

# SCIENTIFIC DATA

OPEN

SUBJECT CATEGORIES

» Water resources

» Hydrology

» Environmental sciences

## A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950–2013

Ben Livneh<sup>1,2</sup>, Theodore J. Bohn<sup>3</sup>, David W. Pierce<sup>4</sup>, Francisco Munoz-Arriola<sup>5</sup>, Bart Nijssen<sup>6</sup>, Russell Vose<sup>7</sup>, Daniel R. Cayan<sup>4</sup> & Levi Brekke<sup>8</sup>

Received: 23 April 2015

Accepted: 24 July 2015

Published: 18 August 2015

A data set of observed daily precipitation, maximum and minimum temperature, gridded to a  $1/16^\circ$  (~6 km) resolution, is described that spans the entire country of Mexico, the conterminous U.S. (CONUS), and regions of Canada south of  $53^\circ$  N for the period 1950–2013. The dataset improves previous products in spatial extent, orographic precipitation adjustment over Mexico and parts of Canada, and reduction of transboundary discontinuities. The impacts of adjusting gridded precipitation for orographic effects are quantified by scaling precipitation to an elevation-aware 1981–2010 precipitation climatology in Mexico and Canada. Differences are evaluated in terms of total precipitation as well as by hydrologic quantities simulated with a land surface model. Overall, orographic correction impacts total precipitation by up to 50% in mountainous regions outside CONUS. Hydrologic fluxes show sensitivities of similar magnitude, with discharge more sensitive than evapotranspiration and soil moisture. Because of the consistent gridding methodology, the current product reduces transboundary discontinuities as compared with a commonly used reanalysis product, making it suitable for estimating large-scale hydrometeorologic phenomena.

<b>Design Type(s)</b>	observation design • time series design
<b>Measurement Type(s)</b>	surface layer temperature • surface layer precipitation
<b>Technology Type(s)</b>	weather stations
<b>Factor Type(s)</b>	
<b>Sample Characteristic(s)</b>	contiguous United States of America • atmospheric boundary layer • Canada • Mexico

<sup>1</sup>Cooperative Institute for Research in Environmental Science (CIRES), University of Colorado, Boulder, 216 UCB, Boulder, Colorado 80309, USA. <sup>2</sup>NOAA Earth System Research Laboratory Physical Sciences Division, 325 Broadway, Boulder, Colorado 80309, USA. <sup>3</sup>School of Earth and Space Exploration, Arizona State University, Tempe, AZ, 781 E. Terrace Mall, Tempe, Arizona 85287-6004, USA. <sup>4</sup>Climate, Atmospheric Sciences, and Physical Oceanography, Scripps Institution of Oceanography, La Jolla, California 92093-0224, USA. <sup>5</sup>Department of Biological Systems Engineering, University of Nebraska-Lincoln, 246L. W. Chase Hall P. O. Box 830726, Lincoln, Nebraska 68583-0726, USA. <sup>6</sup>Department of Civil and Environmental Engineering, University of Washington, Box 352700, Seattle, Washington 98195, USA. <sup>7</sup>NOAA/National Climatic Data Center, 151 Patton Avenue, Asheville, North Carolina 28801, USA. <sup>8</sup>Water and Climate Research Coordinator, U.S. Bureau of Reclamation. Correspondence and requests for materials should be addressed to B.L. (email: ben.livneh@colorado.edu).

## Background & Summary

Observation-based meteorological data sets offer insights into changes to the hydro-climatic system by diagnosing spatio-temporal characteristics<sup>1</sup> and providing a historical baseline for future projections<sup>2</sup>. Spatially consistent transboundary hydroclimatic data have the potential to reduce cross-border water disagreements by enabling equitable allocation and access to water resources<sup>3</sup>. Yet, creating multi-decadal data sets that span international boundaries presents a unique challenge, specifically to avoid discontinuities and artifacts related to data availability, collection protocols, and quality. Such artifacts can subsequently impact key transboundary water-related estimates, such as fish stocks<sup>4</sup>, water sharing agreements<sup>5</sup>, historical hydroclimatic assessments<sup>6</sup>, hydrologic forecasting<sup>7,8</sup>, and regional hydroclimatic phenomena such as the North American Monsoon (NAM<sup>9–11</sup>). Here, we applied a previously published methodology<sup>12</sup> to create a sub continental daily data set of gridded station precipitation and maximum and minimum temperatures with minimal transboundary discontinuity.

Although remote sensing products offer estimates of other hydro-climatic quantities without transboundary discontinuities, e.g., evapotranspiration (ET), terrestrial water, and soil moisture, such estimates are limited in their potential to close the land surface water budget<sup>13,14</sup>, may be unconstrained by available water<sup>15</sup>, and generally have temporal frequencies that may be too coarse for daily applications<sup>16,17</sup>. While satellite precipitation products are available at daily or finer timescales, they tend to have large uncertainties<sup>18</sup> and lack record lengths of more than 20 years needed to assess low recurrence phenomena such as extreme hydrometeorological events and the effects of ocean-atmosphere oscillations.

Precipitation and near-surface temperature records can be used to drive land surface models that explicitly close the water balance<sup>19</sup>. For example<sup>20,21</sup> global gridded observations exist to this end, but at coarse spatial resolution (0.5°). High-resolution datasets are available from re-analyses (0.125° for NLDAS2<sup>22</sup>), although transboundary or temporal discontinuities emerge through incorporation of satellite data (0.3° for NCEP/NCAR reanalysis<sup>23</sup>). Without a single, consistent product, application studies may need to develop merging techniques, which may reveal (or add) biases across precipitation products that manifest as temporal discontinuities in derived model outputs, impairing detection of hydrometeorological phenomena such as drought<sup>24,25</sup>.

Estimating precipitation at ungauged locations is complicated by topographic features that drive orographic precipitation<sup>26–28</sup>. Water resources in the western U.S., Canada, and Mexico depend critically on mountain precipitation<sup>29,30</sup>. Yet most precipitation measurements are made at stations located at lower elevations. Naively interpolating values between lower-elevation stations across the intervening topography would systematically misrepresent the true precipitation field. Adding an orographic adjustment greatly improves the quality of precipitation estimates subject to topographic influences<sup>31</sup>.

The dataset presented in this article was developed for use in climate downscaling applications for the U.S. Bureau of Reclamation, although applications to drive land models and direct analyses are also anticipated. Relative to its methodological predecessors, this data set (herein L15) extends the domain of Livneh *et al.* (2013 (ref. 12): herein L13) into Mexico and Canada, and is both a slightly larger domain (i.e. all of Mexico south of 25° N) and finer spatial resolution (1/16° versus 1/8°) than Maurer *et al.* (2002 (ref. 32): herein M02), and incorporates a topographic adjustment over the entire domain unlike existing Mexico-only products<sup>33,9</sup>. This article discusses: (i) the newly incorporated (Mexican) station data and quality control procedures applied; (ii) implications of topographic adjustments and selection of the climate normal period on precipitation and hydrology, as represented by a hydrologically-based land surface model; (iii) an examination of spatio-temporal discontinuities relative to a reanalysis product over the North American Monsoon (NAM) region.

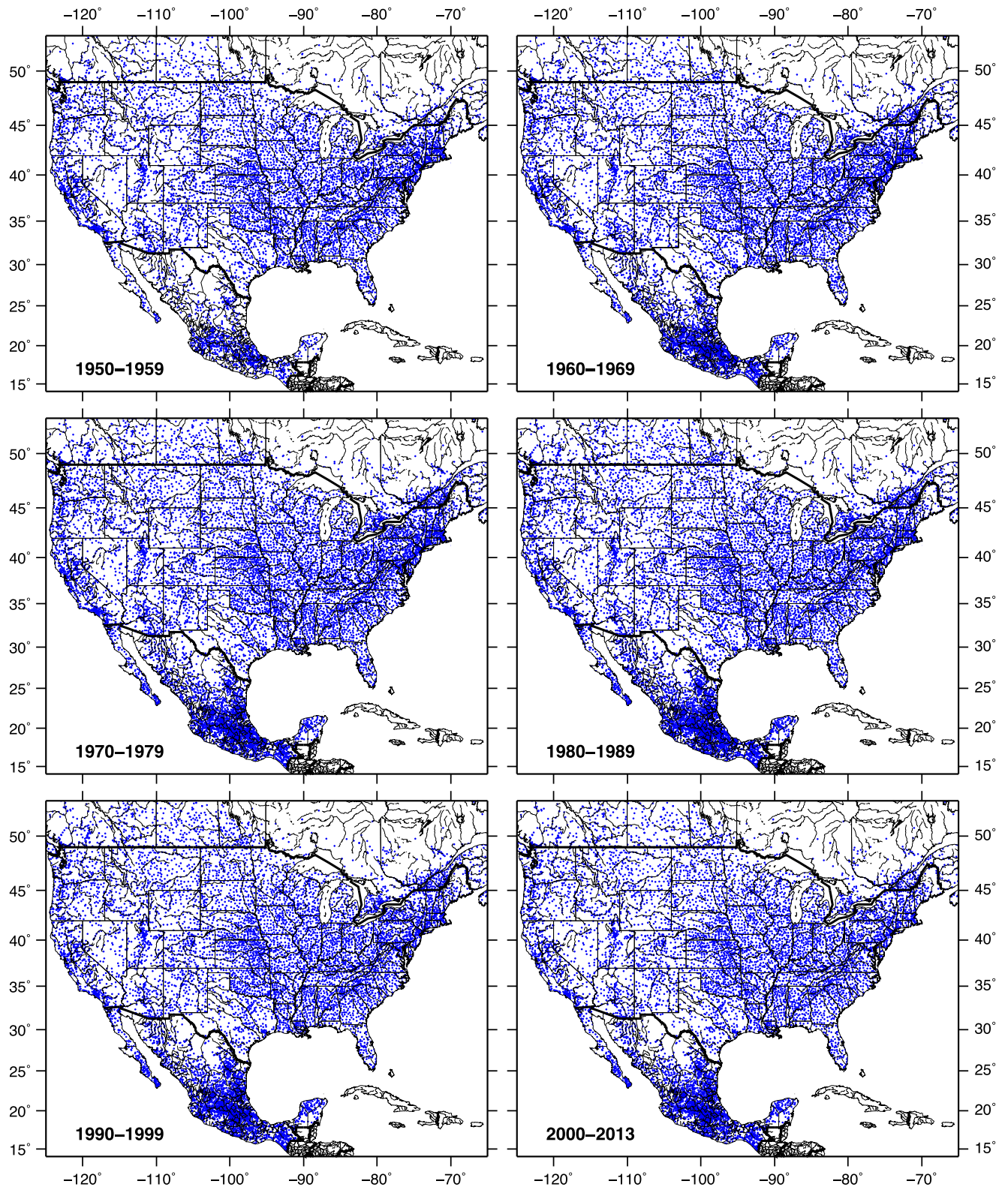
## Methods

### Station data and gridding process

As described in L13, the gridding procedure used the SYMAP algorithm<sup>34</sup> to follow the methodology originally used by M02. As in L13 (and M02), the stability constraint requiring a minimum of 20 years of data was applied only to CONUS and Canadian stations. In contrast, given the relative paucity of station data in Mexico, those data were screened following the procedure put forth in a previous gridded Mexican data set<sup>33</sup> requiring a minimum of 50 valid days of data in any given year for a station to be included. Maps of decadal station density illustrate the temporal evolution of data availability (Fig. 1). Areas of low station density include central and eastern Canada, the U.S. southwest, and northern Mexico.

### Orographic scaling

Consistent with L13, two adjustments were made to the gridded meteorological fields to account for the effects of topography. First, temperatures were lapsed with elevation at a constant rate of 6.5 °C/km. Second, precipitation was scaled based on existing estimates of monthly climatological precipitation that were computed taking topographical effects into account. We used PRISM for this purpose within the CONUS, and the Vose *et al.*<sup>35</sup> extension to an older climatology<sup>36</sup> elsewhere (herein V14). The key steps in precipitation scaling were therefore:



**Figure 1.** Station density illustration. Blue dots show stations included in the gridding process, by decade.

1. Starting with the SYMAP-interpolated data, compute the monthly mean precipitation at every point over a 30-year period (1981–2010, coincident with PRISM and V14 climate normals);
2. For each of the 12 months and for each point, compute the ratio between the topographically-aware PRISM or V14 climatological value and the climatological value from the interpolated data. Both

- PRISM and V14 (1/20° native resolution) were conservatively remapped to the 1/16° gridded resolution prior to taking ratios;
3. Multiply all daily precipitation values at that point and in that month by the ratio computed in step 2. This is done for the entire data record (1950–2013).

Features unique to this analysis relative to L13 were (i) orographic adjustments to locations outside CONUS, and (ii) the climatological reference period chosen for scaling within CONUS.

The older climatology<sup>36</sup> for North America was updated following the methods of V14. Climate normals were computed for each station for the 1981–2010 base period. The irregularly spaced station normals were then interpolated to a 1/20° grid using trivariate thin-plate smoothing splines with latitude, longitude, and elevation as predictors<sup>37</sup>. Thin-plate splines are well suited for a large domain such as North America because the relationship between the dependent and predictor variables can vary in space, which facilitates the reconstruction of complex geographical patterns<sup>38</sup>.

Since the V14 reference period is 1981–2010, the same PRISM normal period was chosen for CONUS for transboundary consistency. This is in contrast to the 1961–1990 period used in L13 (and M02). The selected PRISM period overlaps with the monitoring era (post-1979) of critical precipitation observations at high elevations by the National Resource Conservation Service SNOw TELelemetry (SNOTEL; where temperatures were not used).

### Quality control

Descriptions here are largely restricted to the Mexican data and issues beyond those previously addressed by L13 for CONUS and Canadian stations. The first issue involved spurious precipitation data within the Mexican station record, with periods of near-constant non-zero precipitation, an example shown in Fig. 2. Erroneous values were identified and flagged on a monthly basis for each station by computing the monthly coefficient of variation,  $CV_i$  (the ratio of the standard deviation of daily values to their mean for month  $i$ ) over the entire period of record; and its climatological average value,  $CV_m$  ( $m = 1$  to 12); and removing months with  $CV_i < 0.18 CV_m$ , which was determined empirically from a training set of 25 station records, each longer than 15 years, taken from 7 states. In all cases examined, months flagged as spurious were clear outliers relative to the distribution of all  $CV$  values for the given month, with  $CV_i$  values that fell more than two standard deviations away from the mean  $CV$  for that month,  $CV_m$ . We recognize that this procedure cannot detect spurious data that have similar  $CV$  values to real data; therefore this method should be considered conservative. By removing spurious months of data, we note that the gridding algorithm will subsequently search for another proximal station to estimate gridded precipitation, which may increase or decrease the estimate. Similarly, adding short-lived stations or subnetworks, as is particularly common for Mexico (short-term precipitation network described below), will affect the heterogeneity of the precipitation field, which we do not explicitly quantify here.

Another issue that arose during the gridding process involved the time of observation, TOBS, for Canadian stations, which were initially interpreted as local time readings (as is the case for CONUS) in L13. However, through a lagged correlation analysis it was determined that times were recorded in Greenwich Mean Time (GMT) resulting in an asynchronous offset in meteorological events, which was corrected by adjusting the station TOBS entries to local time.

### Code availability

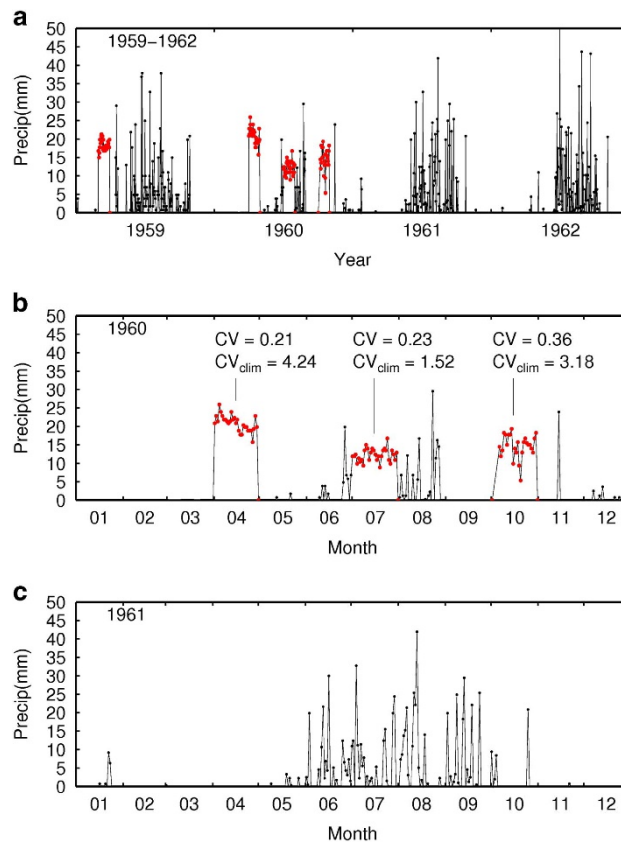
Customized C++ code was used for the major gridding operation in this data set that is publicly available alongside the data set at the Lawrence Livermore National Laboratory (LLNL) accessible here ([ftp://192.12.137.7/pub/dcp/archive/OBS/livneh2014.1\\_16deg/](ftp://192.12.137.7/pub/dcp/archive/OBS/livneh2014.1_16deg/)). Further, processing of the data in network Common Data Form (netCDF) format was done for remapping, aggregating, and scaling of climatologies using open source Climate Data Operators (CDO) and netCDF Operator (NCO) utilities. Hydrologic simulations were performed with publicly available model code (<http://www.hydro.washington.edu/Lettenmaier/Models/VIC/>).

### Data Records

A list of data sources used to build this data set are included in Table 1. Daily Canadian precipitation, maximum and minimum temperature station data were obtained from Environment Canada, and CONUS data from the National Climatic Data Center (NCDC<sup>39</sup>). For Mexico data were provided by the Servicio Meteorológico Nacional (SMN), under the Comisión Nacional del Agua (CONAGUA), for the period 1950–2013. However, since 2000, a sharp decline in Mexican station density was noted in several states, particularly after 2006. To fill gaps for the period 2000–2013, we contacted CONAGUA's regional offices in the states of Chihuahua, Sonora, Coahuila, Durango, Nayarit Puebla, and Yucatan. Additionally, 86 precipitation stations distributed across the Sierra Madre Occidental from the North American Monsoon Experiment Event Rain Gauge Network (NERN)<sup>40</sup> were incorporated, which provide an important sampling of high elevation precipitation absent from the above station sources.

The final data set contains gridded station data for precipitation, daily maximum and minimum temperature, in addition to wind<sup>41</sup> data from National Centers for Environmental Prediction (NCEP) National Centers for Atmospheric Research (NCAR) reanalysis. Also included are hydrologic model outputs over Mexico, CONUS, and southern Canada at a 1/16° resolution for the period 1950–2013. The





**Figure 2.** Demonstration of precipitation station quality control for Mexico. Illustration of spurious precipitation data (red circles) identified during a quality control procedure for Mexican station data—shown for Station ID: 14111, Poncitlan, Jalisco, Mexico. Values of the coefficient of variation (CV) of daily values for each of the months flagged as bad data in 1960 are listed, along with the climatological average CV values for those months. Panel (a) illustrates spurious records alongside records deemed non-spurious for a 4 year period, with respective panels (b) and (c) illustrating these features in greater detail for an individual year.

data are intended to support downscaling and long-term hydroclimatic studies. The meteorological data are archived at the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Information (NCEI), with access details provided in the Data Citation 1. An additional copy of the data, including derived hydrologic output are hosted in a public repository, the LLNL site listed above. Variables (Table 2) are provided in netCDF format.

### Technical Validation

The impact of scaling CONUS precipitation to PRISM climate normals over the period 1981–2010 rather than 1961–1990 (as used in L13 and M02) is illustrated in Fig. 3. Scaling ratios were computed separately for the corresponding 30-year periods for the gridded data relative to the respective normals. Outside CONUS, Fig. 3 shows the difference between scaling precipitation to V14 and using no scaling.

To understand potential hydrologic impacts of the scaling choice, the Variable Infiltration Capacity (VIC) hydrologic model<sup>42</sup> was run with the meteorological forcing data developed here, as well as with required wind data that were interpolated through a bi-cubic approach from a larger (approximately 1.9° grid) reanalysis grid<sup>41</sup>. VIC model parameters in CONUS were obtained from L13, who validated model discharges over major CONUS river basins. 1/16° parameters over Mexico were obtained from previously published work<sup>9</sup>, for which discharges were validated in selected basins<sup>7,33</sup>. Canadian parameters were obtained at 1/8° from M02 and disaggregated to 1/16° using a nearest neighbor approach. We advise users of the simulated hydrologic states and fluxes to judge the fidelity of those simulations by referring to the aforementioned analyses that calibrated and validated VIC over most of CONUS and parts of Canada and Mexico, respectively.

Outside CONUS, precipitation differences between the scaled (V14 1981–2010) and unscaled grids can be substantial, up to 400 mm/yr in mountainous regions, which translate to fractional differences up to 75% although typically less than 25% over these regions. Particularly large differences are found in the Canadian Rocky Mountains and Coastal Range, the Sierra Madre (Occidental and Oriental), and the Baja

Data type	Source	Details
Precipitation, Maximum and Minimum Temperature	CONUS: NCDC ( <a href="ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily">ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily</a> <sup>38</sup> )	Precipitation CONUS: 21137 total stations; 8844 satisfy stability constraint Canada: 7967 total station; 2245 satisfy stability constraint Mexico 5430 stations all satisfy the selection criteria (86 stations from NERN)
	Canada: Environment Canada	Maximum Temperature CONUS: 16919 total stations; 6159 satisfy stability constraint Canada: 7841 total station; 1875 satisfy stability constraint Mexico 5246 total stations; 5220 satisfy the selection criteria
	Mexico: Servicio Meteorológico Nacional, under the Comisión Nacional del Agua <sup>†,‡,§</sup>	Minimum Temperature CONUS: 16983 total stations; 6275 satisfy stability constraint Canada: 7967 total station; 1913 satisfy stability constraint Mexico 5243 total stations; 5220 satisfy the selection criteria
Wind data	NCEP–NCAR reanalysis <sup>41</sup>	

**Table 1.** Summary of daily data sources used to build the present data set. The decadal average and standard deviation of the number of stations reporting for each country within the sampling domain is Canada: 1861, 362; CONUS: 7489,476; Mexico: 2611, 772, respectively. \*Puebla, Boris Isauro Hernandez, Jefe del Centro de Prevision Meteorologica. Email: [boris.hernandez@conagua.gob.mx](mailto:boris.hernandez@conagua.gob.mx) phone: 222-211-8378. †Tlaxcala, Juana Angelica Diaz Trenado, Jefe del Departamento de Hidrometria. Email: [juana.diaz@conagua.gob.mx](mailto:juana.diaz@conagua.gob.mx) phone: 246-468-9305. ‡Nayarit, Ruben Cambero Borrayo, Residente del area de infraestructura hidroagricola, Email: [ruben.cambero@conagua.gob.mx](mailto:ruben.cambero@conagua.gob.mx) phone: 311-214-2385. §Nayarit, Arnulfo Saldierna, Asistencia tecnica y control de calidad, Email: [arnulfo.saldierna@conagua.gob.mx](mailto:arnulfo.saldierna@conagua.gob.mx) phone: 311-210-4346 (director's office).

Variable Name	Units	Source
Precipitation	mm	station
Maximum Temperature	°C	station
Minimum Temperature	°C	station
Wind speed	m/s	reanalysis
Surface Runoff	mm	VIC
Baseflow	mm	VIC
Total ET	mm	VIC
Snow Water Equivalent	mm	VIC
Soil Moisture	mm	VIC
Canopy Moisture	mm	VIC
Latent Heat Flux	W/m <sup>2</sup>	VIC
Sensible Heat Flux	W/m <sup>2</sup>	VIC
Ground Heat Flux	W/m <sup>2</sup>	VIC
Net Radiation	W/m <sup>2</sup>	VIC
Potential ET	mm	VIC

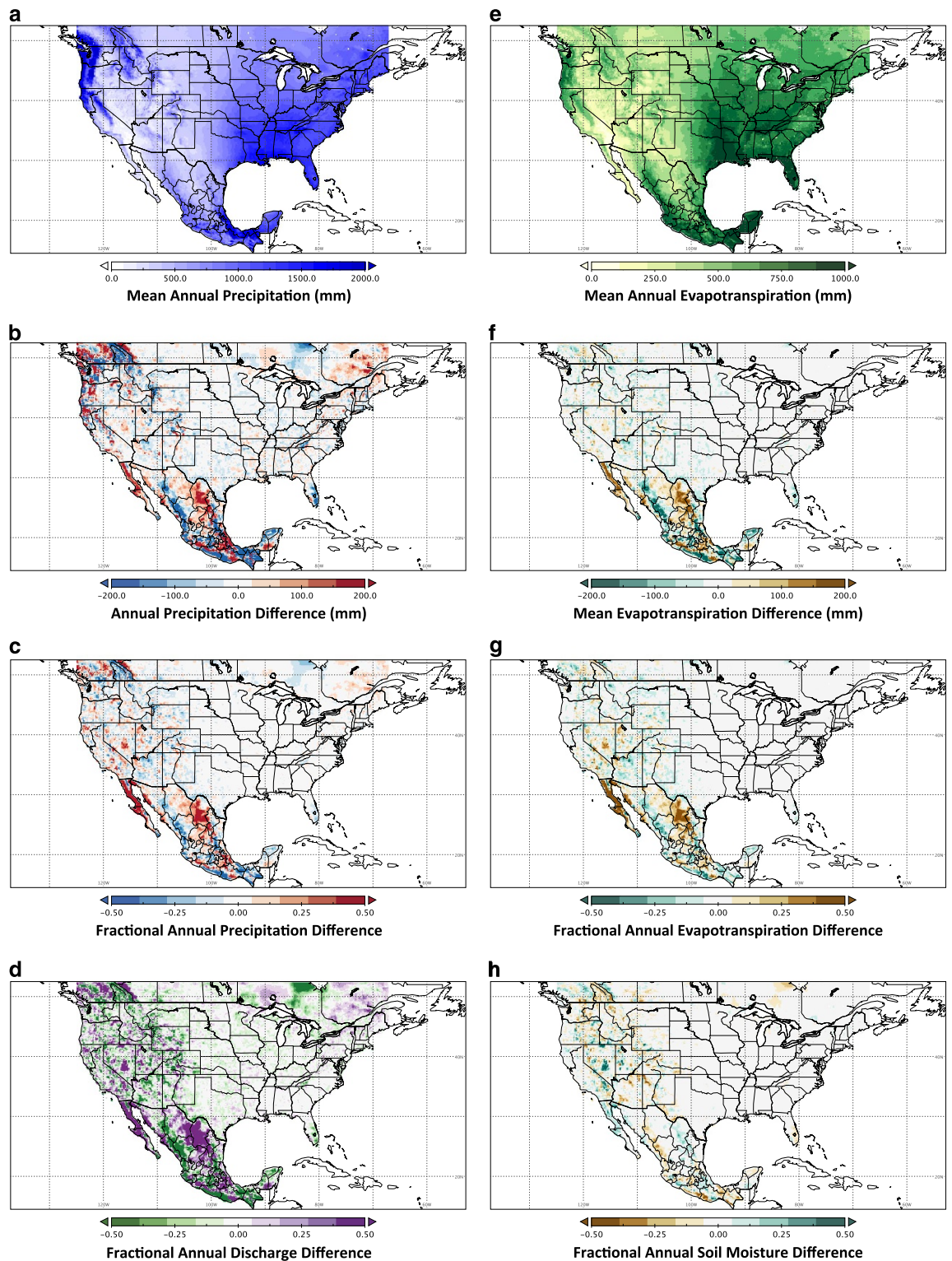
**Table 2.** List of publically available daily hydrometeorological variables.

Peninsula—areas of enhanced topographical relief—with progressively smaller differences over the modest topography of eastern Canada and flatter areas.

Within CONUS, precipitation was scaled to PRISM 1981–2010 for consistency with Mexico and Canada. Differences between this scaling and the 1961–1990 found in L13 and M02 are still notable, although generally less than 200 mm/yr with the exception of wet coastal areas, which altogether translate to fractional differences of less than 50% and typically less than 25%. For all areas, fractional differences in simulated total discharge between the two scaling cases are of slightly larger fractional sensitivity than precipitation, and are relevant for water resource management. Differences in ET and total column soil moisture are less sensitive than discharge, exhibiting sensitivities generally less than 25%/50% inside/outside CONUS between scaling cases.

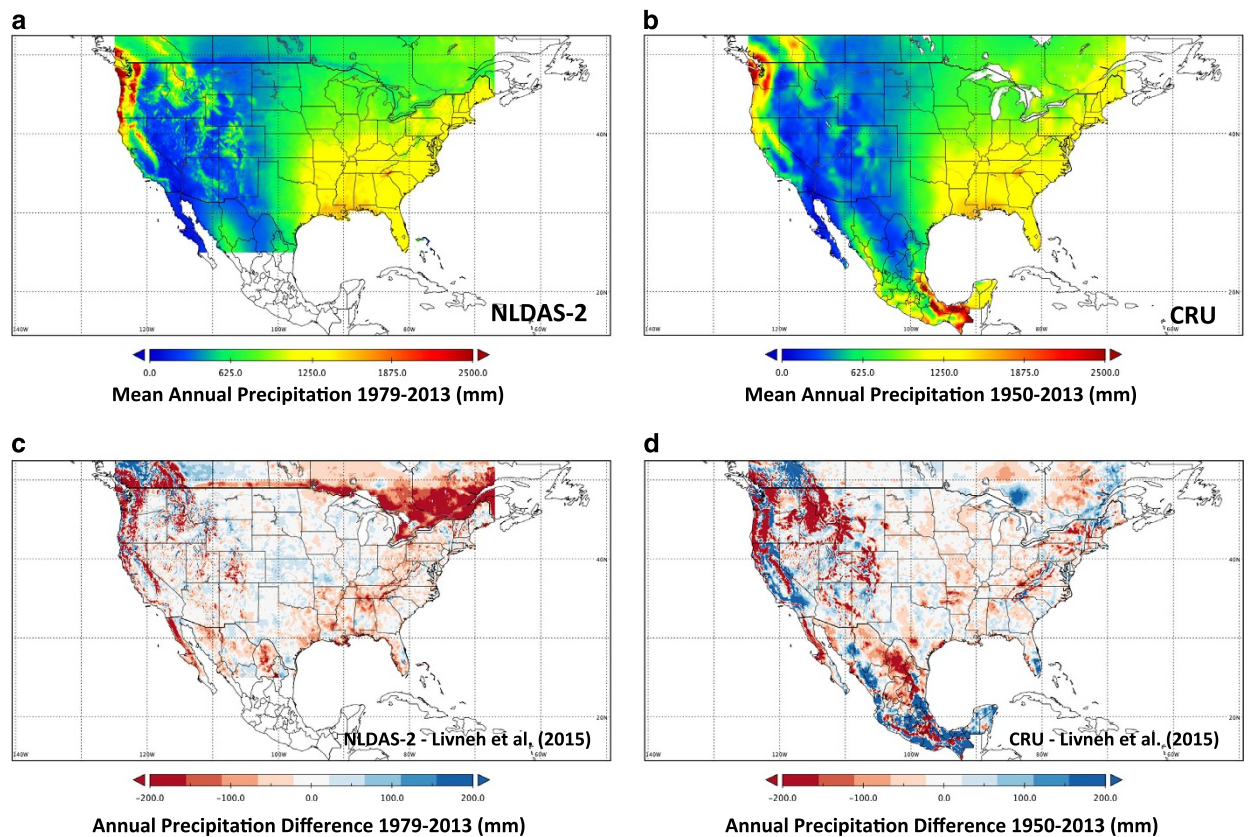
### Spatio-temporal continuity, transboundary consistency, and the north American monsoon

The L15 dataset offers several advantages over other gridded observational products that cross international boundaries (Fig. 4). To illustrate the advantages of L15, we compare it with the commonly



**Figure 3.** Orographic scaling comparisons. Differences in mean annual conditions (1950–2013) between two precipitation scaling scenarios: the first is similar to L13 (i.e., the CONUS scaled to PRISM 1961–1990 normals and no scaling elsewhere), while the second case is the present data set (scaled to PRISM 1981–2010 within CONUS and Vose *et al.*<sup>35</sup> elsewhere). Differences show the former case minus the latter. (a) shows precipitation for the present case, (b) and (c) show precipitation differences, (e) shows ET simulated by the VIC hydrologic model using the present precipitation scaling case; differences between the two scaling scenarios in discharge ( $Q$  = runoff+baseflow) are shown in (d), evapotranspiration (f,g) and soil moisture (h).





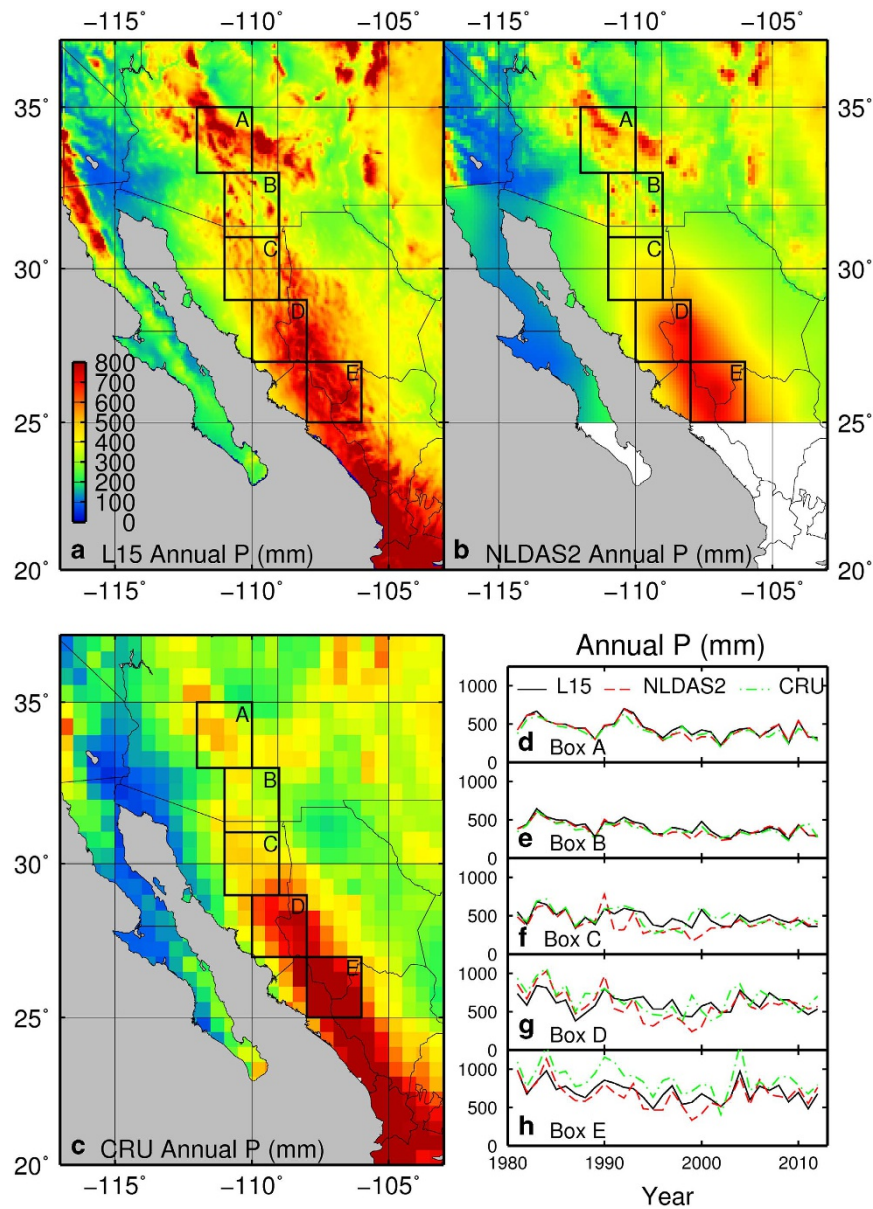
**Figure 4.** Inter-dataset precipitation comparison over the full domain. Mean annual precipitation (1979–2013) in (a) and NLDAS2 and (b) CRU v.3.22, as well difference maps NLDAS2 minus L15 (c) and CRU minus L15 (d). There is a notable transboundary discontinuity in NLDAS2, and NLDAS2 is generally drier than L15.

used 1/8°-resolution North American Land Data Assimilation System phase-2 (NLDAS2) product (104 google scholar citations<sup>21</sup>) as well as the Climate Research Unit (CRU<sup>43</sup>) data (461 google scholar citations). Limitations of NLDAS2 relative to L15 include: (i) shorter duration (NLDAS2 begins in 1979); (ii) lack of coverage south of 25° N; and (iii) transboundary discontinuities in both the topographic correction (derived from PRISM over the US, but not performed over Canada or Mexico) and the source of the precipitation data used. Over CONUS, NLDAS2 uses NCDC gauge data. Over Canada, NLDAS2 uses the North American Regional Reanalysis (NARR/R-CDAS) product, disaggregated from its coarser 32 km resolution. Over Mexico, NLDAS2 uses a hierarchy of data sets, with the first choice being a 1/4° gauge product (1° resolution prior to 2001), temporally disaggregated based on CMORPH satellite-retrieved 8 km data; followed by other data sets when these are not available (for more details, see: <http://ldas.gsfc.nasa.gov/nldas/NLDAS2forcing.php#AppendixC>). The major limitation of CRU relative to L15 is its comparatively coarse spatial resolution (0.5°) and temporal resolution (monthly), while it offers the advantage of being globally available and suitable for trend analyses given its gridding algorithm that computes station-anomalies about a normal period.

Figure 4 compares mean annual precipitation between L15 and NLDAS2 for their common period, as well as L15 and the CRU data, v.3.22 for the full period (1950–2013). The CRU data do not suffer any spatial clear spatial discontinuity, however, they clearly exhibit a smoother surface (Fig. 4, panel b) relative to L15 (Fig. 3). The difference map (Fig. 4, panel d) generally suggests that CRU underestimates precipitation in regions of high precipitation (i.e. mountainous terrain) and slightly underestimates in low precipitation regions relative to L15, which might be expected given its coarser resolution. The two exceptions are on the Yucatan Peninsula (MX) and the Canadian Rocky Mountains, where CRU overestimates precipitation relative to L15. Although the authors of CRU suggest “This dataset should only be used for climate trend analysis, ...”<sup>43</sup>, we feel the comparisons are worthwhile to highlight major differences with that widely used dataset.

Both the raw NLDAS2 (Fig. 4, panel a) and the difference field (Fig. 4, panel c) show a clear discontinuity at the Canadian border, as well as another discontinuity at 50° N, indicating substantial dry biases in NLDAS2 over the Canadian portions of the Columbia and St Lawrence basins. A less dramatic discontinuity at the US-Mexico border leads to similar dry biases in NLDAS2 over the Mexican portions





**Figure 5.** Inter-dataset precipitation comparison for the North American Monsoon region. Annual total precipitation over the period 1981–2012, showing (a–c) maps of mean annual total precipitation from L15, NLDAS2, and CRU, respectively; and (d–h) timeseries of spatial average annual total precipitation from all products for the five boxes labelled A–E in (a–c).

of the Colorado River and Rio Grande/Rio Bravo basins. Overall, L15 is generally wetter than NLDAS2 over areas of topographical complexity.

Discontinuities at the US–Mexico border are particularly important to studies of the North American Monsoon. Summer monsoon rains provide 30–80% of the annual precipitation across northwestern Mexico and the southwestern US between 20° and 35° N (ref. 44). Predicting and understanding this highly variable regional water resource is vital for planning efforts, yet models have difficulty reproducing monsoon precipitation, leading to poor forecast skill<sup>45</sup>.

L15 represents a substantial improvement over existing datasets in the monsoon region. As shown in Fig. 5a–c, L15 and CRU completely cover the monsoon region, but NLDAS2 ends at 25° N. Between 25° N and the US–Mexico border, the lack of an elevation correction in NLDAS2 leads to a precipitation field that has not only a diffuse spatial distribution and unrealistic relationship to topography, but also a substantial discontinuity at the US–Mexico border. L15 and CRU give summer precipitation totals that are more consistent with previous observational studies<sup>46</sup>. However, CRU’s relatively coarse spatial resolution (0.5°) limits its usefulness in basin-scale hydrologic studies within the North American Monsoon region, due to the topographic complexity there. It is worth noting that other available products

not shown here suffer from similar problems to those shown: the North American Regional Reanalysis (NARR<sup>47</sup>) lacks an elevation correction over Mexico, leading to a discontinuity similar to that of NLDAS2; other global gridded meteorology datasets such as the Global Precipitation Climatology Project (GPCP<sup>48</sup>) have coarser spatial resolution than CRU.

Temporal differences among these products also exhibit a discontinuity at the US-Mexico border. Along a transect of  $2 \times 2^\circ$  boxes from central Arizona to southern Sonora (boxes A-E in Fig. 5a-c), the timeseries of annual precipitation from the three products over the period 1981–2012 agree remarkably in the US (boxes A and B; Fig. 5d,e), but differ noticeably over Mexico, particularly during the 1990s (boxes C-E; Fig. 5f-h). All three products capture the 1995–2004 drought<sup>49,50</sup> in the Yaqui River basin (boxes B-D) to varying extents (Fig. 5e-g). L15 bears the strongest resemblance to a previously published timeseries for the region<sup>49</sup> using 97 meteorological stations in Arizona and Sonora, including relatively high precipitation in years 1991–1994, followed by a declining trend through 2004, punctuated by above average precipitation in 1997 and 2000. However, in boxes C and D (Fig. 5d,e), NLDAS2 begins the drought early (in 1991) and CRU ends the drought early (in 2000 in box C and in 2004 in box D). Thus, while differences among precipitation products warrant further investigation, the L15 dataset promises to be of major benefit to studies of hydrology and water resources in the North American Monsoon region.

### Usage Notes

We caution users that similar to predecessor data sets, M02 and L13, the L15 data presented here are not suitable for trend analysis, since they use many stations that do not span the full temporal period 1950–2013. Hence, sampling evolves over time, especially for the country of Mexico over which we used a much less restrictive stability constraint, i.e., >50 days of valid data required for station inclusion, versus >20 years of data for CONUS and Canadian stations. Despite our best efforts to minimize transboundary discontinuities, we identified a discontinuity in daily minimum temperature, T<sub>min</sub>, across the MX/US. In fact, no discontinuities are apparent in seasonal climatologies for T<sub>max</sub>, T<sub>min</sub>, or precipitation, yet when exploring a trend analysis, we note a small downward trend in T<sub>min</sub> in central and northern Mexico, that is opposite in sign north of the US/MX border. We investigated other data sets (not shown) and found a similar discontinuity in trend in the CRU data v.3.0, which is greatly reduced in versions v.3.10 and later (the most recent version is v.3.22 as of 15 July, 2015). The more recent CRU versions switched from gridding T<sub>max</sub> and T<sub>min</sub> separately, in favour of gridding T<sub>mean</sub> and diurnal temperature range (DTR) from which T<sub>max</sub> and T<sub>min</sub> are inferred. We further investigated this issue internally and note a downward trend in station elevations over much of Mexico, suggestive of increased sampling of inversion processes causing the downward trend in T<sub>min</sub> over time.

### References

- Booth, E. L. J., Byrne, J. M. & Johnson, D. L. Climatic changes in western North America, 1950–2005. *International Journal of Climatology* **32**, 2283–2300 (2012).
- Wolock, D. M. & McCabe, G. J. Estimates of Runoff using Water-Balance and Atmospheric General Circulation Models. *Journal of the American Water Resources Association* **35**, 1341–1350 (1999).
- Johnston, R. & Smakhtin, V. Hydrological Modeling of Large river Basins: How Much is Enough? *Water Resources Management* **28**, 2695–2730 (2014).
- Liu, X. & Heino, M. Comparing Proactive and Reactive Management: Managing a Transboundary Fish Stock Under Changing Environment. *Natural Resource Modeling* **26**, 480–504 (2013).
- Patiño-Gomez, C., McKinney, D. C. & Maidment, D. R. Sharing water resources data in the binational Rio Grande/Bravo Basin. *Journal of Water Resources Planning and Management* **133**, 416–426 (2007).
- Buzás, Z. *et al.* in *Transboundary Water resources Management: A Multidisciplinary Approach* (eds Ganoulis J., Aureli A., Fried J.) (Wiley-VCH Verlag GmbH and Co. KGaA, Weinheim, Germany, 2011).
- Muñoz-Arriola, F., Avissar, R., Zhu, C. & Lettenmaier, D. P. 2009) Sensitivity of the water resources of Rio Yaqui Basin, Mexico, to agriculture extensification under multiscale climate conditions. *Water resources research* **45**, W00A20 (2011).
- Fry, L. M. *et al.* The Great Lakes Runoff Intercomparison Project Phase 1: Lake Michigan (GRIP-M). *Journal of Hydrology* **519**, 3448–3465 (2014).
- Tang, Q., Vivoni, E. R., Muñoz-Arriola, F. & Lettenmaier, D. P. Predictability of evapotranspiration patterns using remotely sensed vegetation dynamics during the North American monsoon. *Journal of Hydrometeorology* **13**, 103–121 (2012).
- Robles-Morua, A., Vivoni, E. R. & Mayer, A. S. Distributed Hydrologic Modeling in Northwest Mexico Reveals the Links between Runoff Mechanisms and Evapotranspiration. *Journal of Hydrometeorology* **13**, 785–807 (2012).
- Zhu, C., Cavazos, T. & Lettenmaier, D. P. Role of antecedent land surface conditions in warm season precipitation over northwestern Mexico. *Journal of Climate* **20**, 1774–1791 (2007).
- Livneh, B. *et al.* A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States: Update and Extensions\*. *Journal of Climate* **26**, 9384–9392 (2013).
- Livneh, B. & Lettenmaier, D. P. Multi-criteria parameter estimation for the unified land model. *Hydrology and Earth System Sciences* **16**, 3029–3048 (2012).
- Pan, M. *et al.* Multisource estimation of long-term terrestrial water budget for major global river basins. *Journal of Climate* **25**, 3191–3206 (2012).
- Ferguson, C. R., Sheffield, J., Wood, E. F. & Gao, H. Quantifying uncertainty in a remote sensing based estimate of evapotranspiration over the continental United States. *International Journal of Remote Sensing* **31**, 3821–3865 (2010).
- Mu, Q. *et al.* Evaluating water stress controls on primary production in biogeochemical and remote sensing based models. *Journal of Geophysical Research: Biogeosciences (2005–2012)* **112**, G01012 (2007).
- Swenson, S. C. & Wahr, J. Post-processing removal of correlated errors in GRACE data. *Geophysical Research Letters* **33**, L08402(2006).
- Ebert, E. E., Janowiak, J. E. & Kidd, C. Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bulletin of the American Meteorological Society* **88**, 47–64 (2007).

19. Pechlivanidis, I. G., Jackson, B. M., McIntyre, N. R. & Wheeler, H. S. Catchment scale hydrological modelling: a review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. *Global NEST Journal* **13**, 193–214 (2011).
20. Sheffield, J., Goteti, G. & Wood, E. F. Development of a 50-yr high-resolution global dataset of meteorological forcings for land surface modeling. *Journal of Climate* **19**, 3088–3111 (2006).
21. Adam, J. C., Clark, E. A., Lettenmaier, D. P. & Wood, E. F. Correction of global precipitation products for orographic effects. *Journal of Climate* **19**, 15–38 (2006).
22. Xia, Y. *et al.* Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System Project Phase 2 (NLDAS-2), part 1: intercomparison and application of model products. *Journal of Geophysical Research* **117**, (2012).
23. Sturaro, G. A closer look at the climatological discontinuities present in the NCEP/NCAR reanalysis temperature due to the introduction of satellite data. *Climate Dynamics* **21**, 309–316 (2003).
24. Li, B. & Rodell, M. Evaluation of a model-based groundwater drought indicator in the conterminous US. *Journal of Hydrology* **526**, 78–88 (2014).
25. Zhou, T., Nijssen, B., Huffman, G. J. & Lettenmaier, D. P. Evaluation of real-time satellite precipitation data for global drought monitoring. *Journal of Hydrometeorology* **15**, 1651–1660 (2014).
26. Schermerhorn, V. P. Relations between topography and annual precipitation in western Oregon and Washington. *Water Resources Research* **3**, 707–711 (1967).
27. Livneh, B. *et al.* Filling in the gaps: Inferring spatially distributed precipitation from gauge observations over complex terrain. *Water Resources Research* **50**, 8589–8610 (2014).
28. Basist, A., Bell, G. D. & Meentemeyer, V. Statistical relationships between topography and precipitation patterns. *Journal of Climate* **7**, 1305–1315 (1994).
29. Stewart, I. T., Cayan, D. R. & Dettinger, M. D. Changes in snowmelt runoff timing in western North America under abusiness as usual climate change scenario. *Climatic Change* **62**, 217–232 (2004).
30. Gochis, D. J., Brito-Castillo, L. & Shuttleworth, W. J. Hydroclimatology of the North American Monsoon region in northwest Mexico. *Journal of Hydrology* **316**, 53–70 (2006).
31. Daly, C., Taylor, G. H. & Gibson, W. P. The PRISM approach to mapping precipitation and temperature. in *Proc., 10th AMS Conf. on Applied Climatology* 20–23 (1997).
32. Maurer, E. P., Wood, A. W., Adam, J. C., Lettenmaier, D. P. & Nijssen, B. A Long-Term Hydrologically-Based Data Set of Land Surface Fluxes and States for the Conterminous United States. *Journal of Climate* **15**, 3237–3251 (2002).
33. Zhu, C. & Lettenmaier, D. P. 2007) Long-term climate and derived surface hydrology and energy flux data for Mexico: 1925–2004. *Journal of Climate* **20**, 1936–1946.
34. Shepard, D. S.. *Computer mapping: The SYMAP interpolation algorithm. Spatial Statistics and Models* Gaile G. L. & Willmott C. J. Eds. D. Reidel, 133–145 (1984).
35. Vose, R. S. *et al.* Improved historical temperature and precipitation time series for US climate divisions. *Journal of Applied Meteorology and Climatology* **53**, 1232–1251 (2014).
36. Wehner, M., Easterling, D. R., Lawrimore, J. H., Heim, R. R. Jr. & Vose, R. S. Projections of future drought in the continental United States and Mexico. *Journal of Hydrometeorology* **12**, 1359–1377 (2011).
37. Hutchinson, M. F. *et al.* Development and testing of Canada-wide interpolated spatial models of daily minimum-maximum temperature and precipitation for 1961–2003. *Journal of Applied Meteorology and Climatology* **48**, 725–741 (2009).
38. Daly, C. Guidelines for assessing the suitability of spatial climate data sets. *International Journal of Climatology* **26**, 707–721 (2006).
39. Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E. & Houston, T. G. An overview of the Global Historical Climatology Network-Daily Database. *Journal of Atmospheric and Oceanic Technology* **29**, 897–910 (2012).
40. Gochis, D. J., Leal, J. C., Shuttleworth, W. J., Watts, C. J. & Garatuza-Payan, J. Preliminary diagnostics from a new event-based precipitation monitoring system in support of the North American Monsoon Experiment. *Journal of Hydrometeorology* **4**, 974–981 (2003).
41. Kalnay, E. *et al.* The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society* **77**, 437–471 (1996).
42. Liang, X., Lettenmaier, D. P., Wood, E. F. & Burges, S. J. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research: Atmospheres (1984–2012)* **99**, 14415–14428 (1994).
43. Harris, I., Jones, P. D., Osborn, T. J. & Lister, D. H. Updated high-resolution grids of monthly climatic observations—the CRU TS3.10 Dataset. *International Journal of Climatology* **34**, 623–642 (2014).
44. Higgins, W. & Gochis, D. Synthesis of results from the North American Monsoon Experiment (NAME) process study. *Journal of Climate* **20**, 1601–1607 (2007).
45. Castro, C. L. *et al.* Can a regional climate model improve the ability to forecast the North American Monsoon? *Journal of Climate* **25**, 8212–8237 (2012).
46. Gutzler, D. S. An index of interannual precipitation variability in the core of the North American Monsoon region. *Journal of Climate* **17**, 4473–4480 (2004).
47. Mesinger, F. *et al.* North American Regional Reanalysis. *Bulletin of the American Meteorological Society* **87**, 343–360 (2006).
48. Adler, R. F. *et al.* The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present). *Journal of Hydrometeorology* **4**, 1147–1167 (2003).
49. Hallack-Alegria, M. & Watkins, D. W. Jr. Annual and warm season drought intensity-duration-frequency analysis for Sonora, Mexico. *Journal of Climate* **20**, 1897–1909 (2007).
50. Munoz-Hernandez, A., Mayer, A. S. & Watkins, D. W. Jr. Integrated Hydrologic-Economic-Institutional Model of Environmental Flow Strategies for Rio Yaqui Basin, Sonora, Mexico. *Journal of Water Resource Planning and Management* **137**, 227–237 (2011).

## Data Citations

1. Livneh, B. *et al.* NOAA National Centers for Environmental Information, doi:10.7289/V5X34VF6 (2015).

## Acknowledgements

B. Livneh was funded by United States Bureau of Reclamation (BOR: R11AC81334): Water and Climate Change Research, including Technical Support of Reclamation Programs. T. Bohn was funded by grant 1216037 of the National Science Foundation Science, Engineering, and Education for Sustainability (SEES) Fellows program. We are grateful to CONAGUA’s Servicio Meteorológico Nacional and Regional offices, as well as Dr Agustín Robles-Morua of Instituto Tecnológico de Sonora who was instrumental in obtaining station records for Northwest Mexico.



### Author Contributions

B.L. did the gridding of station data, running of the hydrologic model, and led the manuscript writing. T.B. aided in acquisition and quality control of Mexican station data and performed analysis of precipitation over the North American Monsoon region. D.P. aided in quality control of the data and with coding and memory issues during the gridding process. F.M. was essential in acquiring Mexican station data. B.N. aided in the acquisition and gridding of the wind data. R.V. made his climatological product available and provided description and oversight for its use. D.C. provided guidance throughout the gridding process, involved in key methodological decisions. L.B. provided guidance throughout the gridding process, involved in key methodological decisions.

### Additional Information

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Livneh, B. *et al.* A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950–2013. *Sci. Data* 2:150042 doi: 10.1038/sdata.2015.42 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0>

Metadata associated with this Data Descriptor is available at <http://www.nature.com/sdata/> and is released under the CC0 waiver to maximize reuse.