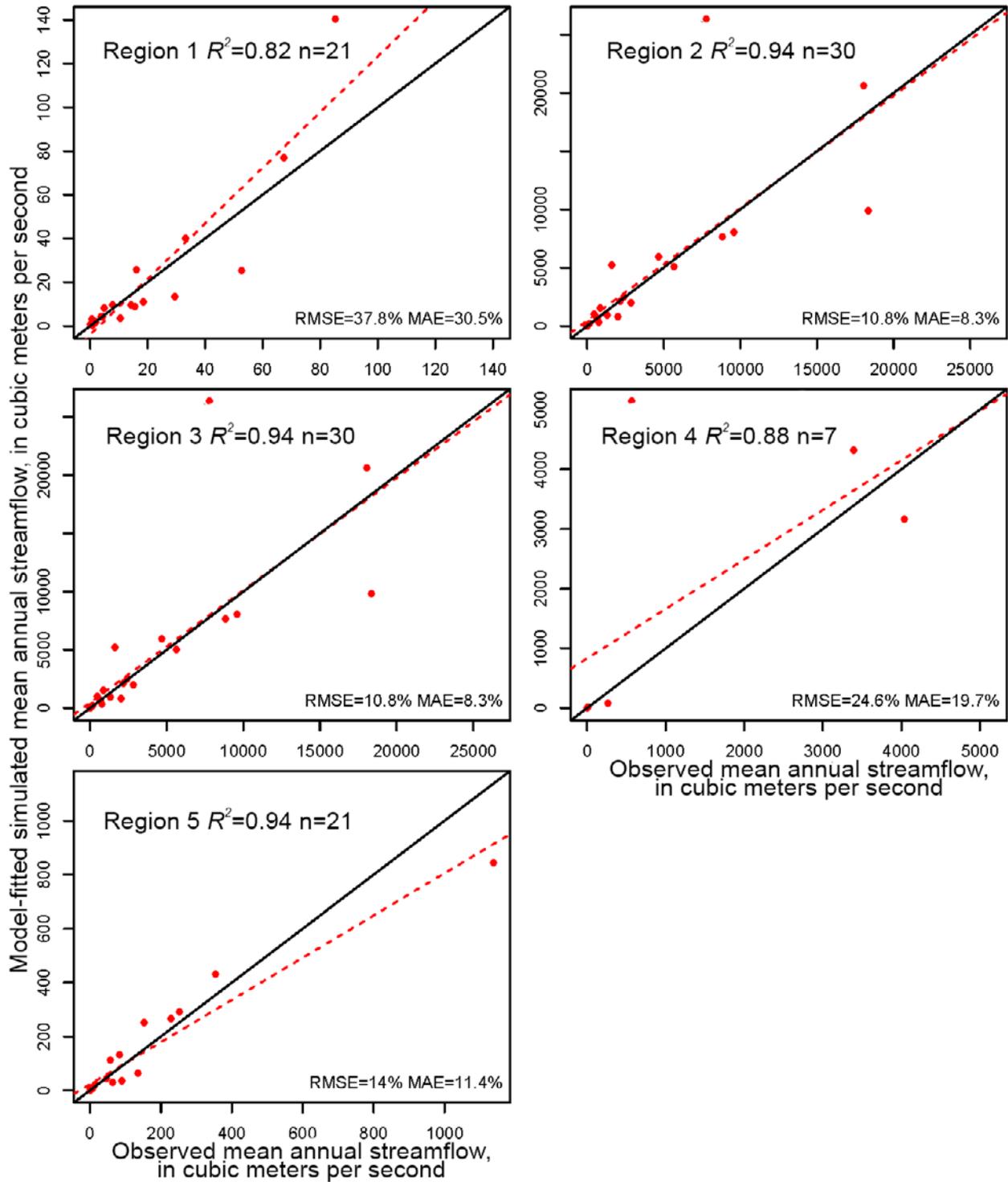
Figure 31. Validation of streamflow estimated from a regionalized regression model by homogeneous region [R2, coefficient of determination; n, number; RMSE, root-mean-square error; %, percent; MAE, mean absolute error].

Table 12.   Fraction of annual mean flow at flow exceedance levels between 5 and 100 percent for the homogeneous regions.

Percent exceedance	Fraction of annual mean flow				
	Region 1	Region 2	Region 3	Region 4	Region 5
5	3.11	1.96	2.08	2.06	3.07
10	2.40	1.60	1.75	1.59	2.33
15	1.65	1.38	1.50	1.32	1.75
20	1.39	1.25	1.32	1.24	1.48
25	1.26	1.18	1.27	1.15	1.32
30	1.16	1.14	1.20	1.10	1.22
35	1.10	1.10	1.13	1.05	1.12
40	1.04	1.06	1.06	1.00	1.04
45	0.97	1.02	0.99	0.95	0.96
50	0.92	0.98	0.92	0.94	0.90
55	0.85	0.95	0.88	0.92	0.86
60	0.79	0.92	0.85	0.90	0.81
65	0.74	0.89	0.82	0.89	0.77
70	0.70	0.86	0.79	0.86	0.73
75	0.66	0.83	0.77	0.81	0.69
80	0.61	0.80	0.75	0.77	0.65
85	0.57	0.77	0.71	0.73	0.59
90	0.52	0.73	0.69	0.68	0.54
95	0.46	0.68	0.59	0.62	0.47
100	0.39	0.57	0.42	0.51	0.40

much interannually. Flow exceedance statistics were calculated for all the homogeneous regions but Region 4 because the mean annual streamflow was not available at the stations in this region by specific year.

The mean annual flow at 40, 60, 70, and 90 percent exceedance levels were then computed as follows:

$$Qi = Q_{avg} * Fraction_i \qquad (12)$$

where

Q_i is the mean flow,

i is the exceedance level (40, 60, 70, or 90 percent),

Q_{avg} is the mean annual flow,

$Fraction_i$ is the ratio of mean annual flow at i exceedance level from table 12.

The computed discharge at above four exceedance levels were incorporated into the geodatabase. Successively, the theoretical potential hydropower was computed using estimated mean annual streamflow as well as the estimated annual streamflow at 40, 60, 70, and 90 percent exceedance level.

Theoretical Hydropower Potential Assessment

This study produced a geodatabase that provides a comprehensive estimate of the magnitude of hydropower potential available in the streams that traverse Bolivia. The geodatabase also contains all the required primary and intermediary geospatial datasets such as in situ rainfall and station discharge datasets, basin/catchment boundaries, homogeneous regions, rainfall climatology datasets, modeled runoff, and hydropower assessment datasets including existing hydropower estimates. The hydropower potential was calculated for each 1-km stream segment; however, some stream segments were less than 1 km because of a stream bifurcation. The total theoretical hydropower potential of the country is defined as the annual energy that is potentially available if all natural runoff at all the streams that traverse Bolivia were to be harnessed without any energy losses.

The hydropower potential of each stream segment was calculated using the hydraulic head and annual mean flow rates at the outlet of the stream segment. The flow was estimated using the regionalization model described in the preceding sections of the report. The hydraulic head associated with each stream segment was obtained from the conditioned SRTM elevation data. Theoretical hydropower potential was calculated for every stream segment in the stream network, considering it as independent from other segments. The schematic of a stream segment and where the drop and flow are calculated for the segment is shown in figure 32. The hydropower for the stream segments was estimated based on this generalized form of the equation. More details on the power equation are provided in the "Available Theoretical Hydropower Potential in Bolivia" section.

$$P = \rho * g * H * Q \qquad (13)$$

where

P is hydropower potential, in watts,

ρ is the density of water, in 1,000 kilograms per cubic meter,

g is gravitational acceleration of 9.81 meters per square second,

H is the hydraulic head in meters at the stream outlet, and

Q is the average flow rate at the outlet, in cubic meters per second.

For computing total theoretical hydropower potential, turbine efficiency was assumed to be 100 percent, no hydraulic head loss was considered in the calculation, and no other limitations were considered; therefore, the resulting hydropower estimates are for gross potential.
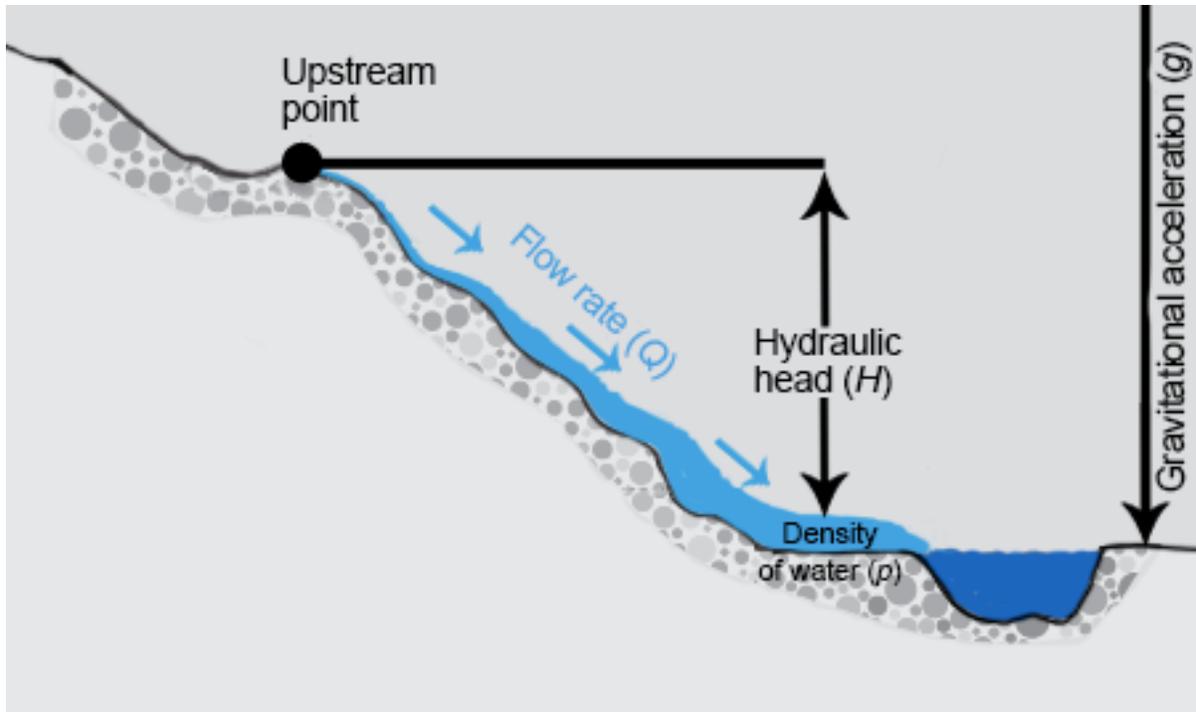
Figure 32.   Schematic diagram of the theoretical potential hydropower computation based on estimates of the streamflow and hydraulic head at the stream outlet.

In addition to the theoretical annual mean hydropower estimates, theoretical potential hydropower estimates with flow rates with exceedance probabilities of 40, 60, 70, and 90 percent are given with the datasets. The flow rate exceedance probabilities were estimated with regionalization models of the flow-duration curves. The final product of the hydropower assessment is a GIS streams vector layer segmented every 1 km with attributes containing estimated flows in cubic meters per second, hydraulic head in meters, and annual mean theoretical power potentials in megawatts and the four above-described levels of probability.

Finally, the theoretical hydropower computed using drop and mean annual discharge, and discharges at four different exceedance levels were incorporated into the geodatabase.

Theoretical Potential Hydropower by Basin

Once the theoretical potential hydropower has been computed for each 1-km segment of the river, we summarized the potential power by three levels (level 1, 2, and 3) of hydrographic units in Bolivia. These hydrographic units were delineated from the Pfafstetter Coding system. The Pfafstetter Coding system is a method for assigning drainage basin identification codes based on the topology of the land surface (Verdin, 1997, Verdin and Verdin, 1999). The system is hierarchical, and drainage basins are delineated from junctions on a river network. For Bolivia, level 1 drainage basins correspond to national-scale drainage basins. Higher levels (levels 2, 3, and so on) represent ever-finer tessellations of the land surface into smaller drainage basins, which are subdrainage basins of lower level drainage basins. Each drainage basin is assigned a specific Pfafstetter Code based on its location within the overall drainage system and on the total drainage area upstream from the drainage basin's outlet. In Bolivia, the level 1 hydrographic units are the drainage areas that contribute flow to another drainage area or to another water body such as a lake. This is the lowest level of the drainage basins. Level 2 hydrographic units are the basins that receive flow from upstream drainage basins; this is the second level of the hydrographic units. Finally, the highest level of the unit is level 3, which contains the headwater of the river reach for which the drainage basin is defined. These hydrographic units were prepared by the Deputy Minister Water Resources and Irrigation and the Sustainable Agricultural Development Program and widely used in Bolivia. These hydrographic units provide the basis for us to summarize the theoretical potential hydropower. The theoretical potential hydropower summary statistics are presented graphically in figures 33, 34, and 35 for hydrographic unit levels 1, 2, and 3, respectively.

Three major hydrographic units under level 1 cover Bolivia. The unit in north-central Bolivia has the highest potential of about 176 gigawatts (GW), followed by the hydrographic unit covering southeastern Bolivia with a potential of about 28 GW, and the hydrographic unit in southwestern Bolivia with a potential of about 8 GW.
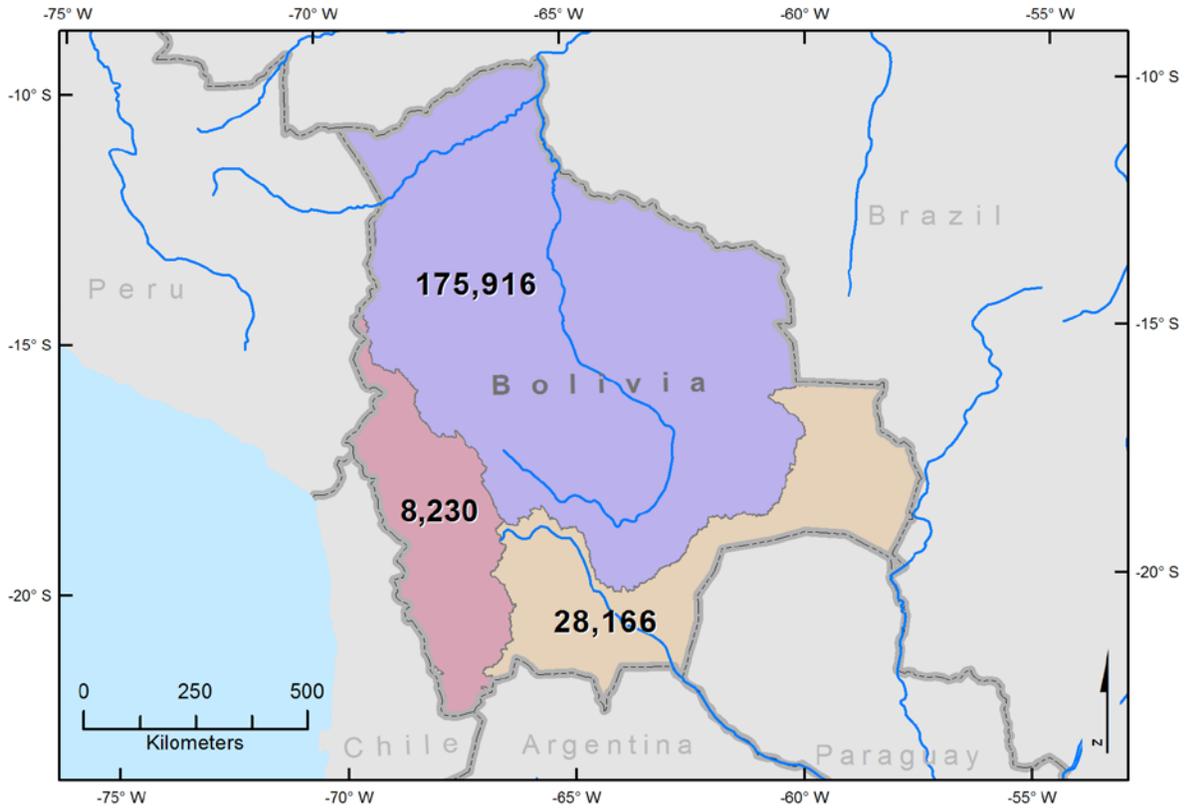
Figure 33.   Total theoretical hydropower potential in megawatts summarized by level 1 hydrographic units for Bolivia.
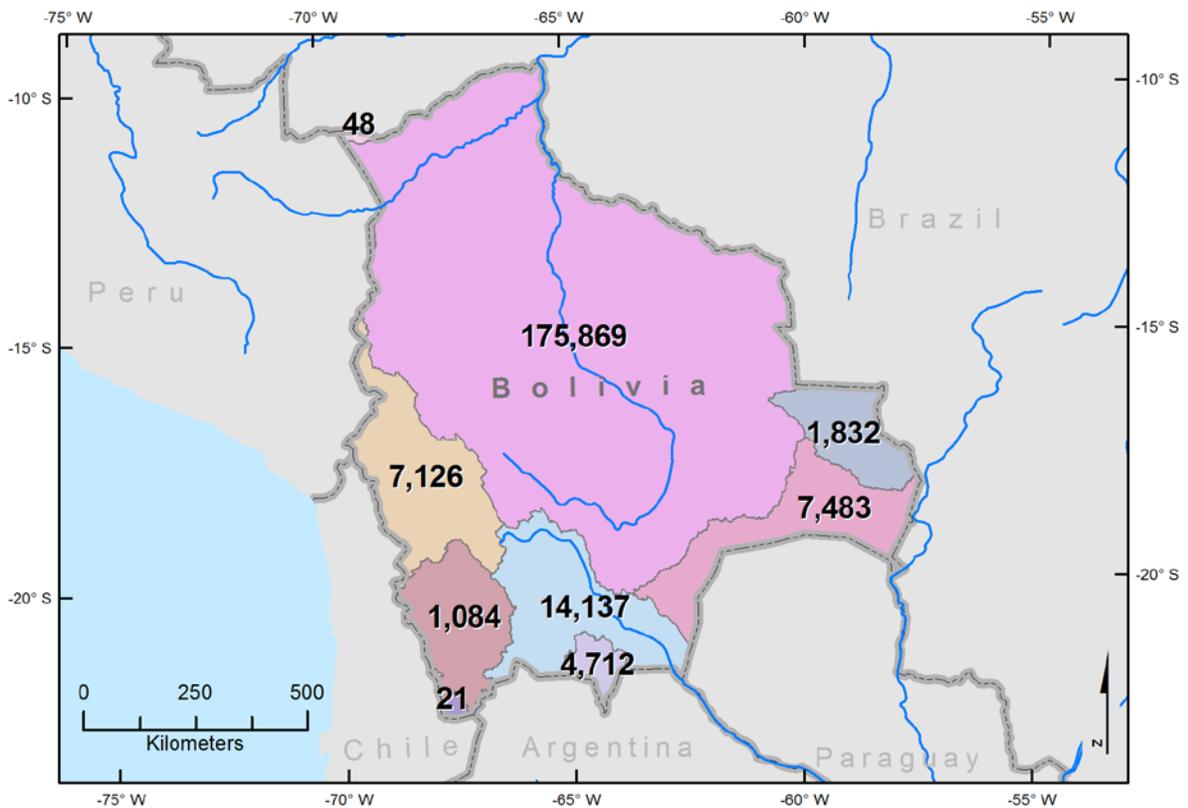
Figure 34.   Total theoretical hydropower potential in megawatts summarized by level 2 hydrographic units for Bolivia.
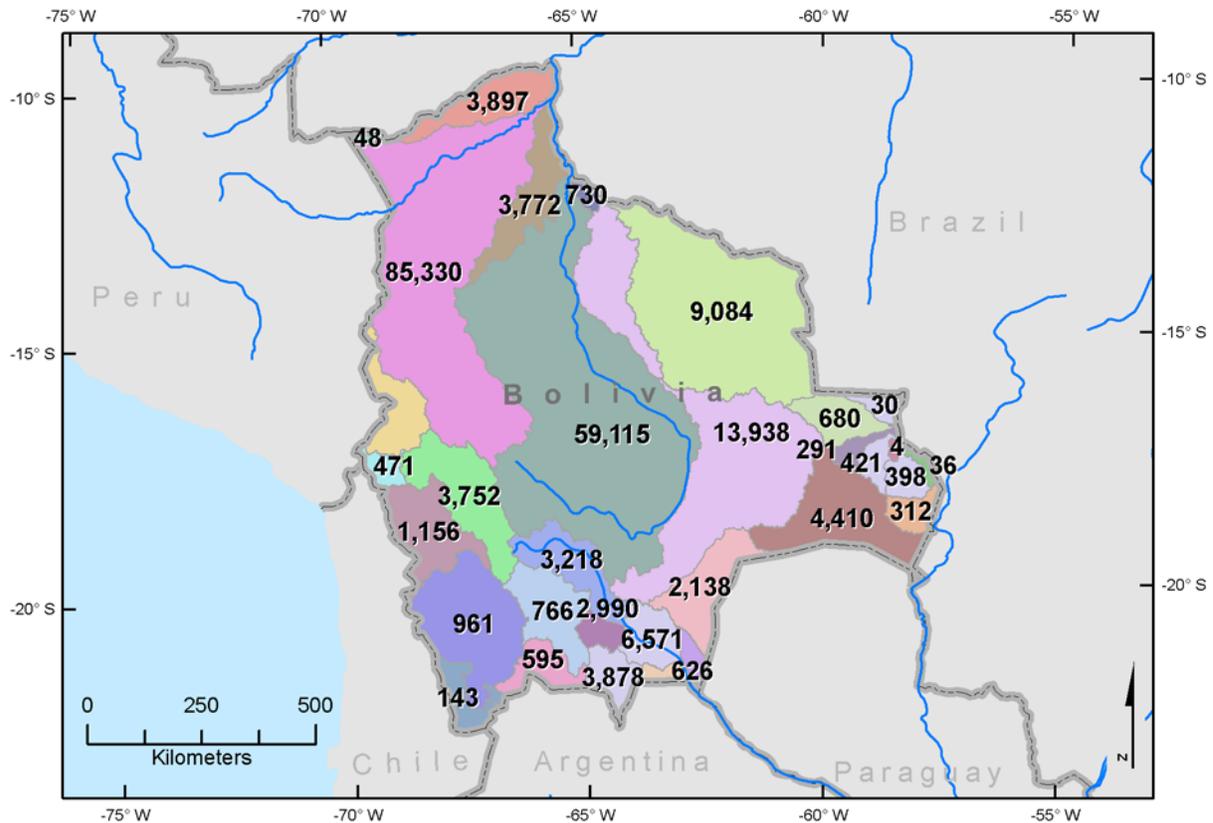
Figure 35. Total theoretical hydropower potential in megawatts summarized by level 3 hydrographic units for Bolivia (not all hydrographic unit's hydropower values are shown, see table 13).

There are 11 level 2 hydrographic units. In level 2, the north-central hydrographic unit of level 1 remains very much the same except one small divide in the north; therefore, it maintains the same potential of about 176 GW of power. The southeastern hydrographic unit of level 1 is divided into four units at level 2, from which one unit at the western part shows the most potential of about 14 GW. The remaining units have less potential in the region because they coincide with relatively lower rainfall areas in Bolivia. The southwestern unit of level 1 is divided into three hydrographic units in level 2 with very little power potential because the region is classified as a dry mountainous region.

Level 3 is the highest level of hydrographic unit. The 9 level 2 hydrographic units are further divided into 67 hydrographic units in level 3. The theoretical potential hydropower for all the units in level 3 is provided in table 13. Out of these units, the unit in central Bolivia has the highest potential for hydropower of about 59 GW.

Protected Areas

The stream segments within and at the periphery of the protected areas were identified and marked in the geodatabase. Theoretical hydropower potential of the streams falling under protected areas should not be considered while computing

total theoretical power potential of drainage basins or for the entire country. Assuming that no further hydropower projects will be constructed that would alter the flows and ecosystems of the protected lands, we summarized the theoretical potential hydropower outside the protected areas by hydrographic units of levels 1, 2, and 3. The summary statistics by hydrographic units are presented in figures 36, 37, and 38 for levels 1, 2, and 3, respectively.

Available Theoretical Hydropower Potential in Bolivia

Theoretical power potential of the streams on which existing hydropower projects are located should be ignored while computing total theoretical power potential in Bolivia. These streams and the streams upstream to the existing hydropower projects have already been harnessed/exploited. This is based on the assumption that no further hydropower developments would take place that would further alter the flows in these streams and affect existing hydropower projects. The total theoretical hydropower potential of the streams falling within the command area of the existing hydropower projects were identified and provided in the geodataset.

Table 13.   Total theoretical hydropower potential in Bolivia and potential hydropower (MW) outside protected areas summarized by level 3 basins. (Basin consolidation occurred in the Uyuni Salt Flats and Grand Chaco areas due to minimum basin area requirements.)

[SL, serial number; km², square kilometer; MW, megawatts; NA, not applicable]

SL number	Basin identification			Basin name	Basin area (km²)	Total potential hydropower (MW)	Potential hydropower outside protected areas (MW)
	Level 1	Level 2	Level 3				
1	0	1	12	Cuenca Rio Lauca	27,755	1,156	1,125
2	0	1	13	Unidad Hidrografica 013	34,850	3,752	3,500
3	0	1	14	Cuenca Rio Mauri	5,942	471	471
4	0	1	15	Unidad Hidrografica 015	19,776	1,717	1
5	0	2	22	Cuenca Salar De Uyuni	50,681	961	920
6	0	2	27	Unidad Hidrografica 027	11,009	143	72
7	4	46	463	Unidad Hidrografica 463	24,325	3,897	2,586
8	4	46	464	Cuenca Rio Beni	169,981	85,330	48,687
9	4	46	465	Unidad Hidrografica 465	25,148	3,772	3,768
10	4	46	466	Cuenca Rio Mamore	234,168	59,115	32,259
11	4	46	467	Unidad Hidrografica 467	2,986	730	730
12	4	46	468	Cuenca Rio Itonamas	124,978	13,938	12,178
13	4	46	469	Unidad Hidrografica 469	127,506	9,084	6,321
14	4	49	462	Cuenca Rio Purus	1,951	48	48
15	8	85	858	Cuenca Rio Bermejo	11,889	3,878	2,745
16	8	86	861	Unidad Hidrografica 871	4,341	626	617
17	8	86	864	Cuenca Rio Carapar?	3,513	834	783
18	8	86	865	Unidad Hidrografica 863	15,083	6,571	5,682
19	8	86	866	Cuenca Rio Pilcomayo	23,519	3,218	3,218
20	8	86	867	Unidad Hidrografica 865	6,095	2,990	2,990
21	8	86	868	Cuenca Rio Tumusla	25,151	766	751
22	8	86	869	Unidad Hidrografica 869	12,653	595	595
23	8	87	872	Cuenca Rio Ca?Ada Ustarez	25,910	2,138	1,518
24	8	87	874	Cuenca Del Rio Negro	44,779	4,410	2,584
25	8	87	879	Unidad Hidrografica 879	8,205	312	107
26	8	89	893	Unidad Hidrografica 893	2,446	36	0
27	8	89	895	Cuenca Rio Correreca	5,975	421	0
28	8	89	895	Cuenca Rio San Miguel	7,271	398	0
29	8	89	895	Unidad Hidrografica 895	1,335	4	0
30	8	89	897	Cuenca Curiche Grande	5,486	30	30
31	8	89	897	Unidad Hidrografica 897	4,932	291	103
32	8	89	899	Unidad Hidrografica 899	14,785	680	523
Totals					**1,084,424**	**212,283**	**134,912**
Totals in gigawatts (GW)					**NA**	**212.3**	**134.9**

There are 24 existing hydropower plants in Bolivia (table 14; fig. 39) grouped under seven major projects. The 24 hydropower plants utilize either natural conditions (natural streamflow and hydraulic head) or engineering design to enhance flow and hydraulic head to compute hydropower; however, we obtained the stream database from Bolivian authorities that identifies the streams as harnessed/exploited under these projects. These streams are labeled as harnessed (not available) in

the geodatabase and excluded from the summary of available hydropower potential.

Details of how each project was analyzed to identify the streams as harnessed and the results corroborated with the dataset obtained from the Bolivian authorities are provided in appendix 5.

The 1-km stream segments identified as harnessed/exploited under each project are provided in table 15.

Results indicate that only 941 out of 316,490 1-km stream segments were harnessed (table 15). This accounts for about 0.3 percent of the total 1-km streams in Bolivia. Similarly, the total theoretical power generated under these projects is about 0.71 percent of the total theoretical hydropower (1.41/197.73 GW).

Computation of Theoretical Technical/Available Hydropower Potential

Total theoretical hydropower potential (*TTHP* or *P*) of the stream is estimated using the following equation:

$$P = (H_{geo} - H_{loss}) \times (Q_{avg} - Q_{eco}) \times g \times \rho \times \epsilon \quad (14)$$

where

H_{geo} is the gross head;
H_{loss} is estimated as a the total head loss because of hydraulic losses and tailrace effect associated with the flow;
Q_{avg} is the average flow for the stream segment;
Q_{eco} is the average discharge left after leaving a minimum environmental flow within the stream segment;
g is the acceleration of gravity (9.81 meters per square second [m/s²]);
ρ is the density of water (1,000 kilograms per cubic meter [kg/m³]); and
ϵ is the overall efficiency (ϵ=0.72).

The parameter H_{loss} is site specific and depends on the amount of discharge that is flowing through the river. This parameter must be estimated using detailed local-based study and analysis (part of Phase II); hence, we do not consider H_{loss} in this study (assumed as zero).

There are several methods to estimate Q_{eco} for a stream/river segment. Detailed literature review of the methods of estimating Q_{eco} is provided in Smakhtin (2001); however, for this study, we used a look-up table generated for identifying ecological flows based on the Tennant method, which is also known as the Montana method (Tennant, 1975) (fig. 40).

The ecological requirement for flow to be retained in the stream ranges from 20 to 40 percent with an average of 30 percent annually; hence, 30 percent of the flows in the streams should always be left in the streams and therefore cannot be harnessed for hydropower production. Because of this ecological requirement, the total theoretical hydropower potential outside of protected areas will reduce by 30 percent.

Another important parameter to be considered while computing available technical hydropower potential is the overall efficiency factor (ϵ). The overall efficiency factor is computed as the product of different efficiencies shown in equation 15.
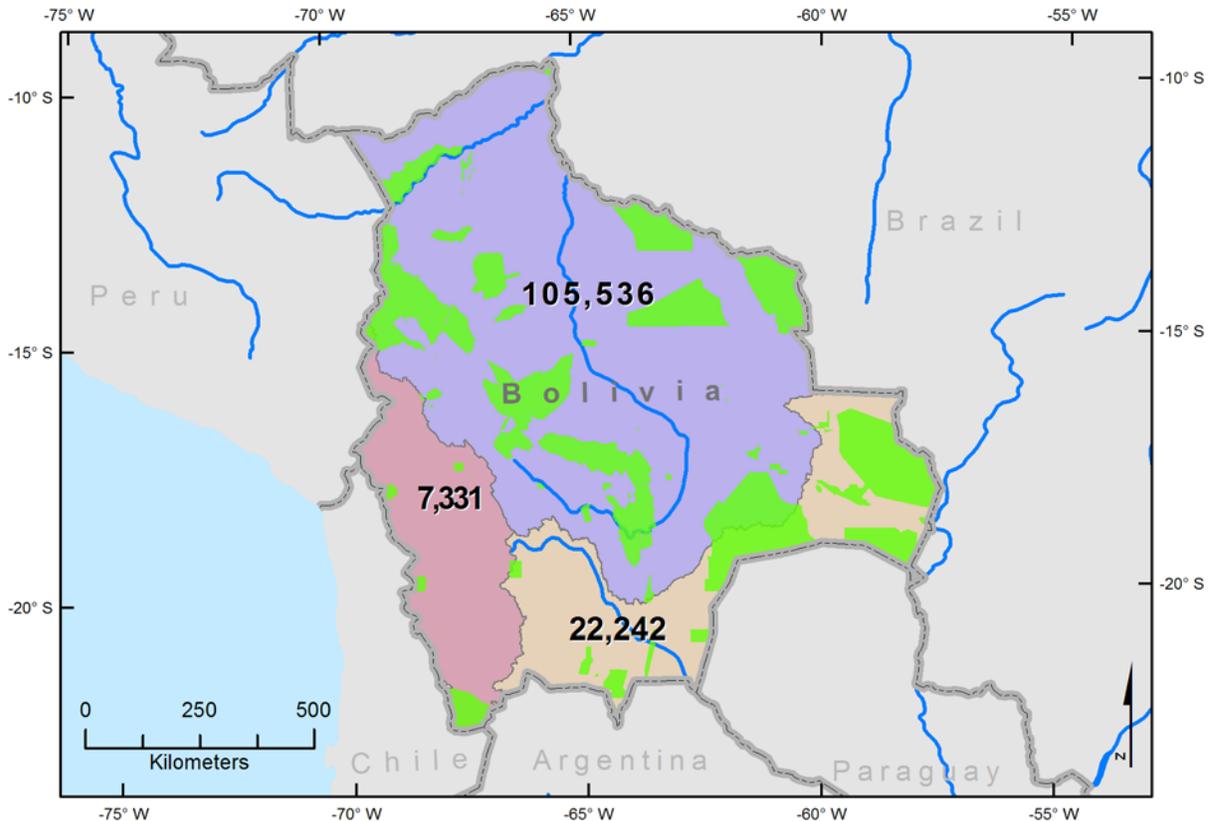
Figure 36.   Total theoretical hydropower potential in megawatts outside the protected areas in Bolivia summarized by level 1 hydrographic unit (not all hydrographic units hydropower values are shown, see table 13).
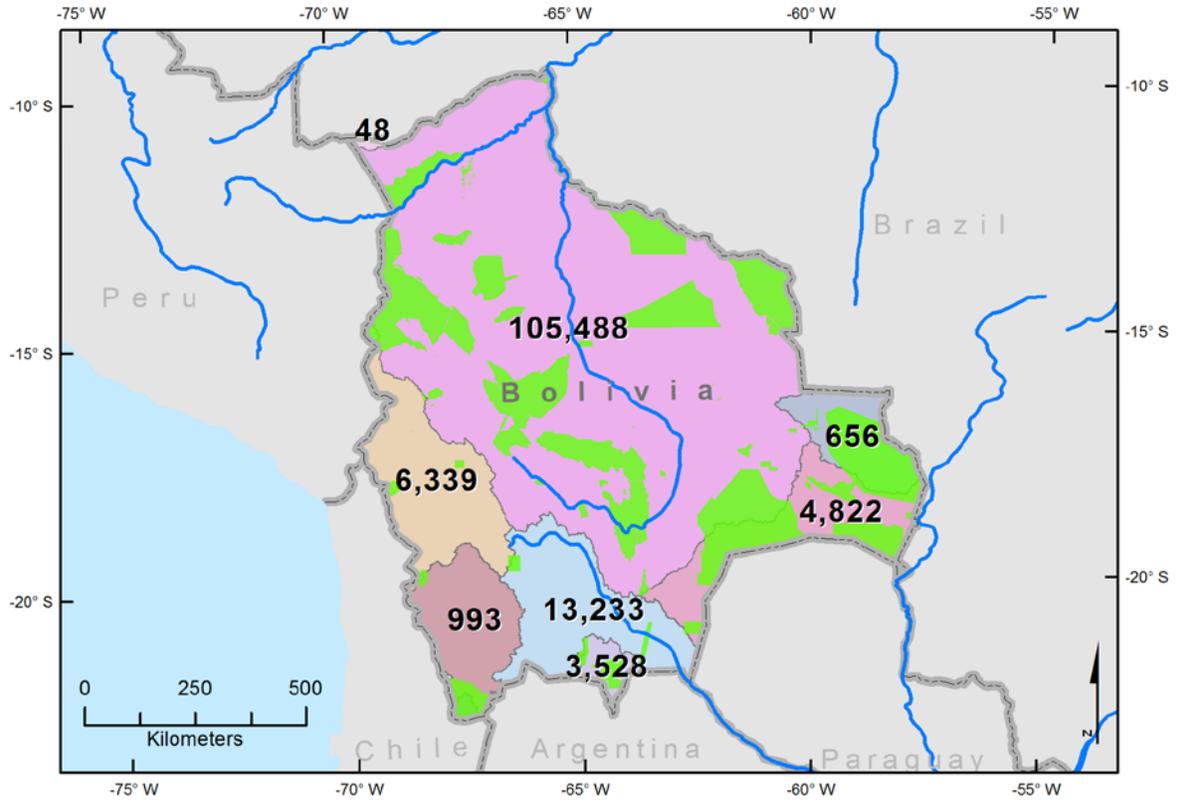
Figure 37.   Total theoretical hydropower potential in megawatts outside the protected areas in Bolivia summarized by level 2 hydrographic unit.
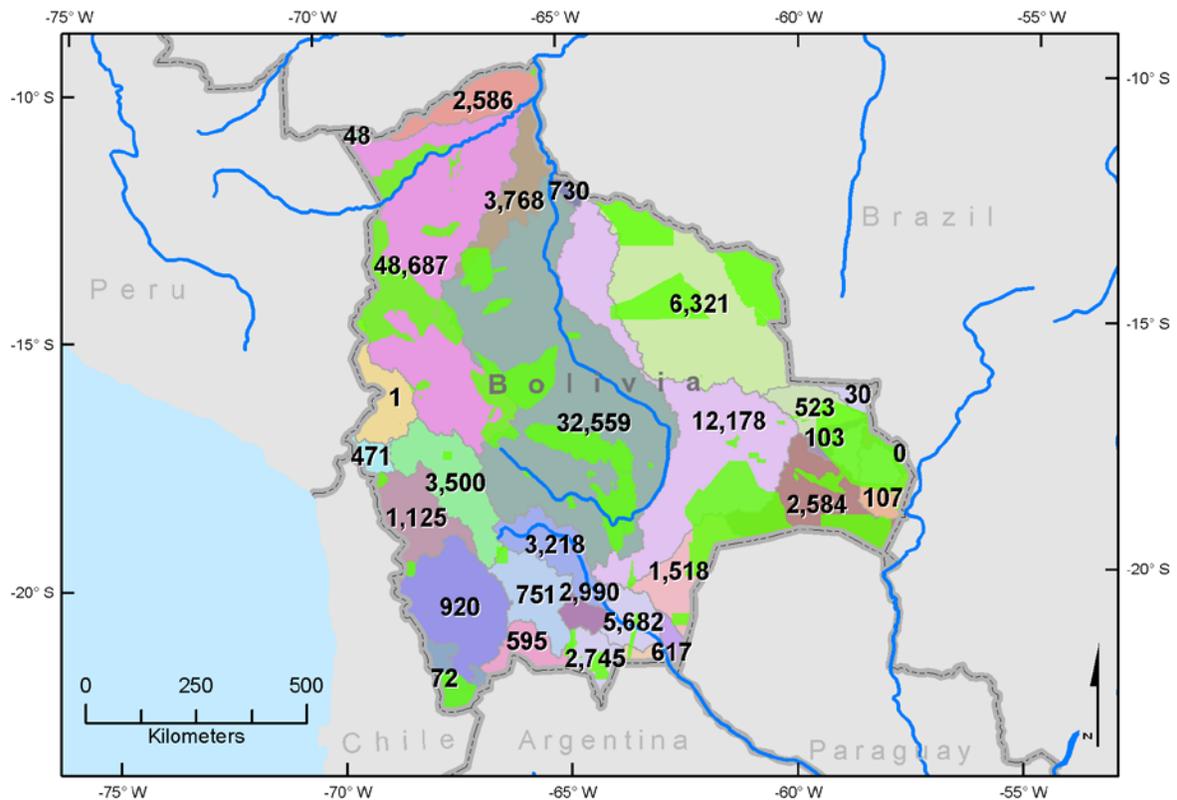
Figure 38.   Total theoretical hydropower potential in megawatts outside the protected areas in Bolivia summarized by level 3 hydrographic unit.

Table 14.   List and characteristics of existing hydropower projects in Bolivia.

[ID, identification number; MW, megawatt; km, kilometer]

ID	Name	Major project group	Description	Power potential (MW)
1	Choquetanga	Miguillas	Not available	6.00
2	Santa Isabel	Corani	A big diversion pipe (0.6 km)	93.30
3	Angostura	Miguillas	Drainage diversion	7.30
4	Corani	Corani	Diverting water from lake to the west	54.00
5	San Jacinto	San Jacinto	Structural Dam (8 MW seem low)	8.00
6	Kilpani	Kilpani	Drainage diversion	11.80
7	Landara	Kilpani	Drainage diversion	5.60
8	Punutuma	Kilpani	Not available	2.50
9	Carabuco	Miguillas	Drainage diversion	6.30
10	Miguillas	Miguillas	Not available	3.00
11	Kanata	Kanata	Not available	0.00
12	Huaji	Zongo	Drainage diversion	30.10
13	Cahua	Zongo	Drainage diversion	28.80
14	Harca	Zongo	Drainage diversion	27.40
15	Chururaqui	Zongo	Drainage diversion	26.10
16	Sainani	Zongo	Drainage diversion	9.90
17	Santa Rosa	Zongo	Not Available	17.50
18	Cuticucho	Zongo	Drainage diversion	21.20
19	Zongo	Zongo	Not Available	10.60
20	Botijlaja	Zongo	Drainage diversion	6.60
21	Tiquimani	Zongo	Drainage diversion	9.40
22	Quehata	Quehata/Chinata	May be run-of-river	2.10
23	Chiãnata	Quehata/Chinata	Diversion or run-of-river	0.30
24	Yura	Kilpani	Not available	0.00

Table 15.   List of installed/operational hydropower projects and the potential of streams classified as harnessed under the limits of these projects.

[No, number; km, kilometer; THPP, total theoretical hydropower potential; MW, megawatts]

No	Project	1-km stream segments	Total THPP (MW)
1	Zongo	94	1,038.0
2	Tauesi	42	59.7
3	Corani	156	142.6
4	Yura eresa	304	22.1
5	Miguillas	34	15.2
6	San Jacinto	110	28.3
7	Kanta	34	22.0
8	Quehata Chinata	167	86.2
Total		**941**	**1,414.3**
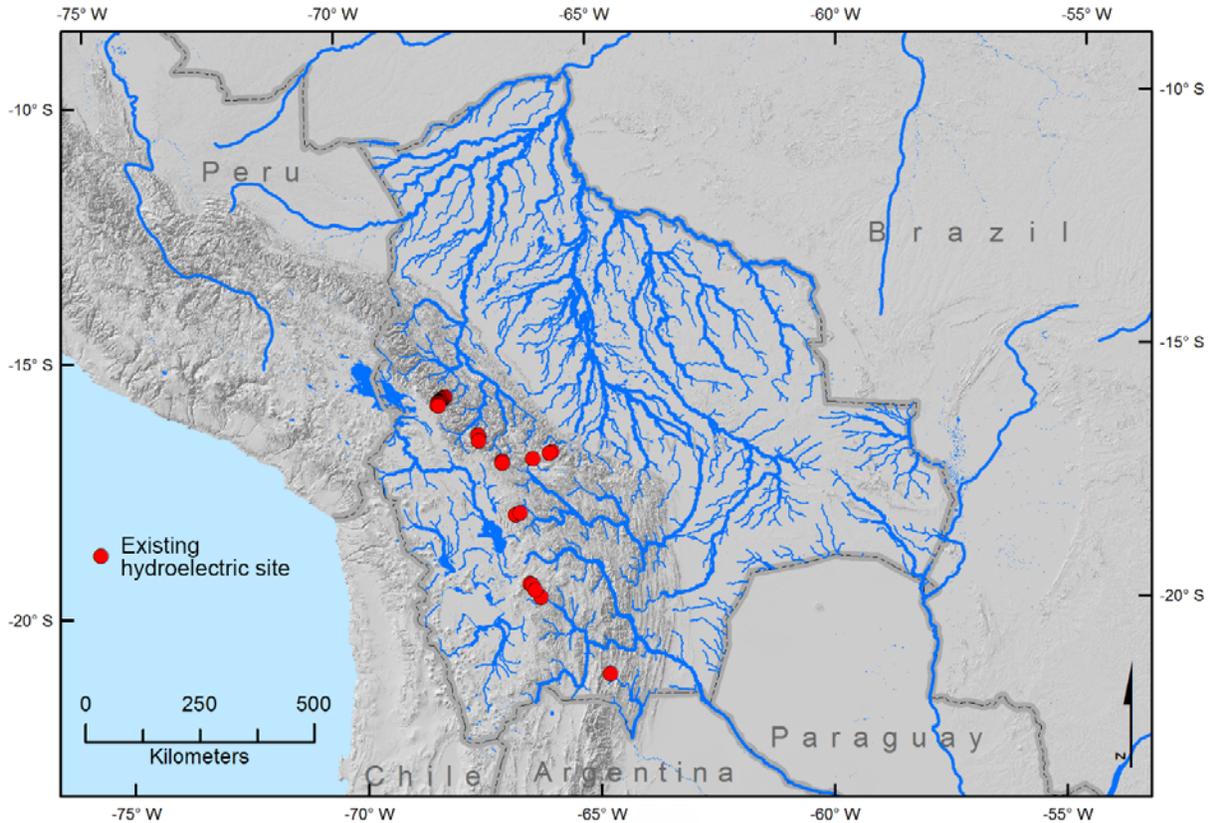
Figure 39.   The distribution of streams and existing hydropower stations in Bolivia.

$$\epsilon = \epsilon_t \times \epsilon_{tr} \times \epsilon_p \times \epsilon_{oth} \qquad (15)$$

where

ϵ_t is defined as turbine efficiency of 0.9 (Intergovernmental Panel on Climate Change, 2011);

ϵ_{tr} is the transmission efficiency of 0.94 (Wong, 2011);

ϵ_p is the parasitic efficiency of 0.9 (Electric Power Research Institute, 2011); and

ϵ_{oth} is the efficiency factor (assumed as 0.95) that accounts for other minor losses.

After applying/considering the overall efficiency factor, the total theoretical hydropower potential minus the protected (36 percent), minus the ecological flows (30 percent), and minus the available (0.72 percent) will be reduced further by 28 percent. This means the total available theoretical technical hydropower in Bolivia would be equal to 72 percent of the potential hydropower after ecological flows have been met.

Summary of Total Theoretical Available/ Technical Hydropower Potential

The summary of TTHP identified from the produced geodatabase in this study is presented in table 16. First, the overall TTHP of all the streams within Bolivia is estimated at 212 GW. This number represents the total capacity of hydropower that can be produced when each and every 1-km stream segment within Bolivian borders is exploited to its full capacity. This number does not represent the true potential that can possibly be extracted because it is not possible to exploit each and every stream.

To arrive at a more realistic number, we first estimated the TTHP of 1-km segments falling within protected areas in Bolivia. Nearly 77.4 GW of potential exists within protected areas. Because hydropower projects cannot be initiated in these protected areas, we estimated the remainder of TTHP outside the protected areas in Bolivia, which was 135 GW.

Care must be taken to understand and interpret the real meaning of the TTHP identified in this study. The TTHP (135 GW) identified in this study represents the total theoretical potential hydropower that can be produced only when all the streams within Bolivia are exploited to their full capacity. In reality, it is impossible to harness 100 percent of the streams in the country to their full capacity. Even though every stream has some potential to produce hydropower, not all streams can be harnessed; for example, about 52 percent of the streams either have head height of less than a meter or residual discharge of 1 cubic meter per second (m^3/s) or less after meeting environmental flow requirements. Together, these streams contribute to a TTHP of nearly 6.7 GW. This combined potential is large (6.7 GW), but it is scattered over several thousands of streams, making it less feasible technically and economically;

DONALD L. TENNANT

U.S. FISH AND WILDLIFE SERVICE

FEDERAL BUILDING

BILLINGS, MONTANA 59101

"Montana Method" for prescribing Instream Flow Regimens for Fish, Wildlife, Recreation and Related Environmental Resources.

Narrative Description of Flows 1/	Recommended Base Flow Regimens Oct-Mar : Apr-Sept	
Flushing or Max.	200% of the average flow	
Optimum Range	60%-100% of the average flow	
Outstanding	40%	60%
Excellent	30%	50%
Good	20%	40%
Fair or Degrading	10%	30%
Poor or Minimum	10%	10%
Severe Degradation	10% of average flow to 0 flow	

1/ Most appropriate description of the streamflow for all the parameters in the title.

Figure 40. Look-up table for prescribing instream flow requirements based Montana method (Maidment, 1993).

hence, the potential of such streams should be excluded from the TTHP identified (135 GW) because they will not be good candidate streams for exploitation of hydropower. Similarly, in several remote locations, it may not be possible to access the streams. In other remote places, it may not be economical to transfer the power produced.

Bolivia has 24 hydropower plants that are in operation and producing power on a daily basis. We computed and classified all the 1-km stream segments falling within the limits of these existing hydropower projects and estimated TTHP of these streams. We estimated that the streams falling within the limits of existing power projects have a potential of 1.41 GW. The remainder of TTHP left to be exploited to the fullest extent is 133.5 GW. This is still not the true potential because this is the number estimated only when all the regions in Bolivia (excluding streams within protected areas and existing power projects) are exploited to their full capacity.

To compute the total theoretical available/technical potential, we must consider several factors, such as environmental flow requirements, efficiency factors, and economic and feasibility factors. If we consider environmental flow

requirements, the total theoretical available/technical power potential reduces by 60–80 percent. Application of the efficiency factor (estimated using equation 15) further reduces the hydropower potential estimate by another 28 percent.

We cannot apply the economic and feasibility factors to calculate a final available potential because they are location-specific factors, and further analysis and research is required to understand the economic and feasibility aspects of hydropower production for each 1-km stream segment. This analysis will be carried out in Phase II of this research.

Identifying Major Contributing Factors to the Hydropower Potential

Theoretical hydropower potential of each 1-km stream segment in Bolivia is identified in this study. To facilitate easy understanding of the potential and to help managers locate the regions for further focused in-depth analysis, we have identified the major contributing factors to the hydropower potential (either discharge or head height) for every 1-km stream

Table 16.   Computation of total theoretical available/technical hydropower potential.

[GW, gigawatt]

Item	Parameter	Equation	Power (GW)
A	Total theoretical hydropower potential (TTHP) of all the streams sharing borders and within Bolivia	--	212.3
B	TTHP within protected areas	--	77.4
C	TTHP outside protected areas	A–B	134.9
D	TTHP from harnessed streams	--	1.4
E	TTHP available after excluding harnessed streams	C–D	133.5
F	TTHP after meeting environmental flow requirement	60–80 percent of E	--
G	Total theoretical available/technical hydropower potential (after considering overall efficiency factor)	72 percent of F	--

segment and added them in the geodatabase. The result from this analysis is a map of Bolivian streams in three distinct classes (fig. 41). Class 0 is where the stream segments do not have any power potential. Classes 1 and 2 indicate streams that have hydropower potential. For Class 1 streams, the major contributing factor for hydropower is streamflow/discharge. For Class 2 streams, the major contributing factor for hydropower is head height.

Prior Hydropower Assessments of Bolivia

During the past 50 years, there have been three regional studies assessing Bolivia's available hydropower potential. The first was conducted in 1964 by the United Nations and was called "The water resources of Latin America III. Bolivia and Colombia" (Los recursos hidráulicos de América Latina III. Bolivia y Colombia). In 1976, "The Hydropower Resources of Bolivia" (Los Recursos Hidroenergéticos, De Bolivia) was published by the National Electricity Company [Empresa Nacional De Electricidad]. Both of these studies shared the same methodologies. The only difference in the two studies was the precipitation records: 15 and 30 years, respectively.

The 1976 study was based on precipitation data using 90 measuring stations, but only 4 exceeded 30 years of records, and 16 had less than 10 years. These precipitation data were used to generate a map of equal precipitation isohyets and combined with a coarser 1:1,000,000 elevation map to generate a precipitation map for Bolivia. This map was translated into runoff using empirical relations and subjective coefficients. Drops in elevation were not used as a variable in the study. The runoff potential was finally translated into hydropower estimations using empirical relations based on experience from other countries (Empresa Nacional De Electricidad, Naciones Unidas, 1976). The final overall results of the 1976 study determined that Bolivia had a TTHP of 36 GW.

The last study, "Evaluation of the hydropower resources of Bolivia" (Evalucion De Los Recursos Hidroenergéticos

De Bolivia), was published in 1984. This study used new information to update the 1976 assessment. The elevation maps were updated to 1:50,000 scale for much of the country, 371 rain gage stations were used, and 69 discharge gages were included. This study implemented a different treatment for streams with discharge information. Classifying power segments was based on inventoried sites and stream reaches. Power was calculated for inventoried sites and those classified as reaches. The final TTHP estimation in this 1984 study also took into account the type of plant. Firm energy was calculated from average energy production considering a discharge coefficient of 0.7 at average capacity. The final result of the assessment indicated that Bolivia had a TTHP of 40 GW (Empresa Nacional De Electricidad S.A, Organizacion Latinoamericana De Energia, 1984).

At the time, each of these studies applied the appropriate methodology given the information and resources available. Since then, the applied science of hydrology has advanced, and the collection of information has grown exponentially. As a result of this progression of science and technology, the methods and information used to derive these results differ vastly from 30 years ago, so the results are no longer comparable.

The methods applied and the technology used for this study allowed the assessment of 316,000 1-km stream segments, covering over 1 million square kilometers (km^2). Compared to the ground-based information collected 50 years ago, this study used satellite information that provided detailed elevation and climate data reaching regions that were inaccessible at the time. Using remotely sensed information in conjunction with ground-based data collected during the past 18 years provides a more comprehensive look at Bolivia's climate and hydrography than previously possible. By applying modern statistical and geospatial analysis to these data, the USGS has provided Bolivia a geographically broad and precise spatial look at this natural resource.
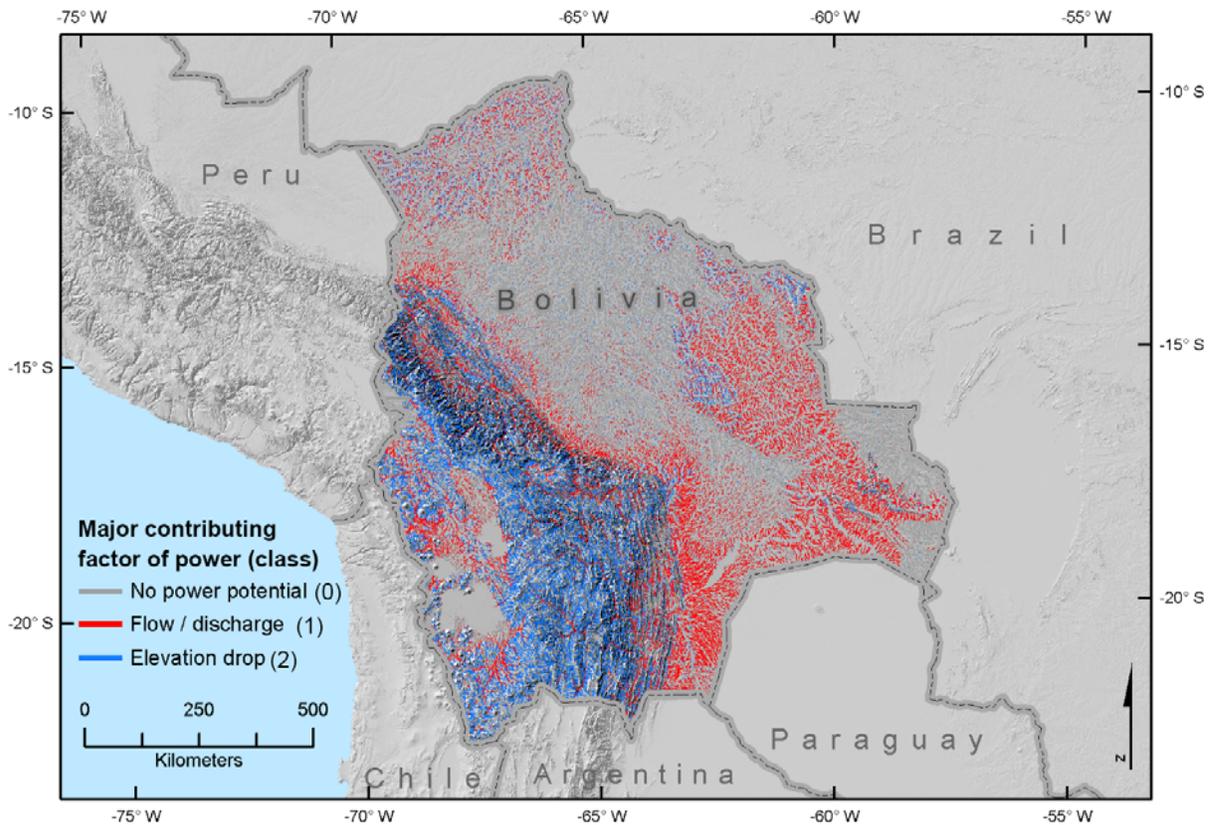
Figure 41. Hydropower contributing factor maps.

Uncertainty in Theoretical Potential Hydropower Estimates

Uncertainty in the hydropower assessment study can be classified into two broad categories: (a) uncertainties related to assumptions and (b) uncertainties arising from the datasets/model. In this section, we discuss these uncertainties in detail.

Uncertainties Related to Assumptions

Parametric Assumptions

Theoretical hydropower potential (P) of the stream is estimated using equations 13 and 14. First, we do not compute H_{loss} in this study and assume it is insignificant/minimal. Second, we assume that the ecological flow requirement of all the stream segments is 30 percent of the average flow. This assumption is based on the look-up table from the Montana method (Tennant, 1975; Maidment, 1993). Also, the overall efficiency is computed based on some of the efficiencies obtained from the literature that may or may not be appropriate in the context of Bolivia or streams in a particular region. Furthermore, we include other losses up to 5 percent that are not accounted for directly in the computation of overall efficiency. These assumptions introduce some degree of uncertainty in the hydropower estimates.

Theoretical Assumptions

Some of the theoretical assumptions that can introduce uncertainties into the final estimation of power are discussed here.

1. *Assumption of natural streamflow condition.*—This study assumes that all the stream segments in Bolivia are flowing based on their natural streamflow condition. Whenever the river discharge is altered (because of construction of a dam or a reservoir or because of diversion of water outside the stream/river), it produces significant differences in the estimates of river discharge (when compared to discharge under natural conditions). This would introduce uncertainty in the final power estimates.

2. *Assumption of run-of-the-river discharge.*—This study is based on the estimation of power potential based on the assumption that discharge of a river that is flowing through the river reach is exploited. In reality, engineering projects can be undertaken to modify the discharge as well as head (drop in elevation). Such modifications can increase the power potential of a river segment.

3. *Assumption of stationarity.*—This study is based on the analysis of historical data (about 30 years, 1981–2013). We assume stationarity in hydrological observations, which means that there are no significant changes in the hydrological observations in this region; that is, the precipitation, stream discharge, and other hydrological parameters have not changed and will not deviate significantly in the near future; furthermore, climate change effects have not been taken into account in the analysis.

Uncertainty Related to Datasets/Models Used

All the input datasets/models used in this study introduce some degree of uncertainty into the final theoretical potential hydropower estimates; hence, it is important to quantify the uncertainty in the individual primary input datasets to understand the potential uncertainty in the final theoretical power estimates. Input data errors and modeling uncertainties, however, are not independent of each other, and it is difficult to tease out the uncertainties resulting from the data and models; moreover, modeling errors likely will be consistent and easy to correct, so we focus here on the uncertainty in the datasets.

The theoretical power potential of the streams in Bolivia is estimated using equation 12. Two key parameters (streamflow and hydraulic head) are used for computing theoretical power estimates; however, because streamflow is modeled using satellite-based precipitation data and hydraulic head is computed directly from SRTM elevation, the uncertainty in precipitation, modeled streamflow, and SRTM elevation directly impacts the overall uncertainty in the power estimates. In this section, we discuss the uncertainty in three parameters (precipitation, streamflow, and hydraulic head) supported by findings from the literature.

Uncertainty in Bias-Corrected Precipitation

The TRMM–D gridded climatology precipitation dataset used in this study was validated using observed rainfall climatology data obtained from 465 stations distributed across Bolivia and neighboring regions of Brazil. The validation result (scatterplot) is presented in figure 18. Results indicate up to 24 percent error (RMSE) in the precipitation model (TRMM–D). A closer look at the validation result also reveals that the error differs with annual rainfall totals. Precipitation is usually overestimated (positive error) when rainfall amounts are low (<1,000 mm/year) and underestimated (negative error) when rainfall amounts are very high (>2500).

Uncertainty in Streamflow

Uncertainty in streamflow is summarized region by region in table 17 below. The error in the streamflow data ranges from as low as 2.09 percent in Region 3 to as high as 33.6 percent in Region 1. The overall combined error in the discharge model was 16 percent.

Our analysis revealed that the error in the discharge estimation contributes to a constant error of up to 14 percent in the power estimates (fig. 42. Regardless of the amount of discharge, there is always an error of up to 14 percent in the power estimates.

Uncertainty in Shuttle Radar Topography Mission (30-meter) Elevation Dataset

The global SRTM dataset was assessed by Rodriguez and others (2006) to measure its performance. The errors in SRTM elevation data were summarized for each continent. Both geolocation and height errors are shown in table 18. Although error estimates for Bolivia are not available, we can generalize error estimates for South America. As indicated in table 18, there could be up to 9 m of absolute geolocation error in the elevation dataset.

Geolocation error in the elevation data often results in the displacement of streams and rivers. The height error in SRTM elevation often leads to inaccurate and discontinuous streams. This displacement and discontinuity of streams would result in inaccurate location of streams and subsequently mismatching of drainage divides. To mitigate this error, a series of conditioning procedures (explained in the "Hydrological Conditioning" section) was applied to reduce the errors. The streams delineated from the raw SRTM elevation data were compared with the stream locations from Landsat imagery. Corrections in the geolocation of the streams were applied based on the actual geolocation of the streams obtained from Landsat; furthermore, the drainage basins delineated from the SRTM elevation were compared with the level 1, 2, and 3 basin boundaries obtained from the Bolivian authorities. When drainage divides indicated any discrepancy, the streams were corrected using Landsat and the drainage basins were redelineated. This iterative process of stream conditioning, delineation and comparison was repeated until the impact of drainage divide and elevation error were negligible.

However, some portion of the height error could propagate into the estimation of hydraulic head computation. This error in hydraulic head could directly impact theoretical power estimates. It is difficult, however, to quantify the error in hydraulic head unless in situ data are available; hence, for this study, we have generalized the height error (from table 18) as anywhere up to 5.5 m.

The contribution/impact of error from SRTM can vary based on the head height used to compute total hydropower. The lower the head height, the higher the impact on the hydropower estimate. On the other hand, when head height is large (>100 m), the impact of a 5.5-m error in SRTM and its contribution to the error power estimate is minimal (<5 percent).
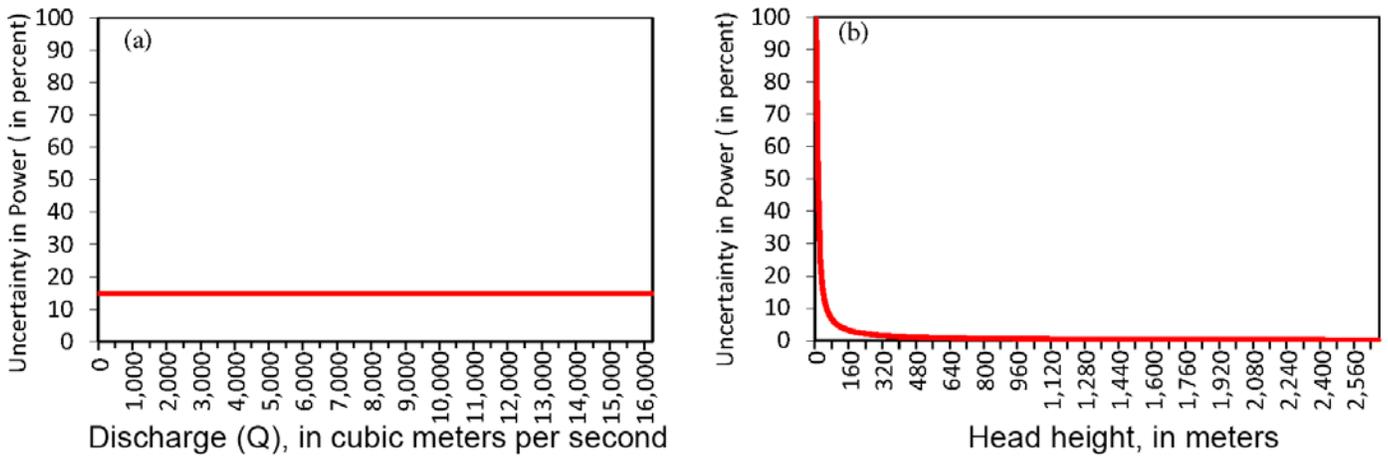
Figure 42. Contribution of (*a*) error in discharge to the final error in power estimation and (*b*) error in Shuttle Radar Topography Mission 30-meter elevation data to the final error in power estimation.

Table 17. Error originating from the streamflow modeling summarized by region.

[R^2, coefficient of determination; *p*-value, probability; RMSE, root-mean-square error; MAE, mean absolute error]

Homogeneous region	R^2	*p*-value	RMSE percent	MAE percent
Region 1	0.86	<0.001	33.6	27.2
Region 2	0.95	<0.001	9.67	7.43
Region 3	0.96	<0.001	2.09	1.75
Region 4	0.98	<0.001	10.3	7.72
Region 5	0.95	<0.001	13.2	10.5

Summary and Conclusions

Hydropower is the world's largest source of renewable energy, accounting for one-fifth of global total electricity generation, and four-fifths of global renewable electricity generation (U.S. Energy Information Administration, 2008). In Bolivia, the contribution of hydropower to total electricity ranges from 25 to 40 percent, and the remaining electricity is produced from burning fossil fuels. In an effort to reduce the carbon footprint, the goal of the Bolivian government during the coming years is to increase the contribution of renewable (hydropower) energy to total electricity produced in the country.

This study produced a geospatial database to be used as a decision support system by the Bolivian authorities for further sustainable hydropower development in Bolivia. The study assessed the theoretical hydropower of all 1-km stream segments in the country using multi-source satellite data and a hydrologic modeling approach. The Shuttle Radar Topography Mission's (SRTM) high-resolution 30-meter (m) digital elevation model (DEM) was used to identify stream segments

with greater than 100,000 flow accumulation and to compute head height for each stream segment. Several preconditioning processes were applied to the 30-m DEM to reduce errors and improve the accuracy of stream delineation and head height estimation. For the convenience of processing and handling large volumes of data, the study area was divided into five processing basins. About 536,264 1-kilometer (km) stream segments were identified covering the five basins and were analyzed to assess the total theoretical hydropower potential (TTHP) of Bolivia. A total of 463 station-based precipitation observations obtained from Bolivian Servicio Nacional de Meteorología e Hidrología (SENAMHI) and the Brazilian Agência Nacional de Águas (ANA) database were used to validate five different gridded precipitation estimates for Bolivia obtained from various sources.

Five gridded precipitation datasets were used in this study: Global Precipitation Climatology Project (GPCP) Version 2.2 combined monthly precipitation dataset (CPDS), Climate Prediction Center Merged Analysis of Precipitation (CMAP) dataset, Global Precipitation Climatology Center (GPCC) monthly precipitation dataset, TRMM Multi-Satellite Precipitation Analysis (TMPA, 3B43) dataset, and Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) precipitation climatology. The validation results indicated that gridded precipitation estimates from National Aeronautics and Space Administration's Tropical Rainfall Measuring Mission (TRMM) reanalysis product (3B43) showed the highest accuracies. However, the coarse resolution (25 km) of TRMM was disaggregated to 5 km using climatology information obtained from CHIRPS rainfall data. A 17-percent bias observed in the disaggregated TRMM estimates was corrected using the station observations. Using this bias-corrected precipitation estimate and SRTM elevation dataset, homogeneous regions for Bolivia were derived using a k-means clustering approach. Furthermore, using data from SENAMHI discharge stations and the bias-corrected, disaggregated TRMM precipitation estimates,

streamflow/discharge was computed using a regionalization approach.

Using the discharge and head height estimates for each 1-km stream segment, we computed hydropower potential for 316,490 stream segments sharing borders and within Bolivia. The TTHP of these segments was found to be 212 gigawatts (GW). Out of this total, 77.4 GW was within the protected areas where hydropower projects cannot be developed. The remaining total theoretical hydropower in Bolivia (outside the protected areas) is estimated as 135 GW. However, nearly 1,000 1-km stream segments were within the boundaries of existing hydropower projects. The TTHP of these stream segments was found to be nearly 1.4 GW. Therefore, the residual TTHP of the streams in Bolivia is estimated as 133.5 GW.

The TTHP identified in this study must be understood and interpreted with utmost care. The TTHP represents the total theoretical potential that can be harnessed when all the streams in Bolivia are exploited to their full capacity. In reality, it is not possible to harness 100 percent of the streams to their full capacity. For example, about 52 percent of the streams have a head height of less than 1 m or have residual discharge of 1 m^3/s or less after meeting environmental flow requirements. Together, these streams contribute to a TTHP of nearly 6.7 GW. Such streams should be excluded from the TTHP identified because these streams will not be good candidates for exploitation of hydropower. Similarly, in several remote locations of Bolivia, it may not be possible to access the streams or transfer the power produced.

Furthermore, factors such as required environmental flows, efficiency, economics, and feasibility need to be considered to better identify a more real-world hydropower potential. If environmental flow requirements of 20–40 percent are considered, the total theoretical power available reduces by 60–80 percent. Additionally, a 0.72-efficiency factor further reduces the estimation by another 28 percent. With this report and accompanying geospatial data providing the base theoretical hydropower potential for Bolivia, the next step is to identify optimal hydropower plant locations and factor in principles to appraise a real-world power potential in Bolivia.

References Cited

Abida, Habib, and Ellouze, Manel, 2006, Hydrological delineation of homogeneous regions in Tunisia: Water Resources Management, v. 20, no. 6, p. 961–977. [Also available at http://dx.doi.org/10.1007/s11269-006-9017-3.]

Adler, R.F.; Huffman, G.J.; Chang, Alfred; Ferraro, Ralph; Xie, Ping-Ping; Janowiak, John; Rudolf, Bruno; Schneider, Udo; Curtis, Scott; Bolvin, David; Gruber, Arnold; Susskind, Joel; and Arkin, Eric, 2003, The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–Present). Journal of Hydrometeorology, v. 4, no. 6, p. 1147–1167. [Also available at http://dx.doi.org/10.1175/1525-7541(2003)004%3C1147:TVGPCP%3E2.0.CO;2.]

Bamler, Richard, 1999, The SRTM mission—A world-wide 30 m resolution DEM from SAR interferometry in 11 days: Berlin, Germany, Wichmann Verlag, Photogrammetric Week, v. 99, p. 145–154.

Burn, D.H., 1997, Catchment similarity for regional flood frequency analysis using seasonality measures: Journal of Hydrology, v. 202, nos. 1–4, p. 212–230. [Also available at http://dx.doi.org/10.1016/S0022-1694(97)00068-1.]

Chiang, Shih-Min, 1996, Hydrologic regionalization for the estimation of streamflows at ungaged sites based on time series analysis and multivariate statistical analysis: New York, Syracuse University, Civil and Environmental Engineering—Dissertations, paper 26.

Ding, Chris, and He, Xiaofeng, 2004, K-means clustering via principal component analysis: New York, The Association for Computing Machinery, p. 29.

U.S. Energy Information Administration, 2008, International energy statistics: U.S. Energy Information Administration, accessed April 25, 2015, at http://www.eia.gov/countries/data.cfm.

Empresa Nacional De Electricidad, Naciones Unidas, 1976, Los Recursos Hidroenergéticos, De Bolivia [The Hydropower Resources of Bolivia]: National Electricity Company - United Nations, v.71, no. 532 BOL. 71/532

Empresa Nacional De Electricidao S.A., Organizacion Latinoamericana De Energia, 1984, Evaluacion De Los Recursos Hidroenergeticos De Bolivia [Hydropower Resource Assessment of Bolivia]: National Electricity Company S.A. and Latin American Energy Organization, no. 995.

Electric Power Research Institute, 2011, Electricity use in the electric sector: Electric Power Research Institute, Program on Technology Innovation, technical report.

Ercan, K., Mehmet, C., Osman, A., 2008. Hydrologic homogeneous regions using monthly streamflow in Turkey. Earth sciences research journal, 12(2), 181–193.

Farr, T.G.; Rosen, P.A.; Caro, Edward; Crippen, Robert; Duren, Riley; Hensley, Scott; Kobrick, Michael; Paller, Mimi; Rodriguez, Ernesto; Roth, Ladislav; Seal, David; Shaffer, Scott; Shimada, Joanne; Umland, Jeffrey; Werner, Marian; Oskin, Michael; Burbank, Douglas; and Alsdorf, Douglas, 2007, The shuttle radar topography mission: Reviews of Geophysics, v. 45, no. 2, 33 p. [Also available at http://dx.doi.org/10.1029/2005RG000183.]

Forgy, E.W., 1965, Cluster analysis of multivariate data—Efficiency versus interpretability of classifications: Biometrics, v. 21, p. 768–769.

Fovell, R.G., and Fovell, M.C., 1993, Climate zones of the conterminous United States defined using cluster analysis: Journal of Climate, v. 6, no. 11, p. 2103–2135. [Also available at http://dx.doi.org/10.1175/1520-0442(1993)006%3C2103:CZOTCU%3E2.0.CO;2.]

Francou, Bernard; Vuille, Mathias; Wagnon, Patrick; Mendoza, Javier; and Sicart, Jean-Emmanuel, 2003, Tropical climate change recorded by a glacier in the central Andes during the last decades of the twentieth century—Chacaltaya, Bolivia, 16°S: Journal of Geophysical Research—Atmospheres, v. 108, no. D5. [Also available at http://dx.doi.org/10.1029/2002JD002959.]

Gesch, D.B., Verdin, K.L., and Greenlee, S.K., 1999, New land surface digital elevation model covers the Earth: Eos, v. 80, no. 6, p. 69–70. [Also available at http://dx.doi.org/10.1029/99EO00050.]

Goyal, M.K., and Gupta, Vivek, 2014, Identification of homogeneous rainfall regimes in Northeast Region of India using fuzzy cluster analysis: Water Resources Management, v. 28, no. 13, p. 4491–4511. [Also available at http://dx.doi.org/10.1007/s11269-014-0699-7.]

Grohman, Greg, Kroenung, George, and Strebeck, John, 2006, Filling SRTM voids—The delta surface fill method: Photogrammetric Engineering & Remote Sensing, v. 72, no. 3, p 213–217.

Guth, P.L., 2006, Geomorphometry from SRTM—Comparison to NED: Photogrammetric Engineering & Remote Sensing, v. 72, no. 3, p 269–277.

Hall, O., Falorni, G., and Bras, R.L., 2005, Characterization and quantification of data voids in the shuttle radar topography mission data. IEEE Geoscience and Remote Sensing Letters 2(2):177–181.

Hartigan, J. A. and Wong, M. A. (1979). A K-means clustering algorithm. Applied Statistics 28, 100–108.

Harvey, C. and Eash, D., 1996. Description of BASINSOFT, a computer program to quantify drainage basin characteristics. Proceedings of the Sixteenth Annual ESRI User Conference. Palm Springs, California. http://gis.esri.com/library/userconf/proc96/to100/pap072/p72.htm, accessed December 2010.

Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Wolff, D.B., Adler, R.F., Gu, Guojun, Hong, Yang, Bowman, K.P., and Stocker, E.F., 2007, The TRMM multisatellite precipitation analysis (TMPA)—Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales: Journal of Hydrometeorology, v. 8, no.1, p. 38–55. [Also available at http://dx.doi.org/10.1175/JHM560.1.]

Interagency Advisory Committee on Water Data, Hydrology Subcommittee (1982), Guidelines for Determining Flood Flow Frequency, Bulletin Number 17B. U.S. Department of the Interior Geological Survey Office ow Water Data Coordination Reston, Virginia 22092.

Intergovernmental Panel on Climate Change, 2011, Special report on renewable energy sources: Intergovernmental Panel on Climate Change Technical Report.

Janowiak, J.E., Joyce, R.J., and Yarosh, Yelena, 2001, A real-time global half-hourly pixel-resolution infrared dataset and its applications: Bulletin of the American Meteorological Society, v. 82, no. 2, p. 205–217. [Also available at http://dx.doi.org/10.1175/1520-0477(2001)082%3C0205:ARTGHH%3E2.3.CO;2.]

Kellndorfer, Josef; Walker, Wayne; Pierce, Leland; Dobson, Craig; Fites, J.A.; Hunsaker, Carolyn; Vona, John; and Cluter, Michael, 2004, Vegetation height estimation from shuttle radar topography mission and national elevation datasets: Remote Sensing of Environment, v. 93, no. 3, p. 339–358. [Also available at http://dx.doi.org/10.1016/j.rse.2004.07.017.]

Lin, Gwo-Fong, and Chen, Lu-Hsien, 2006, Identification of homogeneous regions for regional frequency analysis using the self-organizing map: Journal of Hydrology, v. 324, no. 1–4, p. 1–9. [Also available at http://dx.doi.org/10.1016/j.jhydrol.2005.09.009.]

Maidment, D.R., 1993, Handbook of hydrology: McGraw-Hill Inc., 1424 p.

METI (Ministry of Economy, Trade and Industry)/ERSDAC (Earth Remote Sensing Data Analysis Center), NASA (National Aeronautics and Space Administration)/LP DACC (Land Processes Distributed Active Archive Center), and USGS (U.S. Geological Survey)/EROS (Earth Resources Observation and Science Center), (2009), ASTER Global DEM Validation Summary Report, accessed September 24, 2013 at https://lpdaac.usgs.gov/sites/default/files/public/aster/docs/ASTER_GDEM_Validation_Summary_Report.pdf.

Nash, J.E., and Sutcliffe, 1970. River flow forecasting through conceptual models part I — A discussion of principles: Journal of Hydrology, v.10, no. 3 p. 282-290 [Also available at http://dx.doi.org/10.1016/0022-1694(70)90255-6.]

National Aeronautics and Space Administration, 2005a, Shuttle Radar Topography Mission—Instrument: National Aeronautics and Space Administration, accessed September 24, 2008, at http://www2.jpl.nasa.gov/srtm/instr.htm.

National Aeronautics and Space Administration, 2005b, Shuttle Radar Topography Mission—Instruments—Interferometry explained: National Aeronautics and Space Administration, accessed August 6, 2008, at http://www2.jpl.nasa.gov/srtm/instrumentinterferometry.html

Parks, Bruce, and Robert J. Madison, 1985, Estimation of Selected Flow and Water-quality Characteristics of Alaska Streams, Water-Resources Investigations Report 84-4247, (also available on-line at http://ak.water.usgs.gov/Publications/pdf.reps/wrir84.4247.pdf.

Peel, M.C., Finlayson, B.L., and McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification: Hydrology and Earth System Sciences, v. 11, no. 5, p. 1633–1644. [Also available at http://dx.doi.org/10.5194/hess-11-1633-2007.]

Peterson, P., Funk, C.C., Husak, G.J., Pedreros, D.H., Landsfeld, M., Verdin, J.P., and Shukla, S., 2013, The Climate Hazards Group InfraRed Precipitation (CHIRP) with Stations (CHIRPS)—Development and validation [abs.]: American Geophysical Union Fall Meeting, 2013, abstract H33E–1417.

Pham, D.T., Dimov, S.S., and Nguyen, C.D., 2005, Selection of *K* in *K*-means clustering: Proceedings of the Institution of Mechanical Engineers, Part C, Journal of Mechanical Engineering Science, v. 219, no. 1, p. 103–119. [Also available at http://dx.doi.org/10.1243/095440605X8298.]

Rodriguez, E, Morris, C.S., and Belz, J.E., 2006, A global assessment of the SRTM performance: Photogrammetric Engineering & Remote Sensing, v. 72, no. 3, p. 249–260. [Also available at http://dx.doi.org/10.14358/PERS.72.3.249.]

Saha, Suranjana; Moorthi, Shrinivas; Pan, Hua-Lu; Wu, Xingren; Wang, Jiande; Nadiga, Sudhir; Tripp, Patrick; Kistler, Robert; Woollen, John; Behringer, David; Liu, Haixia; Stokes, Diane; Grumbine, Robert; Gayno, George; Wang, Jun; Hou, Yu-Tai; Chuang, Hui-Ya; Juang, H.H., Sela, Joe; Iredell, Mark; Treadon, Russ; Kleist, Daryl; Van Delst, Paul; Keyser, Dennis; Derber, John; Ek, Michael; Meng, Jesse; Wei, Helin; Yang, Rongqian; Lord, Stephen; Van Den Dool, Huug; Kumar, Arun; Wang, Wanqiu; Long, Craig; Chelliah, Muthuvel; Xue, Yan; Huang, Boyin; Schemm, Jae-Kyung; Ebisuzaki, Wesley; Lin, Roger; Xie, Pingping; Chen, Mingyue; Zhou, Shuntai; Higgins, Wayne; Zou, Cheng-Zhi; Liu, Quanhua; Chen, Yong; Han, Yong; Cucurull; Lidia; Reynolds, R.W.; Rutledge, Glenn; and Goldberg, Mitch, 2010, The NCEP climate forecast system reanalysis: Bulletin of the American Meteorological Society, v. 91, no. 8, p. 1015–1057. [Also available at http://dx.doi.org/10.1175/2010BAMS3001.1.]

Schneider, Udo; Becker, Andreas; Finger, Peter; Meyer-Christoffer, Anja; Rudolf, Bruno; and Ziese, Markus, 2011, GPCC full data reanalysis version 6.0 at 0.5°—Monthly land-surface precipitation from rain-gauges built on GTS-based and historic data. [Also available at http://dx.doi.org/10.5676/DWD_GPCC/FD_M_V6_050.]

Searcy, J. K., Flow-duration curves. US Government Printing Office, 1959.

Smakhtin, V.U., 2001, Low flow hydrology—A review: Journal of Hydrology, v. 240, no. 3, p. 147–186. [Also available at http://dx.doi.org/10.1016/S0022-1694(00)00340-1.]

Stooksbury, D.E., and Michaels, P.J., 1991, Cluster analysis of southeastern U.S. climate stations: Theoretical and Applied Climatology, v. 44, no. 3, p. 143–150. [Also available at http://dx.doi.org/10.1007/BF00868169.]

Tasker, G.D., 1982, Comparing methods of hydrologic regionalization: Journal of the American Water Resources Association, v. 18, no. 6, p. 965–970. [Also available at http://dx.doi.org/10.1111/j.1752-1688.1982.tb00102.x.]

Tennant, D.L., 1975, Instream flow regimes for Fish, Wildlife, Recreation and Related environmental resources, U.S. Fish and Wildlife Services, Federal Buildings, Billings, MT. 30 p.

Vernimmen, R.R.E., Hooijer, A., Mamenun, Aldrian, E., and Dijk, A.I.J.M. van, 2012, Evaluation and bias correction of satellite rainfall data for drought monitoring in Indonesia: Hydrology and Earth System Sciences, v. 16, no. 1, p. 133–146. [Also available at http://dx.doi.org/10.5194/hess-16-133-2012.]

Verdin, K.L., 1997, A system for topologically coding global drainage basins and stream networks: Paper presented at the Esri International GIS User Conference, 17th, San Diego, Calif., July 1997 [Proceedings]. [Also available at http://webgis.wr.usgs.gov/globalgis/metadata_qr/metadata/hydro1k_verdin.htm.]

Verdin, K.L., and Verdin, J.P., 1999, A topological system for delineation and codification of the Earth's river basins: Journal of Hydrology, v. 218, nos. 1–2, p. 1–12. [Also available at http://dx.doi.org/10.1016/S0022-1694(99)00011-6.]

Vogel, R.M., Wilson, Ian, and Daly, Chris, 1999, Regional regression models of annual streamflow for the United States: Journal of Irrigation and Drainage Engineering, v. 125, no. 3, p. 148–157. [Also available at http://dx.doi.org/10.1061/(ASCE)0733-9437(1999)125:3(148).]

Water Resources Council, Hydrology Committee (1967), A Uniform Technique for Determining Flood Flow Frequencies, Bulletin No. 15. Water Resources Council, 1025 Vermont Avenue, N.W. Washington, D.C. 20005

Wong, Lana, 2011, A review of transmission losses in planning studies: California Energy Commission, CEC–200–2011–009, 46 p.

Xie, Pingping, and Arkin, P.A., 1997, Global precipitation—A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs: Bulletin of the American Meteorological Society, v. 78, no. 11, p. 2539–2558. [Also available at http://dx.doi.org/10.1175/1520-0477(1997)078%3C2539:GPAYMA%3E2.0.CO;2.]

Appendixes 1–5

Appendix 1.   Geospatial Data and Processing Methods

Table 1–1   U.S. Geological Survey Stream Segment Theoretical Hydropower Assessment geospatial dataset attributes.

[ID, identification; SRTM, Shuttle Radar Topography Mission; m, meter; km2, square kilometer; m3/s, cubic meter per second; Prob, probability; MW, megawatt; --, not appicable]

Field name	Field alias	Description
ID_REGION	Primary Basin ID	Identification number for primary analysis drainage basins
primary_name	Primary Basin	Name of primary analysis drainage basins
ID	Sub-basin ID	Identification number of sub-basins
subbasin_name	Sub-basin Name	Name of sub-basins
HYDROID	Stream ID	Stream run identification. Unique identification for a stream run within a primary basin. A run is defined as the strength of a stream/river between two confluences.
SEGID	Segment ID	Segment identification. Unique identification for each segment within a stream run
SRTMVDS	SRTM void intersect	A binary value indicating if a stream segment is intersecting an SRTM data void. 0 indicates that the segment does not intersect and a 1 indicates that it does intersect a void.
SELEV	Upstream elevation (m)	Upstream elevation in meters of the stream segment
EELEV	Downstream elevation (m)	Downstream elevation in meters of the stream segment
DROP_	Segment elevation drop (m)	Drop in elevation between the upstream and downstream segment locations
FACKM	Flow accumulation (km^2)	Measurement of upstream drainage area in square kilometers contributing to that segment
PACC	Mean annual surface precipitation	Mean annual surface area weighted precipitation in mm.
Q_AVG	Average Stream Flow (Q)	Average annual modeled stream flow in m^3/sec
QD	Design Flow	Maximum flow a hydropower system can accommodate
Q40	Annual Flow with 40 percent Prob Exceedance	--
Q60	Annual Flow with 60 percent Prob Exceedance	--
Q70	Annual Flow with 70 percent Prob Exceedance	--
Q90	Annual Flow with 90 percent Prob Exceedance	--

Table 1–1 U.S. Geological Survey Stream Segment Theoretical Hydropower Assessment geospatial dataset attributes.—Continued

[ID, identification; SRTM, Shuttle Radar Topography Mission; m, meter; km2, square kilometer; m3/s, cubic meter per second; Prob, probability; MW, megawatt; --, not appicable]

P_AVG	Average Theoretical Power Potential (MW)	Modeled average annual theoretical power output in MW
P40	Theoretical Power Potential (MW) Exceedance 40 percent	--
P60	Theoretical Power Potential (MW) Exceedance 60 percent	--
P70	Theoretical Power Potential (MW) Exceedance 70 percent	--
P90	Theoretical Power Potential (MW) Exceedance 90 percent	--
Available	Defines if segment with hydropower is available for development	Defines if segment with hydropower is available for development (0=No, 1=Yes)
inf1q2d	Influence	Major contributing factor for P (1 = Discharge; 2 = Drop in elevation)
Protareas	Protected Areas	Stream segments falling with a protected area with in Bolivia
Country	Country	Country that the segment is within. If the stream defines a border of Bolivia is it attributed "BOL-BD", Stream with Bolivia are "BOL"

SRTM DEM Hydrologic Conditioning Model

With many hydrologic modeling approaches using digital elevation models (DEM), a level of drainage reinforcement or conditioning is required, particularly with a surface-modeled DEM such as Shuttle Radar Topography Mission (SRTM) (Grohman and others, 2006). One primary objective of conditioning elevation datasets for hydrologic modeling is to remove influential anomalies and manipulate the data to better represent ground conditions. This section walks through the steps (fig. 1–1) of the conditioning process in preparation for a hydropower assessment. This workflow follows four primary steps. Step 1 evaluated the raw drainage network output, modifying stream channels that significantly deviated from reference Bolivian Servicio Nacional de Meteorología e Hidrología's 1:24,000 digital hydrography and Landsat 8 imagery. These modifications were applied to the underlying elevation pixels of the DEM. This manipulation was done using the Agree algorithm (Hellweger, 1997). Step 2 applied these modifications to the drainage model, and the subsequent output was again evaluated for consistency against reference data. If deviations persisted, steps 1 and 2 were repeated until the output was consistent with the reference data. When the drainage network or synthetic streams were consistent, the original DEM was conditioned using this dataset derived from step 3. During this step, the DEM was modified along all the synthetic stream channels, eliminating or reducing the influence of data anomalies. This step in the condition process is often referred to as "stream burning." In the last step (step 4), the hydrologic-improved DEM was used to create the final drainage network used in the hydropower assessment. Also included in the final step was splitting the stream channels into 1-kilometer (km) segments, applying a unique identifier to each segment and calculating its head.
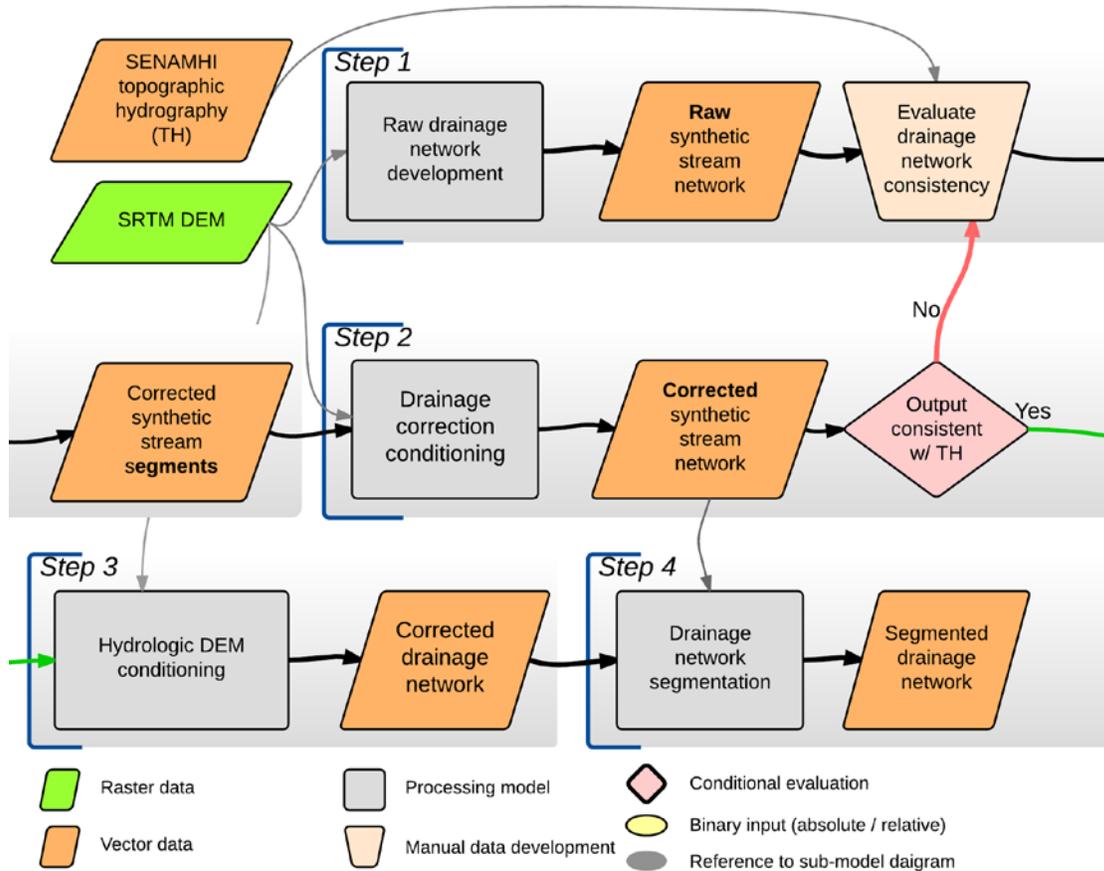
Figure 1–1.   Flowchart of the hydropower potential methodology. [SENAMHI, Servicio Nacional de Meteorología e Hidrología; SRTM, Shuttle Radar Measuring Mission; DEM, digital elevation model; TH, topographic hydrography.

Raw Drainage Network Evaluation (Step 1)

To create the initial drainage network from the raw (unconditioned) DEM, the standard synthetic stream delineation method was processed in Esri's ArcGIS that applies the algorithms defined by O'Callaghan and Mark (1984) and Jenson and Domingue (1988). The workflow used ArcGIS and the following geographic information system processing tools: Hydrologic-Fill, Flow Direction, and Flow Accumulation. Using the resulting Flow Accumulation raster dataset, the steps in figure 1–2 were applied to extract the appropriate stream density and convert from a raster to a vector-based dataset.

Drainage Network Conditioning (Step 2)

Once the synthetic stream network from step 1 is completed, it is evaluated against the reference hydrography dataset and other ancillary data such as Landsat imagery. Where the synthetic streams had substantial deviation from reference sources, a new burn vector dataset was created to digitize the appropriate stream channels for the extent of the deviation. This evaluation is done for the entire synthetic stream

dataset. When the evaluation was completed, the process of stream burning was applied to the raw DEM using the Agree algorithm (Hellweger, 1997), which reduced the DEM's pixel values by 200 meters along the burn line segments. Conditioning the DEM in this manner, the drainage network model was forced to follow a subscribed path, mitigating DEM anomalies along the channel (fig. 1–3). If the subject basin being analyzed was not closed or was farther upstream, then weighting or raising the upstream "edge pixels" was required to enforce the correct flow direction modeling. These upstream edge pixels were elevated to a height that ensured the flow direction model would closely represent the natural downstream flow direction (fig. 1–3). With the weighting applied, new conditioned synthetic streams were created as described in the Raw Drainage Network Evaluation section. This step was an iterative process of manipulating pixels in the DEM, which tended to generate minor anomalies around subject segments, requiring this step to be repeated until a best fit was reached. The frequencies of iterations tended to be related to the number of burn modifications identified. There was no quantifying mechanism to signify when this step was complete; that decision was subjective, and its completion was determined by the person managing the workflow.
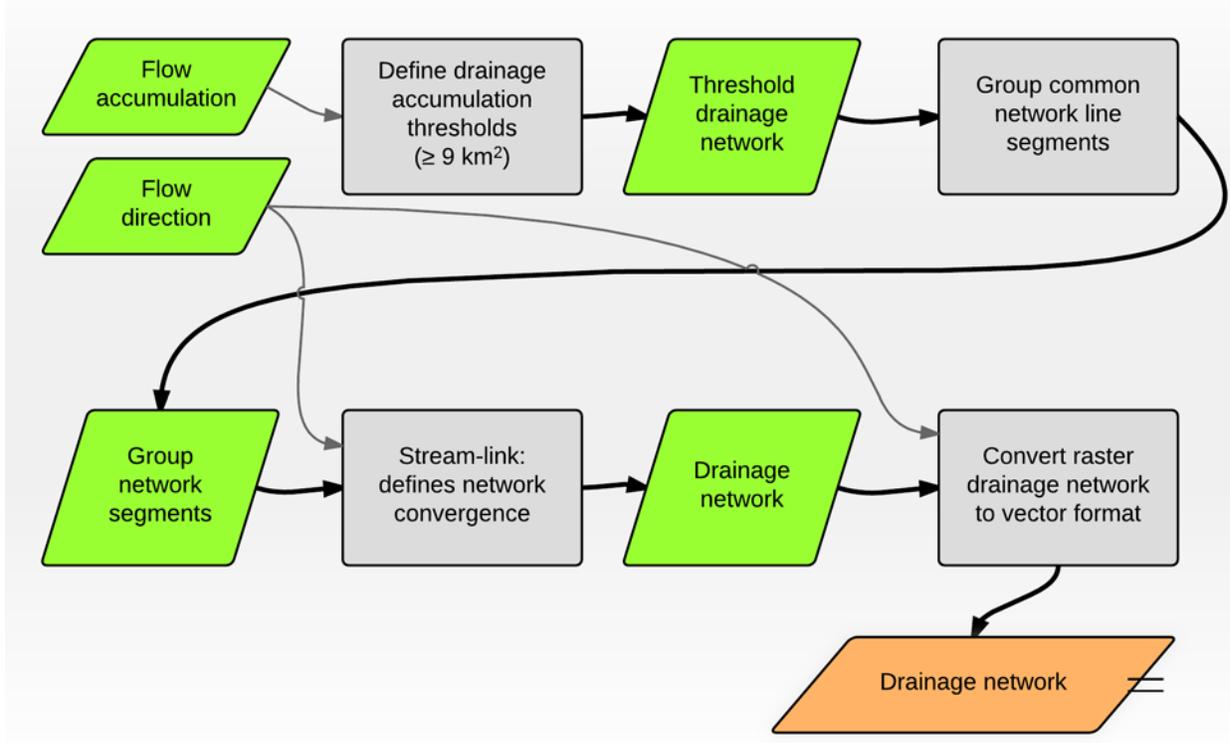
Figure 1–2.   Synthetic drainage network delineation [km², square kilometer].
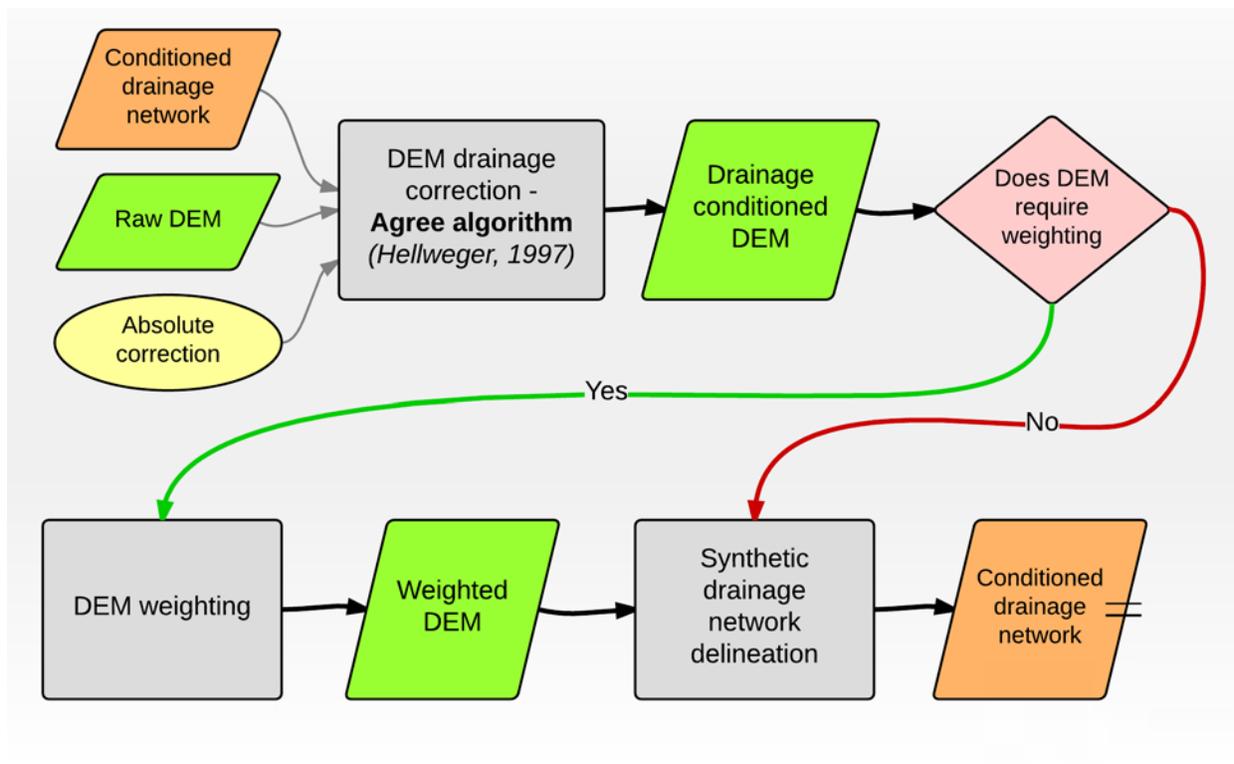

Figure 1–3.   Drainage conditioning model [DEM, digital elevation model].
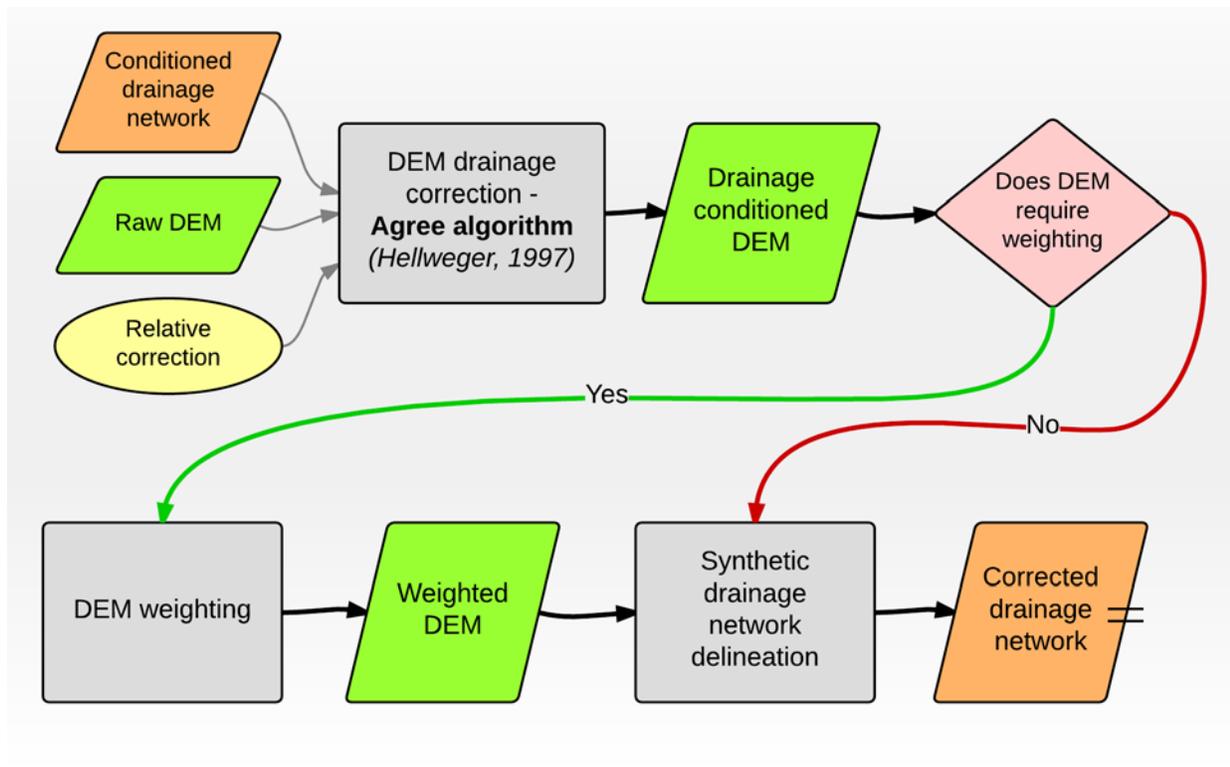
Figure 1–4.   Final drainage conditioning model [DEM, digital elevation model].

## Final Hydrologic Digital Elevation Model Conditioning (Step 3)

When the corrected drainage network was completed, a final burn was applied to the original raw DEM. In the prior burning process, only selected deviating stream segments were burned into the DEM in a deep relative negative depth (greater than 200 meters). In this step, the entire synthetic stream network was burned into the DEM but at a shallower negative depth (between 25 and 75 meters). These two modifications to the process were necessary for the hydropower assessment process. This approach of a shallow relative burn of the entire drainage network allowed for the correction of data anomalies along the stream channels while still maintaining the network's relative elevation. Reducing the elevation of the underlying pixels to a depth from 25 to 75 meters relative to the original value allowed the Fill algorithm to correct sinks along the channel and still enforce the model to follow a defined path when required (fig. 1–4). This relative burn was applied using the same Agree algorithm (Hellweger, 1997) as before. With this newly created conditioned DEM, a new conditioned drainage network was generated using the same method from the Raw Drainage Network Evaluation section.

## Drainage Network Segmentation and Drop Extraction (Step 4)

In the final step of the synthetic stream processing, the stream channels were divided into 1-km segments to prepare for the hydropower assessment. The segmenting process started at the mouth or confluence and progressed upstream (fig. 1–5). This process produced segments less than 1 km in two cases: at the headwaters of the synthetic network and immediately downstream from confluences. Another portion of the segments, less than 0.2 percent, are greater than 1 km. These segments occurred at intersections of SRTM voids. In these cases, the data used to fill the SRTM voids are less accurate than the surrounding native SRTM, so the head measurements were calculated using the upstream and downstream elevations outside these voids. A segment could extend beyond 1 km at 1-km increments until both the upstream and downstream location criteria fell outside a void. The one exception of a head measurement taken inside data voids was when a confluence occurred within a void. With the 1-km growth increment condition, segments spanned several voids and were as long as 16 km (fig. 1–5).

Once the segmenting was completed, each segment was assigned a unique identifier and validated to ensure that its network attributes coincided with its respective neighbors (that is, flow direction). After the validation was completed, the head measurement was calculated by extracting the difference between the segment's farthest up and downstream elevation values. This head or drop value was then applied to the attribute table (fig. 1–6).
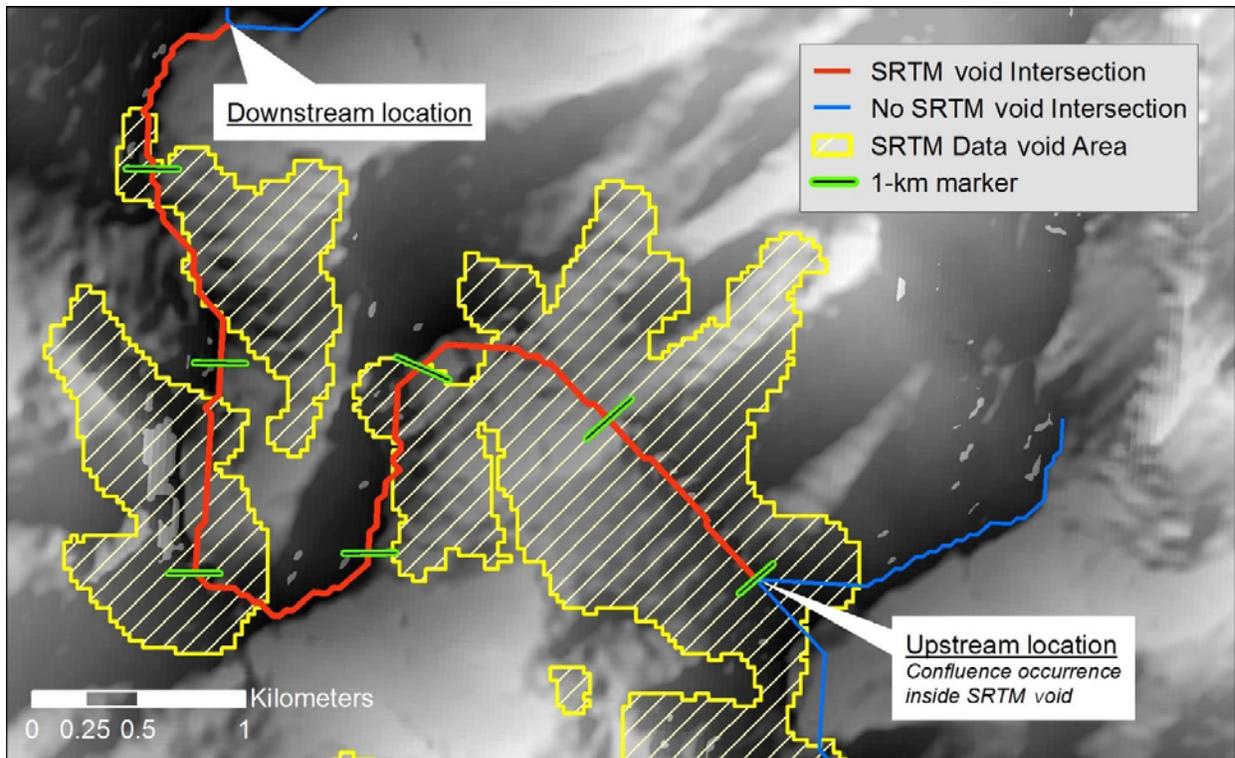
Figure 1–5. The 7-kilometer (km) stream segment highlighted in red transects three Shuttle Radar Measuring Mission (SRTM) void polygons. The green markers show where the stream would have been segmented if the stream had not intersected an SRTM data void; consequently, the segment crosses these voids, each 1-km segment is aggregated in a single 7-km segment with a head calculation determined from the farthest up and downstream locations on the new 7-km aggregated segment.
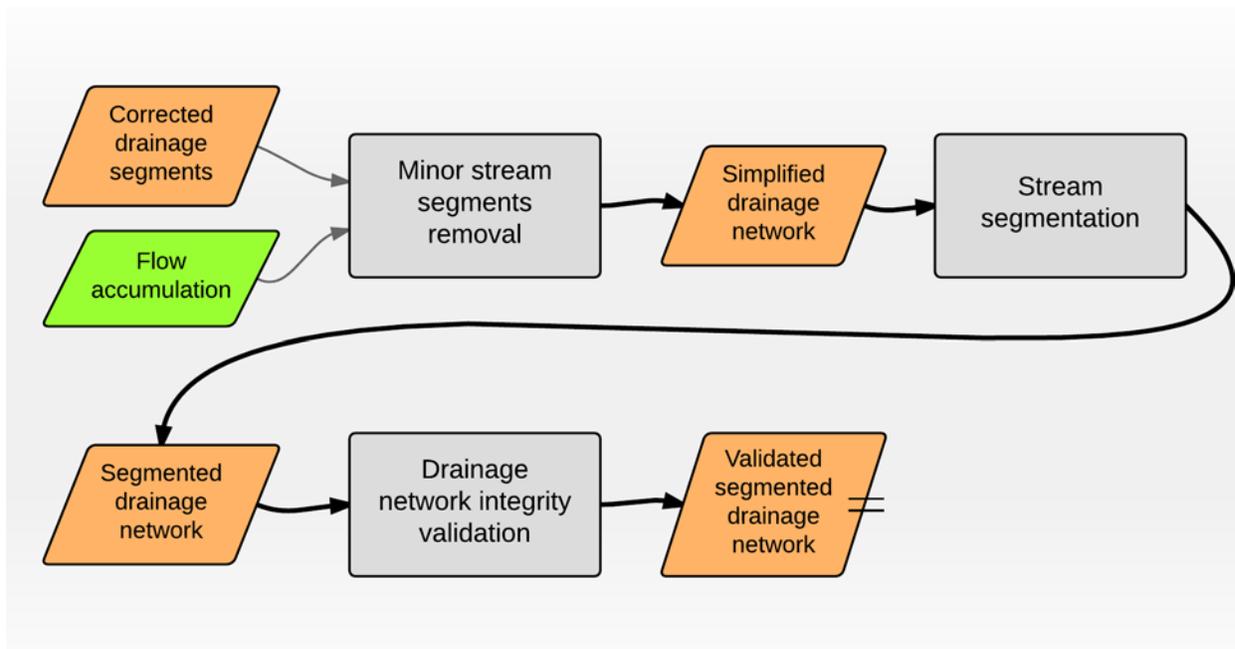
Figure 1–6. Drainage network segmentation model.

References Cited

Grohman, Greg, Kroenung, George, and Strebeck, John, 2006, Filling SRTM voids—The delta surface fill method: Photogrammetric Engineering & Remote Sensing, v. 72, no. 3, p 213–217.

Hellweger, F., 1997, AGREE—Digital Elevation Model Surface Reconditioning System. Accessed July 22, 2016, at http://www.caee.utexas.edu/prof/maidment/gishydro/ferdi/research/agree/agree.html.

Jenson, S.K., and Domingue, J.O., 1988, Extracting topographic structure from digital elevation data for geographic information system analysis: Photogrammetric Engineering & Remote Sensing, v. 54, no. 11, p. 1593–1600.

O'Callaghan, J.F., and Mark, D.M., 1984, The Extraction of Drainage Networks from Digital Elevation Data: Computer Vision, Graphics, and Image Processing, v. 28, no. 3, p. 323–344.

Appendix 2.   Validation and Bias Correction of Rainfall Datasets

Figures 2–1 to 2–25 can be located here: http://dx.doi.org/10.3133/ofr20161156.

Appendix 3.   *K*-Mean Cluster Analysis

Tables 3–1 to 3–2 can ne located here: http://dx.doi.org/10.3133/ofr20161156.

Appendix 4.   Station Flow Exceedance Probability Curve

**Figures 4–1 to 4–5 can be located here:**   http://dx.doi.org/10.3133/ofr20161156.

Appendix 5.   Procedure to Identify and Classify 1-Kilometer Stream Segments in Bolivia into Harnessed/ Exploited and Available/Unharnessed

This annex provides a brief summary of how we identified and classified 1-km stream segments into two classes: harnessed and available segments for hydropower computation. On a case-by-case basis we analyzed all of the 24 existing hydropower sites to understand the engineering aspects to harness streamflow. Based on the understanding developed from this analysis, we identified streams that have already been harnessed. Finally, such harnessed streams were ignored in the computation of available theoretical hydropower in Bolivia. A case by case analysis of the hydropower sites is presented below.

## Existing Hydropower Project—Zongo Hydropower Cascade Schema

The Zongo Hydropower project is a cascade of 10 hydropower projects. We summarized the engineering design based on the information provided in the National Inventory of Dams, 2010. The summary of engineering design for these hydropower projects is provided in figure 5–1. There are several canals that are built to connect several stream segments upstream from Botijlaja hydropower site. After the water is released back into the stream from Botijlaja site, the water is extracted out of the natural stream course into a tunnel. The provision of a tunnel enables continuous slope and flow of water through the series of projects from Zongo site. There are a few more canal systems, such as between Cuticucho and Santa Rosa and into Chururaqiji sites, that divert water from the streams into the tunnel. These diversions often enable increase in flow rate in the tunnel and subsequently facilitate greater power production.

This is a simple case where most the streams and river channels upstream from the final Huaji hydropower site are harnessed. The theoretical hydropower potential of the streams that fall under the Zongo hydropower project is 1,038 megawatts (MW).
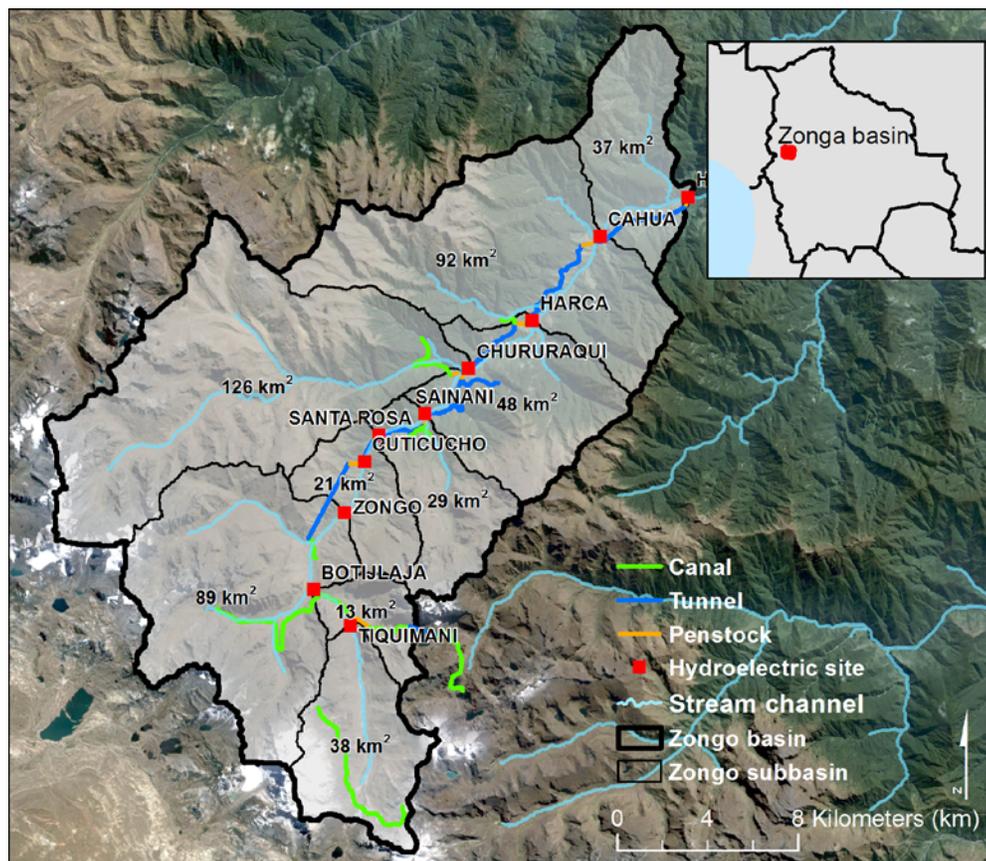
Figure 5–1.   Distribution and engineering design of Zongo cascade of hydropower schema (summarized from the National Inventory of Dams, 2010). All the streams falling within the boundary of the watershed were considered as harnessed [km², square kilometer].
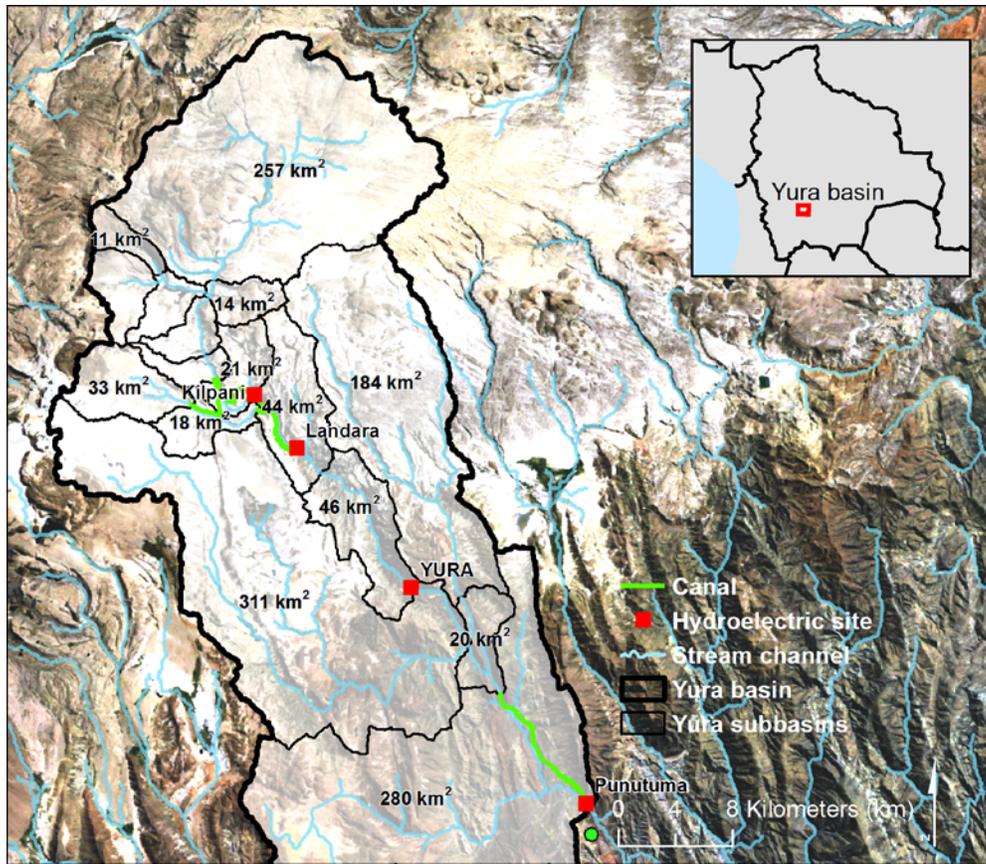
Figure 5–2.   Distribution and engineering design of Kilpani-Punutuma (Yura) cascade of hydropower schema (summarized from the National Inventory of Dams, 2010). All the streams falling within the boundary of watershed were considered as harnessed [km², square kilometer].

Existing Hydropower Project—Kilpani-Punutuma Hydropower Schema

The location and distribution of Kilpani-Punutuma (Yura) hydropower site schema is presented in figure 5–2. Only Kilpani and Landara projects are based on drainage diversions from the upstream segments. These diversions slightly increase the streamflow and hydraulic head to enable power production. Although we do not have information on the design of Yura and Putunuma ("Available Theoretical Hydropower Potential in Bolivia" section, table 14), the National Inventory of Dams (2010) provides information on possible drainage diversion for Punutuma (fig. 5–2). Whatever the case, there seems to be a great deal of engineering of streams within the basin; hence, based on figure 5–2, we identified the streams classified as harnessed. The total theoretical hydropower potential of the harnessed streams is estimated as 22 MW.

Existing Hydropower Project—Corani Hydropower Schema

The Corani hydropower schema (fig. 5–3) represents a complex engineering design. This design involves diverting waters from streams that are otherwise flowing into a different stream network. Figure 5–3 shows the use of canals that reroute water from different stream networks in the upstream area. This rerouted water is diverted to flow into Lake Corani, from where the water flows through its natural outlet into the Corani hydropower project site. This increase in the Corani Lake reserves (possible because of water diversion through canals from the upstream segments) enables increased streamflow and subsequent increase in power production. After the water passes through the Corani hydropower site, water is diverted through canals (along the contour) into the downstream Santa Isabel project site. Based on the information from the engineering design and canal/tunnel network, we have identified corresponding streams as harnessed. The total theoretical hydropower potential of these harnessed streams is 142 MW.
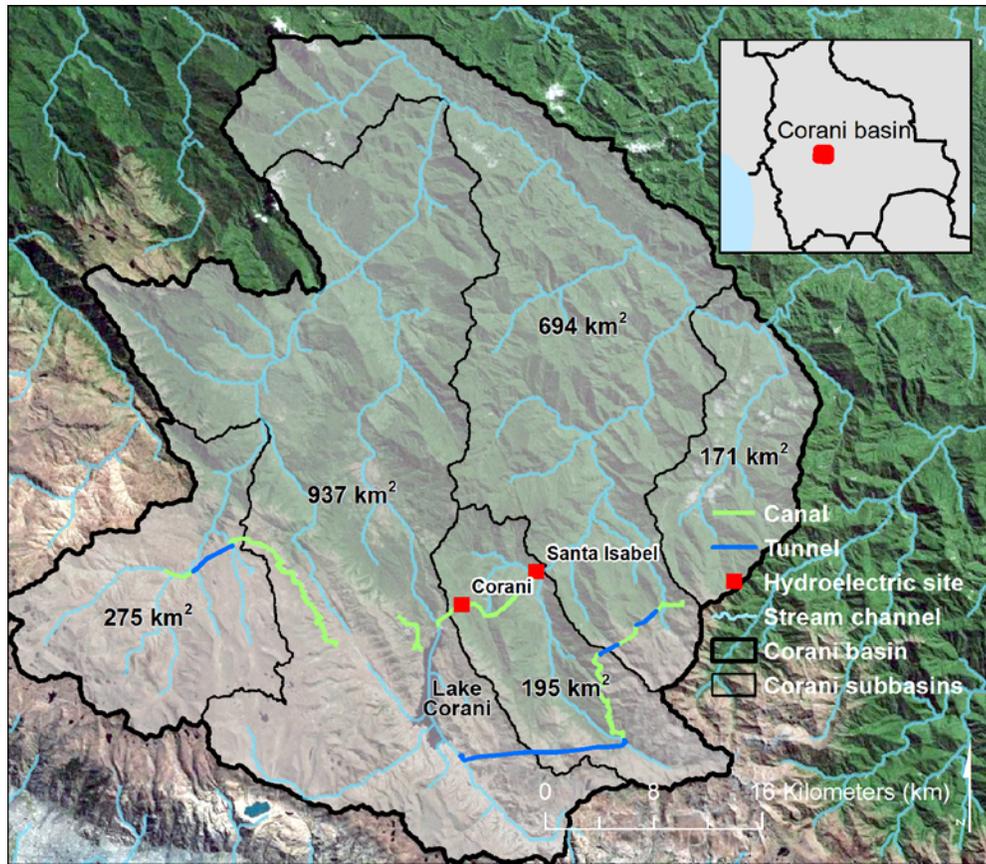
Figure 5–3. Distribution and engineering design of Corani hydropower schema (summarized from the National Inventory of Dams, 2010). All the streams falling within the boundary of the watershed were considered as harnessed [km², square kilometer].

Existing Hydropower Project—Kanata Hydropower Schema

Kanata hydropower is another complex scheme involving intercomplex engineering and interbasin water transfer. Kanata hydropower is actually close to Cochabamba, but the water is drawn from the basin in the north using a series of diversions and water lifting structures. The drainage basin in the north of Kanata hydropower site is shown in figure 5–4. There are no streams identified in the figure that connect the diversion sites because only streams with flow accumulation greater than 10 km² are shown. Based on the information obtained from figure 5–4, we have identified the streams that could possibly be classified as harnessed. The total theoretical hydropower potential of the streams identified as harnessed is only 22 MW.

Existing Hydropower Project—Miguillas Hydropower Schema

The Miguillas hydropower schema is presented in the "Identifying Major Contributing Factors to the Hydropower Potential" section, figure 41. The engineering design appears to be complex and involves a series of canals (green lines in figure 5-5) that interconnect the natural streams. These canals manipulate the natural streamflow favoring hydropower production. Based on the information obtained, we have identified streams as harnessed that are in purple. The total theoretical hydropower potential of the streams identified as harnessed is 15.2 MW

Existing Hydropower Project—Quehata, Chinata and San Jacinto Hydropower Sites

Because of nonavailability of information on engineering designs of Quehata, Chinata, and San Jacinto hydropower projects, we applied a simple approach to identify streams as harnessed where all the upstream segments are identified as harnessed. The total theoretical hydropower potential of these streams was 86 MW.
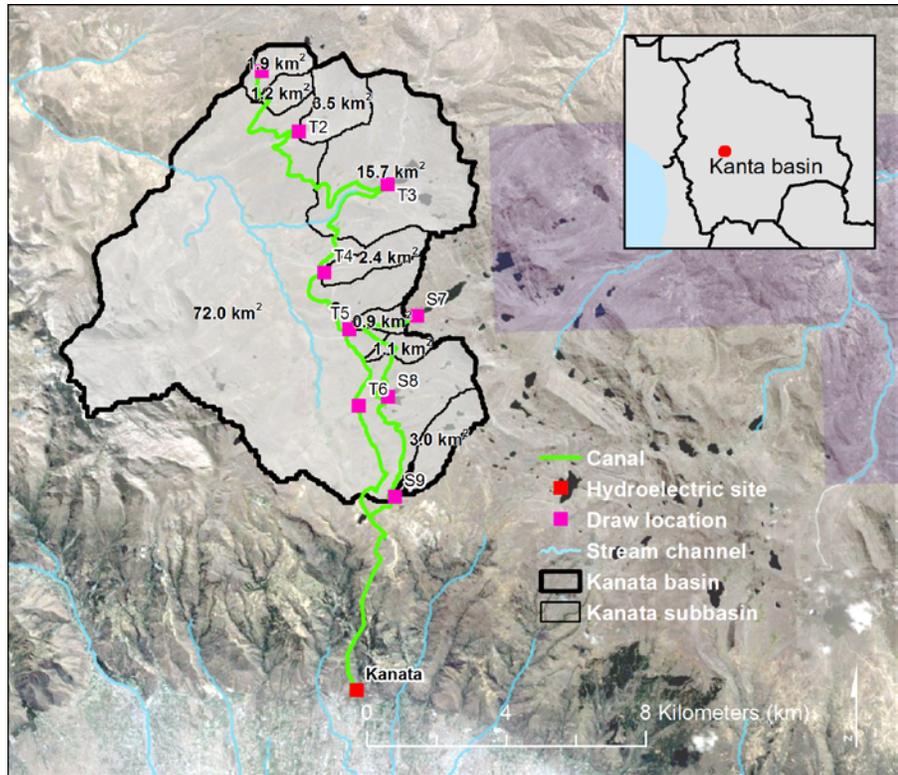
Figure 5–4. Distribution and engineering design of Kanata hydropower schema (summarized from the National Inventory of Dams, 2010). All the streams falling within the boundary of the watershed were considered as harnessed [km², square kilometer].
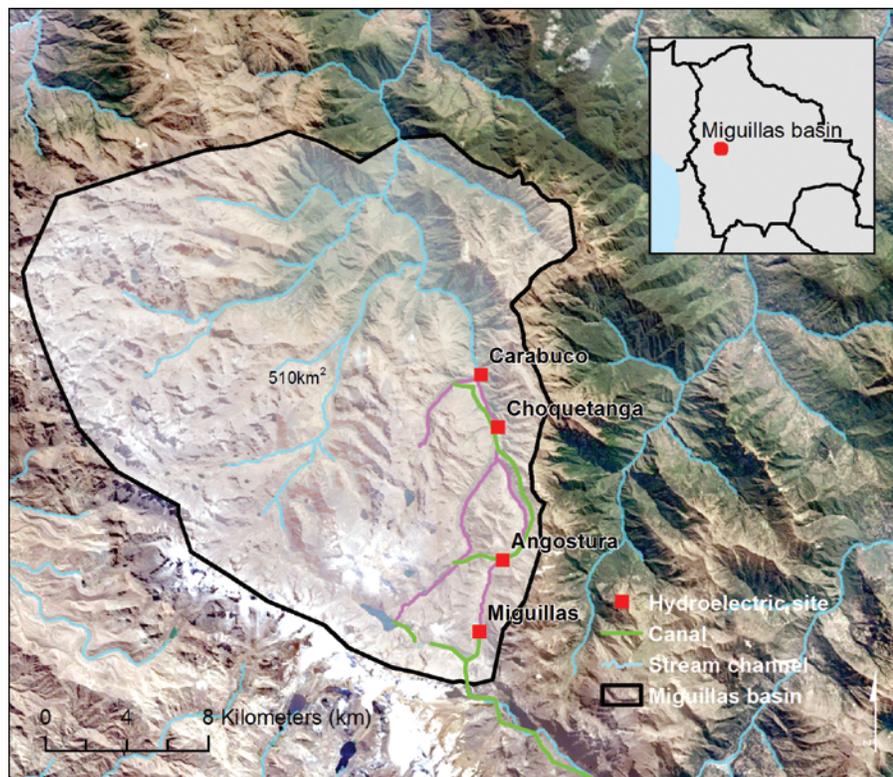
Figure 5–5. Distribution and engineering design of Miguillas hydropower schema (summarized from the National Inventory of Dams, 2010). All the streams falling within the boundary of the watershed were considered as harnessed [km², square kilometer].

Reference Cited

National Inventory of Dams, 2010, (Inventario Nacional de Presas, Bolivia 2010), Ministry of Environment and Water, Vice Ministry of Water Resources and Irrigation, Report available online: http://www.bivica.org/upload/presas-inventario_a.pdf.